Transparency, Governance and Markets

M. Bagella, L. Becchetti and I. Hasan EDITORS

TRANSPARENCY, GOVERNANCE AND MARKETS

This Page is Intentionally Left Blank

TRANSPARENCY, GOVERNANCE AND MARKETS

EDITED BY

PROFESSOR MICHELE BAGELLA

University of Rome
Faculty of Economics,
Rome, Italy

PROFESSOR LEONARDO BECCHETTI

University of Rome
Economics Department
Rome, Italy

PROFESSOR IFTEKHAR HASAN

Lally School of Management and Technology
Rensselaer Polytechnic Institute
Troy, NY, USA

ELSEVIER

Amsterdam • Boston • Heidelberg • London • New York • Oxford
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

First edition 2006

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN-13: 978-0-444-52722-6
ISBN-10: 0-444-52722-2

For information on all Elsevier publications
visit our website at books.elsevier.com

Printed and bound in The Netherlands

06 07 08 09 10 10 9 8 7 6 5 4 3 2 1

Contents

Contributors

Michele Bagella University of Rome II "Tor Vergata", Faculty of Economics, Via Columbia n.2, I-00133 Rome, Italy

Leonardo Becchetti Department of Economics, University of Rome "Tor Vergata", Via Columbia, 2, 00133 Rome, Italy

Wolfgang Bessler Professor of Finance, Center for Finance and Banking, Justus-Liebig-University Giessen, Licher Strasse 74, 35394 Giessen, Germany

Simone Borra University of Rome "Tor Vergata", Department of SEFEMEQ

Stefano Caiazza University of Rome "Tor Vergata", Department of Economics and Institution, Via Columbia n.2, I-00133 Rome, Italy

Andrea Carpentieri NEXTRA

Sungho Choi Lally School of Management & Technology, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, USA 12180-8196

Rocco Ciciretti University of Rome "Tor Vergata", Faculty of Economics, Via Columbia, n.2, I-00133 Rome, Italia

John Cotter Director of Centre for Financial Markets, Department of Banking and Finance, Smurfit School of Business, University College Dublin, Blackrock, Co. Dublin, Ireland

Douglas J. Cumming Director, Severino Center for Technological Entrepreneurship, Lally School of Management & Technology, Rensselaer Polytechnic Institute, Troy, NY, USA 12180

Gerald P. Dwyer, Jr. Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree St. N.E. Atlanta, GA, USA 30309-4470

Stefania Di Giacomo University of Rome "Tor Vergata", Department of Economics and Institution, Via Columbia n.2, I-00133 Rome, Italy

Rifat Gorener Assistant Professor of Finance, Isik University, Istanbul, Turkey

Iftekhar Hasan Lally School of Management & Technology, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, USA 12180

Subal C. Kumbhakar Department of Economics, State University of New York, Binghamton, NY, USA 13902

Robert Lensink University of Groningen, Department of Finance, PO Box 800, 9700 AV Groningen, The Netherlands

François Longin Professor of Finance, Department of finance, ESSEC Graduate Business School, Avenue Berhard Hirsch, B.P. 105-95021, Cergy Pontoise Cedex France

Jeffrey G. MacIntosh Toronto Stock Exchange Professor of Capital Markets, Faculty of Law, University of Toronto, 78 Queen's Park, Toronto, Ontario, Canada M5S 2C5

Donato Masciandaro Professor of Monetary Economics, Paolo Baffi Centre, Bocconi University, Via Sarfatti 25, 20136 Milan, Italy

George Mavrotas Research Fellow & Project Directory, World Institute for Development Economics Research (WIDER) of the United Nations University, Katajanokanlaituri 6 B FIN-00160 Helsinki

James P. Murtagh Froman Professor of Management & Finance, The Sage Colleges, Albany, NY, USA 12208-3425

John R. Norsworthy Professor of Finance, Lally School of Management & Technology, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, USA 12180-8196

Zeno Rotondi Università degli Studi di Ferrara, Department of Economics, Corso Ercole I d'Este 44, Ferrara, 44100, Italy

Margarita Samartin University of Carlos III, Departamento de Economía de la Empresa, Calle Madrid 126, 28903 Getafe, Spain

Heiko Schmiedel European Central Bank, Frankfurt am Main, Germany

Bert Scholtens University of Groningen, Economie, P.O. Box 800, 9700 AV Groningen, The Netherlands

Dona D. Siregar Assistant Professor of Finance, State University of New York – Oneonta, USA

Nazaria Solferino CEIS University of Rome "Tor Vergata", Via Columbia, 2, 00133 Rome, Italy

Sang-Ik Son Incheon Free Economic Zone Authority, Incheon, S. Korea

Gabriele Susinno Computational and Quantitative Finance, Finance and Physics, Finance-and-Physics.org, Piazza Mincio 2, 00198 Rome, Italy

Natalia Utrer-González University Autonoma de Barcelona, 08193 Cerdanyola del Valles, Barcelona, Spain

Cristiano Zazzara Swiss Federeal Institute of Technology, EPFL and Libera Università degli Studi Sociali (LUISS) Guido Carli – Fondo Interbancario di Tutela dei Depositi and Instituto di Studi Economici, Via del Plebiscito, 102, 00162 Rome, Italy

This Page is Intentionally Left Blank

Introduction

Financial markets play a pivotal role in contemporary market economies in several respects. The most important of them is probably represented by the matching realized between people with profitable investment, which can contribute to the creation of economic value at aggregate level, with those having money needed to finance such investment. To perform this specific, task a specialized set of agents called financial intermediaries progressively emerged. The fundamental role of financial intermediaries is that of pooling individual resources, which are insufficient to finance large projects *per se*, and of providing liquidity services, by assuming liquid liabilities toward individual depositors and illiquid assets toward entrepreneurs. In a framework in which the relationship between individual financiers and investors is riddled by informational asymmetries (*ex ante* screening of the most valuable investment project, interim monitoring of investor effort in the entrepreneurial venture and *ex post* verification of the outcome declared by investors in order to avoid strategic default), financial intermediaries also emerged as the most efficient conduit to channel financial resources to investment, given their capacity of realizing economies of scale in the (screening, monitoring) costs required by the above-mentioned informational asymmetries. Finally, well-developed and efficient financial markets are fundamental in that they provide quality asset transformation services and allow individuals with heterogeneous propensity to risk exchanging it both cross-sectionally and intertemporally.

In a world riddled by informational asymmetries and conflicts of interests between different economic agents, these fundamental tasks have never been performed without problems and inherent fragilities. Just to mention some of them, limits in the screening of investment projects has always generated a problem of credit risk and non-performing loans, the latter enhancing the fragility of financial intermediaries implicit in their provision of liquidity services where liquid liabilities toward depositors expose them to the risk of bank runs.

Moreover, in a world in which the high-tech revolution has dramatically reduced the costs of transferring non-physical goods and services, the increased speed of financial transactions has amplified and enhanced the risk that the formation of expectations on values of financial assets generate destabilizing dynamics leading to financial crises or bubbles.

Theoretical and empirical research of these last decades is working on the positive and normative side, in order to deepen its understanding of financial market dynamics and to tackle new and old challenges with the ambitious goal of limiting fragilities and inefficiencies.

Contributions collected in this book represent a valuable and remarkable endeavor in this direction covering different topics.

A first one is related to the aggregate relationship between development of financial markets and economic growth. In this specific field, the focus is on the development of new econometric techniques and in the research of new proxies of financial variables in order to shed light on the well-known nexus between per capita GDP growth and financial deepening. The attempt is to disentangle the direction and the strength of the different causal links that this general result may hide with the help of new methodologies, which try to make it easier the solution of endogeneity problems.

A second topic covered is credit risk. The contribution included in this book directly deals with the issue of improving credit scoring methodologies of financial intermediaries, in order to being able to read and interpret always better signals provided by safe and risky borrowers. The direction of the research is in the refinement of validation techniques capable of measuring with accuracy the out of sample performance of different credit scoring methodologies and trying to increase generality of results in a field in which sample and time specificity is a dominant issue, which limits the extensibility and the significance of successful experiments.

A third important topic is related to the measure of risk in equity and bond markets. The general scope of the research here goes in the direction of extracting different orthogonal risk factors in order to have more precise measure of risk adjusted asset returns. Moreover, an important related issue is the analysis of the dynamics of the "blackbox" of financial markets, risk premia, in a field in which economics and psychology are so closely related to each other. The research here is essential as it can contribute to understanding apparently unexplained dynamics of asset pricing evaluating whether abnormal price movements may be rationally explained by changes in the differential return required by investors for holding a risky asset instead of a risk free one.

Finally, a fourth field covered is the one investigating behavior and efficiency of banking intermediaries. The point here is how to measure banking efficiency, considering the delicate and particular role played by these fundamental financial intermediaries and how transparency and governance rules may help in solving problems generated by informational asymmetries and conflicts of interest, which may limit banking efficiency.

As far as the economy develops and becomes more complex and integrated, the role of financial markets becomes always more essential and central to economic growth and global welfare. In parallel, new and old threats and challenges to its correct functioning and to the performance of its crucial role need to be tackled by empirical and theoretical research.

Overall, contributions collected in the book provide updated evidence and cover new theoretical issues arising in the field. They provide some new solutions but also highlight new and emerging problems and create new questions for further theoretical and empirical research.

Michele Bagella
Leonardo Becchetti
Iftekhar Hasan

This Page is Intentionally Left Blank

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 1

THEORETICAL EXPLANATIONS OF WHY BANKS PROMISE TO PAY PAR ON DEMAND

Gerald P. Dwyer, Jr. and Margarita Samartín

Why do banks promise to pay par on demand? There is a large and growing theoretical literature trying to answer this question, which we survey in this chapter. This theoretical literature can be broadly divided into four strands—liquidity provision, asymmetric information, regulatory restrictions, and a medium of exchange. One strand of the literature argues that banks offer to pay par on demand to provide liquidity insurance to consumers who are uncertain about their future time preferences. These consumers have investment opportunities inconsistent with some of their preferred consumption paths. A common assumption in most of the literature is that demand deposits cannot be traded, which suggests regulatory restrictions that prevent banks and active markets from co-existing. A second strand of the literature argues that banks offer to pay at par as a way to protect un-informed depositors, who would otherwise be disadvantaged relative to better-informed depositors and bank managers if equity contracts were employed instead. The deposit is then on demand to make its value not contingent on states that are not verifiable by the depositor. In this sense, demand deposit contracts are a discipline device because the promise to pay par on demand helps to limit the riskiness of banks' activities. The third strand of the literature argues that banks promise to pay par on demand because of legal restrictions, which prohibit other securities from playing the same role as demand deposits. Finally, other models have been built, based on the observation that bank liabilities function as a medium of exchange.

We conclude that there are sharp predictions by the relevant theories. We assume that it is not zero cost to make a promise to redeem a liability at par value on demand. If so, then the antecedent conditions in the theories are possible explanations of the reasons for banks promising to pay par on demand. If the explanation based on customers' demand for liquidity is correct, payment of deposits at par will be promised when banks hold assets that are illiquid in the short run. If the asymmetric-information explanation based on the difficulty of valuing assets is correct, the marketability of banks' assets determines whether banks promise to pay par. If the legal restrictions explanation of par redemption is correct, banks will not promise to pay par if they are not required to do so. If the medium of exchange explanation is correct, banks will promise to pay par when their liabilities are used as a medium of exchange.

1. INTRODUCTION

Banks promise to pay the par value of certain liabilities on demand in terms of other assets. This has been a long-standing practice, even though it is obvious that, due to gamblers' ruin, no bank can expect to honor this promise forever with less than 100% reserves. Just as significantly, no bank customer can expect it to be honored always either. As time goes to infinity, the probability of breaching this contract goes to 1 under general assumptions. In addition, the consequences—banking panics—are not trivial. In the United States, banking panics happened during the free banking and National Banking periods and at the start of the Great Depression. These are far from unique historically, and financial crises in emerging countries are more recent related events.[1]

Given this situation, why do banks promise what they cannot deliver forever in the first place?

It is possible that banks promise to pay par on demand because depositors want this contractual agreement. There are at least four possible reasons for this desire. Depositors may demand a constant par value because this makes their deposit balances more predictable, thereby increasing the liquidity of deposits compared to assets that have a longer maturity. At many times and in many places, banks have held largely non-marketable assets, which means that customers cannot easily assess the assets' market values. Under these circumstances, deposit values varying with the value of banks' assets may not be a feasible market

[1] For United States history, Dwyer (1996) summarizes some banking panics before the Civil War in the United States, Sprague (1910) summarizes banking panics in the National Banking period, and Friedman and Schwartz (1963) analyze the banking panics at the start of the Great Depression. Banking problems have not ended with the establishment of central banks. Lindgren *et al*. (1996) indicate that 73% of the IMF's member countries suffered banking crises between 1980 and 1996.

equilibrium and redemption on demand can keep the bank from dissipating the depositors' wealth by exploiting superior information. Alternatively, depositors may want a constant par value because it is more convenient when using deposits in transactions. Alternatively, banks may make this promise because they are required to do so and such promises would not occur without that requirement.

In this chapter, we survey theories about banks' promise to pay par on demand, to determine whether these theories can make empirical predictions about when financial intermediaries will promise to pay par on demand. We assume that it is not zero cost to make a promise to redeem a liability at par value on demand. If so, then the antecedent conditions in the theories are possible explanations of the reasons why banks promise to pay par on demand. The theories can be interpreted as making strong predictions, namely that promised payment at par will not be observed unless the theory is relevant. Alternatively, the theories can be interpreted as making weak predictions in the sense that the theory explains some observed promises to pay par on demand. For example, a strong prediction based on customers' demand for liquidity would be that payment of deposits at par will be promised only if banks hold assets that have a longer maturity than deposits, or are otherwise illiquid. If the explanation based on asymmetric information about assets is correct, the marketability of banks' assets determines whether banks' promise to pay par on demand. If the explanation based on deposits' use as a medium of exchange is correct, then financial intermediaries will make such promises only if deposits are used in transactions. If the legal restrictions explanation of par redemption is correct, banks will promise to pay par only if they are required to do so.[2]

The theoretical literature that supports these predictions is summarized in the sections that follow. In this chapter, no attempt is made to examine the consistency of the evidence with the theories; that will be covered in later work.

2. LIQUIDITY PROVISION

One possible explanation for the use of demand deposit contracts is associated with the liquidity insurance provided by financial intermediaries. Diamond and Dybvig (1983), who formalized some of the ideas introduced by Bryant (1980), made a significant contribution by introducing a model of the demand for liquidity and the transformation service provided by banks. They demonstrated that demand deposit contracts, which enable the transformation of illiquid assets into

---

[2] Wallace (1996) does not mention legal restrictions in his analysis of narrow banking and dismissal of the importance of asymmetric information instead of Diamond and Dybvig's model.

more liquid liabilities of the financial intermediary, provide a rationale both for the existence of banks and for their vulnerability to runs.

2.1. Diamond and Dybvig

In the simplest formulation of this class of models, there is a continuum of ex ante identical agents who are risk averse and uncertain about their future time preferences or liquidity needs. These individuals are born with one unit of the good at $T = 0$ and no additional endowment in the subsequent periods.[3] They are subject to privately observed risk at $T = 1$, with probability p_1 of being *early consumers* who derive utility only from consumption in period one and probability $1 - p_1$ of being *late consumers* who derive utility only from consumption in period two. There is an investment technology such that a unit investment at $T = 0$ yields one unit at $T = 1$ or $R > 1$ units at $T = 2$. Consumers can privately store the good from $T = 1$ to $T = 2$. In autarky, early consumers liquidate their investment at $T = 1$ and consume one unit, while late consumers maintain the investment in the technology and receive R units at $T = 2$. There is no aggregate uncertainty, as the fraction p_1 of agents will be early consumers and the fraction $1 - p_1$ will be late consumers.

Diamond and Dybvig show how a financial intermediary can improve consumers' ex ante welfare by offering them a demand deposit contract. This deposit contract can support the full-information risk-sharing equilibrium.

Formally, the first best optimum is obtained by maximizing the ex ante expected utility of agents $p_1 u(c_1) + (1 - p_1)u(c_2)$, where $u(c_1)$ is an early consumer's utility from consumption in period one and $u(c_2)$ is a late consumer's utility from consumption in period two. This expected utility is maximized subject to the two period resource constraints $p_1 c_1 = L$ and $(1 - p_1)c_2 = (1 - L)R$, where L is the amount of the technology to be liquidated at date 1. If the representative agent's relative risk aversion coefficient is greater than 1, i.e., $-cu''(c)/u'(c) > 1$, the optimal contract satisfies $1 < c_1^* < c_2^* < R$, where c_1^* and c_2^* are the optimal consumption in periods one and two of early and late consumers, respectively. This optimal contract insures depositors against being early consumers in the sense that they receive some of the benefits available from the long-term technology, which is more than they would receive in autarky.

A deposit contract can achieve this optimal allocation. The demand deposit contract works as follows. For each unit deposited in the intermediary at $T = 0$,

[3] The model presented is simpler than Diamond and Dybvig's but has the same implications in terms of promised payment and runs.

this deposit contract provides the option of withdrawing either $r_1 = c_1^*$ at $T = 1$ or $r_2 = \frac{(1 - fc_1^*)R}{1 - f}$ at $T = 2$. The second period payment depends on f, the fraction of agents who withdraw at $T = 1$. It is easy to see that if only early consumers withdraw at $T = 1$, i.e. $f = p_1$, the demand deposit contract replicates the optimal allocation.

Implementing this allocation, however, subjects the intermediary to a possible co-ordination problem, because a consumer's type is private information and the intermediary cannot guarantee that only early consumers withdraw at $T = 1$. In fact, late consumers' withdrawals are strategic and depend on what other agents do. If some late consumers withdraw at $T = 1$, then $f > p_1$ and $c_2 < c_2^*$. If enough late consumers withdraw at $T = 1$, then $c_2 < c_1^*$ and everyone withdraws at $T = 1$, which can be interpreted as a bank run.

In the original Diamond and Dybvig model, there are two Pareto-ordered Nash equilibria—a Pareto dominant equilibrium that achieves socially optimal risk sharing in which only early consumers withdraw at $T = 1$; and a second Pareto dominated equilibrium in which all agents withdraw at $T = 1$, an equilibrium which can be interpreted as a bank run. The model can be used to show that there are several measures to prevent the occurrence of the bank run equilibrium.[4]

In a simple framework, this seminal contribution captures three important features of financial intermediaries.[5] First, individuals are uncertain about their future time preferences, which gives rise to a demand for liquidity. Second, projects are irreversible, or at least costly to restart once stopped. Third, the type of the consumer is private information relative to the financial intermediary.[6] This model implicitly assumes a sequential service constraint, that is, depositors are treated on a first-come, first-served basis. This last assumption motivates the papers by Wallace (1988, 1990) and has important implications for the discussion that follows.

---

[4] In the case in which there is no aggregate uncertainty, a suspension of convertibility policy in which withdrawals up to p_1 are allowed would implement the good Nash equilibrium. This policy removes the incentive for late consumers to withdraw early; independent of what other agents do, late consumers always obtain a higher payoff if they wait until the second period than if they withdraw in period one. If there is aggregate uncertainty, though, this measure is not effective for some realizations of p_1. With aggregate uncertainty, Diamond and Dybvig suggest deposit insurance guaranteed by government funds as the effective mechanism that would implement the Pareto dominant equilibrium.

[5] See Gorton and Winton (2002).

[6] In this general model, informational asymmetries are essential to explain the superiority of banks over financial markets in the provision of liquidity insurance. The state of the economy—the number of early consumers—is not publicly observable. As a result, complete contingent markets cannot exist. The financial market outcome is not Pareto optimal. By contrast, a financial intermediary can provide optimal risk sharing with a deposit contract.

Despite the importance of this pioneering contribution, there have been several important criticisms. Jacklin (1987) shows that the optimal deposit contract can also be achieved by trading equity. Instead of investing with the intermediary, Jacklin assumes that agents invest their unit of endowment in stock in a firm, which promises a dividend stream of L units at $T = 1$ and $(1-L)R$ units at $T = 2$, where $L = p_1 c_1^*$. A market for ex-dividend shares opens at date 1. Early agents want to trade their ex-dividend shares $(1-L)R$, for additional consumption, L, in period 1. Late agents are indifferent between consuming in either period so they would trade as long as the price of ex-dividend shares, I, is less than $(1-L)R$. Consumption for each early individual is then $c_1 = L + \frac{(1-L)R}{I}$ and similarly consumption for a type 2 agent is $c_2 = LI + R(1-L)$. Market clearing implies that the equilibrium price is $1 < I = \frac{(1-L)Rp_1}{L(1-p_1)} < R$. This implies that $c_1 = \frac{L}{p_1} = c_1^*$ and $c_2 = \frac{(1-L)R}{1-p_1} = \frac{(1-p_1c_1^*)R}{1-p_1}$, with consumption levels identical to those promised by the deposit contract. This result rules out a positive role for a bank or any other financial intermediary in the economy, because equity markets and well functioning financial intermediaries are perfect substitutes. Arguably a bank is worse than a financial market because a financial market does not have the possibility of the bad equilibrium of a bank run.[7]

In a later paper, Jacklin (1993) extends the Diamond and Dybvig framework to analyze why banking evolved with uninsured demand deposits. Jacklin does this by comparing demand deposits and equity contracts when there is aggregate uncertainty and some depositors have imperfect information about the banks' assets. First, there is aggregate uncertainty regarding the proportion of early

[7] Hellwig (1994) considers a model similar to Diamond and Dybvig's with a stochastic technology from $T = 1$ to $T = 2$ that can be interpreted as technology-induced interest rate risk. He shows that there would still be no role for a bank in this extended framework. Samartín (2001) shows that if individuals have more general preferences, then demand deposits perform better than equity contracts at low enough interest rates.

Recent criticisms of the Diamond and Dybvig model by Green and Lin (1999, 2000) analyze why banking evolved with uninsured demand deposits. They examine the significance of the simple deposit contract and find that it is critical. Confining agents to this type of contract is, in fact, the driving force behind the bank-run equilibrium of the model. Green and Lin show that when agents in the Diamond and Dybvig model are allowed to use a broad class of banking contracts, the bank-run equilibrium disappears, even in the presence of a sequential service constraint. Their results suggest that the banking system might not be inherently unstable and that economists need to attempt to understand the economic and legal environment that produces the simple deposit contract in the real world.

In a later paper, Peck and Shell (2003) show that even when banks can write more sophisticated contracts, bank runs are still possible.

Goldstein and Pauzner (2004) address some of the more fundamental problems with the multiplicity of equilibria in Diamond and Dybvig's model.

consumers in the population. The fraction $\tilde{p}$ of early consumers can take a value p_1 with probability r and p_2 with probability $1-r$. Second, the bank invests in a risky asset that yields a random return $\tilde{R}$. This variable can also take a high value R_h with probability q and a low one R_l with probability $1-q$. A subset of late consumers $p_2 - p_1$ receives perfect information about the future value of the bank's assets. The two random variables $\tilde{p}$ and $\tilde{R}$ can have a nonzero correlation with α_{ij} defined as the probability of $p = p_i$ and $R = R_j$ ($i = 1, 2$ and $j = l, h$) occurring. The paper first considers one source of uncertainty at a time. If there is only aggregate uncertainty about the total number of early consumers in the population, it would be possible to construct a dividend function $L(p)$ and a price of ex-dividend shares $I(p)$, which would fully reveal the value of p and the social optimum is the financial market equilibrium. The same result applies if there is a risky technology and no aggregate uncertainty. In these two situations, equity contracts and demand deposit contracts are equivalent risk sharing instruments. The basic contribution of this paper is to show that unless there is both aggregate uncertainty and bank assets are risky with depositors asymmetrically informed about bank asset quality, then demand deposits and equity contracts can be equivalent risk sharing instruments.

Jacklin's analysis indicates that the use of demand deposit contracts by banks requires an explanation encompassing more than just a need for liquidity transformation. Banking evolved with demand deposit contracts because they included a form of protection to uninformed depositors, who would have otherwise been disadvantaged relative to better informed depositors had equity contracts been used instead. The basic message is that liquidity should be provided using equity contracts when there is little or no potential for asymmetries for information concerning bank asset quality.[8]

In common with Diamond and Dybvig, the above papers assume that individuals have corner preferences, deriving utility from consumption in either period one or period two. As Jacklin (1987) noted, if individuals exhibit more general preferences, then banks and equity contracts are not equivalent risk sharing instruments. In a framework with no aggregate uncertainty and a risk-free technology, demand deposits provide greater risk sharing than equity shares.[9] This important result depends on the assumption that demand deposits cannot

[8] A similar result is obtained by Gorton and Pennachi (1990), discussed below in the section on bank liabilities as a medium of exchange. This result is consistent with the asymmetric information view, which is summarized in the next section.

[9] Jacklin and Bhattacharya (1988) and later Alonso (1996) also consider the relative degree of risk sharing provided by traded and nontraded contracts in a framework in which bank assets are risky and individuals with smooth preferences are informed about bank asset quality. The basic result is that deposit contracts tend to be better for financing low-risk assets.

be traded. In particular, Jacklin argues that the financial intermediary described in the previous models can only exist if trading restrictions limit consumers to demand deposit contracts of the Diamond and Dybvig type.

This highlights the importance of the sequential service constraint and its interpretation. Wallace (1988, 1990) explicitly incorporates a sequential service constraint in the Diamond and Dybvig model. This sequential service feature of the deposit contract is motivated by the fact that agents are isolated from each other. Agents demand liquid assets because they are impatient to spend when they have no access to asset markets in which they can sell any asset at its usual price. An important implication of these models is that some form of isolation of agents is needed in order to motivate illiquid banking arrangements. Otherwise, individuals would in general want to participate in a one period credit market, which is shown to be inconsistent with illiquid banking.

Haubrich and King (1990) explore the role of financial intermediaries in a framework in which individuals have interior preferences—represented by a CES utility function—and are subject to privately observed income shocks. Production opportunities are characterized by a short-term liquid investment technology and a long-term illiquid one. Their main conclusion is similar to Jacklin's (1987), namely that:

> Demand deposits *uniquely* provide insurance only if there are restrictions on financial side exchanges, which may be interpreted as exclusivity provisions or regulations on security markets. If these restrictions cannot be implemented, then our environment does not rationalize banks; other financial institutions can achieve the same real allocations and welfare levels.
>
> (Haubrich and King 1990, p. 362)

Further work in this area has been extended to examine the role of demand deposits when there exists a securities market in which agents can meet and trade (Diamond 1997, Von Thadden 1998).

Von Thadden (1998) presents a continuous-time version of the Diamond and Dybvig model in which depositors can continuously adjust their portfolios, that is, they can join outside coalitions that engage in market activity. In this setting, demand deposits cannot attain the first best allocation. The ability to trade demand deposits in financial markets severely limits liquidity provision by banks. Incentive-compatible deposit contracts are second-best mechanisms for providing liquidity. At the optimum, liquidity provision is negatively correlated with the degree of irreversibility of the investment opportunity. In particular, if the investment is completely reversible, the only incentive compatible contract is the autarky allocation.

Diamond (1997) examines the roles of banks and markets when there is a financial market with limited participation. Such a market has an impact on bank activities but banks remain important. The paper focuses on the interactions between the bank provision of liquidity and the participation in the market. As more agents participate in the market, banks are less able to provide additional liquidity. The paper delivers the Diamond and Dybvig result when there is no participation and the Jacklin result when there is full participation.

In summary, this strand of the literature argues that banks offer to pay par on demand in order to provide liquidity insurance services to individuals who are uncertain about their future time preferences in a framework in which investment opportunities are inconsistent with the possible consumption paths of consumers. These depositors demand liquid assets because they are impatient to spend and they have no access to financial markets in which they can sell any asset at its usual price. These papers try to capture the idea that consumers are isolated from each other and they cannot co-ordinate to go to a security market at the same time and trade. As Wallace (1988) pointed out, the sequential service constraint is an outcome of this isolation assumption. If the trading restriction assumption is dropped from these models, the role of banks is severely limited (Jacklin 1987). A common assumption needed in most of these papers is that demand deposits cannot be traded, which suggests that there are restrictions that impede banks and active markets from co-existing.

2.2. Other explanations of liquidity

Allen and Gale (1997) analyze a different type of intertemporal smoothing role of financial intermediaries in a standard overlapping generations model with two assets: a risky asset that pays a return $\tilde{R}$ at each date, and a safe asset which is represented by a storage technology.[10] The random return is assumed to be i.i.d. and non-negative with a positive and finite expectation and variance. In this context, an economy with incomplete financial markets and no intermediaries yields under-investment in the safe asset. On the other hand, in an economy with financial intermediaries and no financial markets, returns can be smoothed and the non-diversifiable risk can be eliminated by accumulating reserves of the safe asset. In this way, there is an ex ante Pareto improvement compared to

[10] Other attempts to extend the Diamond and Dybvig framework to an overlapping generation context are Qi (1994), Bhattacharya and Padilla (1996) and Fulghieri and Rovelli (1998). These models do not, however, consider intertemporal smoothing.

the previous case. If both financial markets and intermediaries are allowed to co-exist, then intermediaries do not provide any improvement over that obtained by investors in financial markets.

Hölmstrom and Tirole (1998) analyze a different type of liquidity that arises in a framework in which moral hazard limits the effectiveness of transactions between firms with excess liquidity and firms that have a positive demand for liquidity. In this framework, a bank that provides contingent liquidity to those that need it dominates a decentralized market. Their model has three dates. At $T = 0$, the entrepreneur raises outside funds to invest in a project that yields a return at $T = 2$. At $T = 1$, the entrepreneur is subject to a liquidity shock that obliges him to make additional investments in the project. He then has to decide whether or not to continue the project. Because there is moral hazard in inducing the entrepreneur to expend effort, outside investors cannot be promised the full social value of the investment and less financing is raised compared to the social optimum.

Hölmstrom and Tirole show that, if there is no aggregate uncertainty, there is a second-best arrangement that allows firms to hedge against a liquidity shock at $T = 1$, by buying claims on other firms at $T = 0$ and selling them at $T = 1$. Although the private sector provides sufficient liquidity in the aggregate, firms in general will be unable to satisfy their liquidity needs at an individual level. As mentioned before, a financial intermediary that grants liquidity to those that need it may dominate a decentralized market. If there is aggregate uncertainty, the private sector cannot satisfy its own liquidity needs and there may be a role for government-supplied liquidity and its active management.

Kashyap *et al.* (2002) also focus on banks as creators of liquidity. They build on the observation that banks engage in two distinct activities, deposit-taking and lending. In particular, these institutions issue a product that may enable them to distinguish themselves from other lenders such as insurers or finance companies—loan commitments or credit lines. They develop the idea that credit lines and demand deposits can then be seen as two different manifestations of the same function, provision of liquidity on demand. There is a complementarity between these two ways of providing liquidity because they are not perfectly correlated. Once this fact is recognized, it is easy to argue that there may be important synergies in offering both products because the banks hold liquid assets. The paper develops a theoretical and empirical case for this particular synergy.[11]

[11] Similarly, McAndrews and Roberds (1999), also discussed in the section on bank liabilities as a medium of exchange, analyze an extreme version of this complementary—strictly offsetting payments. In their model, the advantage conferred by the complementary is related to the banks' superior ability to enforce debt contracts.

3. ASYMMETRIC INFORMATION

A second explanation for the use of demand deposit contracts is linked to the role of banks in the economy—banks provide valuable services through the creation of non-marketable loans. As a consequence, banks are opaque institutions with loans that are difficult to value and bank managers may be difficult to monitor. The end result is banking arrangements are fraught with moral hazard. A common assumption in all these papers is asymmetric information and the new ingredient compared to Diamond and Dybvig's model is moral hazard.

The general setup of these models is similar to that of Diamond and Dybvig and the subsequent literature summarized in the previous section. There are three dates, a large number of small depositors and a monopoly bank. For simplicity, it is assumed that all agents are risk neutral, have deterministic utility functions and care only about consumption at $T = 2$.[12] The bank has access to two mutually exclusive investment opportunities that require one unit of investment at $T = 0$.[13] The bank has no capital and raises funds by selling deposits to investors, each of whom is endowed with $1/n$ units, so that n investors are needed to finance the project. The deposit contract pays interest r if maintained until date two and no interest if withdrawn before then. The characteristics of the mutually exclusive projects are i) project A pays a high value $R_h > 1$ with probability q and a low value $R_l = 0$ with probability $1 - q$, and ii) project B pays $1 + r < \bar{R} < qR_h$ with probability one. The expected payoff to the bank is always greater with project A, but project B always guarantees depositors their promised interest payment. The choice of the project is not error-free, as there is a small probability (λ) that the bank may make errors in project choice.[14] Depositors can engage in monitoring activities at a cost $K > 0$. By monitoring the bank, depositors may discover the true project choice at $T = 1$, and can force liquidation of the bank by withdrawing their deposits prematurely if they desire.[15] If the bank's projects are liquidated, they are worth only $L < 1$.[16] It can be shown that, if monitoring

[12] In this way, one can focus on the incentive effects of demand deposits and ignore liquidity insurance.

[13] Alternative ways to introduce moral hazard are described below.

[14] This assumption has to be introduced in order to avoid a time consistency problem that would lead to there being no equilibrium.

[15] It is assumed that the payoffs to depositors are such that they will always want to liquidate project A and maintain project B.

[16] This low liquidation value, in combination with the sequential service constraint, ensures that all depositors have an incentive to monitor the bank and be first in line.

costs are not too high, it is an equilibrium for all depositors to monitor the bank and for the bank to choose project B.[17]

This same argument is developed in an early contribution by Calomiris and Khan (1991)—liquid deposits keep the bank's portfolio choice in line with depositors' preferences. Their model has three dates, and the bank has access to an investment technology in which there is a random payoff $\tilde{R}$ at $T = 2$ for each unit invested at $T = 0$. This random payoff can take a high value R_h with probability q and a low one R_l with probability $1 - q$. Moral hazard is introduced by assuming that the bank can abscond with the funds immediately before repayment, thereby reducing the realization of $\tilde{R}$ by A. Depositors are risk neutral and can receive an imperfect signal about the realization of $\tilde{R}$ by paying a cost K. It is shown that uninsured demand deposit contracts discipline bank managers. A deposit contract serves this role due to the combination of two characteristics: the "on demand clause" and the sequential service constraint. The demandable nature of the contract motivates some depositors to monitor the bank, while the sequential service constraint discourages free riding by depositors on others' monitoring.

In a later paper, Jean-Baptiste (1999) also argues that demand deposits can be incentive mechanisms that induce bankers to make efficient monitoring decisions. The model is close in spirit to the previous one.[18] There are three dates, banks have access to an investment technology and they can engage in monitoring activities at a cost $K > 0$. If monitoring is effective, which occurs with a certain probability, the technology generates a value R_h. Otherwise, the value obtained is R_l. Depositors are risk neutral and receive an imperfect, homogenous signal at $T = 1$ about bank quality. In this model, high monitoring costs result in an equilibrium with equity or long-term debt that is inefficient because it does not induce banks to monitor. On the other hand, a demand deposit contract can yield a Pareto superior equilibrium despite the positive probability of inefficient liquidation.[19] In this paper, the sequential service constraint is a commitment technology that adds credibility to the threat of liquidation.

---

[17] A detailed numerical example of the above model can be seen in Greenbaum and Thakor (1995).

[18] As Jean-Baptiste points out, this model differs from the previous one in several ways. Calomiris and Khan's argument is independent of whether the bank has one depositor or a large number of small depositors. Also, Calomiris and Khan's explanation of the sequential service constraint is not completely satisfactory, because the free-rider problem could be solved by the simple expedient of introducing a well-defined priority structure for the bank's liabilities.

[19] The results of the paper also suggest that intermediaries that specialize in financing assets for which information is readily available and monitoring costs are low, can themselves be financed with either equity or long-term debt. This conclusion is related to Gorton and Pennachi (1990) and Jacklin (1993).

Flannery (1994) reaches a similar conclusion: Banks specialize in financing non-marketable, informationally intensive assets and can readily change the composition of their portfolio, which creates a larger moral hazard problem than for non-banking firms. Creditors can form a noisy assessment of bank risk, which implies fair market prices for bank debt and equity. In this setting, short-term debt is employed to control moral hazard associated with asset substitution, because changes in bank risk will be promptly reflected in financing costs.

Most recently, Gorton and Huang (2002a, 2002b) also have a model with asymmetric information that generates demand deposit contracts as an incentive device. They introduce a new ingredient—the industrial organization of the banking system—that is an important determinant of the propensity of the industry to experience banking panics, which are themselves related to the business cycle. Also associated with the likelihood of panics is the existence of certain kinds of private arrangements among banks, private arrangements that can be thought of as precursors of central banks and their role as lenders of last resort. The model has three dates, depositors and bankers. There is a continuum of bankers. Each banker has capital C and measure one of depositors. Each of them also has access to two investment technologies; first, a riskless storage technology—reserves—and second, a risky asset for which each unit invested at $T = 0$ generates $\tilde{R}$ units at $T = 2$ with $\tilde{R} = \tilde{\pi} + \tilde{r}$, where $\tilde{\pi}$ is the systematic component and $\tilde{r}$ the idiosyncratic component. There is asymmetric information in the sense that at $T = 1$, depositors can only observe the state of the economy reflected in π. Finally, there is also moral hazard in that bankers have the opportunity to engage in fraud.

In this context, panics result from depositors monitoring and liquidating deposits in all banks when they anticipate that banks will engage in fraud.[20] At $T = 0$, anticipating what will happen in the different states of the world, banks choose their optimal reserve level. Banks want to maximize their investment in the risky technology, but they also want to hold sufficient reserves to avoid premature liquidation of the risky assets. In this model, a system of large banks is more efficient and less prone to liquidations due to panics than a system of independent unit banks. An incentive-compatible state-contingent bank coalition emerges as a response to the unit bank system's problems. This coalition acts

[20] It should be mentioned, that in contrast with Diamond and Dybvig, the model does not assume a sequential service constraint. However, the assumed form of the utility function implies that depositors will withdraw whenever they expect the bank to engage in fraud.

as a lender of last resort by monitoring and providing insurance to member banks. The resulting banking system is more efficient and less subject to panics than a system of small banks with a lender of last resort. Even so, unit banking is less efficient and more subject to panics than a system of large banks. In order to explain why government central banks and deposit insurance historically replaced private bank coalitions, Gorton and Huang introduce demand deposit's use as a medium of exchange. Panics are costly because they disrupt the use of demand deposits as a medium of exchange, and the government can prevent panics by providing deposit insurance and monitoring banks. In Gorton and Huang's setup, a deposit insurance system can improve welfare if the cost of government monitoring of the banks is low enough.

In summary, this strand of literature argues that banks offer to pay at par as a way to protect uninformed depositors, who would be disadvantaged relative to better-informed individuals if banks offered equity contracts. The deposit is payable on demand because its value is not state contingent. In this sense, demand deposit contracts are a discipline device—bank deposits promise to pay par on demand in order to control the risk taking activities of banks.

This literature suggests that the difficulty of valuing assets and consequent marketability of banks' assets determines whether banks promise to pay at par.[21]

4. LEGAL RESTRICTIONS

A third explanation of why banks promise to pay par on demand is provided by the legal restrictions theory which attempts to explain the co-existence of alternative assets, some of which yield significantly higher yields or returns than others (Wallace 1983 and references therein). As Wallace points out, an example of these paradoxical patterns of returns among assets is the co-existence of U.S. currency and default-free interest bearing securities, such as U.S. savings bonds and Treasury bills. If both deposits and Treasury securities are perfect substitutes, no one would hold non-interest bearing currency instead of Treasury bills. This co-existence can only be explained by the fact that there must be legal restrictions on Treasury bills, which prevent them from playing the same role in transactions as do deposits. If both assets were allowed to be used in transactions

[21] Qi (1998) and Diamond and Rajan (2001a, 2001b, 2003), also study the disciplinary effects of liquid deposits in models that abstract from asymmetric information.

without any legal restrictions, the prediction is that either nominal interest rates would go to zero or government currency becomes worthless.[22]

In summary, Wallace argues that banks promise to pay par on demand because of legal restrictions, which prohibit other securities from playing the same role as demand deposits. If the legal restriction explanation of par redemption is correct, banks will not promise to pay par if they are not required to do so.[23]

5. BANK LIABILITIES AS A MEDIUM OF EXCHANGE

Other models have been built based on the observation that bank liabilities function as a medium of exchange (e.g. Freeman, 1996a, 1996b; Green, 1997; Williamson, 1992; McAndrews and Roberds, 1999). In general, these papers consider a framework in which agents are either spatially separated, so they cannot contract and trade with each other due to their inability to meet at a single location, or there are frictions such as problems of contract enforcement or adverse selection.

In these models banks issue private money to facilitate clearing transactions. One issue that arises is the pricing of these bank notes—if some agents are better informed about the probability of a bank failure, they may be able to gain when trading bank liabilities. An important characteristic of a medium of exchange may be that it entails little or no risk, that is, its value does not depend on the likelihood of the bank failing.

A similar result is obtained by Gorton and Pennachi (1990) in a somewhat different framework in which individuals are risk neutral, and so the demand deposit insurance contract is not explicitly modeled. They argue that financial intermediaries create liquid deposits in response to uninformed depositors. They

---

[22] White (1987) argues that a counter-example to the above theory can be found in the Scottish free banking system from 1716–1844, in which non-interest-bearing currency and interest-bearing securities co-existed and only non-interest-bearing currency was used in transactions. He critiques Wallace's line of argument by suggesting that the liquidity service, or nonpecuniary yield, of currency is important in addition to the pecuniary return and risk. He argues that if technological and computation costs are appropriately considered, interest might not be worth collecting on at least smaller denominations of currency. Hence, non-interest-bearing currency would still survive in the absence of legal restrictions. The legal restriction theory overlooks the costs involved in collecting interest on money, recognizing only the cost of converting large interest bearing assets into smaller liabilities.

[23] The legal restriction theory simply overlooks the costs involved in collecting interest on money. In this respect, the only cost recognized is a cost to intermediation, which converts large interest-bearing assets into smaller liabilities.

define a liquid security as one that entails no private information and model the proposition that trading in liquid securities such as deposit contracts protects uninformed depositors from losses that they would otherwise suffer if they traded illiquid—information-sensitive—securities with informed individuals. Therefore, financial intermediaries' debt should be used for transaction purposes.

Demand deposit contracts are not the unique solution for creating liquid securities that protect uninformed agents. Other risk-free instruments such as government bonds can accomplish the same role.

Even so, it is fair to say that this literature suggests that liabilities of financial intermediaries will promise to pay par if they are used as a medium of exchange.

6. CONCLUSION

There is extensive literature on banks and their promise to pay par on demand. Although there are intersections, the literature can be broken into four different lines. One line of the literature follows Diamond and Dybvig, in whose model banks promise to pay par on demand because households have a demand for the liquidity of such a contract. The greater liquidity of the demand deposit liability is due to a maturity mismatch between the bank's assets and liabilities. A second line of the literature takes a slightly different tack and bases the promised payment at par on information about loan quality known to the bank but not to depositors. Un-informed depositors have less information about loans than do bankers, and non-marketable loans on banks' books cannot be the basis of deposits that are marked to a market value of assets determined by the bank. Hence, the uncertain market value of banks' assets becomes a known value of banks' liabilities by promising to pay the par value of deposits. Because a bank can take actions such as making riskier loans to make itself better off without compensating depositors for the risk, promised payment on demand can reduce the bank's payoff from such strategies. An alternative line of argument takes the simple course—which is not necessarily the wrong one because it is simple. Banks in the United States today are required to pay the par value of "demand deposits" on demand, and the existence of such a promise may reflect nothing other than that legal requirement. A fourth line of argument suggests that liabilities of financial intermediaries, which are used as a medium of exchange, will be characterized by a constant value.

One interpretation of the informativeness of these theories about actual banking arrangements is that they need not say much about anything observed. As one theorist put it, "The real world is a special case, and not a very interesting one at that."

Alternatively, some would claim that these theories are stories with no relation to actual economies and are not interesting. We think that this review of the literature has shown, on the contrary, that these theories can be interpreted as having predictions about when banks will promise to pay par on demand and when they will not make such promises.

Strong predictions from the theories take the form: Promised payment at par will be observed only if certain conditions are met. For example, the legal restriction theory can be interpreted as making the strong prediction that banks will promise to pay par on demand only if they are required to do so. Similar statements can be made for the other theories.

The theories also can be interpreted as partial explanations, explanations that work some of the time. For example, the legal restriction theory can be interpreted as making the weak prediction that banks sometimes will promise to pay par on demand if they are required to do so and for no other reason. In other words, an observation supporting the importance of the legal restrictions theory would be an observation at some time and place that banks promise to pay par on demand and none of the other theories can explain why they would make that promise.

We think that this summary of existing theories indicates that they do have useful predictions and we are currently working on a companion paper, which examines the consistency of actual banking arrangements with the theories.

ACKNOWLEDGEMENT

Dwyer appreciates financial support from the Earhart Foundation. An earlier version of this paper was presented at the XIII International “Tor Vergata” Conference on Banking and Finance. The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any errors are the authors’ responsibility.

BIBLIOGRAPHY

Allen, F. and Gale, D. (1997) “Financial Markets, Intermediaries and Intertemporal Smoothing.” *Journal of Political Economy* **105**: 523–46.

Alonso, I. (1996) “On Avoiding Bank Runs.” *Journal of Monetary Economics* **37**: 73–87.

Bhattacharya, S. and Padilla, J. (1996) “Dynamic banking: a reconsideration.” *Review of Financial Studies* **9(3)**: 1003–32.

Bryant, J. (1980) “A Model of Reserves, Bank Runs, and Deposit Insurance.” *Journal of Banking and Finance* **4**: 335–44.

Bryant, J. (1989) “Interest-Bearing Currency, Legal Restrictions, and the Rate of Return Dominance of Money.” *Journal of Money, Credit and Banking* **21**: 240–45.
Calomiris, C.W. and Kahn, C.M. (1991) “The Role of Demandable Debt in Structuring Optimal Banking Arrangements.” *American Economic Review* **81**: 497–513.
Diamond, D.W. (1997) “Liquidity, Banks and Markets.” *Journal of Political Economy* **105**: 928–56.
Diamond, D.W. and Dybvig, P.H. (1983) “Bank Runs, Deposit Insurance, and Liquidity.” *Journal of Political Economy* **91**: 401–19.
Diamond, D.W. and Rajan, R. (2001a) “Banks and liquidity.” *American Economic Review* **91**: 422–25.
Diamond, D.W. and Rajan, R. (2001b) “Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking.” *Journal of Political Economy* **109**: 287–327.
Diamond, D.W. and Rajan, R. (2003) “Liquidity Shortages and Banking Crises.” NBER Working Paper 10071.
Dwyer, Jr., G.P. (1996) “Wildcat Banking, Banking Panics and Free Banking in the United States.” *Federal Reserve Bank of Atlanta Economic Review* **81**: 1–20.
Freeman, S. (1996) “The Payments System, Liquidity and Rediscounting.” *American Economic Review* **86**: 1126–38.
Freeman, S. (1996) “Clearing House Banks and Banknote Over-issue.” *Journal of Monetary Economics* **38**: 101–15.
Fulghieri, P. and Rovelli, R. (1998) “Capital markets, financial intermediaries and liquidity supply.” *Journal of Banking and Finance* **22**: 1157–79.
Goldstein, I. and Pauzner, A. (2004) “Demand deposit contracts and the probability of bank runs.” *Journal of Finance*, forthcoming.
Gorton, G. and Pennachi, G. (1990) “Financial Intermediaries and Liquidity Creation.” *Journal of Finance* **45**: 49–71.
Gorton, G. and Winton, A. (2002) “Financial Intermediation.” In: Constantinides, G., Harris, M. and Stultz, R. (eds), *Handbook of the Economics of Finance*, Amsterdam.
Gorton, G. and Huang, L. (2002a) *Bank panics and the endogeneity of central banking*. NBER Working Paper 9102.
Gorton, G. and Huang, L. (2002b) *Banking panics and the origin of central banking*. NBER Working Paper 9137.
Green, E. (1997) “Money and debt in the structure of payments.” *Bank of Japan Monetary and Economic Studies*. 15. pp. 63–87.
Green, E. and Lin, P. (1999) *Implementing efficient allocations in a model of financial intermediation.* Working paper, Centre for Public Policy Studies, Lingnan University, Hong Kong.
Green, E. and Lin, P. (2000) “Diamond and Dybvig’s Classic Theory of Financial Intermediation: What’s Missing?” *Federal Reserve Bank of Minneapolis Quarterly Review* **24**: 3–13.
Greenbaum, S. and Thakor, A. (1995) *Contemporary Financial Intermediation*. Dryden Press.
Harris, M. and Raviv, A. (1992) “Financial Contracting Theory.” In: Laffont, J-J (ed.), *Advances in Economic Theory*, Vol. II. Cambridge: Cambridge University Press.

Haubrich, J.G. and King, R.G. (1990) "Banking and Insurance." *Journal of Monetary Economics* **26**: 361–86.
Hellwig, M. (1994) "Liquidity provision, banking and the allocation of interest rate risk." *European Economic Review* **38**: 1363–89.
Holmstrom, B. and Tirole, J. (1998) "Private and public supply of liquidity." *Journal of Political Economy* **106**: 1–40.
Jacklin, C.J. (1987) "Demand Deposits, Trading Restrictions and Risk Sharing." In: Prescott, E.C. and Wallace, N. (eds) *Contractual Arrangements for Intertemporal Trade, University of Minnesota Press*. pp. 26–47.
Jacklin, C.J. (1993) "Market Rate versus Fixed Rate Demand Deposits." *Journal of Monetary Economics* **32**: 237–58.
Jacklin, C.J. and Bhattacharya, S. (1988) "Distinguishing Panics and Information-based Bank Runs: Welfare and Policy Implications." *Journal of Political Economy* **96**: 568–92.
Jean-Baptiste, E. (1999) *Demand Deposits as an Incentive Mechanism*. Unpublished paper, Wharton School, University of Pennsylvania.
Kashyap, A.N., Rajan, R. and Stein, J. (2002) "Banks as liquidity providers: an explanation for the coexistence of lending and deposit taking." *Journal of Finance*. **57**: 33–73.
Lindgren, C., Garcia, G. and Saal, M. (1996) "Bank soundness and macroeconomic policy." International Monetary Fund.
McAndrews, J. and Roberds, W. (1999) "Payment intermediation and the origins of banking." *Working Paper Series of the Federal Reserve Bank of Atlanta*. 99–11.
Peck, J. and Shell, K. (2003) "Equilibrium bank runs." *Journal of Political Economy* **111**: 103–123.
Qi, J. (1994) "Bank liquidity and stability in an overlapping generations model." *Review of Financial Studies* **7(2)**: 389–417.
Qi, J. (1998) "Deposit liquidity and bank monitoring." *Journal of Financial Intermediation* **7**: 198–218.
Rogers, J.S. (1995) *The Early History of the Law of Bills and Notes: A Study of the Origins of Anglo-American Commercial Law*. Cambridge: Cambridge University Press.
Samartín, M. (2001) "Banks Increase Welfare." *Financial Markets, Institutions and Instruments* **10(5)**: 203–34.
Sprague, O.M.W. (1910) *History of Crises under the National Banking System*. U.S. National Monetary Commission. Washington: Government Printing Office.
Usher, A.P. (1934) "The Origins of Banking: The Primitive Bank of Deposit, 1200–1600." *Economic History Review* **4**: 399–428. In: Lane, F.C. and Riemersma, J.C. (eds), *Enterprise and Secular Change: Readings in Economic History*, pp. 262–91. Homewood, Illinois: Richard D. Irwin, Inc., 1953.
Von Thadden, E. (1998) "Intermediated versus Direct Investment: Optimal Liquidity Provision and Dynamic Incentive Compatibility." *Journal of Financial Intermediation*. **7**: 177–197.
Wallace, N. (1983) "A Legal Restrictions Theory of the Demand for 'Money' and the Role of Monetary Policy." *Federal Reserve Bank of Minneapolis Quarterly Review* **7**: 1–7.

Wallace, N. (1988) "Another Attempt To Explain an Illiquid Banking System: The Diamond and Dybvig Model With Sequential Service Taken Seriously." *Federal Reserve Bank of Minneapolis Quarterly Review* **12**: 3–13.

Wallace, N. (1990) "A Banking Model in which Partial Suspension Is Best." *Federal Reserve Bank of Minneapolis Quarterly Review* **14**: 3–16.

Wallace, N. (1996) "Narrow Banking Meets the Diamond-Dybvig Model." *Federal Reserve Bank of Minneapolis Quarterly Review* 20 (Winter), pp. 3–13.

Williamson, S.D. (1992) "Laissez-faire Banking and Circulating Media of Exchange." *Journal of Financial Intermediation* **2**: 134–67.

White, L.H. (1987) "Accounting for Non-interest-bearing Currency: A Critique of the Legal Restrictions Theory of Money." *Journal of Money, Credit, and Banking* **19**: 448–56.

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 2

PRODUCTIVITY AND EFFICIENCY MEASUREMENT USING PARAMETRIC ECONOMETRIC METHODS

Subal C. Kumbhakar

Abstract

In this chapter we survey the econometric approaches to productivity and efficiency measurement. Both primal and dual approaches are considered. More specifically, we examine the production function and distance functions (multiple outputs), cost functions (with single and multiple outputs), standard and alternative profit functions (with single and multiple outputs). Possible extensions of the traditional productivity and efficiency measures in the light of non-traditional inputs are also discussed.

Keywords and Phrases: Partial and total factor productivity; technical change; returns to scale; technical efficiency; technical efficiency change; production function; distance function; cost function; profit function; alternative profit function; mixing models.

JEL Classification No.: C23, C33, D24, O30

> "Productivity cannot be measured directly. Instead, it must be measured indirectly as a relationship between physical outputs and inputs that can be assembled."
>
> (John W. Kendrick 1984, p.9)

> "To measure the productivity change of a firm or an industry, we first have to define what we mean by a productivity change."
>
> (W. Erwin Diewert 1992, p. 163)

1. INTRODUCTION[1]

The most commonly used definition of productivity is the amount of physical output produced by one unit of a given factor of production at a stated period of time. Thus, productivity indicators ordinarily relate output to a single factor of production, creating measures such as labor productivity, capital productivity, etc. These are also defined as partial factor productivity. In contrast, multifactor (total factor) productivity measures output per unit of a set of combined factors of production (such as labor capital, land, etc.) and gives a single overall measure of productivity. These partial and multifactor productivity measures are based on quantities of inputs and outputs and are called primal measures of productivity. Instead of using input and output quantities, productivity can also be defined using cost, profit and price information. For example, average cost can be used to measure productivity. Note that these measures of productivity (average product and average cost) can be measured directly from the data.

Productivity (no matter how it is defined) is likely to change over time. In simple terms, productivity change/growth means more output is produced with a given level of inputs, which occurs when technology improves over time. Since inputs also change over time, productivity change is often defined as output growth net of input growth rate. Accordingly, productivity change can take place without technical change. In a single output single input case, productivity change is simply the rate of change in average product. We will give a formal definition of productivity change with single output and multiple inputs technology. This will be followed by multiple outputs multiple inputs technology. In addition to measuring productivity change, we also consider the sources of productivity growth.

Productivity and change in productivity can be measured using different techniques. Diewert (1992) showed that productivity change can be calculated using an index number approach Fisher (1922) or Törnqvist (1936) productivity index. Both indices require quantity and price information, as well as assumptions concerning the structure of technology and the behavior of producers, but neither requires estimation of anything econometrically. Productivity change can also be calculated using the Divisia index, which is nonparametric. Finally, it can be estimated using econometric techniques.

A disadvantage of index number techniques and the Divisia index is that they do not provide sources of productivity growth, whereas nonparametric and econometric techniques do. Although nonparametric and econometric techniques are capable of measuring productivity change and its sources, only the econometric

---

[1] Parts of this section are drawn from Chapter 8 of Kumbhakar and Lovell (2000).

approach is capable of doing so in a stochastic environment. Here we show how to use econometric techniques to estimate the magnitude of productivity change, and then to decompose estimated productivity change into its various sources.

Productivity is affected in the presence of inefficiency. This is likely to affect productivity growth as well, unless inefficiency is time-invariant. Traditional econometric models of productivity change ignored the contribution of efficiency change to productivity change. Productivity change was decomposed into technical change and scale economies. However, if inefficiency exists, then efficiency change provides an independent contribution to productivity change. If efficiency change is omitted from the analysis, its omission leads to an erroneous allocation of productivity change to its sources. Accordingly, it is desirable to incorporate the possibility of efficiency into econometric models of productivity and productivity change.

To get a flavor of the issues involved, we begin with a quantity-based (primal) approach for the estimation and decomposition of productivity change. In doing so we allow for the possibility that production plans can be technically inefficient. The general structure of the primal approach is illustrated in Figure 2.1, in which a single input is used to produce a single output. Assume that for a producer at time t the production plan is given by the input-output combination (x^t, y^t) and the production frontier (the maximum possible output function given input quantities) is $f(x, t)$. Productivity for this input-output combination is defined by the ratio of output to input, viz., y^t/x^t, which can be easily measured from input and output data. Note that $y^t < f(x^t, t)$, which means that the production plan is technically inefficient. We define technical efficiency as $y^t/f(x^t, t)$, which is at most unity. Since the production frontier, $f(x, t)$, is not known,

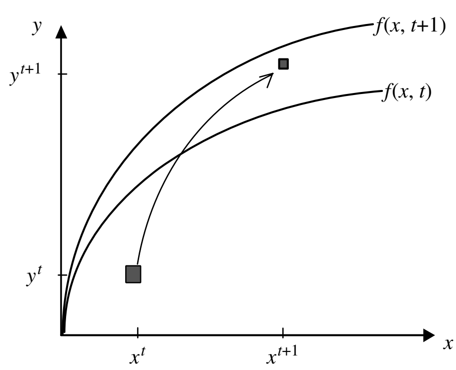


Fig. 2.1. The Primal Approach to the Estimation and Decomposition of Productivity Change

the measurement of technical efficiency requires estimation of the frontier. Furthermore, productivity is reduced in the presence of technical inefficiency, i.e., $y^t/x^t < f(x^t, t)/x_t$. Alternatively, the more efficient a firm is, the higher is its productivity, *ceteris paribus*. If we assume that the producer expands the production plan from (x^t, y^t) to (x^{t+1}, y^{t+1}), and technical progress has occurred between periods t and $t+1$, this would imply that $f(x, t+1) > f(x, t)$. Ignoring noise, it is clear that production is technically inefficient in both periods, since $y^t < f(x^t, t)$ and $y^{t+1} < f(x^{t+1}, t+1)$, and that technical efficiency has improved from period t to period $t+1$, since $[y^t/f(x^t, t)] < [y^{t+1}/f(x^{t+1}, t+1)]$. It is also clear that productivity growth has occurred, since $(y^{t+1}/x^{t+1}) > (y^t/x^t)$. The estimated rate of productivity growth can then be decomposed into returns to scale, technical change and change in technical efficiency.

It is clear from the above discussion that in a single-output single-input case, productivity at a point in time is measured by y^t/x^t, which is nothing more than the average product of x at time t. Productivity change (treating time as a continuous variable) is measured by the rate of change in the average product of x, i.e., $\partial \ln(y^t/x^t)/\partial t = \dot{y}^t - \dot{x}^t$, where the dot over a variable indicates its rate of change. Thus both productivity and productivity change can be measured from observed data without estimating anything. If inefficiency is present, then the productivity of an input is lower than its maximum possible value. The degree to which productivity is lowered depends on the degree of inefficiency, which has to be estimated. Similarly, the effect of inefficiency on productivity change will depend on the temporal behavior of inefficiency.

If there are multiple inputs, one can obtain the partial factor productivity of each input j, defined as y/x_j and the partial factor productivity change defined as $\partial \ln(y/x_j)/\partial t = \dot{y} - \dot{x}_j$. Observing that both the partial factor productivity and its change (with or without technical inefficiency) depend on the usage of other factors, which are likely to differ among inputs, a single measure of productivity (i.e., total factor productivity, TFP) and productivity change is needed. In practice, labor productivity is often used as the measure of productivity, although its level depends on the amount of other factors being employed, such as capital, energy, and materials used, in the production process.

The objective of this paper is to survey some issues related to productivity measurement and the decomposition of TFP change in the context of a panel data framework. The focus is on the parametric econometric models based on production, distance, cost, and profit functions. TFP change is decomposed into technical change, scale economies, and technical efficiency components. Several alternative strategies in modeling technical change and production technology (in both primal and dual contexts) are considered.

We begin by using a production frontier approach to obtain estimates of productivity change, and to decompose estimated productivity change into a technical change, a returns to scale, and a component associated with change in technical efficiency. We also consider the factor augmenting approach to technical change, which allows the researcher to decompose technical change into contributions due to each individual input. To allow for the possibility of multiple technologies and to measure the technology gap across regions, countries, etc., we discuss mixing (latent class) models and estimation of a best practice (meta) frontier. Furthermore, we use the input and output distance functions as an alternative to the production function approach. We repeat this investigation using a dual cost function approach, which has the advantage of accommodating multiple outputs. With multiple outputs, an additional component of productivity growth, viz., markup in output prices, is added. We also use a profit function approach with both single and multiple outputs. The profit function approach is extended to accommodate markups in output prices. Finally, we consider the alternative profit function in which output prices (instead of output quantities) are treated as endogenous.

2. THE PRODUCTION FUNCTION APPROACH

2.1. Time Trend Representation of Technical Change

We summarize the technology of the firm in time period t by its period t production function $f^t(.)$. Assuming that there is one output, the production function can be represented as $y_{it} = f^t(x_{1it}, \ldots, x_{Jit})$, where y_{it} is the output of the i^{th} firm $(i = 1, \ldots, N)$ in period t $(t = 1, \ldots, T)$, $f^t(\cdot)$ is the production technology, and x is a vector of J inputs. In order to estimate the parameters of such a production function econometrically, it is necessary to relate the production function in period t to the corresponding production function for other periods. A common approach is to assume that the production function is atemporal, meaning that its form and the parameters do not depend on the time index t. Exploiting this fact and accommodating technical inefficiency, the production function can be represented by

$$y_{it} = f(x_{it}, t)\exp(-u_{it}), \tag{1}$$

where $u_{it} \geq 0$ is output-oriented technical inefficiency. Technical inefficiency, u_{it}, measures the proportion by which actual output (y_{it}) falls short of maximum possible output $f(x, t)$. Technical efficiency is then defined by $y_{it}/f(x_{it}, t) = \exp(-u_{it}) \leq 1$. The time trend variable t in (1) represents technical change (a shift in the production function over time, *ceteris paribus*).

When input quantities change, productivity change is measured by what is popularly known as TFP change (or the Divisia index of productivity change, denoted by $T\dot{F}P$) and is defined as[2]

$$T\dot{F}P = \dot{y} - \dot{x} \equiv \dot{y} - \sum_j S_j^a \dot{x}_j, \tag{2}$$

where $S_j^a = w_j x_j / C^a$ and $C^a = \Sigma_j w_j x_j$, w_j being the price of input x_j. Here the rate of change in the composite input (x) is defined as the weighted average of the rate of change on individual inputs. By using this definition, the TFP index can be constructed from $TFP_t = TFP_{t-1}(1 + T\dot{F}P_t)$, $t = 2, \ldots, T$ when, for example, *TFP* for the base year ($t = 1$) is 100, i.e., $TFP_1 = 100$.

To decompose $T\dot{F}P$ we differentiate (1) totally and use the definition of *TFP* change in (2) to obtain

$$\begin{aligned} T\dot{F}P &= TC - \frac{\partial u}{\partial t} + \sum_j \left(\frac{f_j x_j}{f} - S_j^a \right) \dot{x}_j \\ &= (RTS - 1) \sum_j \lambda_j \dot{x}_j + TC + TEC + \sum_j \{\lambda_j - S_j^a\} \dot{x}_j, \end{aligned} \tag{3}$$

where $TC = \frac{\partial \ln f(x,t)}{\partial t}$, $TEC = -\frac{\partial u}{\partial t}$ and $RTS = \Sigma_j \frac{\partial \ln y}{\partial \ln x_j} = \Sigma_j \frac{\partial \ln f(.)}{\partial \ln x_j} = \Sigma_j f_j(.) x_j / f(.) \equiv \Sigma_j \varepsilon_j$ is the measure of returns to scale. Finally, $\lambda_j = \{f_j x_j / \Sigma_k f_k x_k\} = \varepsilon_j / RTS$ when f_j is the marginal product of input x_j. The relationship in (3) decomposes TFP change into scale $\left((RTS-1)\Sigma_j \lambda_j \dot{x}_j\right)$, technical change $\left(\frac{\partial \ln f(x,t)}{\partial t}\right)$, technical efficiency change $(-\partial u / \partial t)$, and price effect $(\Sigma_j \{\lambda_j - S_j^a\} \dot{x}_j)$ components. This last component captures either deviations of input prices from the value of their marginal products, i.e., $w_j \neq p f_j$, or the departure of the marginal rate of technical substitution from the ratio of input prices ($f_j / f_k \neq w_j / w_k$). Thus, the last component can be dropped from the analysis if one assumes that firms are allocatively efficient.[3]

If technical inefficiency is time-invariant (i.e., $-\partial u / \partial t = 0$), the decomposition in (3) shows that TEC does not affect TFP change. Under the assumption of

[2] Subscripts i and t are omitted to avoid notational clutter.

[3] TFP growth is computed using the Divisia index. Therefore, the sum of the TFP growth components obtained from the parametric model (for example, from (3)) has to be equal to the TFP growth obtained using the Divisia index. But in practice, a wide gap between the two measures is often observed, especially when TFP growth (which is the sum of the TFP growth components) is estimated using a parametric model. This problem can be avoided by using the TFP growth equation as an additional equation in the system. See Kumbhakar and Vivas (2004a) for details on this issue in the context of a dual cost function estimation.

constant returns to scale, the TFP change formula in (3) is identical to the one derived in Nishimizu and Page (1982), viz.:

$$T\dot{F}P = TC - \frac{\partial u}{\partial t} + \sum_j \{\varepsilon_j - S_j^a\}\dot{x}_j. \qquad (4)$$

The TFP growth formula with input-oriented technical inefficiency (for which the production function is written as $y = f(x\exp(-\tau), t)$, $\tau \geq 0$ being input-oriented technical inefficiency) is presented in Appendix A. Technical inefficiency can also be non-neutral. The TFP growth formulae for non-neutral (input- and output-oriented) technical inefficiency are presented in Appendices B and C.

2.2. The Factor-Augmenting Model of Technical Change

Since technical progress is defined in terms of shifts in the production function or the isoquants, it means that either a given level of output can be produced with fewer inputs, or more output can be produced using the same amount of inputs. This, in turn, implies that technical progress increases the productivity of at least one input. Does technical progress increase productivity of all or some of the inputs? Is it possible that the productivity of some inputs increases while those of other inputs remain constant or decline? What is the contribution of a particular input to overall technical change? The standard time trend model (discussed in Section 2.1) gives an estimate of the overall effect of technical change on output but it cannot answer the above questions. Answers to these questions can be obtained by specifying technical change in factor augmenting (FA) form (Beckmann and Sato (1969), Sato and Beckmann (1969), Kumbhakar (2002, 2003), viz.:

$$\begin{aligned} y = f(Ax)\exp(-u) &= f(A_1(t)x_1, \ldots, A_J(t)x_j)\exp(-u) \\ &\equiv f(\tilde{x}_1, \ldots, \tilde{x}_J)\exp(-u) = f(\tilde{x})\exp(-u), \end{aligned} \qquad (5)$$

where $\tilde{x}_j = A_j(t)x_j$ is the j^{th} variable input measured in efficiency units, and $f(\cdot)$ is the production technology. $A_j(t) > 0$ is the efficiency factor associated with input j $(j = 1, \ldots, J)$.[4] It can also be viewed as an input-specific productivity/efficiency index. If $A_j(t)$ increases over time, then the productivity of input j rises. Thus, $A_j(t) - A_j(t-1)$ measures the productivity change of input j from period $(t-1)$ to period t. Consequently, productivity growth in input

[4] Although we are making A (.) a function of time, in principle, it can depend on other exogenous variables.

j (i.e., $A_j(t) - A_j(t-1) > 0$) implies an increase in partial factor productivity (y/x_j) but not vice-versa.

Using the same definition of technical change as before, TC in the FA model can be expressed as

$$TC_p = \sum_j \frac{\partial \ln f(\tilde{x})}{\partial \ln \tilde{x}_j} \frac{\partial \ln \tilde{x}_j}{\partial t} = \sum_j \frac{\partial \ln f(\tilde{x})}{\partial \ln \tilde{x}_j} \dot{A}_j \equiv \sum_j TC_p^j. \tag{6}$$

TC_p^j represents the contribution of the j^{th} input to the aggregate (overall) primal technical change TC_p. It is clear from (6) that TC_p^j depends on the rate of change of input productivity $(\dot{A}_j)$ and $\frac{\partial \ln f(\tilde{x})}{\partial \ln \tilde{x}_j}$, which under competitive market conditions, is the cost share of input j in total revenue.

The essential difference between the specifications of the production technology in (1) and (5) is that technical progress in (1) shifts the production function over time, *ceteris paribus*, whereas in (5) it enhances input quantities in efficiency units $(\tilde{x})$, thereby causing a movement along the production function $y = f(\tilde{x})$. Such a movement may be caused by factors other than time.

The input-specific productivity indices (A_j) can be functions of time as well as variable and quasi-fixed inputs.[5] Thus, depending on the specification of A_j, technical change can be purely exogenous (functions of only time), purely endogenous (functions of choice/decision variables), or a mixture of the two. However, in the context of estimating a single equation production function, endogeneticity of inputs is typically not taken into account. In such a case the distinction between endogenous and exogenous technical change is not clear. TFP growth in this setup is

$$T\dot{F}P = (RTS - 1)\sum_j \lambda_j \dot{x}_j + TC_P + \sum_j \{\lambda_j - S_j^a\}\dot{x}_j - \partial u/\partial t, \tag{7}$$

To examine these components in detail, we assume a translog functional form to represent the underlying production technology, viz.:

$$\ln y = \alpha_0 + \sum_j \alpha_j \ln \tilde{x}_j + \frac{1}{2}\sum_j \sum_k \alpha_{jk} \ln \tilde{x}_j \ln \tilde{x}_k - u, \tag{8}$$

where $\tilde{x}_j = A_j(t)x_j$. It is necessary to specify $A_j(t)$ in order to estimate the above model. To illustrate this, we consider the following two forms for $A_j(t)$. First, we specify $A_j(t)$ as a function of time with the following parameterization

$$\ln A_j(t) = a_j t + b_j t^2 \tag{9}$$

[5] Variables such as R&D expenditure that is considered a choice variable in the dynamic model also affects technical change. Technical change caused by choice variables is often labeled endogenous technical change.

where a_j and b_j are parameters which are to be estimated along with the parameters of the production function. In the second specification, the A_js are functions of time as well as other x variables, viz.:

$$\ln A_j = t\left(a_j + \sum_k b_{jk} \ln x_k\right) \tag{10}$$

where a_j and b_{jk} are parameters to be estimated.

From the above specifications one can easily test whether the rate of change in efficiency factors are constant or not, by restricting $b_j = 0 \ \forall \ j$ in (9) and $b_{jk} = 0$ in (10). Similarly, the hypothesis of no change in productivity of inputs can be tested from the restrictions $a_j = b_j = 0 \ \forall \ j$ in (9) and $a_j = b_{jk} = 0 \ \forall \ j, k$ in (10). Specification (10) assumes that productivity of an input at a point in time depends not only on time but also quantities of inputs being used at that point of time. That is, the productivity gain associated with an input is not purely exogenous. For example, productivity of labor in an environment with computers (higher level of capital) might be higher than a similar person working in an environment with fewer computers. Thus, labor productivity might depend on the stock of capital, and efficiency of capital might depend on the quantity of labor. Using this specification we can test whether one input affects the productivity of another input (i.e., whether some inputs are complementary to others from an efficiency point of view).

The measure of overall technical change (TC_p) using the above production function is

$$TC_p = \frac{\partial \ln y}{\partial t} = \sum_j R_j \dot{A}_j \equiv \sum_j TC_p^j, \tag{11}$$

where

$$R_j = \frac{\partial \ln y}{\partial \ln \tilde{x}_j} = \alpha_j + \sum_k \alpha_{jk} \ln \tilde{x}_j, \tag{12}$$

$$TC_p^j = R_j \dot{A}_j(t). \tag{13}$$

2.3. Latent Class Model

So far it has been assumed that there is a unique technology employed by every firm. However, firms in particular industries may use different technologies. In such a case, estimating a single frontier function encompassing every sample observation may not be appropriate in the sense that the estimated technology is not likely to represent the "true" technology. That is, the estimate of the underlying technology may be biased. Furthermore, if the unobserved technological differences are not taken into account during estimation, the effects of these omitted unobserved technological differences might be inappropriately labeled as inefficiency.

To reduce the likelihood of such misspecification, the sample observations are often classified into certain categories using exogenous sample separation information, and a separate technology is estimated for each group. For example, Mester (1993) and Grifell and Lovell (1997) grouped banks into private and savings banks. Kolari and Zardkoohi (1995) estimated separate costs functions for banks grouped in terms of their output mix. Mester (1997) grouped sample banks in terms of their location. In the above studies, estimation of the technology using a sample of firms is carried out in two stages. First, the sample observations are classified into several groups. This classification is based on either some *a priori* sample separation information (e.g., ownership of firms (private, public and foreign), location of firms, etc.) or applying cluster analysis to variables such as output and input ratios. In the second stage, separate analyses are carried out for each class/sub-sample.

To exploit the information contained in the data more efficiently, a latent class model approach (hereafter LCM) is often used, in which technological heterogeneity can easily be incorporated by estimating a mixture of production functions.[6] In the standard finite mixture model, the proportion of firms (observations) in a group is assumed to be fixed (a parameter to be estimated), see, for example, Beard and Gropper (1991) and Caudill (2003). However, the probability of a firm using a particular technology can be explained by some covariates and may vary over time, and these technologies along with the (prior) probability of using them (that might vary across firms) can be estimated simultaneously. Once the technologies are estimated, one can define the best practice technology (metafrontier) by taking the outer envelope of the individual technologies (see Battese *et al*. (2004)). The advantage of defining the best practice technology is that it can be compared to the individual technologies to measure technology gaps among countries, industries, firms, etc.

Assume that there are j technologies that can be represented by the density functions $f_j(v_j)$, $j = 1, \ldots, J$. Each producer in the sample uses one of these technologies (that is each firm at a point in time belongs to a particular class). The analyst does not know who is using what technology. Sometimes the analyst

---

[6] Finite mixture/LCMs are widely used in marketing, biology, medicine, sociology, psychology, and many other disciplines. For applications in social sciences, see, Hagenaars, J.A. and McCutcheon (2002). Statistical aspects of the mixing models are dicussed in detail in Mclachlan and Peel (2000).

follows some *ad hoc* criteria such as the size of firm, region in which the firm is located, etc., to group the sample firms. In the LCM one assumes that each firm has a probability of belonging to any group. Reintroducing the firm subscript i, these probabilities ($\pi_{ij} \geq 0, j = 1, \ldots, J \ \forall \ i$ and $\Sigma_j \pi_{ij} = 1 \ \forall \ i$) can be either constants or functions of covariates. Thus, in a LCM the technology for the j^{th} class is specified as

$$\ln y_i = \ln f(x_i, t)|_j + v_i|_j \tag{14}$$

where $j = 1, \ldots, J$ stand for class. For each class, the stochastic nature of the frontier is modeled by adding a two-sided random error term $v_i|_j$. The noise term for class j is assumed to follow a normal distribution with mean zero and constant variance (σ_{vj}^2). Thus, the conditional likelihood function for firm i given that it belongs to class j is

$$f(i|j) = \phi\left[\frac{v_i|j}{\sigma_{vj}}\right] \tag{15}$$

where $\phi(\cdot)$ is the pdf of a standard normal variable. Consequently, the unconditional likelihood function for firm i is

$$l(i) = \sum_j \pi_{ij}\phi\left[\frac{v_i|j}{\sigma_{vj}}\right] \tag{16}$$

where π_{ij} is the prior probability assigned by the analyst on firm i to be using the technology of type j. Sometimes firm-specific covariates are used to explain

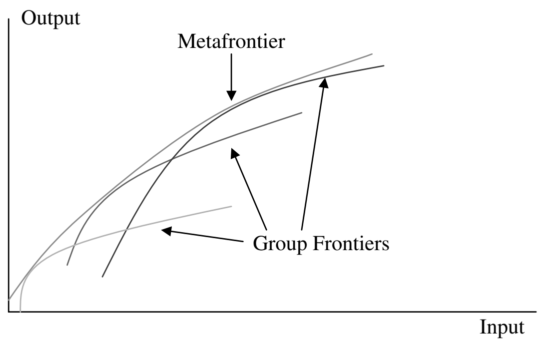


Fig. 2.2. The Metafrontier

[AU1

the prior probabilities. To impose the constraint that these probabilities are non-negative and sum to unity, the multinomial form is often used. That is

$$\pi_{ij} = \frac{\exp(z_j'\eta_j)}{\sum_j \exp(z_j'\eta_j)},\ \eta_j = 0 \tag{17}$$

where z_j are covariates explaining the prior probabilities and η_j are the associated parameters.

The log likelihood function is then

$$\log L = \sum_{i=1}^{n} \log l(i) \tag{18}$$

which is maximized with respect to β_j (which is the vector of parameters associated with the class j technology), η_j and σ_{vj} $(j = 1, \dots, J)$. Maximization of this likelihood function can be done with conventional gradient methods. The main problem is the possibility of multiple optima. Thus, one should use different starting values to make sure that the global maximum is achieved. The EM algorithm is a particularly useful device in this setting.[7] Furthermore, the EM algorithm is simple and intuitive (see Caudill (2003) for details). The above approach can easily be extended to incorporate technical inefficiency (Caudill (2003), Orea and Kumbhakar (2004), Kumbhakar and Tsionas (2003)).

Given that there are J groups (industries, regions, countries, etc.), the LCM estimates J different technologies.[8] The estimated parameters can be used to compute the conditional posterior class probabilities. Using Bayes' theorem, the posterior group probabilities can be obtained from $P(j|i) = l(i, j) \cdot \pi_{ij} / \sum_{j=1}^{J} l(i, j) \cdot \pi_{ij}$. The highest posterior probability can be used to assign a group for each producer, i.e., firm i is classified into group k $(= 1, \dots, J)$ if $P(k|i) = \max_j P(j|i)$. The technology of class k is then used as the reference technology to estimate technical efficiency of firm i. The outer envelope of these group technolo-

[7] See Dempster, Laird and Rubin (1977) and McLachlan and Peel (2000).

[8] In estimating a latent class model one has to address the problem of determining the number of groups/classes. The AIC and BIC (Schwartz's criterion) are the most widely used criterion in standard latent class models to determine the appropriate number of classes. Both statistics measure the model's goodness of fit but penalize by the complexity (number of parameters) of the model. Hence, they can be used to compare models with different numbers of classes. The best model is the one with lowest AIC or highest BIC.

gies (frontier for each group) is used to define the metafrontier (best practice technology), $f(x) = \max_j\{f(x)|_j \forall\ j\}$. Finally, the group technologies can be compared with the best practice technology to obtain measures of technology gap, defined as the ratio of $f(x)|_j$ to $f(x)$.

2.4. Multiple Outputs

One can express the multiple output production technology either in input or output augmenting form, viz., (i) $F(y, \tilde{x}) = a$ and (ii) $F(\tilde{y}, x) = a$, where $\tilde{y} = D(t)y$ is a vector of M outputs and $D(t)$ is the corresponding vector of efficiency factors while a is a constant. These models can be further extended to accommodate technical inefficiency by attaching the one-sided inefficiency term to either inputs or outputs. One way to interpret the formulation of technical change in (i) is that technical progress shifts the isoquants inward to produce a given level of output (thereby meaning that less inputs are needed) because input quantities in efficiency units are higher ($\dot{A}_j(t) > 0$). In formulation (ii) technical progress shifts the production possibility frontier (PPF) outward, thereby meaning that more outputs are produced (which implies that $\dot{D}_m < 0$), given the input quantities.

If the specification in (ii) is chosen, then the production function can be written as

$$y_1 = f(D_2 y_2, \ldots, D_M y_M, x_1, \ldots, x_J)\exp(-u) \tag{19}$$

where output 1 plays an asymmetric role. This is a major disadvantage because estimation of the production technology is not invariant to the choice of output 1. Furthermore, there is the endogeneticity problem. If output 1 is endogenous, why not output 2, etc? Because of these problems, multiple output production functions are not estimated econometrically. These functions will be considered again in the dual set-up where behavioral assumptions (cost minimization and profit maximization behaviors) are introduced explicitly into the model.

A convenient way of modeling multiple outputs in a primal framework is to use distance functions. In this framework the homogeneity property (degree one) is used to solve the asymmetry problem that was encountered in the multiple output production functions. Both input and output distance functions are specified as $F(x, y) = a$. The homogeneity property separates one from the other. The input (output) distance function is homogeneous of degree one in inputs (outputs). We now discuss the distance function approach.

3. DISTANCE FUNCTIONS

When many inputs are used to produce many outputs, Shephard's (1953, 1970) distance functions provide a functional characterization of the structure of the production technology. Input distance functions characterize input sets, and output distance functions characterize output sets. Not only do distance functions characterize the structure of production technology, but also they are intimately related to the measurement of technical efficiency.

An *input distance function* is a function $D_I(y, x, t) = \max\{\lambda : x/\lambda \in L(y)\}$, where L(*y*) describes the sets of input vectors that are feasible for each output vector. It adopts an input-saving approach to the measurement of the distance from a producer to the boundary of production possibilities. It gives the maximum amount by which an input vector can be radially contracted while still being able to produce the same output vector. In Figure 2.3, the scalar input x is feasible for output y, but y can be produced with smaller input (x/λ^*), and so $D_I(y, x, t) = \lambda^* > 1$.

An *output distance function* is a function $D_O(x, y, t) = \min\{\mu : y/\mu \in P(x)\}$, where $P(x)$ describes the sets of output vectors that are feasible for each input vector x. It takes an output-augmenting approach to the measurement of the distance from an observed input bundle to the boundary of production possibilities. It gives the minimum amount by which an output vector can be deflated and still remain producible for the same vector of inputs. In Figure 2.4 scalar output y can be produced with input x, but so can larger output (y/μ^*), and so $D_O(y, x, t) = \mu^* < 1$.

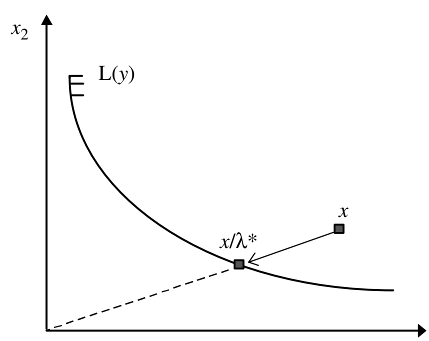


Fig. 2.3. An Input Distance Function (N = 2)

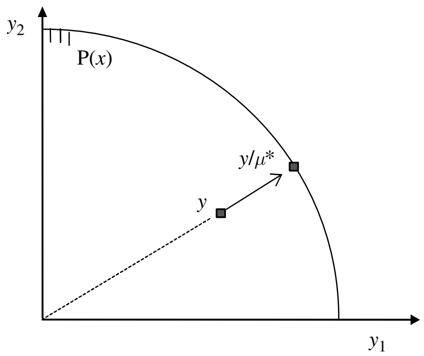


Fig. 2.4. An Output Distance Function (J = 2)

Fare and Primont (1995) discussed properties of both output and input distance functions. They show how to derive returns to scale (RTS) using these distance functions. These results are used next in the derivation of the TFP growth formulae.

3.1. Output Distance Function

In the presence of technical inefficiency, the output distance function is written as, $D_O(x, y, t) \leq 1 \Rightarrow D_O(x, y, t)\exp(u) = 1$, where $u \geq 0$ is output-oriented technical inefficiency. Differentiating totally, we get

$$\sum_m \frac{\partial \ln D_O(x, y, t)}{\partial \ln y_m}\dot{y}_m + \sum_j \frac{\partial \ln D_O(x, y, t)}{\partial \ln x_j}\dot{x}_j + \frac{\partial \ln D_O(x, y, t)}{\partial t} + \frac{\partial u}{\partial t} = 0$$

$$\Rightarrow \sum_m R_m \dot{y}_m - \sum_j S_j \dot{x}_j (RTS) + \partial \ln D_O/\partial t + \partial u/\partial t = 0$$

$$\Rightarrow T\dot{F}P = (RTS - 1)\sum_j S_j \dot{x}_j - \partial \ln D_O/\partial t - \partial u/\partial t \tag{20}$$

where

$$T\dot{F}P = \sum_m R_m \dot{y}_m - \sum_j S_j \dot{x}_j, \quad RTS = -\sum_{j=1}^{J} \frac{\partial \ln D_O(x, y, t)}{\partial \ln x_j},$$

$$S_j = \frac{w_j x_j}{\sum_j w_j x_j} = \frac{\partial \ln D_O/\partial \ln x_j}{\sum_j \partial \ln D_O/\partial \ln x_j}, \; R_m = \frac{p_m y_m}{\sum_m p_m y_m} = \frac{\partial \ln D_O/\partial \ln y_m}{\sum_m \partial \ln D_O/\partial \ln y_m} = \frac{\partial \ln D_O}{\partial \ln y_m}$$

since $\sum_m \partial \ln D_O/\partial \ln y_m = 1$ (which follows from the homogeneity property of D_O).[9] To estimate the above model, we use the homogeneity property and write the output distance functions as $D_O(x, y, t)/y_1 = D_O(x, \tilde{y}, t)$ where $\tilde{y} = (y_2/y_1, \ldots, y_J/y_1)$. Finally, we rewrite it as $\ln D_O(x, y, t) - \ln y_1 = \ln D_O(x, \tilde{y}, t) \Rightarrow -\ln y_1 = \ln D_O(x, \tilde{y}, t) + u$ where $\ln D_O = -u \leq 0$. After this, a parametric function is assumed for $\ln D_O(x, \tilde{y}, t)$ and a stochastic noise term is added prior to estimation. The standard frontier technique of maximum likelihood method can then be used to estimate the output distance function.

3.2. Input Distance Function

In the presence of technical inefficiency the input distance function is written as, $D_I(x, y, t) \geq 1 \Rightarrow D_O(x, y, t)\exp(-\tau) = 1$, where $\tau \geq 0$ is input-oriented technical inefficiency. Total differentiation of the input distance function yields

$$\begin{aligned}
&\sum_m \frac{\partial \ln D_I(x, y, t)}{\partial \ln y_m}\dot{y}_m + \sum_j \frac{\partial \ln D_I(x, y, t)}{\partial \ln x_j}\dot{x}_j + \frac{\partial \ln D_I(x, y, t)}{\partial t} - \frac{\partial \tau}{\partial t} = 0 \\
&\Rightarrow (-RTS^{-1})\sum_m R_m \dot{y}_m + \sum_j S_j \dot{x}_j + \partial \ln D_I/\partial t - \partial\tau/\partial t = 0 \\
&\Rightarrow T\dot{F}P = (1 - RTS^{-1})\sum_m R_m \dot{y}_m + \partial \ln D_I/\partial t - \partial\tau/\partial t \qquad (21)
\end{aligned}$$

where

$$T\dot{F}P = \sum_m R_m \dot{y}_m - \sum_j S_j \dot{x}_j, \;\; RTS^{-1} = -\sum_{m=1}^{M} \frac{\partial \ln D_I(x, y, t)}{\partial \ln y_m},$$

$$R_m = \frac{p_m y_m}{\sum\limits_m p_m y_m} = \frac{\partial \ln D_I/\partial \ln y_m}{\sum_m \partial \ln D_I/\partial \ln y_m}, \;\; S_j = \frac{w_j x_j}{\sum\limits_j w_j x_j} = \frac{\partial \ln D_I/\partial \ln x_j}{\sum\limits_j \partial \ln D_I/\partial \ln x_j} = \frac{\partial \ln D_I}{\partial \ln x_j}$$

since $\sum_j \partial \ln D_I/\partial \ln x_j = 1$ (which follows from the homogeneity property of D_I). Using the same property, we can express $D_I(x, y, t)$ as $D_I(x, y, t)/x_1 = D_I(\tilde{x}, y, t)$ where $\tilde{x} = (x_2/x_1, \ldots, x_J/x_1)$. Thus, $\ln D_I - \ln x_1 = \ln D_I(\tilde{x}, y, t)$ which in turn implies that $-\ln x_1 = \ln D_I(\tilde{x}, y, t) - \tau$ where $\ln D_I = \tau \geq 0$. Therefore, the second component in (21) (i.e., $\partial \ln D_I(x, y, t)/\partial t$) encapsulates technical change, where $\partial \ln D_I(x, y, t)/\partial t = \partial \ln D_I(\tilde{x}, y, t)/\partial t \geq 0$ (≤ 0) implies technical progress (regress). The last component of (21) (i.e., $-\partial\tau/\partial t$) represents the effects of technical efficiency change. Once τ is estimated from the

[9] See Brummer *et al.* (2002) for details on the decomposition.

model $-\ln x_1 = \ln D_I(\tilde{x}, y, t) - \tau$, after a parametric function is assumed for $\ln D_I(\tilde{x}, y, t)$ and a stochastic noise term is added, $\partial\tau/\partial t$ can be easily estimated. Thus, by estimating an input distance function, all three components of TFP growth can be obtained. For example, if a translog form is assumed for $\ln D_I(\tilde{x}, y, t)$, viz.:

$$\begin{aligned} -\ln x_1 = {} & \alpha_0 + \sum_{j=2}^{J} \alpha_j \ \ln(x_j/x_1) + \sum_{m=1}^{m} \beta_m \ln y_m + \frac{1}{2} \sum_{j=2}^{J} \sum_{h=1}^{J} \alpha_{jh} \ln(x_j/x_1) \ln(x_h/x_1) \\ & + \frac{1}{2} \sum_{m=1}^{M} \sum_{l=1}^{M} \beta_{ml} \ln y_m \ln y_l + \sum_{j=1}^{J} \sum_{k=1}^{J} \gamma_{jk} \ \ln(x_j/x_1) \ln y_k \\ & + \psi_0 t + \frac{1}{2} \psi_{00} t^2 + \sum_{j=2}^{J} \xi_{jt} t \ln(x_j/x_1) + \sum_{m=1}^{M} \zeta_{mt} t \ln y_m + v - \tau \end{aligned} \tag{22}$$

Standard frontier techniques can be used to estimate the above model from which RTS, R_m, S_j and $\partial\tau/\partial t$ can be computed. These are then used to obtain the components of TFP growth in (21).[10]

4. THE COST FUNCTION APPROACH[11]

4.1. Single Output

While modeling inefficiency in a cost model, we can argue that inefficiency increases cost and therefore the cost function can be written as $C^a = C^0 \times CE$ where C^a is the observed cost, C^0 is the minimum (frontier) cost, and $1/CE \leq 1$ is cost efficiency. While this argument is true for input-oriented technical inefficiency, it is worthwhile to derive the above result from the firm's optimization problem, which is

$$\min_x \ w'x \text{ subject to } y = f(xe^{-\tau}, t)$$

where $\tau \geq 0$ is the input-oriented technical inefficiency. Assuming that firms are allocatively efficient, $f_j/f_1 = w_j/w_1$, $j = 2, \ldots, J$, where f_j, is the marginal product (MP) of input j. The solution of the above problem gives inefficiency

[10] See Karagiannis *et al.* (2004) for details.
[11] See Kumbhakar (2000) for details on this model with allocative inefficiency.

adjusted input demand functions, $x_j(.)e^{-\tau}$ as a function of (w, y, t). These input demand functions can be used to define actual (observed) cost, C^a, which is

$$C^a = \sum_j w_j x_j \Rightarrow C^a e^{-\tau} = \sum_j w_j x_j e^{-\tau} = C^0(x, y, t) \tag{23}$$

where C^0 (which is the minimum cost function without technical inefficiency). The above cost function can be written as

$$\ln C^a = \ln C^0(w, y, t) + \tau, \tag{24}$$

where, $\tau \geq 0$ can be interpreted as the percentage increase in cost due to technical inefficiency.[12]

Following Denny and Waverman (1981), it can be shown that the TFP growth formula in a cost minimizing set up is

$$\begin{aligned} T\dot{F}P \equiv \dot{y} - \sum_j S_j^a \dot{x}_j &= \dot{y}\left(1 - \frac{\partial \ln C}{\partial \ln y}\right) - \dot{C}_t - \frac{\partial \tau}{\partial t} \\ &= (1 - 1/RTS)\dot{y} - \dot{C}_t - \frac{\partial \tau}{\partial t}, \end{aligned} \tag{25}$$

which decomposes *TFP* growth into scale, *TC*, and *TEC* components.

The cost function, which is dual to the factor augmenting production function, augmented to accommodate input-oriented technical inefficiency (τ) can be written as

$$C^a = C(\tilde{w}, y)\exp(\tau), \tag{26}$$

where $\tilde{w}_j = B_j(t)w_j$. If A_j depends only on time, then $B_j(t) = 1/A_j(t)\ \forall\ j$. Thus, an increase in efficiency of an input is equivalent to a decrease in its effective price ($\tilde{w}$). The overall technical change in a cost model is expressed as $TC_c = -\frac{\partial \ln C}{\partial t} = -\sum_j \frac{\partial \ln C(.)}{\partial \ln \tilde{w}_j}\dot{B}_j(t)$, which can be expressed as $TC_c = \Sigma_j TC_c^j$ where $TC_c^j = -\frac{\partial \ln C(.)}{\partial \ln \tilde{w}_j}\dot{B}_j(t)$. Thus, TC_c is a weighted average of input productivity change ($\dot{A}_j(t) = -\dot{B}_j(t)$), where the weights are the cost shares (S_j^a).

Thus, the TFP growth formula can be written as

$$\begin{aligned} T\dot{F}P \equiv \dot{y} - \sum_j S_j^a \dot{x}_j &= \dot{y}\left(1 - \frac{\partial \ln C}{\partial \ln y}\right) + TC_c - \frac{\partial \tau}{\partial t} \\ &= (1 - 1/RTS)\dot{y} + TC_c - \frac{\partial \tau}{\partial t} \end{aligned} \tag{27}$$

[12] See Kumbhakar (1996) and Kumbhakar and Wang (2005) for the derivation of the cost function with output-oriented technical inefficiency.

The TFP growth formulae with output-oriented technical (along with and without allocativity) inefficiency are given in Appendix D. In Appendix E, we derive the TFP growth formulae with non-neutral output-oriented technical inefficiency.

4.2. Multiple Output Cost Function

If there are multiple outputs, TFP growth is defined as $T\dot{F}P = \sum_m R_m \dot{y}_m - \sum_j S_j^a \dot{x}_j$, where $R_m = p_m y_m / \sum_l p_l y_l$ with p_m representing the price of output y_m $(m = 1, \ldots, M)$. It can be shown (Denny *et al.* (1981)) that the components of TFP are

$$\begin{aligned} T\dot{F}P &= TC + (1 - RTS^{-1})\dot{Y}_C + (\dot{Y}_P - \dot{Y}_C) - \frac{\partial \tau}{\partial t} \\ &= TC + Scale + Markup + TEC, \end{aligned} \tag{28}$$

where $\dot{Y}_C = \sum_l \left(\frac{\varepsilon_{cy_l}}{\sum_l \varepsilon_{cy_l}} \right) \dot{y}_l$, $\dot{Y}_P = \sum_l \frac{p_l y_l}{R} \dot{y}_l$, and $\varepsilon_{cy_l} = \partial \ln C / \partial \ln y_l$. TC and RTS in (28) can be obtained from a parametric cost function $(C = C(w,\ y,\ t))$ when RTS is defined as $RTS^{-1} = \sum_l \partial \ln C / \partial \ln y_l$. Thus, the first component of TFP growth is TC and the second component is the *Scale* component (related to RTS), which is zero if RTS is unity. The third component is non-zero if output markets are non-competitive. That is, if output prices depart from their respective marginal costs then $\dot{Y}_P \neq \dot{Y}_C$. Finally, the last component is due to technical efficiency change $(TEC = -\partial \tau / \partial t)$. It is positive (negative) if efficiency improves (deteriorates), i.e., τ declines (increases) over time.

4.3. The Factor-Augmenting Model of Technical Change

Here we consider the multiple output production technology $F(y, \tilde{x}) = a$ with input-oriented technical inefficiency, for which the cost function is $C^a = C(\tilde{w}, y) \exp(\tau)$. Following the procedure for the time trend model, the TFP growth formula for the FA model can be written as

$$\begin{aligned} T\dot{F}P &= TC_c + (1 - RTS^{-1})\dot{Y}_C + (\dot{Y}_P - \dot{Y}_C) - \frac{\partial \tau}{\partial t} \\ &= TC_c + Scale + Markup + TEC, \end{aligned} \tag{29}$$

where as before $TC_c = -\frac{\partial \ln C}{\partial t} = -\sum_j \frac{\partial \ln C(.)}{\partial \ln \tilde{w}_j} \dot{B}_j(t)$.

To pursue this decomposition in detail, we assume a translog form for $C(\tilde{w}, y)$ and write it as

$$\ln C^a = \alpha_0 + \sum_j \alpha_j \ln \tilde{w}_j + \sum_m \beta_m \ln y_m + \sum_j \sum_m \alpha_{jm} \ln \tilde{w}_j \ln y_m + \frac{1}{2}\left\{\sum_j \sum_k \alpha_{jk} \ln \tilde{w}_j \ln \tilde{w}_k + \sum_l \sum_m \beta_{lm} \ln y_l \ln y_m\right\} + \tau, \qquad (30)$$

where $\tilde{w}_j = B_j(t)w$. The cost function in (30) is assumed to satisfy symmetry and linear homogeneity (in $\tilde{w}$) restrictions. We specify the B_j as quadratic functions of time, i.e.:

$$\ln B_j = t(a_j + b_j t)$$

We then use formula $TC_c = -\sum_j \frac{\partial \ln C(.)}{\partial \ln \tilde{w}_j} \dot{B}_j(t)$ to estimate technical change where

$$\frac{\partial \ln C}{\partial \ln \tilde{w}_j} = \alpha_j + \sum_k \alpha_{jk} \ln \tilde{w}_k + \sum_m \beta_{jm} \ln y_m$$

with the appropriate forms for $\dot{B}_j$ derived from $\ln B_j(t)$. Furthermore, technical change as defined above can be decomposed into pure, scale and input price components.

In banking applications, multiple outputs are almost always used. One common problem with multiple outputs is the presence of zero values for some outputs (meaning that some banks do not produce all the outputs). Zero outputs create problems for the Cobb Douglas and translog cost functions, since the log of zero is not defined. To avoid this problem, researchers often replace the zero values with a small positive number. Some replace them by unity so that log values are zero, while others add the small positive number to all observations. For example, if output m has zero values for some banks, $\ln y_m$ is redefined as $\ln\, y_m = \ln(y_m + c)$, where c is a positive number supplied by the user. This procedure, although widely used in practice, has two problems. First, zero values for output(s) for a bank might be due to the fact that the bank specializes in a few outputs, and adding small numbers for outputs that are never produced puts the specialized banks in the same group as others. In fact, this procedure does not recognize the fact that some banks can be specialized in certain outputs. Thus, no matter how small c is, this procedure is not innocuous. Second, the output values are changed (either from zero to a positive constant c or from y_m to $(y_m + c)$) without changing the cost. That is, the cost of the extra output (c) is zero. Looking at it from a purely mathematical point of view, either $C = C(w, y_1, \ldots, y_m, \ldots, y_M, t)$, or $C = C(w, y_1, \ldots, y_m, (y_{m+1} + c), \ldots, (y_M + c), t)$ is true, but not both at the same time.

Replacing the zero values by unity makes sense if the technology of the banks that are specialized in fewer outputs is the same as those that produce all the outputs. For example, if the specialized banks produce Q outputs while the other banks produce M $(M > Q)$ outputs and the production technology for both types of banks is Cobb-Douglas, i.e.:

$$\ln C_i = \alpha_0 + \sum_{j=1}^{J} \alpha_{\mathrm{j}} \ln w_{ji} + \sum_{m=1}^{Q} \beta_m \ln y_{mi} + v_i, \quad i = 1, \ldots, N_1$$

$$\ln C_i = \alpha_0 + \sum_{j=1}^{J} \alpha_{\mathrm{j}} \ln w_{ji} + \sum_{m=1}^{M} \beta_m \ln y_{mi} + v_i, \quad i = N_1 + 1, \ldots, N$$

then we can simply combine them and write the technology as

$$\ln C_i = \alpha_0 + \sum_{j=1}^{J} \alpha_j \ln w_{ji} + \sum_{m=1}^{M} \beta_m \ln y_{mi} + v_i, \quad i = 1, \ldots, N$$

where $y_{mi} = 1$ for $m = Q + 1, \ldots, M$ and $i = 1, \ldots, N_1$. Since the hypothesis, that the technology for the specialized and non-specialized banks is the same, is testable, there is no point in making the assumption *a priori*. Thus the problem is not as innocuous as it seems.

This zero value problem is not endemic to the multi-output cost functions. It applies equally to output and input distance functions, as well as the latent class models discussed before.

4.4. Latent Class Models

The existence of multiple technologies can easily be accommodated in each of the above cases. Orea and Kumbhakar (2004) used a multiple output cost function approach and found four separate technologies used by the Spanish banks. Their approach can easily be extended to construct the metacost function, which is the inner envelope of the group cost frontiers. Such a metacost function can then be used to measure technology gaps among banks using different technologies. The Orea-Kumbhakar approach also accommodates technical inefficiency that varies across banks and differs among technologies. Furthermore, these inefficiencies can vary over time in a flexible manner. Because of the similarity of the approach with the production function approach discussed earlier, no further detail is given.

5. THE PROFIT FUNCTION APPROACH

5.1. Single Output

When modeling inefficiency by using a profit model, we start with the argument that inefficiency increases cost and therefore the profit function can be written as $\pi^a = \pi^0 \times PE$, where π^a is the observed profit, π^0 is the maximum (frontier) profit, and $PE \leq 1$ is profit efficiency, which is modeled as a one-sided error term. Furthermore, it is assumed to be independent of the regressors (input and output prices and time). Here it is shown that this assumption is generally untrue.

The profit function, in the presence of technical inefficiency corresponding to the production function in (1), can be written as $\pi(p, w, t, u) = \pi(w, pe^{-u}, t)$ (Lau (1978), Theorem II-3, p. 154). The rationale behind the above result is as follows. First, the first-order conditions of profit maximization are $f_j = w_j/pe^{-u}$. Second, the production function in (1) can be rewritten as $ye^u = f(x, t)$. Thus, if one substitutes pe^{-u} for p and ye^u for y, the above first-order conditions and the production function look like a standard neo-classical production function. Consequently, the solutions of input demand and output supply functions (adjusted for inefficiency) can be expressed as $x_j = x_j(w, pe^{-u}, t)$ and $y = y(w, pe^{-u}, t)$. Therefore, the profit function, conditional on u is defined as

$$\pi(w, pe^{-u}, t) = \max_{ye^u, x}[\{py - w'x\}|y = f(x, t)e^{-u}]$$

which also equals the actual profit $\pi^a = py - w'x$. It means that profit, when price is p and output equals y, is the same as profit when output equals $f(x)$ but price equals pe^{-u}. That is, a 10% reduction in output given inputs has the same effect on profit as a 10% reduction in output price holding output constant.

Using $p^s \equiv pe^{-u}$, the profit function can be implicitly written as $\pi(w, p^s, t)$. Then the Hotelling's lemma, $\partial\pi/\partial p^s = ye^u$ and $\partial\pi/\partial w_i = -x_j$ can be applied to obtain the input demand and technical inefficiency adjusted output supply functions. Note that the argument of the profit function is p^s, which in turn implies that technical inefficiency may not appear additively in the log profit function. For example, in the case of the translog profit function

$$\begin{aligned}\ln \pi = {} & \alpha_0 + \sum_j \alpha_j \ln w_j + \beta_p \ln p^s + \alpha_t t \\ & + \frac{1}{2}\left\{\beta_{pp} \ln p^s \ln p^s + \sum_j \sum_k \alpha_{jk} \ln w_j \ln w_k + \alpha_{tt} t^2\right\} \\ & + \sum_j \gamma_{jm} \ln w_j \ln p^s + \beta_{pt} \ln p^s t + \sum_j \delta_{jt} \ln w_j t, \qquad (31)\end{aligned}$$

which can be rewritten as

$$
\begin{aligned}
\ln \pi = & \alpha_0 + \sum_j \alpha_j \ln w_j + \beta_p \ln p + \alpha_t t \\
& + \frac{1}{2}\left\{ \beta_{pp} \ln p \ln p + \sum_j \sum_k \alpha_{jk} \ln w_j \ln w_k + \alpha_{tt} t^2 \right\} + \sum_j \gamma_{jm} \ln w_j \ln p \\
& + \beta_{pt} \ln pt + \sum_j \delta_{jt} \ln w_j t - u\left(\beta_p + \sum_j \gamma_{jm} \ln w_j + \beta_{pt} t \right) + \frac{1}{2}\beta_{pp} u^2 \quad (32) \\
\equiv & \ln \pi^0 - g(u, \ln p, \ln w, t)
\end{aligned}
$$

where $\ln \pi^0$ is the translog profit frontier, and $g(u, \ln p, \ln w, t) = u(\beta_p + \sum_j \gamma_{jm} \ln w_j + \beta_{pt} t) - \frac{1}{2}\beta_{pp} u^2 \geq 0$ is profit inefficiency (profit loss due to technical inefficiency). It is clear from the profit function above that $g(u, \ln p, \ln w, t)$ is a non-linear function of u and it cannot assumed that $g(u, \ln p, \ln w, t)$ is an independently and identically distributed random variable. As a result, the standard tools used to estimate the production frontier cannot be used here.[13] A simple sign change is not enough to model technical inefficiency in the profit function, unless the underlying production function is homogeneous.[14]

This will also affect the TFP growth formulation. Differentiating the profit function $\pi = \pi(w, p^s, t)$ totally, we get

$$
\frac{d \ln \pi}{dt} = \frac{\partial \ln \pi}{\partial \ln p^s} \dot{p}^s + \sum_j \frac{\partial \ln \pi}{\partial \ln w_j} \dot{w}_j + \frac{\partial \ln \pi}{\partial t} = \frac{1}{\pi^s} \left\{ py \; \dot{p}^s - C^a \sum_j S_j \dot{w}_j \right\} + \frac{\partial \ln \pi}{\partial t},
$$

using $\partial \pi / \partial p = y$ and $\partial \pi / \partial w_j = -x_j$ (Hotelling's lemma).

From the definition of profit $\pi^a = py - \Sigma_j w_j x_j$ we get

$$
\frac{d \ln \pi}{dt} = \frac{py}{\pi} \{ \dot{y} + \dot{p} \} - \frac{C^a}{\pi} \sum_j \{ S_j \dot{x}_j + S_j \dot{w}_j \}
$$

Equating the above two equations gives (after some algebraic manipulations),

$$
\begin{aligned}
\pi \left\{ \frac{\partial \ln \pi}{\partial t} \right\} &= py\left(\dot{y} - \sum_j S_j^a \dot{x}_j \right) + py \sum_j S_j^a \dot{x}_j - C^a \sum_j S_j^a \dot{x}_j + py\{\partial u / \partial t\} \\
&= py(T\dot{F}P) - py(RTS^s - 1) \sum_j S_j^a \dot{x}_j + py\{\partial u / \partial t\}
\end{aligned}
$$

---

[13] See Kumbhakar (2001) and Kumbhakar and Tsionas (2005) for details on estimation issues.

[14] There are numerous banking papers that use translog profit functions as $\ln \pi^a = \ln \pi^0 - u$ (for example, see Berger and Mester (2003) and the references cited in there), which is clearly inappropriate.

The above equation can be rearranged to give

$$T\dot{F}P = \frac{\pi}{py}\left\{\frac{\partial \ln \pi}{\partial t}\right\} + (RTS^s - 1)\sum_j S_j^a \dot{x}_j - \{\partial u/\partial t\}. \tag{33}$$

where $RTS^s = 1 - \{\partial \ln \pi/\partial \ln p^s\}^{-1}$. Furthermore, it follows from the envelope theorem that $\frac{\partial \pi}{\partial t} = p\frac{\partial f}{\partial t}$, which in turn implies that $\frac{\pi^a}{py}\frac{\partial \pi}{\partial t} = \frac{\partial \ln f}{\partial t} = TC$, which is used in (3). There is one important difference between (3) and (33). In deriving (3) we did not rely on any behavioral assumptions explicitly, whereas in (33) we used the profit maximizing conditions to determine the allocation of inputs and the production of output. Consequently, input quantities in (33) (and therefore RTS) are affected by the presence of technical inefficiency.[15]

Using the following results, the components of TFP growth can be expressed in terms of the profit function, viz.:

$$\begin{aligned}
&\frac{py}{\pi^s} = \frac{\partial \ln \pi}{\partial \ln p^s},\\
&\{RTS^s - 1\} = -\{\partial \ln \pi/\partial \ln p^s\}^{-1},\\
&S_j^a = \frac{\{\partial \ln \pi^s/\partial \ln w_j\}}{\sum_k\{\partial \ln \pi/\partial \ln w_k\}},\\
&\dot{x}_j = \left\{\frac{\partial^2 \ln \pi}{\partial t\ \partial \ln w_j}\right\}\left\{\frac{\partial \ln \pi}{\partial \ln w_j}\right\}^{-1} + \frac{\partial \ln \pi}{\partial t} - \frac{\partial \ln w_j}{\partial t}.
\end{aligned} \tag{34}$$

The problem with the above decomposition is that each component depends on technical inefficiency. This can be avoided by using the relation $\ln \pi = \ln \pi^0 - g(u, \ln p, \ln w, t)$ in the above formula. In doing so, we can capture both direct and indirect effects (via input and output prices as well as time) of technical inefficiency on TFP growth. However, if we specify (erroneously) the translog profit function as $\ln \pi = \ln \pi^0 - u$, then (i) the estimated parameters of the profit function are likely to be biased, and (ii) the indirect effect (via input and output prices and time) of technical inefficiency TFP growth would not be captured.

5.2. Multiple Output

As before, we define $T\dot{F}P$ as $T\dot{F}P = \Sigma_m R_m \dot{y}_m - \Sigma_j S_j^a \dot{x}_j$ when $R_m = p_m y_m/R$ and $R = \Sigma_m p_m y_m$. In the standard case, where firms are assumed to be both

[15] See Kumbhakar (2000) for the TFP growth formula that takes into account both technical and allocative inefficiency. The decomposition without any inefficiency is given in Kumbhakar (2000b).

technically and allocatively efficient (so that the multiple output profit function is $\pi = \pi(w, p, t)$), the TFP change formula can be expressed as[16]

$$T\dot{F}P = \frac{\pi}{R}\frac{\partial \ln \pi}{\partial t} + (RTS - 1)\sum_j S_j \dot{x}_j, \tag{35}$$

where $RTS - 1 = -\{\Sigma_m \partial \ln \pi / \partial \ln p_m\}^{-1}$.

To derive the TFP change in the present case, we denote $p^s = e^{-u}(p_1, \ldots, p_M)'$ and start from the profit function $\pi = \pi(w, p^s, t)$, which is defined as $\pi(w, p^s, t) = \sum_m p^s_m y_m e^u - \sum_j w_j x_j$. However, the actual (observed) profit is $\pi^a = \sum_m p_m y_m - \sum_j w_j x_j = \sum_m p^s_m y_m e^u - \sum_j w_j x_j = \pi(w, p^s, t)$. Thus, the TFP change formula is

$$T\dot{F}P = \frac{\pi^s}{R}\frac{\partial \ln \pi^s}{\partial t} - \frac{\partial u}{\partial t} + (RTS^s - 1)\sum_j S_j \dot{x}_j, \tag{36}$$

in which we used the results $\partial \pi / \partial p_m = y_m$ (Hotelling's lemma) and $\partial \pi / \partial w_j = -x_j$. All the components in (36) can be expressed in terms of profit. For example, $R = \pi^s\{\Sigma_m \partial \ln \pi^s / \partial \ln p^s_m\}$, $RTS^s - 1 = -\{\Sigma_m \partial \ln \pi / \partial \ln p^s_m\}^{-1}$, which gives $(RTS^s - 1) = -\frac{\pi}{R}$. Using these results, coupled with those in (34) for S^a_j, everything in (36) can be expressed in terms of π and rates of change in input and output quantities.

The first term on the right-hand side of (36) can be expressed as

$$\frac{\pi^s}{R}\frac{\partial \ln \pi^s}{\partial t} = \sum_m R_m \frac{\partial \ln y_m(.)}{\partial t}, \tag{37}$$

where $y_m(.)$ is the supply function of output y_m. Since the expression in (37) is a weighted average (weights being the revenue share of each output) of rates of change of individual outputs ($\partial \ln y_m / \partial t$), holding everything else constant, it can be viewed as a measure of output technical change. Note that this measure of output technical change is different from profit technical change, defined as $\partial \ln \pi^s / \partial t$.

It can be seen from (36) that contribution of technical change and returns to scale depends on both technical inefficiency via p^s. Since technical change, returns to scale, etc., are usually defined in terms of the profit frontier, it is possible to rewrite (36) in terms of the profit frontier so that the effects of technical

[16] See Karagiannis (2000) for TFP growth decomposition with multiple outputs in a profit function framework with quasi-fixed inputs.

inefficiency on TFP change are separated. For example, in the translog case we can express actual profit as $\ln \pi = \ln \pi^0 - g(u, \ln p, \ln w, t)$, where $\ln p$ and $\ln w$ are vectors of output and input prices. This result can be used in the TFP growth formula to separate technical inefficiency effects in the TFP growth components.

5.3. Alternative Profit Functions

Use of the profit function approach is based on the assumption that prices are exogenous, and that producers seek to maximize profit (or variable profit) by selecting outputs and inputs under their control. One justification for exogeneticity of prices is that producers operate in competitive markets. If producers have some degree of monopoly power in their product markets, then demand would be exploited to determine output prices and quantities jointly, and only input prices would be exogenous.

Recently Humphrey and Pulley (1997), among others, have introduced the notion of an "alternative" profit frontier to bridge the gap between a cost frontier and a profit frontier. For example, Berger and Mester (1997) suggest that the alternative profit approach may be helpful when (i) there are substantial unmeasured differences in the quality of banking services; (ii) outputs are not completely variable; (iii) output markets are not perfectly competitive; and (iv) output prices are not accurately measured.

An alternative profit frontier is defined as

$$\begin{aligned} \pi^A &= \max_{p,x}\{p'y - w'x | F(x, y, t) = 0, g(p, y, x, t) = 0\} \\ &\Rightarrow \pi^A = \pi(y, w, t), \end{aligned} \tag{38}$$

where the endogenous variables are (p, x) and the exogenous variables are (y, w, t). $F(x, y, t) = 0$ is the production function (it can also be specified by an output distance function), and $g(p, y, w, t) = 0$ represents what Humphrey and Pulley refer to as the producer's "pricing opportunity set," which captures the producer's ability to transform exogenous (y, w, t) into endogenous product prices p. However, $\pi^A(.)$ is not dual to the production function because it incorporates both the structure of production technology and the structure of the pricing opportunity set. Moreover, without specifying the properties satisfied by the function $g(p, y, w, t) = 0$, it is not possible to specify the properties satisfied by $\pi^A(.)$.

The main problem with the alternative profit function is that it has no theoretical foundation such as with the standard profit and/or cost functions. Consequently, the approach has no use other than measuring efficiency, which is also problematic, as will be shown later. Since the whole idea is based on the pricing opportunity function $g(p, y, w, t) = 0$, it is worth exploring the issue further. First, the single output case is considered. The profit maximization problem is partitioned into two steps. In step 1 the cost is minimized, given the production function, to obtain the cost function, $C^a = w'x = C(w, y, t)$. In step 2 the profit is maximized, viz., $\max_p\{py - C(w, y, t)|g(p, y, w, t) = 0\}$. Note that $g(p, y, w, t) = 0$ can be expressed implicitly, as $p = p(w, y, t)$, which is nothing but the inverse demand function. Thus, there is no need to solve the optimization problem in step 2 for p. Consequently, there is no meaning to the alternative profit function because producers do not maximize profit to obtain p. The so-called optimal price p is not related to the production technology (either the production or the cost function).

A similar result is obtained when multiple outputs are considered. The cost function from step 1 can still be written as $C^a = w'x = C(w, y, t)$. The first-order conditions from step 2 are: $p_m = \mu\, \partial g(.)/\partial p_m \;\; m = 1, \ldots, M$ where μ is the Lagrange multiplier associated with the pricing opportunity function. Thus, the output prices can be solved from the above first-order conditions and the pricing opportunity function $g(p, y, w, t) = 0$, without any reference to the technology (production or cost functions). These solutions are, in implicit form, $p_m = p_m(w, y, t)$. Thus, cost of production (namely, marginal cost) does not play any role in determining optimal output prices.

Now we examine the case with (input-oriented) technical inefficiency and write the single output production function as $y = f(x \exp(-\tau), t)$, for which the cost function is $C^a = w'x = C(w, y, t) \exp(\tau)$. The first-order conditions from the second step are exactly the same as before. Thus, the solution of p would not be affected by the presence of technical inefficiency. Consequently, revenue will be unaffected by the presence of inefficiency. Improvement in technical inefficiency would not be transmitted to revenue through output prices. Using this optimal price $p = p(w, y, t)$ in the definition of profit we get

$$\pi^a = p(w, y, t)y - C(w, y, t) \exp(\tau) \neq \pi^A(w, y, t) \exp(-\tau) \tag{39}$$

Thus, the above relationship cannot be expressed as $\ln \pi^a = \ln \pi^0 - \tau$, as is common in the literature (e.g., Berger and Mester (2003), DeYoung and Hasan (1998), among others). That is, the observed profit is not necessarily reduced by the same proportion by which cost is increased. For example, if profit without inefficiency is $\pi_0^a = p(w, y, t)y - C(w, y, t) = 100 - 80 = 20$ and the profit with technical inefficiency is $\pi_0^a = p(w, y, t)y - C(w, y, t) \exp(\tau) = 100 - 80(1.10) = 12$, profit is reduced by 40% although cost has increased by only 10%.

In the presence of multiple outputs, the cost function is $C^a = w'x = C(w, y, t) \exp(\tau)$. Maximization of profit subject to $g(p, y, w, t) = 0$ gives the solutions of output prices that can be implicitly expressed as $p_m = p_m(w, y, t)$. Note that these prices are not affected by the presence of technical inefficiency. Thus, given output quantities, revenue is not affected by the presence of inefficiency. Consequently, the argument that alternative profit function takes into account the effect of inefficiency on both cost and revenue does not hold. Furthermore, actual profit is

$$\pi^a = \sum_m p_m(w, y, t)y_m - C(w, y, t)\exp(\tau) \neq \pi^A(w, y, t)\exp(-\tau). \tag{40}$$

That is, actual profit is not reduced by the same percent by which cost is increased. Therefore, the efficiency estimates based on the alternative profit function cannot be related to technical inefficiency based on the production technology. Since we cannot draw a meaningful interpretation of the inefficiency term in the model $\ln \pi^A = \ln \pi^0(w, y, t) - \tau$, and the whole idea of estimating the alternative profit function is to estimate profit inefficiency, the alternative profit function approach seems vacuous. However, this is not the case with the standard profit function, although care has to be taken in modeling profit inefficiency (as was shown in the previous section).

This brings us back to the standard profit function that can be extended to accommodate non-competitive behavior in the output markets. This issue is explored next (see Kumbhakar and Lozano-Vivas (2004) for details).

5.4. Modeling Markups in Variable Profit Functions

Assuming that the objective of the banks is to maximize profit, the bank's optimization problem can be formulated with the following two steps. In step 1, a bank solves the following problem, given the output vector y, to determine the least cost (variable) input quantities, i.e.:

$$\text{Min}_x \; w'x \text{ subject to } F(y, x, z, t) = 0$$

where z is the vector of Q quasi-fixed inputs and output attributes. The solution to the above problem gives the conditional input demand functions $x_j = x_j(w, y, z, t)$ that are then substituted into the objective function to obtain the minimum cost function $C(w, y, z, t)$. In step 2, the bank's problem is to maximize profit, i.e.:

$$\text{Max}_y \; \pi = p'y - C(w, y, z, t)$$

in which the choice variables are outputs. If the output markets are competitive, the first-order conditions (FOC) for profit maximization can be expressed as

$$p_m = \frac{\partial C}{\partial y_m} \equiv MC_m, m = 1, \ldots, M,$$

where MC_m is the marginal cost associated with output y_m. However, if the output markets are not competitive and the banks possess monopoly power, the FOC of the above problem become

$$p_m^* = MC_m \tag{41}$$

where $p_m^* = p_m \theta_m$, with $\theta_m \geq 1$ representing the markup factor. Thus, the presence of markups can be tested by restricting $\theta_m = 1 \forall\, m$. In the presence of markups, the relevant output prices are $p_m^* = p_m \theta_m$ and the output supply functions are $y_m = y_m(w, p^*, z, t)$. Consequently the variable profit function can be expressed as $\pi^* = \pi^*(w, p^*, z, t)$ from which the input demand and output supply functions can be obtained by using Hotelling's lemma, viz.:

$$\frac{\partial \pi^*}{\partial p^*} = y_m \text{ and } \frac{\partial \pi^*}{\partial w_j} = -x_j.$$

The profit function $\pi^*(.)$ is often labeled as the shadow variable profit function.

However, it should be noted that π^* is not observed and $\pi^* \neq \pi^a$ where π^a is the actual profit, defined as $\pi^a = \sum_m p_m y_m - \sum_j w_j x_j$. Invoking Hotelling's lemma it can be shown that

$$\pi^a = \pi^* \left\{ \sum_m SR_m^*/\theta_m + \sum_j SC_j^* \right\} \equiv \pi^* H, \tag{42}$$

when $H = \left\{ \sum_m SR_m^*/\theta_m + \sum_j SC_j^* \right\}$, $\frac{\partial \ln \pi^*}{\partial \ln p_m^*} = SR_m^*$, and $\frac{\partial \ln \pi^*}{\partial \ln w_j} = -SC_j^*$. The shadow shares SC_m^* and SC_j^* are not observed. These shadow shares are related to the actual (observed) shares as follows:

$$SR_m^a = \frac{p_m y_m}{\pi^a} = \frac{SR_m^*}{H} \left\{ \frac{1}{\theta_m} \right\} \tag{43}$$

$$SC_j^a = \frac{w_j x_j}{\pi^a} = -\frac{SC_j^*}{H}.$$

The above model can be estimated using either the variable profit function in (42) along with the associated share equations in (43), or using the share equations strictly by themselves. Before proceeding we need to specify the behavior of the markup factors (θ_m) over time. First, θ_m may be specified as

$$\theta_m = \exp(b_m + c_m t), \quad m = 1, \ldots, M \tag{44}$$

where b_m and c_m are parameters to be estimated. The exponential function is often chosen to guarantee $\theta_m \geq 0$. Depending on the signs of b_m and c_m, θ_m can decrease or increase over time. We are primarily interested in testing whether the θ s approach unity over time.

The system of equations in (42) and (43) can be operational only when a parametric form of the shadow variable profit function $\ln \pi^*$ is assumed. To minimize *a priori* restrictions on the underlying production technology, we use a translog form of the shadow variable profit function, $\ln \pi^*$, and write it as

$$\begin{aligned}
\ln \pi^* = {} & \alpha_O + \sum_j \alpha_j \ln w_j + \sum_q \beta_q \ln z_q + \alpha_t t + \frac{1}{2}\Bigg\{ \sum\sum \alpha_{jk} \ln w_j \ln w_k \\
& + \sum_q \sum_l \beta_{ql} \ln z_q \ln z_l + \sum_m \sum_n a_{mn} \ln p_m^* \ln p_n^* + \alpha_{tt} t^2 \Bigg\} + \sum_j \sum_q b_{jq} \ln w_j \ln z_q \\
& + \sum_j \sum_m c_{jm} \ln w_j \ln p_m^* + \sum_m \sum_q d_{mq} \ln p_m^* \ln z_q + \sum_j \alpha_{jt} \ln w_j t \\
& + \sum_m a_{mt} \ln p_m^* t + \sum_q b_{qt} \ln z_q t
\end{aligned} \tag{45}$$

The above shadow profit function is assumed to satisfy the symmetry and convexity conditions. Furthermore, it is homogeneous of degree one in input and shadow output prices. The symmetry and homogeneity restrictions are imposed in estimating the parameters of the model. The symmetry restrictions on (45) are

$$\alpha_{jk} = \alpha_{kj}, \;\; \beta_{ql} = \beta_{lq}, \;\; a_{mn} = a_{nm},$$

and the homogeneity restrictions are

$$\begin{aligned}
& \sum_j \alpha_j + \sum_m a_m = 1, \sum_k \alpha_{jk} + \sum_m c_{jm} = 0 \;\forall\; j, \;\; \sum_n a_{mn} + \sum_j c_{jm} = 0 \;\forall\; m \\
& \sum_j b_{jq} + \sum_m d_{mq} = 0 \;\forall\; q, \;\; \sum_m a_{mt} + \sum_j \alpha_{jt} = 0 \;\forall\; t.
\end{aligned}$$

Normalizing the shadow profit function in (45) will impose the homogeneity restrictions with respect to one price. Since the objective is to estimate markups for both outputs, we use the J^{th} input price to normalize the other prices. Then we calculate SR_m^* and for $m = 1, 2, \ldots, M$, and SC_j^* for $j = 1, 2, \ldots, J-1$.

From the shadow profit function in (45), SR_m^* and SC_j^* can be derived as

$$SR_m^* = \frac{\partial \ln \pi^*}{\partial \ln p_m^*} = a_m + \sum_n a_{mn} \ln p_n^* + \sum_j c_{jm} \ln w_j + \sum_q d_{mq} \ln z_q + a_{mt} t, \quad (46)$$

and

$$-SC_j^* = \frac{\partial \ln \pi^*}{\partial \ln w_j} = \alpha_j + \sum_k \alpha_{jk} \ln w_k + \sum_q b_{jq} \ln z_q + \sum_m c_{jm} \ln p_m^* + \alpha_{jt} t. \quad (47)$$

Finally, using (46) and (47) above, $H = \{\Sigma_m SR_m^*/\theta_m + \Sigma_j SC_j^*\}$ can be expressed in terms of the unknown parameters and observed data. Consequently, the profit system (after adding classical error terms to each equation) becomes

$$\begin{aligned} \ln \pi^a &= \ln \pi^* + \ln H + v \\ SR_m^a &= \frac{p_m y_m}{\pi^a} = \frac{SR_m^*}{H} \left\{ \frac{1}{\theta_m} \right\} + v_m \\ SC_j^a &= \frac{w_j x_j}{\pi^a} = -\frac{SC_j^*}{H} + \varepsilon_j \end{aligned} \quad (48)$$

which can be estimated using the iterative non-linear seemingly unrelated regression technique. One of the share equations has to be dropped because the shares sum to unity.[17]

It should be noted here that the above approach does not recognize endogeneticity of output prices. If one assumes that producers have monopoly power in the output market, then w have to add the demand (or inverse demand) functions to take care of the endogeneticity of output prices. Adding M such inverse demand functions (one for each output price) will make the above system consistent in the sense that the number of equations equals the number of endogenous variables $(x, \ y, \ p)$.

One endemic problem in estimating profit functions (standard or alternative), using a translog or Cobb-Douglas functional form, is that profit has to be positive. This is, however, not the case in reality. In many studies, especially in

[17] Productivity decomposition formula for this problem is left to the readers.

banking, negative profits force researchers to use other functional forms that allow negative profits. A majority of the banking studies, however, avoid this problem by adding a positive number large enough to make profit for every bank positive (see, for example, Berger and Mester (1997, 2003), Berger and DeYoung (2001), and many others). This is clearly inadmissible, both economically and econometrically. Profit used in the profit function should be defined as revenue minus cost (which may be different from accounting profit). So any change in profit has to be reflected in the profit function. That is, if we assume no inefficiency and start from the profit function $\pi^a = \pi(p, w, t)$ then $\pi^a + c = \pi(p, w, t) + c$ and, therefore, $\ln(\pi^a + c) = \ln\{\pi(p, w, t) + c\}$. However, what is used in practice is $\ln(\pi^a + c) = \ln \pi(p, w, t)$, i.e., no adjustment is made on the right-hand side of the equation. This *ad hoc* procedure cannot be economically justified. Adding something to the dependent variable and then taking the log of it changes the intercept as well as the coefficients of all the right-hand side variables (regressors) in the profit function. Thus, the estimates of technological parameters can easily be manipulated (and therefore, estimates of returns to scale, input substitutability, technical change, etc.) simply by changing the positive constant that is added to profit. Since the alternative profit function approach uses the same procedure to avoid taking log of negative profits, it is subject to the same criticism.

6. SOME NEW ISSUES AND CHALLENGES

6.1. Competitiveness, Deregulation and Efficiency

Economists since Adam Smith have argued in favor of the virtues of competitive markets. The competitive framework is used as a benchmark because it leads to socially efficient outcomes. Any departures from competitive input and/or output markets lead to the inefficient allocation of resources and production of outputs, resulting in a deadweight loss to society. There might be many reasons why firms in some markets may not be competitive. For example, firms might posses some degree of market power in selling their products and/or buying their inputs. Consequently, the main driving force behind deregulatory effort is to increase competition in the hope of reducing the deadweight loss to the society. Here we examine the competitive and efficiency issues in the context of banking.

Like many other countries, the banking industry in Europe has faced important changes. These changes were specifically designed to liberalize the provision of services, bringing increased competition to the banking industry. Additionally, the establishment of the economic and monetary union (EMU), along with developments in information technology and removal of entry barriers are other important changes faced by the European banking markets. Many of these changes have important implications for the competitive structure of the banking and financial sectors. Some of the techniques proposed in this survey can be used to examine whether output markets are competitive or not, as well as the contribution of non-competitive output markets on TFP growth. We can also estimate the degree and the temporal behavior of non-competitiveness from the estimates of markup factors. In a cross-country study the information on markups may be useful to the policymakers (at the EU level), in the sense that they can examine whether banks from the new EU member (or to-be EU member) countries can survive if such markups are eliminated due to competition. More work is needed in this area.

Given that the banking system plays a unique role in an economy, the efficiency of that system can hardly be ignored. Studying the efficiency of the individual banks, as well as the banking system, can provide feedback on ways to improve policy-making decisions. Models of efficiency studies discussed in this survey can answer many questions, some of which are: Does regulation increase efficiency of banks? Are *de novo* banks more efficient than existing banks? Can banks from the countries joining the EU survive in competition with the foreign banks, if left without subsidy or other assistance? Can efficiency and/or scale economies explain bank mergers? These questions can be addressed using micro data from the EU countries. One can also use the aggregate industry level data from the EU member countries to examine differences in efficiency, productivity, technology, and competitiveness of the banking industry. If the industry is not competitive and one erroneously makes the assumption that it is competitive, efficiency and productivity measures are likely to be biased (especially when an econometric technique is used).

If one uses value added per worker as the measure of productivity and the sole indicator of performance, the measure cannot separate the contribution of competitiveness to productivity improvement. That is, if one compares value added per worker from two different industries, the difference would not tell us anything about the contribution of competitiveness, even if we know that one industry is more competitive than the other. This is because difference in value added per worker captures the effect of many other things such as difference in technology, size and scale differences, price differences, etc. Contribution of these factors to the overall productivity growth (whether using a partial or total factor productivity index) cannot be separated without estimating the technology directly or indirectly. For example, if cost-minimizing behavior is used, the contribution of competitiveness on *TFP* growth can be computed once the cost function is estimated. The benchmark is the competitive output markets. The main advantage of the cost function approach is that we do not have to assume a particular form of market structure to estimate this effect. It can also handle multiple outputs, and can disentangle the effect of non-competitive behavior in each output market. This is especially useful if the industries under investigation produce more than one output and the degree of competition varies across outputs.

The cost function approach cannot test any hypotheses about competitive output markets. For this, a profit function approach has to be used that can allow distortions in output prices due to regulation, non-competitive behavior, etc. Thus, if we want to model price distortion (and decomposes its effect on productivity growth), it is necessary to use a profit function approach.

Another challenge is to incorporate quality differences in outputs that can be confounded with non-competitive behavior, because quality difference is reflected in the prices. If output qualities differ across firms and these are not taken into account, deviations in the $p = MC$ rule might be thought of as non-competitive behavior. Similar to quality differences, there might be heterogeneity in products, behaviors, etc., among firms/industries operating in different countries. Such differences are to be taken into account. Furthermore, cross-country data may not be compared directly, unless the outputs, inputs, behavior, etc., are homogeneous. Another complication is that efficiency/performance of firms is likely to change over time and a model should be used that allows for his possibility.

Multi-dimensional indices of productivity and efficiency measures, such as labor productivity, employment growth, competitiveness, globalization, foreign investment, etc., are often constructed to examine performance. Some firms may be champions based on, for example, one or some of these criteria. This does not, however, mean that the same firms will be champions when the other criteria are used. Thus, an overall productivity index has to be constructed from these individual indices to make a meaningful comparison. That is, to get a macro outlook, we have to dig down in to the microanalysis.

In all the approaches discussed so far, the effect of competition or lack of it is examined from the producers' point of view. However, it is well-known that non-competitive markets contribute to deadweight loss to the society meaning that the society would be better off under competitive markets. By focusing only on the production side, we fail to see the compete picture and miss the deadweight loss arising out of, for example, monopolistic markets. The deadweight loss can be measured from the markups in output prices and difference in the outputs produced with and without competitive market conditions. Both the output differential and markups can be obtained from the estimates of shadow profit function. No one needs to look into the deadweight loss component in TFP growth decomposition.

6.2. Data Requirement for Improving the Efficiency and Productivity Measures

To address the issues mentioned above, for example in banking, extensive data is required, some of which are not reported in the balance sheets and income statements. For a meaningful efficiency analysis, for example, data is needed on different types of loans (consumer loans, business loans, real estate loans, etc.), different types of deposits, assets, capital, foreign exchange reserves, labor, wages and salaries, interest paid on different types of deposits, and rates charged on different types of loans. Quality measure of loans might be important for disentangling efficiency and productivity measures from output quality. Variables on quality of inputs, R&D expenditure, expenditure on information technology, etc., are important and useful in disentangling productivity differential. Depth of the study will depend on the type and extent of data made available to do the analysis.

In this survey we discussed several methods, some of which can be used to crosscheck results from competing models. Data requirements of these models also vary, and so does behavioral assumptions. For example, the distance functions require data on input and output quantities, while estimation of profit functions require data on input and output prices as well as profit. While discussing alternative profit functions in Section 5.3, we mentioned that profitability might be affected through revenue and cost. For example, loan quality and non-interest net income affect revenue, and cost inefficiency increases cost. Thus, excluding quality of loans from the analysis is likely to mis-measure revenue, which is likely to bias the results of the profit function approaches. If information on the quality of output variables is obtained and these quality variables are used in the cost function, estimates of the efficiency results would be much more reliable. If the quality variables are not included (due to non-availability of data), estimates of efficiency scores are likely to be contaminated. High efficiency score might be due to bad output quality and vice versa.

To put some of these in a broader perspective, consider some of the simpler non-econometric performance measures that are widely used in practice. For example, partial factor productivity, such as value added per worker, is often used as an indicator of performance. Another indicator is employment growth, although comparison between growth rates of value added and employment across industries, sectors, regions, etc., might not give similar performance measures. For example, value added per worker can increase without increasing employment growth if workers become more productive by using more capital (substituting labor with capital) and/or better technology. In this respect, growth in value added per worker is a better measure than growth in employment. However, even with this simple measure, we would like to decompose growth in value added per worker into returns to scale and technical change components. To do so, we have to estimate the underlying technology using a primal or dual approach. In estimating the technology and using it to measure growth rates of value added per worker, we can control for employment growth (indirectly in the regression equation), instead of using it as a separate measure of productivity. Again, we can use the metafrontier approach to explain differences in productivity growth across regions, industries, and over time. The advantage of the metafrontier approach, as mentioned before, is that we can examine the role of differences in technology (labeled as technology gap) and efficiency change in the overall productivity change. This will be helpful in examining whether the top performing firms (champions) are using the best practice technology (metafrontier) and are also technically efficient. For some industries/regions, the champion firms might be technically efficient and may also be using the best practice technology. But this may not be true for every sector/region. The typical growth accounting procedure (although useful and simple) cannot capture the sources of productivity growth differentials across regions, industries, size of firms in an industry, etc. Therefore, the simple value added approach cannot separate the effects of technical efficiency from the technology.

6.3. Use of Aggregated and Micro Data

Since the number of EU member countries is relatively small, we have to pool time series data on the member countries to estimate efficiency, productivity, etc., especially if we want to use the aggregated macro data. This puts a limit on allowing technological heterogeneity among EU member countries, especially in econometric models. If we assume a single technology for all member countries, the estimated inefficiency will be relative to the single frontier defined for all member countries. Thus, we cannot address the question of technology gap. That is, technology gap and inefficiency will be lumped together, yielding higher estimates of inefficiency. Furthermore, the estimated technology might be biased, if more than one technology is used in practice. This problem can be avoided if we use either a mixing model approach or somehow group the countries based on some *a priori* information, and estimate the technology for each group separately.

Another potential problem in using cross-country data is consistency of the data. That is, all the input and output variables, prices, and other control variables (if any) need to be defined in the same way. Sometimes the variables are indeed defined in the same way but certain components might be missing. This is hardly the case in reality. For example, data on loans might have ten classifications in one country but for another country there might be five categories. Thus, the analyst has to know whether similar procedures are used in constructing the variables that are used to estimate the technology. This problem is avoided if we use bank level data and estimate a separate frontier function for each country.

The data definition/construction problem mentioned above is absent when we uses cross-sectional (panel) data on banks from each country because the same procedure is used to construct the variables for each bank. Thus, we can estimate the frontier function for each country and measure efficiency and productivity relative to the estimated frontier. However, sometimes we want to compare bank efficiency across countries. Since each country has a frontier of its own, for a valid comparison of efficiency we have to take into account the fact that the technologies are different. Spanish banks might be efficient relative to their own frontier, but they might not be using the best practice technology. Thus, there might be a technology gap. To estimate the technology gap, we have to construct a metafrontier using the country-specific frontiers. The deviation of country frontiers from the metafrontier is then viewed as a technology gap. Thus, for example, Swiss banks might be technically efficient (relative to their own frontier) but when the technology gap is taken into account, they may not be as so efficient. That is, if a technology gap exits, efficiency of Swiss banks relative to the metafrontier (global frontier) would not be as efficient. In other words, to get a full picture we have to take into account both the country-specific frontiers and the metafrontier. This would be a better indicator of bank performance than that which is currently used in the banking literature.

While working with aggregate country level data, we often look at things beyond the traditional production/cost/profit function approach. For example, a new literature on productivity measurement (productivity and the new economy (Nordhouse, 2001)) has recently emerged. Following this approach, variables can be included, such as technological capacity, human capital, financial capacity, enterprise and innovation, openness, adaptability, etc. to enrich the productivity measures. These variables (for which indices are to be constructed) can affect productivity either indirectly by enhancing efficiency of traditional inputs such as capital and labor, or directly through improving efficiency, i.e., increasing outputs holding inputs constant.

7. CONCLUSIONS

In this survey we focused on the parametric models to estimate and decompose TFP growth into scale, technical change, and technical efficiency change components. For modeling inefficiency we concentrated on output-oriented technical inefficiency. The TFP growth formulas for input-oriented technical inefficiency (for some selected models) are given in the appendix. Throughout the analysis, we assumed that firms are allocatively efficient to make things simple. For modeling technical change we used both time trend and factor-augmenting approaches. In the primal (quantity based) models we discussed the single output production function approach, as well as input and output distance functions (that accommodates multiple outputs). To allow for the possibility that sample firms might use more than one technology and the analysts might not know who is using which technology, we considered the latent class modeling approach. In addition to estimating technical inefficiency for each firm, this approach is capable of measuring technology gap (distance between the individual/group frontiers from the metafrontier).

In the dual cost function approach, our TFP decomposition analysis was done for single and multiple output cost functions. We briefly discussed the modeling issues for the latent class models. Finally, we considered the profit function approach in both single and multiple output frameworks. We also examined the alternative profit function and discuss problems associated with this approach. Finally, we discussed extensions of the standard profit function to accommodate markups in the output markets.

Estimation issues are not addressed in this survey. Since one standard technique would not fit all the models, and some of the techniques (especially when the cost and profit function models that require use of a system approach) are quite involved, we decided not to discuss estimation techniques in this survey. However, econometric techniques are available to estimate every model discussed in this survey.

BIBLIOGRAPHY

Battese, G., D.S.P. Rao, and C.J. O'Donnell (2004) A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis* **21**: 91–103.

Beard, T., S. Caudill, and D. Gropper (1991) Finite mixture estimation of multiproduct cost functions. *Review of Economics and Statistics* **73**: 654–64.

Beard, T., S. Caudill, and D. Gropper (1997) The diffusion of production processes in the US. banking industry: A finite mixture approach. *Journal of Banking and Finance* **21**: 721–40.

Beckmann, M.J. and R. Sato (1969) Aggregate Production Functions and Types of Technical Progress: A Statistical Analysis. *American Economic Review* **59**: 88–101.

Berger, A.N and R. DeYoung (2001) The effects of geographic expansion on bank efficiency. *Journal of Financial Services Research* **19**: 163–94.

Berger, A.N., L.J. Mester (1997) Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of Banking and Finance* **21**: 895–947.

Berger, A.N. and L.J. Mester (2003) Explaining the dramatic changes in performance of US banks: Technological change, deregulation, and dynamic changes in competition. *Journal of Financial Intermediation* **12**: 57–95.

Brummer, B., T. Glauben, and G. Thijssen (2002) Decomposition of productivity growth using distance functions: The case of dairy farms in three European countries. *American Journal of Agricultural Economics* **84**: 628–44.

Caudill, S. (2003) Estimating a mixture of stochastic frontier regression models via the EM algorithm: A multiproduct cost function application. *Empirical Economics* **28**: 581–98.

Dempster, A.P., N.M. Laird, and D.B. Rubin (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society*, Series B **39**: 1–38.

Denny, M., M. Fuss, and L. Waverman (1981) The measurement and interpretation of total factor productivity in regulated industries: An application to Canadian telecommunications. in *Productivity Measurement in Regulated Industries*, edited by Cowing, T. and R. Stevenson. NY: Academic Press.

DeYoung, Robert and I. Hasan (1998) Performance of de novo commercial banks: A profit efficiency approach. *Journal of Banking and Finance* **22**: 565–87.

Diewert, W.E. (1992) The measurement of productivity. *Bulletin of Economic Research* **44(3)**: 163–98.

Fare, R. and D. Primont (1995) *Multi-Output Production and Duality: Theory and Applications*. Boston: Kluwer Academic Publishers.

Fisher, I. (1922) *The Making of Index Numbers*. Boston: Houghton Mifflin.

Grifell, E. and Lovell, C.A.K. (1997) The sources of productivity change in Spanish banking, *European Journal of Operational Research* **98**: 364–80.

Hagenaars, J.A. and McCutcheon, A.L. (2002) *Applied Latent Class Analysis*. New York, NY: Cambridge University Analysis.

Hulten, C.R. (2000) Total factor productivity: A short biography and new directions in productivity analysis, in Hulten, C.R., Dean, E.R. and Harper, M.J. (eds), *Studies in Income and Wealth*. Chicago: the University of Chicago Press for the National Bureau of Economic Research.

Humphrey, D.B. and L.B. Pulley (1997) Banks' responses to deregulation: Profits, technology and efficiency. *Journal of Money, Credit and Banking* **29**: 73–93.

Jondrow, J., K. Lovell, I. Materov, and P. Schmidt (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* **19**: 233–8.

Karagiannis, G. (2000) Total factor productivity growth and technical change in a profit function framework. *Journal of Productivity Analysis* **14**: 31–51.

Karagiannis, G., P. Midmore, and V. Tzouvelekas (2004) Parametric decomposition of output growth using a stochastic input distance function. *American Journal of Agricultural Economics* **86**: 1044–57.

Kendrick, J.W. (1984) *Improving Company Productivity: Handbook with Case Studies*. Baltimore, MD: Johns Hopkins University Press.

Kolari, J. and A. Zardkoohi (1995) Economies of scale and scope in commercial banks with different output mixes. Working Paper, College Station, Texas: Texas A&M University.

Kumbhakar, S.C. (1996) Efficiency measurement with multiple outputs and multiple inputs. *Journal of Productivity Analysis* **7**: 225–56.

Kumbhakar, S.C. (2000) Estimation and decomposition of productivity change when production is not efficient: A panel data approach. *Econometric Reviews* **19**: 425–60.

Kumbhakar, S.C. (2001) Estimation of profit functions when profit is not maximum, *American Journal of Agricultural Economics* **83**: 1–19.

Kumbhakar, S.C. (2002a) Decomposition of technical change into input-specific components: A factor augmenting approach. *Japan and the World Economy* **14**: 243–64.

Kumbhakar, S.C. (2002b) Productivity measurement: A profit function approach. *Applied Economics Letters* **9**: 331–4.

Kumbhakar, S.C. (2003) Factor productivity and technical change. *Applied Economics Letters* **10**: 291–97.

Kumbhakar, S.C. and C.A.K. Lovell (2000) *Stochastic Frontier Analysis*. New York: Cambridge University Press.

Kumbhakar, S.C. and A. Lozano-Vivas (2004a) Does deregulation make markets more competitive? Evidence of mark-ups in Spanish savings banks. *Applied Financial Economics* **14**: 507–15.

Kumbhakar, S.C. and A. Lozano-Vivas (2004b) Deregulation and productivity: The case of Spanish banks. *Journal of Regulatory Economics* (forthcoming).

Kumbhakar, S.C. and E.G. Tsionas (2005) Estimation of stochastic frontier production functions with input-oriented technical inefficiency. *Journal of Econometrics* (forthcoming).

Kumbhakar, S.C and H.-J. Wang (2005) Estimation of technical and allocative inefficiency in a stochastic frontier production model: A system approach. *Journal of Econometrics* (forthcoming).

Lau, L.J. (1972) Profit functions of technologies with multiple inputs and outputs. *Review of Economics and Statistics* **54**: 281–89.

Lee, Y. and P. Schmidt, (1993) A production frontier model with flexible temporal variation in technical efficiency, in Fried, H., Lovell, C.A.K. and Schmidt, S. (eds). *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford: Oxford University Press, pp. 3–67.

Malmquist, S. (1953) Index numbers and indifference surfaces. *Trabajos de Estadistica* **4**: 209–42.

McLachlan, G.J. and D. Peel (2000) *Finite Mixture Models*. New York: John Wiley & Sons.

Mester, L. (1997), Measuring efficiency at US banks: Accounting for heterogeneity is important. *European Journal of Operational Research* **98**: 230–424.

Nishimizu, M. and J.M. Page Jr. (1982) Total factor productivity growth, technological progress and technical efficiency change: Dimensions of productivity change in Yugoslavia, 1965–78. *Economic Journal* **92**: 920–36.

Nordhouse, W. (2001) Productivity and the new economy, NBER working paper.

Orea, L. (2002) Parametric decomposition of a generalized Malmquist productivity index. *Journal of Productivity Analysis* **18**: 5–22.

Orea, L. and S.C. Kumbhakar (2004) Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics* **29**: 169–83.

Sato, R. and M.J. Beckmann (1968) Neutral inventions and production functions. *Review of Economic Studies* **35**: 57–66.

Shephard, R.W. (1953) *Cost and Production Functions*. Princeton, NJ: Princeton University Press.

Shephard, R.W. (1970) *The Theory of Cost and Production Function*. Princeton, NJ: Princeton University Press.

Törnqvist, L. (1936) The Bank of Finland's consumption price index. *Bank of Finland Monthly Bulletin* **10**: 1–8.

Appendix A

TFP growth formula with input-oriented technical inefficiency

Write the production function as

$$
\begin{aligned}
& y = f(\hat{x}, t) \equiv f(x \exp(-\eta), t) \\
& \Rightarrow \dot{y} = \frac{1}{f} \sum_j f_j \hat{x}_j \left(\dot{x}_j - \frac{d\eta}{dt} \right) + \frac{\partial \ln f}{\partial t} \qquad \text{(A.1)} \\
& \Rightarrow T\dot{F}P = (RTS - 1) \sum_j \lambda_j \dot{x}_j + \sum_j (\lambda_j - s_j^a) \dot{x}_j - RTS \sum_j \lambda_j \frac{\partial \eta}{\partial t} + \dot{f}_t
\end{aligned}
$$

where $\hat{x}_j = x_j \exp(-\eta)$, $RTS = \sum_j \frac{\partial \ln f}{\partial \ln x_j} = \frac{1}{f} \sum f_j x_j e^{-\eta}$, $\lambda_j = f_j x_j / \sum_k f_k x_k$, and the dot over a variable represents its rate of change.

Since RTS depends on η (so is λ_j), the above formula can be rewritten as

$$
T\dot{F}P = (RTS^0 - 1) \sum \lambda_j^0 \dot{x}_j + \dot{f}_t^0 - \frac{d\eta}{dt} + \text{misc} \qquad \text{(A.2)}
$$

where RTS^0 $(= RTS | \eta = 0)$, TC^0 $(= TC | \eta = 0)$ and $\dot{f}_t^0$ $(= \dot{f}_t | \eta = 0)$. Finally, the miscellaneous component can be further decomposed into deviations of RTS from RTS^0, TC from TC^0, etc.

Note: The OO and IO measures are identical if the production function is homogeneous. In other words, one is a constant multiple of the other and there is no difference in estimating these models.

Appendix B

TFP growth formula with output-oriented non-neutral technical inefficiency

The production function is

$$y = f(x,t)\exp(-u(z,t)) \text{ where } z \text{ does not include } x.$$

$$\Rightarrow \dot{y} = \frac{1}{f}\sum f_j x_j \dot{x}_j + \sum_k \frac{\partial \ln e^{-u}}{\partial \ln z_k}\dot{z}_k - \frac{\partial u}{\partial t} + \dot{f}_t$$

$$= RTS\sum_j \lambda_j \dot{x}_j + \sum_k I_k z_k - \frac{\partial u}{\partial t} + \dot{f}_t$$

where $\frac{\partial \ln e^{-u}}{\partial \ln z_k} = I_k$ is efficiency change induced by z_k

$$\text{Thus, } T\dot{F}P = (RTS-1)\sum_j \lambda_j \dot{x}_j + \dot{f}_t - \frac{\partial u}{\partial t} + \sum_k I_k \dot{z}_k + \sum_j \left(\lambda_j - s_j^a\right)\dot{x}_j \quad \text{(B.1)}$$

$= \text{Scale} + \text{TC} + \text{TEC} + \text{induced by the } z \text{ variables} + \text{price/allocative}$

Appendix C

TFP growth formula with input-oriented non-neutral technical inefficiency

The production function is:

$$y = f(\hat{x}, t) \equiv f(x \exp(-\eta(z, t), t)$$
$$\Rightarrow \dot{y} = \frac{RTS}{\sum_k f_k \hat{x}_k} \left[\sum_j f_j \hat{x}_j \left\{ \dot{x}_j + \sum_k I_k \dot{z}_k - \frac{d\eta}{dt} \right\} \right] + \dot{f}_t$$

where $RTS = \sum_j \dfrac{\partial \ln y}{\partial \ln x_j} = \frac{1}{f} \sum f_j \hat{x}_j$ (depends on η)

$$\text{Thus,}\ T\dot{F}P = (RTS - 1) \sum_j \lambda_j \dot{x}_j + \dot{f}_t + RTS \sum_k I_k\ \dot{z}_k \sum_j \lambda_j$$
$$- RTS \sum_j \lambda_j \partial\eta/\partial t + \sum_j (\lambda_j - s_j^a) \dot{x}_j \qquad \text{(C.1)}$$
$$= \text{Scale} + TC + \text{Induced by z} + TEC + \text{Price/Allocative}$$

Note: Each component of TFP growth depends on η. The formula can be rewritten so that the above components, except a residual component, are free from η. The residual term will contain η, which can be further decomposed into scale, TC, etc.

Appendix D

TFP growth formula with output-oriented technical inefficiency

Without allocative inefficiency

$$
\begin{aligned}
&\mathrm{Min}_x\ w'x \text{ s.t. } y = f(x,t)e^{-u} \\
&\Rightarrow C^a = C(w, ye^u) \\
&\Rightarrow T\dot{F}P = \dot{Y}(1-E_{cy}) - \dot{C}_t - E_{cy}\frac{\partial u}{\partial t}
\end{aligned} \tag{D.1}
$$

where $E_{cy} = \partial \ln C/\partial \ln y$.

With allocative inefficiency

Production function: $ye^u = f(x,t)$

$$
\text{The first-order conditions are}: \frac{f_j}{f_1} = \frac{w_j e^{\xi j}}{w_1} \equiv \frac{w_j^s}{w_1^s}\ (\xi_1 = 0)
$$

$$
\begin{aligned}
&\Rightarrow x_j = x_j(w^s, ye^u) \\
&\Rightarrow C^s = \sum(w_j^s x_j) = C^s(w^s, ye^u, t)
\end{aligned}
$$

Shephard's Lemma: $\frac{\partial \ln C^s}{\partial \ln w_j^s} = S_j^s = \frac{w_j^s x_j}{C^s}$.

$$
\text{Thus, } C^a = \sum w_j x_j = C^s \sum_j w_j S_j^s / w_j^s \equiv C^s \cdot G(\cdot)
$$

$$
\begin{aligned}
&\Rightarrow \ln C^a = \ln C^s + \ln G(\cdot). \\
&\Rightarrow T\dot{F}P = \dot{y}(1-E_{cy}^s) + \sum_j (S_j^a - S_j^s)\dot{w}_j - \dot{C}_t^s - E_{cy}^s \frac{\partial u}{\partial t} - \frac{\partial \ln G(\cdot)}{\partial t}
\end{aligned} \tag{D.2}
$$

where $E_{cy}^s = \partial \ln C^s/\partial \ln y$.

Appendix E

TFP growth formula with non-neutral output-oriented technical inefficiency

With Allocative Inefficiency

Production function: $y = f(x, t) \exp(-u(z, t))$

$$\text{The first-order conditions}: \; f_j/f_1 = w_j e^{\xi_j}/w_1 = w_j^s/w_1$$

$$\Rightarrow \ln C^a = \ln C^s \left(w^s, ye^{u(z,t)}, t\right) + \ln G(\cdot)$$

$$T\dot{F}P = \dot{y}(1 - E_{cy}^s) + \sum (S_j^a - S_j^s)\dot{w}_j - \dot{C}_t^s - E_{cy}^s \frac{\partial u}{\partial t} + \sum I_k \dot{z}_k E_{cy}^s - \frac{\partial \ln G}{\partial t} \quad \text{(E.1)}$$

where $I_k = \left\{ \frac{\partial \ln e^{-u(z,t)}}{\partial \ln z_k} \right\}$

Note: (i) As before, the above formula can be expressed in terms of scale, TC, TEC, etc., defined at the frontier ($u = 0$). (ii) Cost function with non-neutral IO technical inefficiency will be similar to the neutral case except for one extra term involving $(I_k \dot{z}_k)$. (iii) Every component depends on u.

Rewrite $T\dot{F}P$ as

$$T\dot{F}P = \dot{y}(1 - E_{cy}^0) - \dot{C}_t^0 - E_{cy}^0 \frac{\partial u}{\partial t} + \rho(\xi, w, u, y) \quad \text{(E.2)}$$

where the miscellaneous component, $\rho(\xi, w, u, y)$, can be further decomposed into deviations of technical change, RTS, TEC, etc., from their respective values at the frontier.

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 3

FIRM, MARKET AND STRATEGIC FACTORS IN VENTURE CAPITAL EXITS

Douglas J. Cumming and Jeffrey G. MacIntosh

Abstract

There are five principal types of exit in venture capital (VC) finance: IPOs, acquisitions, secondary sales, buybacks and write-offs. While prior work has typically focused on market and firm-specific factors in affecting exit outcomes, the role of strategy-specific factors – namely, pre-planned exit strategies versus unsolicited offers – has received scant attention. We build upon a prior dataset to test the importance of strategy relative to other factors that could impact exit outcomes, such as proxies for entrepreneurial firm quality, firm stage of development at time of VC investment, VC investment duration, VC fundraising and MSCI public market returns. Our hand-collected data (survey data collected in conjunction with the Canadian Venture Capital Association and Venture Economics) enables a richer analysis of exit than that which comparable industry data would be able to provide. We also provide a comparative analysis of Canada and the United States to illustrate the impact of legal and institutional constraints on exit strategies.

Keywords and Phrases: Venture Capital; Exits; IPOs; Acquisitions; Regulation

JEL Classification No.: G24, G28, G32, G38, K22

1. INTRODUCTION

Venture capitalists typically invest in young entrepreneurial companies over a period of 3–8 years, with a view to selling or "exiting" their investment for the purpose of capital gain. The ability to make a profitable exit lies at the heart of

VC investing (Sahlman, 1990; Bergmann and Hege, 1998; Gompers and Lerner, 1999). There are five principle types of venture capital (VC) exits:

1. *an initial public offering (IPO)*, in which a significant portion of the firm is sold into the public market;
2. *an acquisition exit*, in which the entire firm is bought by a third party (typically a strategic corporate acquiror);
3. *a secondary sale*, in which only the VC's shares are sold to a third party (again, typically a strategic acquiror);
4. *a buyback*, in which the VC's shares are repurchased by the entrepreneurial firm; and
5. *a write-off*, in which the VC walks away from the investment.

IPOs and acquisitions are relatively more "desirable" forms of exit and have been the subject of increasing attention in the academic literature (see e.g., Gompers and Lerner, 1999; Smith and Smith, 2000; Cumming and MacIntosh, 2003a,b). Other forms of exit such as secondary sales, buybacks and write-offs are relatively less profitable (Cumming and MacIntosh, 2003a,b; Cochrane, 2005), typically involve lower quality entrepreneurial firms, and have received far less attention in the literature (although see, e.g., Wright *et al.*, 2001, re management buyouts). In this chapter, we present empirical results relating to all five types of exit vehicle.

In addition, empirical investigations of venture capital investing have often been based on large datasets, compiled either by venture capital associations or industry trackers that include relatively few explanatory variables. In this chapter, we use the results of a survey sent to both American and Canadian venture capitalists (VCs) that provide us with a greater richness of explanatory variables. We are thus able to test a number of hypotheses bearing on the VC's choice of exit vehicle that do not appear elsewhere in the literature, including our own prior work on this topic. These include whether the form of exit was pre-planned from the outset of the investments, whether it was made in response to an unsolicited offer, and whether it was a response to extant market conditions. Our survey results also indicate (*inter alia*) the stage at which the VC's first investment was made, the duration of the VC's investment, and the industry of the investee firm. Moreover, because we have comparable data from two countries, we are able to draw comparisons between the results in the two countries.

Some of our more important findings are summarized as follows. We show that higher book values and higher market/book values increase the probability of an IPO exit. These factors assume greater economic significance in the United States than in Canada, which is consistent with the presence of institutional distortions in Canada. We also show that, in the United States, when investments are made at earlier stages in the entrepreneurial firm's development, the likelihood of an IPO exit increases. However, this does not seem to be the case in Canada, suggesting that U.S. VCs add more value to their investee firms than Canadian VCs. The evidence is also consistent with the view that U.S. VCs are more skilled at timing their IPOs than are Canadian VCs.

With regard to specific information that is not available in industry-wide samples, such as the Venture Economics database or the VentureOne database used in other studies (e.g., Das *et al.*, 2003; Cochrane, 2005), we show that preplanned strategies and unsolicited offers can impact the exit outcome. These particular effects are not only statistically significant, but also among the most economically significant variables that determine the exit outcome.

This chapter is organized as follows. Testable hypotheses are discussed in Section 2. Legal and Institutional differences between Canada and the United States are discussed in Section 3. Section 4 describes the data. Empirical tests are provided in Section 5. Section 6 concludes.

2. TESTABLE HYPOTHESES

Factors that can impact the means by which venture capitalists exit their investments may be grouped into four broad categories:

1. the information asymmetries between the entrepreneurial firm and the new owner(s);
2. the value-added assistance provided to the entrepreneurial firm;
3. market conditions; and
4. deal-specific strategy factors.

There are more specific factors within each category, as discussed below.

2.1. Information Asymmetries

The exiting VC will be motivated to secure the highest exit price possible for its investment. Its ability to do so will depend in part on the degree of informational asymmetry that arises between the entrepreneurial firm and its new owners (i.e. the purchasers of the VC's share). This is because information risk affects the discount rate that a purchaser will apply to the expectation of future cash flows. In particular, greater information risk will lead to a higher discount rate, adversely impacting the exit price. Moreover, as we have noted in previous work (MacIntosh, 1997; Cumming and MacIntosh, 2001, 2003a,b) different forms of exit are typically associated with different degrees of information asymmetry. In general, IPOs are associated with the greatest degree of information asymmetry. This is because the buyers (institutional and retail investors who subscribe to the offering) are subject to a free rider problem that creates an incentive to allow other investors to expend the effort to examine and price the offering. Moreover, many such investors lack the expertise to appropriately evaluate many technology-based venture-backed offerings. Thus, public investors must rely to a considerable extent on financial intermediaries such as underwriters and auditors to appropriately price an initial public offering (see, e.g., Pagano *et al*., 1998).

By contrast, an acquisition exit generally involves unrestricted access to firm information by a sophisticated buyer (typically a strategic corporate acquiror) who is not subject to free rider problems. Thus, acquisition exits substantially mitigate the information asymmetry problem, which can (*ceteris paribus*) lead to a higher price for an acquisition exit. Buybacks, in which the entrepreneur is the buyer,[1] essentially eliminate the information asymmetry problem, subject to a potential lack of sophistication on the part of the entrepreneur in evaluating the information in his or her possession. Secondary sales, in which the VC alone sells its shares, will sometimes provide the purchaser with inferior access to firm-specific information than will acquisition exits, although once again the purchaser will probably be a sophisticated party untainted by free rider problems (again, often a strategic corporate acquiror). Thus, a rank ordering of the five exit vehicles by the degree to which information asymmetries are most likely (*ceteris paribus*) to lead to information risk and hence price discounting are: IPOs, secondary sales, acquisitions, and then buybacks.[2]

We also hypothesize that the degree of information asymmetry will vary with the duration of the VC's investment in the entrepreneurial firm. That is, the longer the duration of the investment, the less pronounced the information asymmetry between the entrepreneurial firm and its new owners. This intuitively appealing conjecture was initially developed and tested by Megginson and Weiss (1991), who show that IPO underpricing is less pronounced the longer the duration of the VC's investment. We couple this with the observation that the marginal value of a reduction in information asymmetry will be greatest when information asymmetry is otherwise most likely to be pronounced. Thus, Hypothesis 1 is that the longer the duration of VC investment, the more likely IPOs (which have

---

[1] While the buyback may formally be effected by the entrepreneurial firm, the entrepreneur is the "real" purchase in the sense that the entrepreneur will own a large portion, if not all of the equity of the firm after the VC is cashed out.

[2] See MacIntosh (1997) and Cumming and MacIntosh (2001, 2003a,b).

comparatively high information asymmetry) will be selected relative to other exit outcomes.

2.2. Scale of Offering and Growth Potential

Relatively large entrepreneurial firms, and those with significant growth potential, are not only likely to command a high price in the market, but to have a relatively high probability of returning to the market for additional funds in the future. This makes an IPO offering particularly attractive, since the public market is the deepest capital pool from which the firm may draw. It is thus most likely to be able to digest the aggregate price of a large firm offering, in addition to *future* capital needs. In addition, firms with a significant current scale of operations are more likely to meet stock exchange listing requirements.

Acquisition exits will also be attractive for high value and high growth firms. Such exits are often effected by relatively large public firms with considerable financial resources, and hence the ability to meet present and future funding requirements. However, they are not as likely as public markets to be able to absorb the present and future funding needs of the largest entrepreneurial firms or those with the highest growth trajectories. Nor are they as likely to spread risk as efficiently as the public market, which is characterized by a large number of highly diversified investors.[3]

Because of entrepreneurial wealth constraints, buyback exits are likely to be restricted to relatively small-scale acquisitions. Moreover, as explored more fully in Cumming and MacIntosh (2003a), because buybacks create significant firm-level debt, they almost certainly limit future financing options for the entrepreneurial firm. Thus, buybacks are not suited to facilitating VC exit from either large or high growth entrepreneurial firms. Secondary sales do not supply new capital, but neither appear to enhance or diminish future financing possibilities, although unilateral VC exit may be a signal of low current value and/or growth opportunities.

We thus conjecture (Hypothesis 2) that, in comparison to small entrepreneurial firms, VCs will exit large entrepreneurial companies more frequently by IPOs. We also conjecture (Hypothesis 3) that firms with high growth potential (as

---

[3] In theory, an acquisition exit by a liquid public company will efficiently spread risk to its own public shareholders. However, managers of public companies are often under-diversified, and thus have private incentives to limit the commitment of capital to any single acquisition. In turn, these private costs will enhance the compensation demanded by the managers, thus affecting the acquiror's bottom line. Thus, increasing firm-specific risk is a cost to potential acquirors. See Coffee (1986).

proxied by a higher market/book value) are more likely to be exited by (from highest to lowest probability) via IPOs, acquisitions, secondary sales, buybacks and write-offs.

2.3. VC Value-Added and the Stage of First Investment

Gompers and Lerner (1999), Sapienza (1992), Sapienza *et al.* (1992), Sahlman (1990), among others, show that VCs are active value-added investors. When VCs invest in a firm in the earliest stages of its development, they add the greatest value to the firm, and cultivate a more highly trained and capable management cadre. This in turn raises the opportunity, at future stages of development, of replacing the existing management team. Because IPO exits often cede pre-exit management, a much more significant managerial roll than acquisition exits in the post-exit enterprise (Black and Gilson, 1998), we hypothesize (Hypothesis 4) that the earlier the stage of entrepreneurial firm development at first VC investment, the more likely the VC will exit via an IPO instead of an acquisition.

2.4. Market Conditions and Other Firm-Specific Factors

Gompers and Lerner (1999) show that favorable market conditions give rise to more frequent VC-backed IPOs. We test the role of market conditions (Hypothesis 5) through the use of two primary variables:

1. average MSCI public market returns in the year of exit; and
2. through the use of a survey variable, in which the VC stated that the reason for exit was due to “market conditions”.

In addition, we use dummy variables to control for different industry effects, which could be related to market factors that affect exit outcomes.

2.5. Deal-Specific Strategy Factors

Idiosyncratic factors associated with a particular transaction may give rise to different exit outcomes. Such idiosyncratic factors include, for example, the pre-exit configuration of ownership in the entrepreneurial firm (which will determine, *inter alia*, whether the VC or the entrepreneur controls the exit decision), whether the deal was syndicated, and whether the VC and the entrepreneur were able to forge an amicable working relationship, etc. A variety of such factors are enumerated in Busenitz *et al.* (2003), Manigart *et al.* (1996, 2002a,b) and Wright and Lockett (2004). Such idiosyncratic features are typically absent in commonly used datasets compiled by venture capital associations and statisticians around the world.[4]

Our data set, while of course not completely comprehensive on all factors, nevertheless allows us to account for two very important idiosyncratic variables that have not been included in any other exit study: whether the VC contemplated a specific exit strategy when buying the investment, and whether exit was prompted by an unsolicited offer from the purchaser of the VC's investment. In respect of the first, if there are cases (as our survey data suggests) in which the VC pre-plans a specific exit strategy from the outset of the investment, this clearly needs to be accounted for in any multivariate study of the factors that determine exit choice.[5] Thus, Hypothesis 6 is that the pre-planning of an exit strategy has an impact on actual exit strategy. In respect of the second, at any given point in time the VC will maintain (at least at some rudimentary level) some understanding of the vector of opportunity costs associated with maintaining a given investment (Cumming and MacIntosh, 2001). The vector of opportunity costs visualized by the VC will determine whether the VC continues the investment or divests. More specifically, if the opportunity costs exceed the anticipated gains, then exit will occur. Otherwise, the VC will stay invested.

Importantly, the vector of opportunity costs will depend on the likelihood, at any given time, that a high-valuing acquiror will make an offer to purchase either the VC's interest of the entire firm. Sale to an acquiror will take place when the acquiror is willing to pay more than the capitalized value of the investment to the VC. However, knowledge of this portion of the vector of opportunity costs will inevitably be based on imperfect information, since the degree to which there is a strategic fit between the entrepreneurial firm and any given potential acquiror will often depend on private information not available to the VC. The occurrence of an unsolicited offer to purchase the firm will frequently update the VC's visualization of the vector of opportunity costs, causing exit to occur in a manner and at a time that had not been anticipated. An unsolicited offer will often consist of an offer from a strategic acquiror to buy either the whole firm or just the VC's interest. It may also, however come from the entrepreneur. Thus, Hypothesis 7 is that the receipt of an unsolicited offer will raise the probability of acquisition, secondary sale and buyback exits.

---

[4] The most commonly used dataset to study VC exits is Venture Economics and VentureOne in the U.S.; see, e.g., Cochrane (2005) and Das *et al.* (2003).

[5] Further research along this dimension will undoubtedly explore those factors that lead a VC to contemplate a pre-planned exit strategy when embarking upon an investment.

We are able to control for these variables through the use of an innovative dataset described below in Section 4. As the data are derived from a sample of Canadian and U.S. VC exits, we begin by noting some legal and institutional differences between Canada and the United States that may condition the array of exit possibilities.

3. LEGAL AND INSTITUTIONAL DIFFERENCES BETWEEN CANADA AND THE UNITED STATES

Previous research has documented regulatory differences across Canada and the United States, and stressed the impact of such differences on small- and medium-sized enterprises (MacIntosh, 1994; see also Gillen, 1992, and Halpern, 1997, on Canadian regulation; see Levin, 1995, and Gompers and Lerner, 1999a, on U.S. regulation). This subsection only briefly highlights some of the more important differences (pertaining to securities regulation and government sponsorship of venture capital in Canada[6]) – as they pertain to the choice of an IPO versus an acquisition exit.

The important differences between Canada and the United States can be summarized as follows.[7]

Factors that Suggest a Lower Frequently of IPOs in Canada[8]

(1) VCs encounter greater difficulty in disposing of their investments following an IPO, owing to a more onerous hold period and escrow requirements than those applying in the United States.
(2) Canadian secondary markets are less liquid than those in the United States, making it more difficult and costly for a Canadian VC to exit its investment following an IPO.
(3) A lesser degree of underwriter specialization exists in Canada. This will raise the comparative cost of an IPO in Canada.

---

[6] By and large, taxation factors do not appear to color the relative selection of exit strategies in either Canada or the United States, nor the relative comparative selection of exit strategies between the two countries (Cumming and MacIntosh, 2003a).

[7] Taken together, these factors suggest that it is better to treat the Canadian and U.S. data as distinct sub-samples than to pool the data. Nonetheless, in our empirical tests below, we both segregate and pool the data.

[8] In prior work, we suggest that Canadian and U.S. markets are not sufficiently integrated that a U.S. IPO is a ready alternative for all Canadian firms. See Cumming and MacIntosh (2003a).

Factors that Suggest a Higher Frequently of IPOs in Canada

(1) The direct costs of an IPO (including underwriting commissions, legal and accounting costs, filing fees and printing costs) in Canada are lower. As reported by Schutt and Williams (2000), the direct costs of going public on the Toronto Stock Exchange (versus NASDAQ [versus NYSE]) as a per cent of total proceeds from January 1998 to September 1999 were approximately 12% (17%) for offerings of less than US$ 10 million, 9% (11%) [13%] for offerings between $US 10 million and $US 50 million, 8% (9%) [9%] for offerings between $US 50 million and $US 200 million, 5% (6%) [6%] for offerings between $US 200 million and $US 500 million, and 4% (4.5%) for offerings of more than $US 1 billion.

(2) Costs in terms of underpricing in Canada are lower. For the period January 1998 to September 1999, Schutt and Williams (2000) indicate that the average underpricing for the first day of trading was approximately 10% for the TSE, 40% for NASDAQ and 11% on NYSE. The weighted average underpricing was 5.8% on the TSE, 49.6% on the NASDAQ and 10.9% on the NYSE.

(3) The TSE, and other regional exchanges that existed during the period of time covered by specified less demanding listing criteria than did either the NYSE or the NASDAQ.

In short, there are factors that would suggest both a higher and a lower frequency of IPOs in Canada relative to the United States.

Fewer Strategic Acquirors in Canada

There are fewer strategic acquirors in Canada than in the United States, lowering the likelihood of an acquisition exit as compared to other forms of exit.[9]

Factors that Suggest Less Fit with any Theoretical Model Based on Purely Economic Factors

(1) In Canada, tax subsidization of "Labour Sponsored Venture Capital Corporations" (LSVCCs) has led LSVCCs to dominate the Canadian venture capital industry (Cumming and MacIntosh, 2006). Statutory constraints on LSVCC behaviour has distorted investment and exit behaviour, making it less likely that theoretical predictions based on purely economic factors will hold.

[9] In prior work, we suggest that Canadian and U.S. markets are not sufficiently integrated that the U.S. acquisition market is fully available to Canadian firms. See Cumming and MacIntosh (2003a).

(2) There is good evidence that LSVCC managers are significantly less skilled than the managers of private venture capital limited partnerships in both Canada and the United States, again introducing noise into the Canadian exits data and making it less likely that the theoretical model will hold (Cumming and MacIntosh, 2006).

It is also noteworthy that different restrictive covenants and other constraints imposed on venture capitalists may exist across countries. For example, in the United States, VC partnership agreements specify a number of restrictive covenants on the actions of general partners (VC managers) (Gompers, 1996). These restrictions include covenants relating to the management of the fund (e.g., the size of investment in any one firm, the use of debt, coinvestment, reinvestment of capital gains), covenants relating to the activities of general partners (e.g., coinvestment by general partners, sale of partnership interests, fundraising, the addition of other general partners), and covenants relating to the types of investment (e.g., investments in other venture funds, public securities, leveraged buyouts, foreign securities and other asset classes). In the United States, the frequency with which these restrictions are used changes over time and also with changes in economic conditions. These restrictions may differentially impact on VC exits as well as the comparative risk and return of venture capital activity in different countries. We thus provide both segregated Canadian and U.S. data (in the spirit of Black and Gilson, 1998; Jeng and Wells, 2000, and Armour and Cumming, 2006) and integrated data.

The data used to test these propositions and the comparative effects of regulation in the two countries is described in Section 4. Empirical tests follow in Section 5.

4. DATA

We make use of a survey data (collected with the assistance of Venture Economics in the United States and Macdonald & Associates in Canada). The data comprise 112 observations from Canada and the United States on IPO and acquisition exits between 1992 and 1995 (and 246 exits in total, including secondary sales, buybacks and write-offs). Other VC research using industry-wide data (e.g., Barry *et al*., 1990; Megginson and Weiss, 1991; Gompers and Lerner, 1999; Cochrane, 2005) invariably involves some sacrifice of detail in the data set. Our data are similar in scope to related VC research involving custom-generated data sets, which typically use between 50 and 200 observations (e.g., Gompers, 1997; Trester, 1998). Other noteworthy research on IPOs not specifically focused on VCs also employs data of similar scope (e.g. Pagano *et al.*, 1998, 69 observations). The data are summarized in Tables 3.1, 3.2 and 3.3.

The data for all exit types in the sample are presented in Tables 3.1 and 3.2. We present summary statistics for full and partial exits for each exit type.[10] On average, VC real returns (internal rates of return, or IRRs) are greater in the United States than in Canada.[11] The average annual real returns across all exit vehicles are low in our sample, given the relatively large frequency of write-offs in our data. In the United States, returns for IPOs (54.9%) and acquisitions (57.8%) are greater than those observed in Canada (27.8% for IPOs and 13.3% for acquisitions).

Tables 3.1 and 3.2 provide additional details regarding investment duration (time in years from the VC's first investment to exit), the reason for exit (pre-planned strategy, unsolicited offer, market conditions), and high-tech firms versus other firms (details on specific industries are provided in Table 3.3).

Additional summary statistics and univariate comparison tests for the combined sample of the Canadian and U.S. data are described in Table 3.3.[12] Consistent with our expectations (Hypotheses 1 and 2), Table 3.3 indicates that IPOs tend to have higher market values, book values and market/book values relative to the other exits (albeit some of the difference tests are not statistically significant, in light of the very high variability of returns). The univariate comparison tests indicate that when a VC first invests at an early stage in the entrepreneurial firm's development, investments are more often exited by an IPO, which is consistent with Hypothesis 4. Table 3.3 further indicates that a greater proportion of IPOs are for reasons of pre-planned strategies and market conditions than are other forms of exit. This result is not unexpected, given that a VC will presumably not embark upon an investment if it anticipates making an

[10] An exit may be full or partial (Cumming and MacIntosh, 2003b). A full exit for an IPO involves a sale of all of the venture capitalist's holdings within one year of the IPO; a partial exit involves sale of only part of the venture capitalist's holdings within that period. A full acquisition exit involves the sale of the entire firm for cash; in a partial acquisition exit, the venture capitalist receives (often illiquid) shares in the acquiror firm instead of cash. In the case of a secondary sale or a buyback exit (in which the entrepreneur or entrepreneurial firm buys out the venture capitalist), a partial exit entails a sale of only part of the venture capitalist's holdings. A partial write-off involves a write down of the investment.

[11] The data from the U.S. comprise private limited partnership funds. The data from Canada comprise both private funds and LSVCCs. The particular characteristics of LSVCCs are described in detail by Macdonald (1992), MacIntosh (1994), Halpern (1997), and Cumming and MacIntosh (2001, 2003a). The presence of LSVCCs in the Canadian data may account for the comparatively low returns.

[12] The data are aggregated across the Canadian and U.S. sample for the purpose of the comparison of means and proportion tests in order to provide a sufficient number of degrees of freedom to carry out comparison tests for each exit vehicle.

Table 3.1 U.S. Venture Capital Exits Data Summarized by Exit Vehicle*

	Number of	Stated Reasons for Exit			Average	Technology		Extent		Full Sample Including Partial Exits					Sample Excluding Partial Exits				
	Portfolio	Pre-	Market	Unsolicited	Duration	Industry		of Exit		Average	Average	Average Gross	Average Annual	Variance in	Average	Average	Average Gross	Average Annual	Variance in
Exit Vehicle	Companies	Planned	Conditions	Offer	(Years)	No	Yes	Partial	Full	Investment**	Exit Value**	Real Return(%)	Real Return(%)	Real Return(%)	Investment**	Exit Value**	Real Return(%)	Real Return(%)	Real Return(%)
IPO	30	17	8	0	4.70	12	18	8	22	2,035,036	12,565,880	464.64	54.92	51.15	2,052,934	13,058,260	506.16	44.49	14.93
Acquisition	30	8	9	7	5.17	9	21	6	24	1,720,349	3,859,077	143.04	57.83	754.75	1,640,617	3,017,889	119.87	67.26	943.07
Secondary Sale	9	4	4	0	6.33	2	7	3	6	519,931	1,005,871	54.88	–7.57	6.69	428,671	242,123	–42.39	–9.76	4.38
Buyback	6	4	0	2	4.00	5	1	5	1	784,397	2,687,449	145.04	24.79	3.27	2,634,352	10,265,360	289.67	40.50	0.00
Write-off	33	1	17	0	4.36	15	18	2	31	1,984,068	92,500	–97.85	–90.01	4.88	2,032,191	98,468	–97.71	–89.36	5.13
Other	4	0	0	1	2.75	2	2	1	3	1,112,445	1,539,990	35.28	34.02	83.94	1,355,079	1,712,773	–8.19	–9.87	10.36
Total	112	34	38	10	4.75	45	67	25	87	1,714,030	4,706,597	147.38	5.61	256.60	1,568,089	3,796,201	109.94	–2.58	304.84

* Source: Venture Economics

** Real U.S. Dollars (base year = 1990). CPI data source: International Financial Statistics, Label 11/64; available at www.chass.utoronto.ca.

Partial exit market values are adjusted to reflect full values. Real returns are calculated assuming investment at the beginning of the year, and exit at the end of the year, reflecting the lowest possible estimate.

Table 3.2 Canadian Venture Capital Exits Data Summarized by Exit Vehicle*

	Number of	Stated Reasons for Exit			Average	Technology		Extent		Full Sample Including Partial Exits					Sample Excluding Partial Exits				
	Portfolio	Pre-	Market	Unsolicited	Duration	Industry		of Exit		Average	Average	Average Gross	Average Annual	Variance in	Average	Average	Average Gross	Average Annual	Variance in
Exit Method	Companies	Planned	Conditions	Offer	(Years)	No	Yes	Partial	Full	Investment**	Exit Value**	Real Return(%)	Real Return(%)	Real Return(%)	Investment**	Exit Value**	Real Return(%)	Real Return(%)	Real Return(%)
IPO	36	16	18	2	5.86	3	33	20	16	1,464,087	5,170,185	1385.85	27.83	9.82	1,520,666	3,536,233	187.58	21.51	8.67
Acquisition	16	3	0	12	6.94	9	7	1	15	1,945,386	3,271,514	84.58	13.31	2.95	1,998,818	3,353,662	85.00	13.80	3.12
Secondary Sale	12	1	10	0	3.08	0	12	5	7	402,144	968,181	165.70	54.90	90.28	304,109	581,065	157.81	18.09	8.09
Buyback	41	15	2	14	6.34	30	11	7	34	668,245	808,686	66.97	3.82	1.50	729,096	879,255	41.72	2.57	1.52
Write-off	27	0	3	0	4.07	18	9	1	26	332,038	3,821	–97.10	–92.04	4.38	338,780	3,968	–96.99	–91.74	4.53
Other	2	0	0	1	6.00	2	0	1	1	2,412,731	3,687,627	60.15	9.53	0.17	2,766,431	3,492,890	42.17	12.44	0.00
Total	134	35	33	29	5.53	62	72	35	99	969,012	2,169,579	399.08	–3.25	33.87	937,431	1,459,018	43.64	–16.24	24.92

* Source: Canadian Venture Capital Association.

** Real Canadian Dollars (base year = 1990) converted to U.S. Dollars. CPI data source: CANSIM, Label P700000; available at www.chass.utoronto.ca. Foreign exchange rates from CANSIM, label B3400. Values expressed in U.S. dollars for comparative purposes only. Returns were computed in Canadian dollars and do not reflect exchange rate changes.

Partial exit market values are adjusted to reflect full values. Real returns are calculated assuming investment at the beginning of the year, and exit at the end of the year, reflecting the lowest possible estimate.

Table 3.3 Descriptive Statistics for the Canadian and U.S. Data (Exit Years: 1992–1995)

	Combined Canada–U.S. Sample All Exits	Combined Canada–U.S. Sample IPOs	Combined Canada–U.S. Sample Acquisitions	Combined Canada–U.S. Sample Secondary Sales	Combined Canada–U.S. Sample Buybacks	Combined Canada–U.S. Sample Write-offs	Difference Tests Between IPOs and Acquisitions	Difference Tests Between IPOs and Secondary Sales	Difference Tests Between IPOs and Buybacks	Difference Tests Between IPOs and Write-offs
Average Duration (years)	4.17	4.33	4.78	3.48	5.04	3.23	–0.75	1.28	–1.07	2.38**
Average Market/Book	3.84	10.67	2.23	2.18	1.77	0.02	1.45	1.46	1.53	1.84*
Average Market ('000)	3721.36	9514.32	3980.63	1190.27	1288.06	53.17	3.03***	4.74***	4.77***	5.57***
Average Book ('000)	1450.63	1914.29	2013.35	515.75	852.24	1276.34	–26.47	5.74***	4.18***	2.25**
# Seed Investments	33	9	3	6	2	10	1.57	0.54	1.95*	–0.47
# Start-up Investments	63	15	10	3	17	18	0.82	2.45**	–0.70	–0.93
# Early Stage Investments	40	15	12	1	3	9	0.36	3.05***	2.72***	1.10
# Expansion Investments	95	24	18	11	22	17	0.74	2.06***	–0.03	0.96
# Buyout, Turnaround Investments	15	3	3	0	3	6	–0.11	1.40	–0.11	–1.19
# Exits for Reasons of Market Conditions	71	26	9	14	2	20	2.95***	1.81*	4.68***	0.71
# Preplanned Exits	68	33	11	5	19	0	3.63***	4.71***	2.06**	6.38***
# Exits for Reasons of Unsolicited Offer	39	2	19	0	16	0	–4.11***	1.13	–3.62***	1.36
Average MSCI Index Return in Year of Exit	0.0066	0.0066	0.0058	0.0092	0.0065	0.0066	0.32	–0.75	0.07	0.00
Avg Industry Fundraising in Exit Year ($m)	2950.05	3024.69	2899.81	2837.7370	2937.4543	2961.0121	0.80	1.04	0.78	0.48
# in Multimedia Industries	14	4	2	3	3	1	0.68	0.22	0.25	1.26
# in Computers – Hardware	29	10	5	3	3	9	1.12	1.60	1.78*	0.02
# in Computers – Software	39	13	7	7	4	9	1.18	1.07	2.04**	0.69
# in Manufacturing	28	1	4	0	2	8	–1.40	0.80	–0.63	–2.57**
# in Industrial Produts	48	11	10	0	1	17	0.00	2.81***	2.73***	–1.57
# in Agriculture	9	3	1	0	13	3	0.87	1.40	–2.75***	–0.12
# in Medical	40	17	8	6	9	3	1.67*	2.00**	1.43	3.18***
# Other Industries	39	7	9	2	10	10	–0.70	1.34	–0.95	–0.99
Number of Observations	246	66	46	21	47	60	–	–	–	–

Difference of means and proportions tests: *, **, *** Significant at the 10%, 5% and 1% levels, respectively. Total number of observations includes a total of 6 other/unknown exits, as indicated in Tables 3.1 and 3.2.

inferior form of exit (such as a buyback or a secondary sale). Some differences across the various industries are also observed in Table 3.3.

5. MULTIVARIATE EMPIRICAL TESTS

Our multivariate empirical tests make use of the standard binomial logit model. Cumming and MacIntosh (2003a) use a multinomial logit model to consider all five exit vehicles simultaneously. However, that approach involves a number of significant limitations. Most importantly, most of the variables of potential interest cannot be included in a multinomial logit specification across each exit outcome, as such variables give rise to singular Hessian matrices. By examining in depth each exit vehicle separately in a series of binomial logit models, we are able to consider a much richer array of variables that are pertinent to the hypotheses developed above.

A correlation matrix for all of the variables that we consider in the multivariate analysis is provided in Table 3.4. Importantly, the estimates in the logit models (presented in Table 3.5) are not biased by collinearity across the different variables. Additional specifications are available upon request.

Our empirical estimates of the factors that affect the exit outcomes are provided in Table 3.5. The left-hand-side variable is a dummy variable equal to 1 if the exit was as indicated in the particular column, and 0 otherwise. The right-hand-side variables are as follows:

- *Duration*: the log of the time (in years) from first investment to time of exit;[13]
- *Market/Book*: the log of the value the VC receives upon exit divided by the cost of the investment;
- *Dummy variables for the stage of entrepreneurial firm development (from earliest to latest) at first investment*: i.e. seed, start-up, early-stage and expansion stages (note that dummy variables for later stage buyout and turnaround investments were suppressed to avoid perfect collinearity);
- *Dummy variables for particular factors affecting exit*: i.e. pre-planned exits, market conditions, and unsolicited offers (a dummy for "other reasons" was suppressed to avoid perfect collinearity);

---

[13] Note that natural logs are used for duration, book value, market/book, fundraising and 1+MSCI return in the specifications in order to account for nonlinearities that would be expected for these variables. For example, an increase in the book value by $US 1 million would have a more pronounced affect if the book value went from $1 million to $2 million as compared to a change from $100 million to $101 million. Note that we added one to the duration and market/book values to avoid taking logs of a few observations with the value of "0" (which is undefined).

Table 3.4 Correlation Matrix

	IPO	Acquisition	Secondary Sale	Buyback	Write–off	Log(Duration)	Log(Book Value)	Log(Market/Book)	Seed Stage	Start-up Stage	Other Early Stage	Expansion	Market Conditions Reason for Exit	Preplanned Exit	Unsolicited Offer	Log(Average MSCI Index in Year of Exit)	Log(Industry Fundraising in Year of Exit)	Multimedia Industry	Computer Hardware Industry	Computer Software Industry	Manufacturing Industry	Industrial Products Industry	Agricultural or Biotechnology Industry	Medical Industry	Canada Dummy Variable
IPO	1.00																								
Acquisition	**–0.29**	1.00																							
Secondary Sale	**–0.18**	**–0.15**	1.00																						
Buyback	**–0.29**	**–0.23**	**–0.15**	1.00																					
Write–off	**–0.34**	**–0.27**	**–0.17**	**–0.28**	1.00																				
Log(Duration)	0.04	0.10	–0.08	0.10	**–0.15**	1.00																			
Log(Book Value)	**0.21**	**0.22**	**–0.24**	**–0.26**	–0.04	0.06	1.00																		
Log(Market/Book)	**0.53**	0.06	0.02	–0.01	**–0.60**	**0.14**	0.00	1.00																	
Seed Stage	0.00	–0.10	**0.14**	**–0.13**	0.05	0.02	–0.11	–0.01	1.00																
Start-up Stage	–0.04	–0.04	–0.08	0.12	0.06	0.00	**–0.16**	–0.06	**–0.23**	1.00															
Other Early Stage	0.11	**0.13**	–0.10	**–0.13**	–0.02	0.05	0.09	0.02	**–0.17**	**–0.26**	1.00														
Expansion	–0.03	0.01	0.09	0.08	–0.12	–0.01	**0.21**	0.01	**–0.31**	**–0.47**	**–0.35**	1.00													
Market Conditions Reason for Exit	**0.14**	–0.10	**0.25**	**–0.26**	0.06	–0.05	0.07	–0.03	0.01	0.02	0.01	0.01	1.00												
Preplanned Exit	**0.30**	–0.04	–0.03	**0.13**	**–0.33**	0.12	0.07	**0.35**	–0.03	–0.03	–0.01	0.04	**–0.40**	1.00											
Unsolicited Offer	**–0.21**	**0.33**	**–0.13**	**0.24**	**–0.25**	0.11	0.07	0.05	**–0.14**	–0.03	**0.14**	0.04	**–0.28**	**–0.27**	1.00										
Log(Average MSCI Index in Year of Exit)	0.00	–0.03	0.05	0.00	0.00	0.10	0.05	–0.07	–0.05	–0.04	0.09	0.04	0.11	–0.04	0.00	1.00									
Log(Industry Fundraising in Year of Exit)	0.07	–0.06	–0.05	0.03	0.00	**0.15**	0.03	0.05	–0.05	–0.02	–0.02	0.05	0.01	0.04	0.05	**0.17**	1.00								
Multimedia Industry	0.01	–0.03	0.11	0.06	–0.10	–0.01	0.04	0.05	–0.05	–0.06	–0.06	**0.17**	0.08	0.00	–0.06	–0.08	0.01	1.00							
Computer Hardware Industry	0.06	–0.01	0.02	–0.11	0.06	–0.03	0.03	0.02	0.00	0.10	–0.02	–0.06	–0.01	0.00	–0.06	–0.02	–0.01	–0.09	1.00						
Computer Software Industry	0.06	–0.01	**0.15**	**–0.18**	–0.01	0.00	0.00	**0.18**	**0.19**	**–0.13**	**0.17**	**–0.16**	0.09	–0.05	–0.10	0.04	0.02	–0.11	**–0.16**	1.00					
Manufacturing Industry	**–0.19**	–0.04	–0.11	**0.25**	0.03	0.02	**–0.13**	–0.07	–0.10	0.05	–0.12	0.06	**–0.14**	–0.05	0.12	–0.05	–0.03	–0.09	**–0.13**	**–0.16**	1.00				
Industrial Products Industry	–0.04	0.03	**–0.15**	0.00	**0.13**	0.00	0.11	–0.10	–0.10	–0.03	–0.11	**0.16**	0.00	0.06	0.01	0.11	**0.14**	–0.12	**–0.18**	**–0.21**	**–0.18**	1.00			
Agricultural or Biotechnology Industry	0.03	–0.04	–0.06	0.02	0.04	0.02	–0.03	–0.08	–0.01	–0.06	0.03	0.02	–0.08	0.07	–0.03	–0.03	0.02	–0.05	–0.07	–0.08	–0.07	–0.10	1.00		
Medical Industry	**0.16**	0.01	0.10	–0.05	**–0.17**	–0.02	–0.03	0.06	–0.01	0.04	**0.19**	–0.12	0.08	0.07	–0.04	–0.01	0.01	–0.11	**–0.16**	**–0.19**	**–0.16**	**–0.22**	–0.09	1.00	
Canada Dummy Variable	0.00	**–0.19**	0.02	**0.32**	–0.11	0.09	**–0.29**	0.03	**–0.29**	**0.16**	**–0.15**	**0.16**	–0.10	–0.05	**0.17**	0.06	0.12	0.05	–0.05	–0.12	**0.20**	**–0.19**	0.09	0.03	1.00

Correlations are presented for the full sample of all 246 exits in Canada and the United States. Correlations significant at the 5% level are highlighted in bold and underline font

Table 3.5 Logit Estimates of the Likelihood of Different Exit Outcomes
Panel A. Full Sample Combining the Canadian and U.S. Data

Explanatory Variable	IPOs		Acquisitions		Secondary Sales		Buybacks		Write–offs	
	Coefficient	t–statistic	Coefficient	t–statistic	Coefficient	t–statistic	Coefficient	t–statistic	Coefficient	t–statistic
Constant	–11.20	–1.47	2.64	0.48	25.05	1.90	1.20	0.13	–1.31	–0.26
Log (Duration)	–0.38	–1.00	0.44	1.28	–0.65	–1.16	0.36	1.01	–0.53	–2.17**
Log (Book Value ('000))	0.95	4.25***	0.45	2.24**	–1.32	–4.17***	–0.70	–3.74***	–0.07	–0.52
Log (Market / Book)	2.92	6.21***	–0.08	–0.25	–0.80	–1.62	–0.52	–1.37	—	—
Seed Dummy	2.14	1.58	–2.06	–2.02**	2.79	1.98**	–0.73	–0.54	–0.01	–0.02
Start–up Dummy	1.17	0.94	–0.99	–1.12	1.57	1.02	0.51	0.54	–0.12	–0.18
Early Stage Dummy	2.18	1.66*	–1.14	–1.26	—	—	–1.18	–0.95	–0.20	–0.27
Expansion Dummy	0.49	0.41	–1.06	–1.29	3.48	2.46**	0.62	0.66	–0.81	–1.25
Market Conditions Dummy	1.43	1.81*	–0.17	–0.29	4.20	3.28***	–1.09	–1.17	0.50	1.42
Preplanned Exit Dummy	0.96	1.22	0.29	0.46	3.31	2.57**	2.31	3.23***	—	—
Unsolicited Offer Dummy	–2.97	–2.34**	2.72	3.83***	—	—	2.43	3.45***	—	—
Log (1+MSCI Return in Exit Year)	13.16	0.82	0.43	0.03	39.90	1.43	6.08	0.40	0.54	0.05
Log (Industry Fundraising in Exit Year ($m))	–0.18	–0.18	–0.80	–1.16	–3.30	–1.97**	–0.07	–0.06	0.23	0.37
Multimedia Dummy	–1.08	–1.02	–0.08	–0.08	4.05	2.87***	0.98	1.04	–1.38	–1.22
Computers – Hardware Dummy	0.25	0.27	–0.39	–0.50	3.24	2.51**	–1.70	–1.61	0.15	0.26
Computers – Software Dummy	–1.38	–1.50	0.05	0.07	4.32	3.36***	–1.51	–1.22	–0.42	–0.73
Manufacturing Dummy	–2.24	–1.65*	–0.61	–0.76	—	—	0.59	0.87	0.38	0.63
Industrial Products Dummy	0.33	0.39	–0.65	–0.95	—	—	0.06	0.09	0.48	0.89
Agriculture Dummy	0.51	0.42	0.003	0.002	—	—	–0.76	–0.63	0.58	0.70
Medical Dummy	0.14	0.18	0.22	0.31	3.09	2.56**	–0.21	–0.27	–1.64	–2.25**
Canadian Dummy	1.72	2.75***	–1.89	–3.51***	–0.42	–0.45	1.343	2.05**	–0.385	–0.98
Number of Observations	246		246		246		246		246	
Loglikelihood	–68.92		–89.23		–34.72		–74.10		–121.90	
Chi–square	148.29***		58.61***		74.07***		91.77***		29.53**	
	Predicted Outcomes		Predicted Outcomes		Predicted Outcomes		Predicted Outcomes		Predicted Outcomes	
Actual Outcomes	Other Exit	IPO	Other Exit	Acquisition	Other Exit	Secondary Sale	Other Exit	Buyback	Other Exit	Write–off
Dependent Variable = 0	165	15	193	7	221	4	189	10	181	5
Dependent Variable = 1	18	48	32	14	9	12	25	22	5	9

Dependent variable = 1 if exit outcome as indicated, and = 0 otherwise.
*, **, ***, Significant at the 10%, 5%, and 1% level, respectively.
"—": Variable exclude to avoid collinearity problems.

- *Control variables*: the log of the industry fundraising in the year of exit,[14] the log of one plus MSCI index returns in the year of exit, a dummy variable

[14] In the full sample estimates, we rescaled the fundraising variable so that the Canadian values had the same means as the U.S. values. This was done to avoid a problem of collinearity between the Canadian dummy and the (Canadian) fundraising dummy variables, which otherwise arises in view of the fact that industry fundraising in Canada is of course much lower than in the U.S. We did not use re-scaled fundraising values in the subsamples.

Table 3.5 Logit Estimates of the Likelihood of Different Exit Outcomes
Panel B. Subsample of Canadian Data Only

Explanatory Variable	IPOs Coefficient	IPOs t–statistic	Acquisitions Coefficient	Acquisitions t–statistic	Secondary Sales Coefficient	Secondary Sales t–statistic	Buybacks Coefficient	Buybacks t–statistic	Write–offs Coefficient	Write–offs t–statistic
Constant	4.71	0.38	–9.54	–1.31	–20.77	–1.19	2.98	0.61	0.39	0.08
Log (Duration)	3.29	2.51**	0.14	0.24	–2.99	–1.97**	0.38	0.94	–1.23	–2.76***
Log (Book Value ('000))	1.93	3.16***	0.57	1.80*	–1.71	–2.66***	–0.56	–2.73***	–0.20	–1.05
Log (Market / Book)	5.26	2.83***	0.13	0.20	–0.89	–0.86	–0.55	–1.28	–	–
Seed Dummy	–0.84	–0.02	—	—	4.01	0.73	–0.42	–0.24	–0.80	–0.48
Start–up Dummy	2.55	0.59	–0.20	–0.13	2.49	0.46	0.85	0.79	–0.74	–0.77
Early Stage Dummy	7.04	1.39	–0.56	–0.34	—	—	–0.69	–0.51	–1.40	–0.95
Expansion Dummy	2.34	0.54	0.35	0.22	1.93	0.38	0.67	0.61	–1.65	–1.62
Market Conditions Dummy	11.80	3.35***	—	—	6.24	1.94*	–0.90	–0.92	–1.41	–1.79*
Preplanned Exit Dummy	8.15	3.12***	1.49	1.17	3.92	1.22	2.15	2.56**	—	—
Unsolicited Offer Dummy	—	—	3.86	3.24***	—	—	1.95	2.67***	—	—
Log (1+MSCI Return in Exit Year)	47.80	1.34	9.55	0.33	48.51	0.87	15.36	0.90	–23.84	–1.31
Log (Industry Fundraising in Exit Year ($m))	–6.50	–2.39**	0.16	0.17	3.77	1.42	–0.30	–0.44	0.50	0.69
Multimedia Dummy	1.12	0.58	—	—	3.41	1.40	0.71	0.68	–0.72	–0.56
Computers – Hardware Dummy	2.57	1.36	0.06	0.04	—	—	–1.58	–1.47	–0.08	–0.09
Computers – Software Dummy	0.91	0.44	0.04	0.02	1.87	0.86	–1.21	–0.92	–1.74	–1.34
Manufacturing Dummy	–2.49	–1.33	–0.04	–0.04	—	—	0.44	0.63	–1.09	–1.29
Industrial Products Dummy	–1.06	–0.45	0.46	0.41	—	—	0.01	0.02	0.77	0.99
Agriculture Dummy	0.08	0.04	0.84	0.58	—	—	–0.70	–0.59	0.05	0.04
Medical Dummy	5.75	2.51**	–0.72	–0.52	–0.09	–0.06	–0.17	–0.20	–2.58	–2.01**
Number of Observations	134		134		134		134		134	
Loglikelihood	–21.56		–30.50		–14.70		–61.38		–50.32	
Chi–square	112.83***		37.01***		51.40***		42.29***		34.01***	

Actual Outcomes	Predicted Outcomes		Predicted Outcomes		Predicted Outcomes		Predicted Outcomes		Predicted Outcomes	
	Other Exit	IPO	Other Exit	Acquisition	Other Exit	Secondary Sale	Other Exit	Buyback	Other Exit	Write–off
Dependent Variable = 0	94	4	113	5	120	2	85	8	101	6
Dependent Variable = 1	5	31	10	6	4	8	19	22	17	10

Dependent variable = 1 if exit outcome as indicated, and = 0 otherwise.
*, **, ***, Significant at the 10%, 5%, and 1% level, respectively.
"—": Variable exclude to avoid collinearity problems.

for entrepreneurial firm industry (the sample was broken into seven industrial categories, with dummies for the remaining industries excluded to avoid perfect collinearity), and a dummy variable equal to 1 for the subset consisting of Canadian data. We do not include a right-hand-side variable for the extent of exit (i.e. full versus partial exits; see Note 10). The extent of exit is determined by, among other things, the choice of exit vehicle, and not vice versa (see Cumming and MacIntosh, 2003a,b).

Table 3.5 Logit Estimates of the Likelihood of Different Exit Outcomes
Panel C. Subsample of U.S. Data Only

Explanatory Variable	IPOs		Acquisitions		Secondary Sales		Buybacks		Write–offs	
	Coefficient	t–statistic	Coefficient	t–statistic	Coefficient	t–statistic	Coefficient	t–statistic	Coefficient	t–statistic
Constant	–23.99	–2.30**	3.14	0.53	79.76	2.29**	2.06	0.83	–4.87	–0.83
Log (Duration)	–2.68	–2.87**	1.29	2.08**	3.67	1.95*	0.01	0.01	–0.82	–1.98**
Log (Book Value ('000))	1.85	2.50**	0.28	0.89	–2.83	–2.53**	–1.11	–2.32**	0.33	1.16
Log (Market / Book)	4.31	4.35***	–0.31	–0.81	0.22	0.19	0.88	1.19	—	—
Seed Dummy	5.32	2.46**	–3.71	–2.54**	—	—	—	—	0.98	0.86
Start–up Dummy	3.72	1.81*	–1.90	–1.46	1.06	0.41	—	—	–0.09	–0.08
Early Stage Dummy	5.47	2.42**	–2.33	–1.73*	—	—	—	—	0.41	0.36
Expansion Dummy	3.09	1.57	–3.01	–2.25**	6.39	2.22**	—	—	–0.06	–0.06
Market Conditions Dummy	2.74	2.15**	–0.01	–0.02	3.73	1.94*	—	—	1.42	2.70***
Preplanned Exit Dummy	2.10	1.95*	0.51	0.63	3.58	1.40	1.91	1.51	—	—
Unsolicited Offer Dummy	—	—	2.59	2.46**	—	—	—	—	—	—
Log (1+MSCI Return in Exit Year)	78.74	1.94*	–29.77	–1.35	38.48	0.59	–70.66	–1.00	36.24	1.91*
Log (Industry Fundraising in Exit Year ($m))	0.38	0.31	–0.75	–1.02	–10.22	–2.30**	—	—	0.13	0.19
Multimedia Dummy	–3.75	–1.59	1.65	1.24	7.09	2.17**	—	—	—	—
Computers – Hardware Dummy	–1.82	–0.91	0.46	0.42	–0.38	–0.13	—	—	0.97	1.07
Computers – Software Dummy	–2.32	–1.30	0.81	0.81	5.46	2.25**	—	—	0.88	1.11
Manufacturing Dummy	—	—	–0.09	–0.05	—	—	4.43	2.17**	2.22	1.81*
Industrial Products Dummy	0.21	0.15	–0.24	–0.23	—	—	–	—	0.68	0.85
Agriculture Dummy	4.09	1.24	—	—	—	—	—	—	2.09	1.31
Medical Dummy	–2.55	–1.47	1.47	1.43	4.70	1.81*	—	—	–0.78	–0.77
Number of Observations	112		112		112		112		112	
Loglikelihood	–26.81		–51.46		–12.57		–15.16		–55.49	
Chi–square	76.56***		27.26**		37.49***		16.46**		24.81**	

Actual Outcomes	Predicted Outcomes		Predicted Outcomes		Predicted Outcomes		Predicted Outcomes		Predicted Outcomes	
	Other Exit	IPO	Other Exit	Acquisition	Other Exit	Secondary Sale	Other Exit	Buyback	Other Exit	Write–off
Dependent Variable = 0	77	5	75	7	101	2	105	1	73	6
Dependent Variable = 1	7	23	18	12	4	5	4	2	17	16

Dependent variable = 1 if exit outcome as indicated, and = 0 otherwise.
*, **, ***, Significant at the 10%, 5%, and 1% level, respectively.
"—": Variable exclude to avoid collinearity problems.

Information criteria were used to infer the appropriateness of the included right-hand-side variables. Our empirical specification gives rise to concern regarding two potentially endogenous variables: the market/book variable and the duration variable. We tested for endogeneity with these variables using Durbin-Wu-Hausman tests. We did not find the presence of endogeneity sufficient to warrant the use of instrumental variables.

The estimates are presented in Tables 3.5, Panels A, B and C. Panel A provides the estimates using the combined sample of Canadian and U.S. data (and a dummy variable for Canada as discussed above). Panel B considers the sub-sample of the Canadian data only. Panel C considers the sub-sample of the U.S. data only. We compare the estimates to shed light on the role of institutional factors across countries, as discussed in Section 3.

The data provide strong support for most of the hypotheses. The market/book coefficient (see Hypothesis 3 and accompanying text in Section 2) is positive in the IPO regressions for the combined sample (Table 3.5, Panel A), the Canadian sub-sample (Panel B), and the U.S. sub-sample (Panel C). In terms of the estimated economic significance,[15] an increase in the market/book value from 1 to 2 increases the probability of an IPO by 9.7% in the combined sample (Panel A), 0.2% in the Canadian sub-sample (Panel B), and 7.5% in the U.S. sub-sample (Panel C). By comparison, an increase in market/book value from 5 to 6 increases the probability of an IPO by 3.7% in the full sample, 0.6% in the Canadian subsample, and 2.9% in the U.S. subsample. (Recall that the market/book variable is expressed in logs, so that the effect decreases at higher values.) That the economic significance is greater in the United States than Canada may be consistent with the fact that U.S. stock exchanges (and in particular NASDAQ and NYSE) specify more demanding minimum capitalization requirements as a condition for listing than does the Toronto Stock Exchange (and other regional Canadian exchanges that existed in the period covered by our data). This is also consistent with evidence from other research suggesting lower VC managerial skill in Canada (see the discussion above in Section 3, and similar evidence discussed below).

The Canadian dummy included in the full sample suggests that there is a significantly lower probability of an exit via an acquisition in Canada. This is consistent with the view that there are fewer strategic acquirors in Canada, and that the Canadian and U.S. acquisition markets are not fully integrated. The Canadian dummy variable in the full sample regressions for IPOs is positive and significant, which is consistent with the finding that the costs of going public (in terms of direct fees and underpricing) are lower in Canada (as reported by Schutt and Williams, 2000, and discussed above in Section 3). It also suggests that these factors outweigh in importance those factors enumerated above, which would suggest a lower frequency of IPOs in Canada.

The book value coefficients also have a positive effect on IPOs in each of the three sub-samples, offering support to Hypothesis 2. In terms of economic significance, an increase in the book value of the VC's investment from $US 1,000,000 to $US $2,000,000 (and a change in book value from $US 11,000,000

---

[15] The tables report the standard logit coefficient estimates. The marginal effects were computed separately and are not shown in the tables, but are discussed throughout the text.

to $US 12,000,000) increases the probability of an IPO by 5.4% (0.7%) in the full sample (Panel A), 0.1% (0.05%) in the Canadian sub-sample (Panel B), and 5.5% (0.7%) in the U.S. sub-sample (Panel C). Higher book values also increase the probability of acquisition exits in the full sample (Panel A) and in Canada (Panel B), but this effect is insignificant in the U.S. sub-sample (Panel C). Lower book values (i.e. lower investments made by the VC) raise the likelihood of secondary sales and buybacks in all three sub-samples, consistent with the view that these exits tend to be used more often for smaller investments. The probability of a write-off is not significantly affected by the book value of the VC investment.

There is support for the hypothesis (Hypothesis 1) that the longer the investment duration, the greater the likelihood of an IPO, but only in the Canadian subsample (Panel B). Curiously, in the U.S. sub-sample, longer duration was associated with a greater likelihood of an acquisition, but a lower likelihood of an IPO. However, there is a significantly negative relationship between duration and the likelihood of a write-off in all three sub-samples. For example, an increase in duration from 1 to 2 years reduces the probability that the investment will be written off by 8.7% in the full sample (Panel A), 5.3% in Canada (Panel B), and by 6.1% in the United States. While this may provide some support to the VC's oft-repeated homily that "the lemons ripen quickly while the plums take longer to mature", our evidence suggests that there is a more complicated relationship between duration and quality. It may be, for example, that some of the most promising investments are divested relatively quickly via IPOs (at least in the United States), while good investments of slightly less than home-run quality are sheltered under the VC's wing until they can be sold through acquisition exits. In any case, the fact that U.S. VCs are quicker to write-off their bad investments is consistent with the view that U.S. VCs are relatively more skilled than their Canadian counterparts (see Section 3) and better able to winnow the wheat from the chaff.

In the IPO regressions, the stage of investment variables are mostly positive and significant in the combined sample (Panel A) and in the U.S. sub-sample (Panel C), but not in the Canadian sub-sample (Panel B). This suggests that the results in Panel A are driven by the U.S. results. In the U.S. sub-sample, investing in a relatively early stage in the entrepreneurial firm's development increases the probability of an IPO by about −23% (the marginal effects showed some differences depending on the stage, with the largest marginal effects at the seed and early stage and the lowest marginal effects at the expansion stage). These results are consistent with Hypothesis 4, which was that the earlier the stage at which the first investment takes place, the greater the ability of the VC to add value to the entrepreneurial firm. The insignificance of the Canadian results in Panel B results is also consistent with the view that Canadian VCs add less value to their portfolio firms than their U.S. counterparts (see Section 3).

With respect to Hypothesis 5, market conditions were assessed by means of two variables: a dummy variable equal to one if the reason for exit was due to market conditions, and the MSCI public market return. Note that use of these two variables did not produce collinearity problems. In the full sample, the market conditions dummy raises the probability of an IPO by 11.8%, while in the Canadian sub-sample it raises the probability by 1.3% (Panel B). In the U.S. subsample, it raises the probability by 11.7%. Finally, note that the MSCI index returns are significant in the U.S. sub-sample only (for example, an increase in the average MSCI index return in the year of exit from 5% to 10% increases the probability of an IPO by 15.8%[16]).

These results suggest two things. First, there appears to be little correspondence between the market conditions dummy (reflecting our survey results) and general market conditions. This is likely because when our survey respondents indicated that they were exiting due to market conditions, these market conditions might have been either favourable or unfavourable. This view is given credence by the fact that in all samples (Panels A – C in Table 3.5), the market conditions dummy was associated with a significantly elevated probability not only of an IPO, but of a secondary sale as well – an inferior form of exit. Second, the fact that increases in the MSCI return produced a significantly elevated probability of an IPO in the United States but not in Canada suggests that U.S. VCs are more willing or more able to time IPOs to coincide with upswings in the market.[17]

Another interesting result is that in the full sample (Panel A), the manufacturing dummy was associated with a significantly lower probability of an IPO exit. This is consistent with the view (Pagano *et al.*, 1998) that the public market views manufacturing firms as less "sexy" than technology firms, with a correspondingly diminished appetite for such offerings. The manufacturing dummy was necessarily excluded from the U.S. subsample because none of the U.S. VC-backed IPOs involved manufacturing firms. However, in the U.S. sub-sample, manufacturing firms were 4.2% more likely to be exited as buybacks. This is an intuitive result insofar as buybacks, in which debt is substituted for equity, are more easily effected by "cash cows"; i.e. firms with a steady and reliable cash flow. Manufacturing firms are more likely to meet this description than are high growth technology companies.

---

[16] See Lerner (1994) and Gompers and Lerner (1999) for similar U.S. evidence.

[17] Regulatory hurdles in Canada may impede the ability of Canadian VCs to time IPOs to correspond to market upswings; see Section 3.

A result that may be less intuitive is that industry fundraising was not associated with any significant affect on the frequency of IPOs, except in Canada (Panel B), where it was significantly negative, such that, for example, an increase in fundraising from $US 1 billion to $US 2 billion in the year lowers the probability of an IPO by 0.5%. Gompers' (1996) theory of grandstanding (see also Gompers and Lerner, 1999) suggests that in periods of rapid industry growth, younger VC firms will prematurely exit some of their investments as IPOs in order to establish a track record that will assist them in further fundraising, thus pointing to a *positive* relationship. It may be that the Canadian result reflects an agency problem whereby, in periods when abundant new funds are made available to the venture capital industry, venture capital managers shift their focus away from their portfolio firms and to their fundraising activities.

Our results offer confirmation to Hypothesis 6. where an exit is pre-planned, this increases the probability of an IPO by 7.9% in the full sample, 0.9% in the Canadian sub-sample, and by 9.0% in the U.S. sub-sample. It also increases the probability of both a secondary sale and a buyback. Our evidence thus strongly suggests that pre-planning is an important determinant of exit type.

Note, however, that the probability of a secondary sale is increased by only 1.4% in the full sample. By comparison, the probability of a buyback is increased by 13.1% in the full sample and 36.3% in the Canadian sub-sample. Thus, the two forms of exit that are most affected by pre-planning are the IPO and the buyback. This is curious, insofar as IPOs are the generally the most desirable form of exit, and buybacks the least desirable (short of a write-off) (Cumming and MacIntosh, 2003a). We interpret the totality of evidence relating to pre-planning in the following manner. First, VCs pre-plan their exits only when they envision exiting via an IPO, rather than by other means. Second, in a significant number of cases, a pre-planned exit does in fact result in an IPO; i.e. the VCs in our sample were successful in identifying a set of investee firms for which an IPO exit was feasible. Third, however, not all pre-planned exits can be realized, and when an IPO fails to occur, the VC will then enforce contractual redemption rights (i.e., the right to put its shares back to the entrepreneur) to exit its investment.

It is also noteworthy that VCs in our sample typically do *not* pre-plan acquisition exits, which are often a highly profitable form of exit. In other words, VCs do not attempt to function as "brokers" who identify an attractive target and purchase it with a view to "flipping" the firm to a higher valuing strategic corporate acquiror. This may be because, in order to "flip" the entire firm, the VC would have to secure the cooperation of the entrepreneur. Such cooperation may not be forthcoming if, as Black and Gilson (1998) suggest, the entrepreneur frequently places value on remaining involved in the firm in a managerial capacity in the long term (and would therefore prefer an IPO to an acquisition exit).

Our results also offer support to Hypothesis 7, insofar as the receipt of an unsolicited offer increased the probability of an acquisition exit by 26.7% in the full sample, 13.0% in the Canadian sub-sample, and by 42.3% in the U.S. sub-sample. The making of an unsolicited offer also raised the likelihood of a buyback exit in both regressions (Panels A and B) in which it was included (only one unsolicited offer resulted in a buyback exit in the US subsample, and therefore it was not econometrically feasible to include the unsolicited offer variable in the U.S. buyback regression). The increase in the probability of a buyback from an unsolicited offer is 13.8% in the full sample and 32.8% in the Canadian subsample. This suggests not only that unsolicited offers impact on exit vehicle, but that such offers typically come from a strategic acquiror who wishes to purchase the whole firm (rather than just the VC's interest) or from the entrepreneur. It is also noteworthy that the differences between the Canadian and U.S. buyback estimates is consistent with the much greater frequency of buybacks in Canada than in the United States (see also Tables 3.1 and 3.2).

6. CONCLUSIONS

In this chapter, we tested a number of hypotheses concerning factors that impact upon the VC's choice of exit vehicle. As expected, as book values and growth rates rise, the probability that the VC will exit its investment via an IPO increases. Higher book values also raised the likelihood of an acquisition exit, while reducing the probability of a secondary sale or buyback exit. This offers support to the view that the latter two types of exits are used more often in respect of relatively small entrepreneurial firms.

We found support in the Canadian sub-sample for the view that longer investment duration (a proxy for information asymmetry) enhanced the likelihood of an IPO. However, in the U.S. sub-sample, longer duration was associated with a *greater* likelihood of an acquisition and a *lower* likelihood of an IPO. At the same time, we observed a significantly negative relationship between duration and the likelihood of a write-off in both the United States and Canada (and the combined sample). This suggests that the VC's oft-repeated homily that "the lemons ripen quickly while the plums take longer to mature" is something of an oversimplification. It would appear that investments of "home run" quality (i.e. those that are exited via IPOs) can often be matured quickly, while successful investments of something less than home run quality are matured over a relatively long period of time before being sold in acquisition exits.

Our results show surprisingly little relationship between general economic conditions (as proxied by MSCI public market returns) and exit vehicle. Nonetheless, when VCs indicated that their exits were inspired by deal-specific market conditions, as indicated in survey responses, this significantly raised the probability not merely of an IPO exit, but of other inferior exit types as well.

We also found evidence that VCs sometimes pre-plan their exits. When this is the case, the probability of both an IPO and a buyback exit is significantly enhanced (and that of a secondary sale slightly enhanced), indicating that pre-planning is a factor that should be taken into account in examining exit type. We interpreted our evidence as suggesting that when an IPO fails to materialize as planned, the VC will often exit the investment by enforcing contractual rights of redemption against the entrepreneur. Pre-planning does not elevate the probability of an acquisition exit, probably because it is more difficult to pre-plan an exit when the cooperation of the entrepreneur is a necessary condition to effecting this type of exit. This is consistent with Black and Gilson (1998), who suggest that entrepreneurs tend to prefer IPO exits, in order to remain involved in the enterprise in the longer term.

In addition, our results suggest that the receipt of an unsolicited offer can result in an unanticipated exit, which will usually take the form of either an acquisition exit or a buyback. This is a novel result is the literature on venture capital.

Finally, we found evidence that acquisition exits are less likely in Canada, confirming the view that there are comparatively few strategic acquirors in Canada, and that Canadian and U.S. acquisition markets are not fully integrated. We also found evidence that IPOs are more likely in Canada, which is consistent with lower costs (in terms of fees and underpricing) of going public in Canada (Schutt and Williams, 2000). However, we also showed that the economic significance of variables leading to an IPO exit (such as book and market/book values, as well as the stage of VC investment) is much lower in Canada relative to the United States.

In the United States, as predicted, earlier stage investments are more likely to result in an IPO exit. This is consistent with the view that when VCs are involved with a firm at an early stage, they elevate the firm to a higher level of professionalism. In Canada, however, no such relationship existed. We suggested that this is consistent with evidence from other research suggesting lower VC managerial skill in Canada, such that Canadian VC managers are less able to add value to their portfolio companies. The lesser economic significance in the Canadian subsample of all of the variables that gave rise to an IPO is consistent with this interpretation.

ACKNOWLEDGEMENT

In this chapter we follow up issues raised in our earlier work (Cumming and MacIntosh, 2003a) on "Venture Capital Exits in Canada and the United States". The authors thank Varouj Aivazian, Paul Halpern, Vijay Jog, Aditya Kaul, Frank Mathewson, Vikas Mehrotra, Michael Patry, Tom Ross, Corrine Sellars, and especially Ralph A. Winter for helpful comments and suggestions. We are particularly indebted to an anonymous referee for helpful comments. We are also grateful for comments from the workshop and seminar participants at the University of Toronto (1997), Canadian Law and Economics Association (1998), University of Alberta (1999), Northern Finance Association (2000), the Law & Entrepreneurship Research Conference Sponsored by the Northwestern School of Law of Lewis & Clark College (2000), the Multinational Finance Society (2001), the Financial Management Association (2001), and the Babson Entrepreneurship Conference (2002).

BIBLIOGRAPHY

Armour, John, and Douglas J. Cumming (2006) "The Legislative Road to Silicon Valley." Oxford Economic Papers, forthcoming.

Barry, Christopher B., Chris J. Muscarella, John W. Peavy III, and Michael R. Vetsuypens (1990) "The Role of Venture Capitalists in the Creation of Public Companies: Evidence from the Going Public Process." *Journal of Financial Economics* **27**: 447–71.

Bascha, Andreas, and Uwe, Walz (2001a) "Convertible Securities and Optimal Exit Decisions in Venture Capital Finance." *Journal of Corporate Finance* **7**: 285–306.

Berglöf, Eric (1994) "A Control Theory of Venture Capital Finance." *Journal of Law, Economics, and Organization* **10**: 247–67.

Bergmann, Dirk, and Hege, Ulrich (1998) "Venture Capital Financing, Moral Hazard, and Learning." *Journal of Banking & Finance* **22**: 703–35.

Black, Bernard S., and Ronald J. Gilson (1998) "Venture Capital and the Structure of Capital Markets: Banks versus Stock Markets." *Journal of Financial Economics* **47**: 243–77.

Busenitz, Lowell W., James O. Fiet and Douglas D. Moesel (2003) "Reconsidering the Venture Capitalists' 'Value Added' Proposition: An Interorganizational Learning Perspective." *Journal of Business Venturing*, forthcoming.

Cochrane, John (2005) "The Risk and Return to Venture Capital." *Journal of Financial Economics*, forthcoming.

Coffee, John C. (1986) "Shareholders versus Managers: The Strain in the Corporate Web." *Michigan Law Review* **85**: 1–109.

Cumming, Douglas J., and MacIntosh, Jeffrey G. (2001) "Venture Capital Investment Duration in Canada and the United States." *Journal of Multinational Financial Management* **11**: 445–63.

Cumming, Douglas J., and MacIntosh, Jeffrey G. (2003a) "Venture Capital Exits in Canada and the United States." *University of Toronto Law Journal* **55**: 101–200.

Cumming, Douglas J., and MacIntosh, Jeffrey G. (2003b) "A Cross-Country Comparison of Full and Partial Venture Capital Exits." *Journal of Banking and Finance* **27**: 511–48.

Cumming, Douglas J., and MacIntosh, Jeffrey G. (2006) "Crowding Out Private Equity: Canadian Evidence" *Journal of Business Venturing*, forthcoming.

Das, Sanjiv, Murali Jagannathan and Atukya Sarin (2003) "Private Equity Returns: An Empirical Examination of the Exit of Venture-Backed Companies." *Journal of Investment Management* **1**: forthcoming.

Davidson, Russell and James G. MacKinnon (1993) *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Gompers, Paul A. (1996) "Grandstanding in the Venture Capital Industry." *Journal of Financial Economics* **42**: 133–56.

Gompers, Paul A., and Josh Lerner (1999) *The Venture Capital Cycle*, Cambridge: MIT Press.

Halpern, Paul (1997) *Financing Growth in Canada* (ed.). Calgary, AB: University of Calgary Press.

Jeng, Leslie A., and Philippe C. Wells (2000) "The Determinants of Venture Capital Fundraising: Evidence Across Countries." *Journal of Corporate Finance* **6**: 241–89.

Lerner, Josh (1994) "Venture Capitalists and the Decision to Go Public." *Journal of Financial Economics* **35**: 293–316.

MacIntosh Jeffrey G. (1994) *Legal and Institutional Barriers to Financing Innovative Enterprise in Canada*, monograph prepared for the Government and Competitiveness Project, School of Policy Studies, Queen's University, Discussion Paper 94–10.

MacIntosh, Jeffrey G. (1997) "Venture Capital Exits in Canada and the United States." in Paul J.N. Halpern, ed., *Financing Growth in Canada*. Calgary, AB: University of Calgary Press. 279–356.

Manigart, S., H. Sapienza, and W. Vermeir (1996) "Venture Capital Governance and Value-Added in Four Countries." *Journal of Business Venturing* **11**: 439–69.

Manigart, S., A. Lockett, M. Meuleman, M. Wright, H. Landstrom, H. Bruining, P. Desbrieres, and U. Hommel (2002a) "Why Do European Venture Capital Companies Syndicate?" Working Paper, Vlerick Leuven Gent Management School.

Manigart, S., K. DeWaele, M. Wright, K. Robbie, P. Desbrieres, H.J. Sapienza, and A. Beekman (2002b) "The Determinants of the Required Returns in Venture Capital Investments: A Five-Country Study." *Journal of Business Venturing* **17**: 291–312.

Megginson, William L., and Kathleen A. Weiss (1991) "Venture Capitalist Certification in Initial Public Offerings." *Journal of Finance* **46**: 879–903.

Pagano, Marco, Fabio Panetta, and Luigi Zingales (1998) "Why Do Companies Go Public? An Empirical Analysis." *Journal of Finance* **53**: 27.

Sahlman, William A. (1990) "The Structure and Governance of Venture Capital Organizations." *Journal of Financial Economics* **27**: 473–521.

Sapienza, Harry (1992). "When Do Venture Capitalists Add Value?" *Journal of Business Venturing* **7**: 9–27.

Sapienza, Harry, Sophie Manigart, and Wim Vermeir (1996) "Venture Capital Governance and Value-Added in Four Countries." *Journal of Business Venturing* **11**: 439–69.

Shutt, Theresa, and Hugh Williams (2000) "Going to Market: The Costs of IPOs in Canada and the United States." Conference Board of Canada (June 2000). http://www.tse.com/en/pdf/GoingToMarket.pdf

Smith, Janet K., and Richard L. Smith (2000) *Entrepreneurial Finance*. New York: Wiley.

Trester, Jeffrey J. (1998) "Venture Capital Contracting Under Asymmetric Information." *Journal of Banking & Finance* **22**: 675–99.

Wright, M., and A. Lockett (2004) "The Structure and Management of Alliances: Syndication in the Venture Capital Industry." *Journal of Management Studies*, forthcoming.

Wright, M., R.E. Hoskisson, L.W. Busenitz, and J. Dial (2001) "Finance and Management Buyouts: Agency versus Entrepreneurship Perspectives." *Venture Capital: An International Journal of Entrepreneurial Finance* **3**: 239–62.

This Page is Intentionally Left Blank

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 4

FINANCIAL DEVELOPMENT, INFLATION UNCERTAINTY AND GROWTH VOLATILITY

Robert Lensink* and Bert Scholtens

Abstract

We investigate whether the financial system dampens or exacerbates shocks of inflation uncertainty to the economy. Our GMM-estimates for 88 countries over a period of 25 years show that inflation uncertainty has a positive and significant impact on the volatility of economic growth. More importantly, we find that financial development significantly dampens the negative effects of inflation uncertainty on the volatility of economic growth. This confirms the importance of a well-developed financial sector.

Keywords: financial system, inflation uncertainty, growth volatility, financial development, Friedman's hypothesis.

JEL Classification No: E44, E62, F18, O42.

1. INTRODUCTION

This chapter contributes to the discussion on financial development and economic growth, as well as to that on the impact of inflationary shocks on the economy. There are many papers that examine the growth effects of financial development.

*Corresponding author University of Groningen, Department of Finance, PO Box 800, 9700 AV Groningen, The Netherlands; email B.W. Lensink@RUG. NL; Phone ++31 50 363 3712; fax ++31 50 363 8252.

In such literature, financial institutions emerge to lower transaction and information costs, to exert corporate control and to mobilize savings (see Levine, 1997; Allen and Santomero, 2001). From this, it is assumed that economies with more developed financial institutions may enjoy higher economic growth.

The aim of our communication is different. We set out in a new direction in financial development research. In line with other work, we analyze how inflation uncertainty affects per capita economic growth volatility for 88 countries over a period of 25 years. More importantly, we examine to what extent financial development dampens or exacerbates the effects of inflation uncertainty on the volatility of economic growth. A special feature that we employ is a system general methods of moments (GMM) estimator that can better control for endogeneity and measurement problems than, for instance, the ordinary least squares (OLS) method that is mostly used.

We hypothesize that inflation uncertainty has a positive effect on the volatility of economic growth. Moreover, we argue that in countries with a poorly developed financial system, this negative effect of inflation is stronger than in those with a more developed financial system. In our view, financial development improves the hedging and insurance capacity of the private sector by offering liquidity as well as allowing the purchase of various contingent financial claims (see Holmström and Tirole, 1998; Allen and Santomero, 2001). As such, it promotes efficient investment and consumption spending over time. Furthermore, financial development implies that more information is being gathered and processed within the economy. This might improve the allocation of funds within the economy (Diamond, 1984; Von Thadden, 1995). In both ways, financial development mitigates the impact of uncertainty about inflation on the volatility of economic growth. Finally, a more developed financial system provides more and better possibilities for banks to borrow, and hence to neutralize monetary policies. The result of this is that monetary shocks, often proxied by inflation uncertainty, will be dampened in countries with a well-developed financial system.

In line with our hypotheses, we find that inflation uncertainty has a positive and significant impact on the volatility of economic growth. Financial development as such does not have a significant effect on the volatility of economic growth. However, we find a significant negative effect of financial development interacted with inflation uncertainty on such volatility. This strongly suggests that financial development dampens the negative effects of inflation uncertainty on growth.

The structure of the chapter is as follows. Section 2 provides a survey of relevant literature. Section 3 is a description of our data. The estimation methodology is given in Section 4. The results are in Section 5 and Section 6 concludes.

2. LITERATURE SURVEY

There is voluminous literature on the economic effects of inflation and inflation uncertainty. Friedman's hypothesis is that changes in inflation induce erratic responses by monetary authorities, which may lead to more uncertainty about future inflation (Friedman, 1977). Much literature argues that inflation uncertainty rises with inflation, and that inflation uncertainty is one of the most important costs of inflation (see, e.g., Ball, 1992; Hess and Morris, 1996). Ungar and Zilberfarb (1993) argue that there is a negative association between inflation and its variance in the case of low inflation or more investments in forecasting inflation. An excellent survey of the literature is provided by Golob (1994). He distinguishes *ex ante* and *ex post* effects of inflation uncertainty.

The *ex ante* effects of inflation uncertainty refer to situations where companies or households make economic decisions that differ from those they would make otherwise. This could be a result of the interest-rate increasing effects of higher inflation uncertainty and the fact that higher inflation uncertainty leads to higher uncertainty in other economic variables, such as future wages, rents and taxes. Moreover, inflation uncertainty may be costly if firms and households spend resources to avoid the risks of future inflation, for example, by using financial instruments.

The *ex post* effects of inflation uncertainty refer to the situation where companies or households learn that inflation differs from what was expected after the decisions had been made. An unexpected increase in inflation will lead to wealth transfers and is costly in cases where contracts are specified in nominal terms. On the basis of a survey of the literature, Jansen (1989) comes up with two explanations for the negative impact of inflation uncertainty on economic efficiency. First is that increased volatility in inflation makes long-term contracts more costly because the future value of the monetary unit is more uncertain. Second is that increased inflation volatility reduces the ability of markets to convey information to market participants about relative movements.

While inflation and inflation uncertainty might be costly in theory, it may not be empirically confirmed. Many papers use OLS techniques to deal with the effects of inflation, or inflation uncertainty, on real gross domestic product (GDP) growth. Recently, Apergis (2004) used univariate GARCH models and a panel set for the G7 countries to analyze causality between inflation, output growth, and inflation uncertainty. He finds that inflation affects output growth, while inflation causes inflation uncertainty. This empirical literature is not conclusive as some papers show that inflation or inflation uncertainty have negative effects, while some do not find significant results. However, others suggest an inverted U curve relationship between inflation and economic growth (see, e.g., Cukierman and Meltzer, 1986; Holland, 1995; Hwang, 2001; Apergis, 2004). There is also literature describing the effects of inflation uncertainty on the variability of economic growth. Theoretically it seems clear that higher inflation uncertainty leads to increased output fluctuations, since such uncertainty spreads over into higher variability in all types of economic variables. There are, however, only a few empirical studies that have tested the relationship between inflation uncertainty and the variability of growth. Katsimbiris (1985) finds no significant relationship between inflation uncertainty and output growth. Tomassi (1994) and Grier and Perry (2000) find a negative association between inflation uncertainty and the variability of output growth. Hess and Morris (1996), Dotsey and Sarte (2000) and Beck *et al.* (2001) come up with support for a positive association between the two. Thus, again, the empirical results are inconclusive.

Another strand of the economic literature investigates the interaction between financial development and growth volatility. First, Kiyotaki and Moore (1997) argue that imperfections in the capital market may amplify the effects of productivity shocks. The effect of these imperfections on the net wealth (constrained) borrowers is to be held responsible for the amplifications. Then, fewer capital market imperfections, i.e. more developed financial intermediaries and financial markets, would suggest a reduced impact of shocks. As such, financial development could have a dampening effect on the volatility of economic growth. Second, we may derive arguments for a negative, but also a positive, relationship between financial development and growth volatility from the literature that studies the credit channel of monetary policy transmission (the so-called credit view). Bernanke and Blinder (1992) and Bernanke and Gertler (1995) argue that monetary policy impacts on the economy through both the bond market and the credit market. Bonds and credit are imperfect substitutes.

Private banks, as the main providers of credit, play a crucial role in the transmission of monetary policy. Interest rate changes will affect profitability, asset values and collateral. As such, they directly affect the borrowing capacity within the economy. Furthermore, if banks cannot easily manage their deposits and if their assets are not perfect substitutes, the supply of bank credit can also be affected. In that case, monetary shocks can be magnified by the banking sector. However, it can also be argued that a more developed financial system provides better opportunities for banks to borrow and hence to neutralize monetary policies. If this is the case, monetary shocks will be dampened in countries with a well-developed financial system. Aghion *et al.* (2004) assess the macro-economic effects of specific shocks to the financial sector as well as the effects of financial liberalization on the stability of the macro-economy. They introduce a framework to analyze these effects but do not put it to the test.

Previous papers that empirically investigate the impact of financial development on macro-economic volatility are inconclusive. Some find that financial development reduces macro-economic volatility (e.g., Gavin and Hausmann, 1995; Denizer *et al*., 2000; Easterly *et al*., 2000). However, the transmission channel is left unaccounted for in these studies. Carranza and Galdon-Sanchez (2004) build a model of financial intermediation that explains GDP volatility during the development process. They find that per capita output in middle-income economies is more volatile than in both low and high-income economies. Beck *et al*. (2001) find no robust relation between financial development and growth volatility. Furthermore, they assess that financial development magnifies the impact of inflation volatility in low- and middle-income countries as financial intermediaries may act as a conduit for monetary policy propagation. However, they use a simple OLS-regression technique. This may bias the results because of endogeneity and measurement problems.

3. THE DATA

Our dataset includes 88 countries in all income ranges (see Appendix 1 for a list of countries). We employ a five-period panel (1976–80, 1981–85, 1986–90, 1991–95, and 1996–2000). In all estimates, the same time periods and the same set of countries are used. However, the number of observations differs somewhat per estimate due to missing observations for some of the variables (see Appendix 2 for precise information on the number of observations per variable). We construct a dataset that is constituted on the basis of data availability, variation in time, and limited number of independent variables, as otherwise we would have too many instruments in our GMM-analysis.

Almost all of the data are derived from the 2002 online version of the World Bank Development Indicators. The dependent variable is the standard deviation of per capita real GDP growth (*STDGROW*). *STDGROW* is constructed by taking the standard deviation of real per capita growth figures (constructed from constant 1995 US$ GDP per capita figures, market rates) within each time period.

Since there is no measure for inflation uncertainty directly available, we have to derive a proxy for it. The literature distinguishes several methods to measure inflation uncertainty. In many papers, uncertainty is simply proxied by the variance, or standard deviation, of inflation. A somewhat more sophisticated method uses the standard deviation of the unpredictable part of a stochastic process (see e.g., Aizenman and Marion, 1993 and 1999). We follow the latter procedure. This method of measuring the volatility, or uncertainty, of inflation can be summarized as:

1. set up a forecasting equation for inflation
2. estimate the forecasting equation to obtain the unpredictable part of the fluctuations of inflation, i.e., the estimated residuals; and
3. compute the conditional standard deviations of the estimated residuals as the uncertainty measure of inflation.

In particular, for all countries in the dataset, we first estimate a forecasting equation for inflation (π) by using a second-order autoregressive process, extended with a time trend (T) and a constant (a_1):

$$\pi_{i,t} = a_{i,1} + a_{i,2}T + a_{i,3}\pi_{i,t-1} + a_{i,4}\pi_{i,t-2} + \varepsilon_{i,t}$$

where $\varepsilon_{i,t}$ is an error term for country i in period t. The subscripts i and t refer to countries and time, respectively. We inserted a trend term into the forecasting equation to deal with the problem of a stationary distribution of the unpredictable part of the stochastic process (see Ghosal and Loungani, 1996, 2000). We have yearly observations for the estimation period of 1970–2001. Since we do not expect clustering, we use a simple OLS estimator instead of a GARCH procedure that would have been more appropriate for high frequency data. Next, we calculate for each country the standard deviation of the residuals of the forecasting equation for π within each time period distinguished in our panel. This gives per country, and per sub-period, a proxy for inflation uncertainty.

We have two measurements for financial development. The logarithm of domestic credit provided by the banking sector as a percentage of GDP (*BANK*) and the logarithm of bank credit to the private sector as a percentage of GDP (*PRIV*). For both indicators, we use averages over the periods in the estimates. Both measures are widely used in studies about financial development and economic growth (see Levine, 1997). Ideally, we would have liked more measurements, such as for the role of non-bank financing, but due to numerous omitted observations (especially in the 1970s and 1980s) we decided against their use. Other variables used in the estimates are:

- the logarithm of the beginning of the period real GDP per capita (*GDPPC*);
- the logarithm of the period averages of general government final consumption expenditures as a percentage of GDP (*GOV*);
- the average annual growth rate of real GDP per capita per period (*GROW*);
- the average inflation rate (*INFL*); and
- the logarithm of the period averages of trade as a percentage of GDP (*TRADE*).

These variables are the "usual suspects" that are being used in the economic literature that assesses the relationship between growth and shocks.

Table 4.1 gives the descriptive statistics of the variables used in the estimates, whereas Table 4.2 gives the correlation matrix of the variables. Table 4.1 shows that our shock measures are indeed "shocking and shaking". Table 4.2 reveals high correlations between inflation and inflation uncertainty, as well as between bank credit to GDP and private credit in relation to GDP.

Table 4.1 Descriptive Statistics

	TRADE	PRIV	INFL	BANK	GDPPC	GOV	INFU	GROW	STDGROW
Mean	4.04	3.33	33.95	3.73	7.62	2.623	37.02	1.32	0.032
Median	4.04	3.38	8.58	3.81	7.37	2.64	4.05	1.39	0.027
Maximum	5.94	5.30	2846	5.71	10.72	4.03	5296.5	10.88	0.254
Minimum	2.39	−5.18	−3.19	−5.14	4.90	1.43	0.18	−7.47	0.002
Std. Dev.	0.57	1.08	189.89	1.00	1.67	0.39	296.4	2.69	0.024
Skewness	0.20	−2.52	11.37	−3.56	0.27	−0.12	14.38	−0.01	2.609
Kurtosis	4.01	18.63	145.99	28.27	1.76	3.26	237.80	3.84	18.646
Jarque-Bera	21.66	4909	384332	12488	33.41	2.23	1025909	12.85	4986.85
Probability	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.0016	0.000000
Observations	440	437	440	435	440	439	440	440	440

Table 4.2 Correlation Matrix

	TRADE	PRIV	INFL	BANK	GDPPC	GOV	INFU	GROW	STDGROW
TRADE	1								
PRIV	0.20	1							
INFL	−0.15	−0.04	1						
BANK	0.11	0.88	−0.01	1					
GDPPC	0.21	0.56	0.02	0.47	1				
GOV	0.31	0.28	−0.08	0.29	0.36	1			
INFU	−0.11	−0.03	0.95	−0.02	0.004	−0.08	1		
GROW	0.12	0.28	−0.16	0.17	0.19	−0.06	−0.14	1	
STDGROW	−0.05	−0.23	0.08	−0.17	−0.28	−0.04	0.06	−0.33	1

4. ESTIMATION METHODOLOGY

We specify equations of the following form:

$$STDGROW_{it} = \sum_a \alpha_{1,a} X_a + \alpha_2 FIN_{i,t} + \alpha_3 INFU_{it} + \alpha_4 FIN_{i,t} \times INFU_{i,t} + \sum_z \alpha_{5,z} T_z + \sum_h \alpha_{6,h} R_h + \alpha_7 STDGROW_{i,t-1} + n_i + e_{i,t}$$

where X_a is a vector of explanatory variables. In the base regressions, $a \in (INFL_{i,t},\ GOV_{i,t},\ TRADE_{i,t})$. In alternative regressions $a \in (INFL_{i,t},\ GDPPC_{i,t},\ GOV_{i,t}, TRADE_{i,t})$, or $a \in (GROW_{i,t},\ GDPPC_{i,t},\ GOV_{i,t},\ TRADE_{i,t})$. We ignore *INFL* in one set of regressions because of the high multicollinearity between *INFL* and *INFU*. *FIN* is our proxy for financial development (*BANK* or *PRIV*). *T* is a vector of time dummies, with a one if t = z and a zero otherwise, z ∈ (1976–80, 1981–85, 1986–90, 1991–95, 1996–2000). These time dummies are used as additional instruments. R_h is a vector of "region" dummies. The dummy gets a one if a country i is in region h and a zero otherwise, and $h \in$ (high income, upper middle-income, lower middle-income, lower income).[1] η is an unobserved country-specific effect (a country-specific error term) and ε is an overall error term.

Our aim is to examine the effects of financial development on the volatility of growth, and more specifically to consider whether financial development dampens or increases the impact of inflation uncertainty on the volatility of growth. We focus on the volatility of growth since greater volatility of growth reduces consumer welfare and economic efficiency. The most important reason for this is that the economy is less likely to produce at its full potential if the variability of growth rises under the assumption that potential output grows steadily so that a highly variable real growth causes actual output to deviate more often from potential.

The overall effect of financial development on the volatility of growth is given by

$$\frac{dSTDGROW}{dFIN} = \alpha_2 + \alpha_4 INFU.$$

The direct effect of financial development on the volatility of growth is given by α_2. The way in which shocks are transmitted via financial development is reflected by α_4.

[1] The classification of countries is based on the World Bank classification.

Before examining the estimates, some remarks on the estimation methodology are needed. There are several problems with estimating the above equation by OLS. First, OLS assumes that the regressors are uncorrelated with the error term. However, as can be shown, the lagged dependent variable is correlated with the country-specific error term.[2] The second problem is that OLS assumes that the regressors are exogenous. However, it is difficult to justify why some of the regressors, especially our indicators for financial development, are not determined simultaneously with the standard deviation of per capita growth. If these regressors were treated as exogenous, when they are not, then this would result in biased parameter estimates.

Estimating our models using OLS might also be problematic due to measurement problems (we use constructed proxies). Therefore, we estimate our panel-based models using an instrumental variable approach. The instrumental variable estimation technique controls for the fact that the explanatory variables are likely to be correlated with the error term and the firm-specific effect, and deals with possible endogeneity problems. More specifically, we estimate the models with the system GMM estimator, using DPD98 for Gauss (see Arellano and Bond, 1998). A method of moments estimator derives the coefficients from the so-called moment restrictions, i.e. restrictions on the covariances between regressors and the error term.

The system GMM estimator combines the differenced equation with a levels equation to form a system GMM. Blundell and Bond (1998) show that, under certain conditions, the system estimator provides more efficient estimators than a regression in first differences. Lagged levels are used as instruments for the contemporaneous differences and lagged differences as instruments for the contemporaneous levels. If the error terms are not serially correlated, Arellano and Bond argue that, starting from t – 2, the whole history of the series (in levels) can be used as instruments for the first differences. With respect to the levels equations, valid instruments for the regressions are the lagged differences of the corresponding variables. Here, only the most recent difference is used as the instrument. Additional lagged differences would be redundant, since the instruments for the first differences already cover them.

The system GMM estimator is a two-step GMM estimator. In the first step, homoscedasticity and independent error terms are assumed. In the second step, these assumptions are relaxed by using a consistent variance-covariance matrix

[2] Consider a simple version of our equation to be estimated: $STDGROW_{it} = \alpha_2 FIN_{i,t} + \alpha_7 STDGROW_{i,t-1} + n_i + e_{i,t}$. Since $E(n_i^2) \neq 0$, $E[\eta_i(STDGROW_{it-1})] = E[\eta_i(\alpha_2 FIN_{i,t-1} + \alpha_7 STDGROW_{i,t-2} + n_i + e_{i,t-1})] \neq 0$. Therefore, the error term is correlated with the lagged dependent variable.

that is constructed from the first step residuals. However, the two-step estimator has weak small sample properties, i.e. the standard errors are biased downwards. The estimator becomes problematic, especially when there are a small number of cross-section units, in relation to the number of instruments, i.e. the number of time series units. In our case this might be problematic, although we have 88 cross-section units (countries) in our dataset. This might result in biased asymptotic inference. We address this problem by presenting coefficients and t-values using two-step GMM estimates, based on robust, finite sample corrected standard errors. Windmeijer (2000) shows how the two-step standard estimates can be corrected. We followed this approach.

The reliability of the system GMM estimation procedure depends on the validity of the instruments, which we consider by presenting a Sargan test, a test on over-identifying restrictions. It is asymptotically distributed as a χ^2 variable and tests the null hypothesis of validity of the (over-identifying) instruments. P-values report the probability of incorrectly rejecting the null hypothesis, so that a p-value above 0.05 implies that the probability of incorrectly rejecting the null hypothesis is above 0.05. In this case, a higher p-value makes it more likely that the instruments are valid.

The consistency of the estimates also depends on the absence of serial correlation in the error terms. This will be the case if the differenced residuals display significant negative first-order serial correlation and no second-order serial correlation. We present tests for first-order and second-order serial correlation related to the estimated residuals in first differences. The test statistics are asymptotically distributed as standard normal variables. Here, the null hypothesis relates to "insignificance" so that a low p-value for the test on first-order serial correlation and a high p-value for the test on second-order serial correlation suggest that the disturbances are not serially correlated. The serial correlation tests (M1 and M2 in the Table) refer to the one-step GMM estimates.

We also present Wald tests. These test statistics are also asymptotically distributed as χ^2 variables. As such, we test for joint significance of all parameters (or for a subset of parameters). The null hypothesis refers to "insignificance", implying that low p-values suggest joint significance. Wald tests for the joint significance of the time dummies and the region dummies are presented.

5. RESULTS

The results of our analysis are in Table 4.3. We find that the direct effect of financial development on the volatility of per capita economic growth is positive, although never significant. We also find that the direct effect of the shocks

Table 4.3 Financial Development, Inflation Uncertainty and the Volatility of Growth

	1	2	3	4	5	6
INFL (*1000)	0.066 (2.62)	0.058 (2.58)		0.037 (2.47)	0.033 (2.11)	
GROW			−0.003 (−1.91)			−0.003 (−2.03)
GOV	0.0146 (2.02)	0.0124 (2.22)	0.0060 (1.04)	0.0123 (1.76)	0.012 (1.92)	0.0067 (1.37)
TRADE	−0.0128 (−2.13)	−0.0028 (−0.40)	−0.0029 (−0.36)	−0.0137 (−2.24)	−0.0041 (−0.64)	−0.0042 (−0.47)
STD-GROW(−1)	0.0138 (0.22)	0.089 (1.50)	0.119 (1.83)	0.0385 (0.55)	0.088 (1.52)	0.117 (1.86)
GDPPC		0.018 (2.53)	0.016 (2.17)		0.018 (3.21)	0.013 (1.67)
BANK	0.0036 (1.14)	0.0023 (0.81)	0.0012 (0.49)			
PRIV				0.0037 (1.22)	0.0017 (0.76)	0.002 (1.11)
INFU (*1000)	0.132 (3.87)	0.120 (3.70)	0.063 (2.56)	0.129 (3.35)	0.122 (3.20)	0.0075 (4.83)
INFU*BANK (*1000)	−0.047 (−3.58)	−0.043 (−3.49)	−0.015 (−2.42)			
INFU*PRIV (*1000)				−0.0048 (−3.44)	−0.044 (−3.14)	−0.021 (−4.61)
M1	−2.08 $p = 0.04$	−2.15 $p = 0.03$	−2.32 $p = 0.02$	−2.106 $p = 0.035$	−2.179 $p = 0.03$	−2.331 $p = 0.02$
M2	0.540 $p = 0.59$	0.003 $p = 0.99$	0.488 $p = 0.63$	0.705 $p = 0.481$	0.218 $p = 0.83$	0.517 $p = 0.61$
SARGAN	60.02 $p = 0.33$	65.81 $p = 0.41$	74.40 $p = 0.18$	60.86 $p = 0.31$	71.52 $p = 0.24$	65.08 $p = 0.44$
WTEST TIME	7.42 $p = 0.06$	13.79 $p = 0.003$	10.56 $p = 0.014$	10.88 $p = 0.012$	15.19 $p = 0.002$	13.54 $p = 0.004$
WTEST REG	20.11 $p = 0.00$	35.63 $p = 0.00$	18.73 $p = 0.00$	22.82 $p = 0.00$	30.29 $p = 0.00$	25.15 $p = 0.00$

Note: In all regressions, starting from t−2, the entire history of the series in levels are used as instruments for the first differences. For the levels equations, the one period lagged differences of the corresponding variables are used as instruments. The t-values are between brackets.

from unexpected inflation is positive, as expected. Furthermore, this effect is highly significant. Most importantly, in all regressions the interactive terms between inflation uncertainty and financial development are negative and highly significant. This holds for both our financial development proxies. From this, we infer that financial development dampens the negative effects of inflation uncertainty on the volatility of economic growth on a per capita basis.

As to the "usual suspects", we find that increased government consumption positively and significantly affects the volatility of growth. Furthermore, more trade – although not always significantly – reduces growth volatility. The results for financial development are not significantly affected by the inclusion of the (logarithm of the) begin of period real per capita income at market rates (GDPPC) and the average annual real GDP growth (GROW) rate, respectively. Again, the direct effects of financial development are positive, but insignificant, and the interactive terms are significantly negative. In all, Our results are robust.

For the statistical diagnostics of our results, we find that all equations seem to be reasonably good. The SARGAN tests show for all regressions that we cannot reject the null hypothesis of the validity of the instruments. In addition, the M1 and M2 statistics show that the equations do not suffer from first- or second-order serial correlation. Finally the WALD tests (WTEST) show that the time dummies, as well as the region dummies, are jointly significant.

6. CONCLUSION

The aim of this chapter is to examine the impact of inflation uncertainty on the volatility of economic growth. More importantly, it provides some first evidence on the question as to whether financial development dampens or strengthens the effects of inflation uncertainty on the volatility of per capita growth.

We hypothesize that a negative impact of inflation uncertainty on economic efficiency is reduced by financial development. The recent literature about financial intermediation and finance and growth offers some reasons why this might be the case. A well-developed financial system offers instruments and mechanisms to absorb shock and produces additional information about relative movements. As such, it may help to improve economic efficiency, especially in the case of an uncertain inflation environment.

We investigate the impact of financial development on the effect of inflationary uncertainty on per capita growth variability for 88 countries in all income ranges for a period of 25 years (1976–2001). We estimate the relationships on the basis of GMM and employ five five-year period panels. We find that inflation uncertainty has a positive and significant impact on the volatility of per capita economic growth. That is, more uncertainty about the inflation level increases this volatility. We also find empirical evidence for our hypothesis that financial development has a dampening effect on the impact of inflation uncertainty on this growth volatility. This is because we have a significant negative effect of financial development interacted with inflation uncertainty on the volatility of per capita economic growth. The reason behind our findings can be, first, that financial development offers hedging/insurance against uncertainty. As such, financial development mitigates the *ex ante* effect of inflation uncertainty on economic efficiency. Second, financial development improves the information production and information revelation to market participants. A well-developed financial system offers a lot of liquidity to the banking sector. In that case, they are less vulnerable to monetary shocks and monetary policies. Financial development implies that banks are fit to perform their information function in the modern economy (Boot and Thakor, 1997). As such, we establish that financial structure and development does indeed play an important role in the transmission of shocks to the economy.

Our findings are in line with most of the existing literature on the relationship between inflation and inflation uncertainty. However, we take a somewhat distinct position with respect to the relationship between inflation uncertainty and growth volatility from the perspective of financial development. While in our study financial development dampens the positive effect of inflation uncertainty on the volatility of per capita growth for the entire panel of countries, Beck *et al.* (2001) come to the opposite result, at least for developing countries. The precise reasons for the different results are unclear. It may be a result of the other estimation techniques they have used or it may also be caused by differences in the estimation period and/or estimation sample. It is clear that the debate on how financial development affects shocks is not yet resolved. We feel that our efforts provide a first empirical contribution to this debate, but also realize that more future theoretical and empirical research is needed to better understand how financial development amplifies or mitigates shocks.

ACKNOWLEDGEMENT

We wish to thank Linda S. Goldberg, Kjersti-Gro Lindquist, E. Philip Davis and C. Maxwell Watson for their comments and suggestions. The usual disclaimer applies.

BIBLIOGRAPHY

Aghion, P., Bacchetta, P., and Banerjee, A. (2004) "Financial Development and the Instability of Open Economies." NBER Working Paper 10246.

Aizenman, J., and Marion, N.P. (1993) "Macroeconomic uncertainty and private investment." *Economic Letters* **41**: 207–10.

Aizenman, J., and Marion, N.P. (1999) "Volatility and investment: interpreting from developing countries." *Economica* **66**: 157–79.

Allen, F., and Santomero, A.M. (2001) "What do financial intermediaries do?" *Journal of Banking and Finance* **25**: 271–94.

Apergis, N. (2004) "Inflation, output growth, volatility and causality: evidence from panel data and the G7 countries." *Economics Letters* **83**: 185–91.

Arellano, M., and Bond, S.R. (1988) "Dynamic panel data estimation using DPD – a guide for users." Institute for Fiscal Studies Working Paper No. 88/15.

Ball, L. (1992) "Why does high inflation raise inflation uncertainty?" *Journal of Monetary Economics* **29**: 371–88.

Beck, T., Lundberg, M., and Majnoni, G. (2001) "Financial Intermediary Development and Growth Volatility: Do Intermediaries Dampen or Magnify Shocks?" World Bank Policy Research Working Paper 2707.

Bernanke, B.S., and Blinder, A.S. (1992) "The Federal Funds rate and the channels of monetary transmission." *American Economic Review* **82**: 901–21.

Bernanke, B.S., and Gertler, M. (1995) "Inside the Black Box: The credit channel of monetary policy transmission." *Journal of Economic Perspectives* **9**: 27–48.

Blundell, R., and Bond, S.R. (1998) "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* **87**: 115–43.

Boot, A.W.A., and Thakor, A.V. (1997) "Financial system architecture." *Review of Financial Studies* **10**: 693–733.

Carranza, L., and Galdon-Sanchez, J.E. (2004) "Financial intermediation, variability and the development process." *Journal of Development Economics* **73**: 27–54.

Cukierman, A., and Meltzer, A. (1986) "A theory of ambiguity, credibility, and inflation under discretion and asymmetric information." *Econometrica* **54**: 1099–128.

Denizer, C., Iyigun, M.F., and Owen, A.L. (2000) "Finance and Macroeconomic Volatility." World Bank Policy Research Working Paper 2487.

Diamond, D. (1984) "Financial intermediation and delegated monitoring." *Review of Economic Studies* **51**: 393–414.

Dotsey, M., and Sarte, P. (2000) "Inflation uncertainty and growth in a cash-in-advance economy." *Journal of Monetary Economics* **45**: 631–55.

Easterly, W., Islam, R., and Stiglitz, J.E. (2000) "Shaken and Stirred: Explaining Growth Volatility." Paper presented at the ABCDE Conference World Bank.

Friedman, M. (1977) "Nobel lecture: inflation and unemployment." *Journal of Political Economy* **85**: 451–72.

Gavin, M., and Hausmann, R. (1996) "Securing Stability and Growth in a Shock Prone Region: the Policy Challenge for Latin America. Inter-American Development." Office of the Chief Economist, Working Paper 315.

Ghosal, V., and Loungani, P. (1996) "Product market competition and the impact of price uncertainty on investment: some evidence from U.S. manufacturing industries." *Journal of Industrial Economics* **44**: 217–28.

Ghosal, V., and Loungani, P. (2000) "The differential impact of uncertainty on investment in small and large businesses." *Review of Economics and Statistics* **82**: 338–49.

Golob, J.E. (1994) "Does inflation uncertainty increase with inflation?" Economic Review Federal Reserve Bank of Kansas City. Third Quarter, 27–38.

Grier, K., and Perry, M. (2000) "The effects of real and nominal uncertainty on inflation and output growth: some GARCH-M evidence." *Journal of Applied Econometrics* **15**: 45–58.

Hess, G.D., and Morris, C.S. (1996) "The long-run costs of moderate inflation." Economic Review Federal Reserve Bank of Kansas City. Second Quarter, 71–88.

Holland, S. (1995) "Inflation and uncertainty: tests for temporal ordering." *Journal of Money, Credit, and Banking* **25**: 514–20.

Holmstrom, B., and Tirole, J. (1997) "Financial intermediation, loanable funds, and the real sector." *Quarterly Journal of Economics* **52**: 663–91.

Hwang, Y. (2001) "Relationship between inflation rate and inflation uncertainty." *Economics Letters* **73**: 179–86.

Jansen, D.W. (1989) "Does inflation uncertainty affect output growth? Further evidence." *Economic Review Federal Reserve Bank of St. Louis*, July/August, 43–54.

Katsimbris, G.M. (1985) "The relationship between the inflation rate, its variability, and output growth variability." *Journal of Money, Credit, and Banking* **17**: 179–88.

Kiyotaki, N., and Moore, J. (1997) "Credit cycles." *Journal of Political Economy* **105**: 211–48.

Levine, R. (1997) "Financial development and economic growth: views and agenda." *Journal of Economic Literature* **35**: 688–726.

Thadden, E.-L. von (1995) "Long-term contracts, short-term investment, and monitoring." *Review of Economic Studies* **62**: 557–75.

Tomassi, M. (1994) "The consequences of price stability on search markets: toward understanding the effects of inflation." *American Economic Review* **84**: 1385–96.

Ungar, M., and Zilberfarb, B. (1993) "Inflation and its unpredictability – theory and empirical evidence." *Journal of Money, Credit, and Ba*nking **25**: 709–20.

Windmeijer, F. (2000) "A finite sample correction for the variance of linear two-step GMM estimators." IFS Working Paper 00/19. London: The Institute for Fiscal Studies.

Appendix 1

Variables used in the estimates

If not indicated otherwise, variables are derived from data published in the on-line version of the 2002 World Bank development indicators.

BANK:	The logarithm of the period averages of domestic credit provided by the banking sector as a percentage of GDP. Number of observations: 435. Missing observations for Hong Kong (2), Hungary, Lesotho, and Trinidad and Tobago
GDPPC:	The logarithm of the begin of period real GDP per capita. Number of observations: 440.
GOV:	The logarithm of the period averages of general government final consumption expenditures as a percentage of GDP. Number of observations: 439. Missing observation for Argentina.
GROW:	The average annual growth rate of real GDP at market rates per capita per period. This proxy is calculated by using figures on constant 1995 US$ GDP per capita data. Number of observations: 440.
INFL:	The average inflation rate for a period. Constructed by taking the average of annual inflation rates, based on GDP deflators. Number of observations: 440.
INFU:	Inflation uncertainty. Constructed by taking the standard deviation of the error terms from a second order autoregressive forecasting equation for inflation (based on annual GDP deflators). Number of observations: 440.
PRIV:	The logarithm of the period averages of credit to the private sector as a percentage of GDP. Number of observations: 437. Missing observations for Hong Kong (2) and Hungary.
STDGROW:	Standard deviation of real per capita growth. Per capita growth is constructed from constant 1995 US$ figures, GDP per capita. Number of observations: 440.
TRADE:	The logarithm of the period averages of trade as a percentage of GDP. Number of observations: 440.

Economies are divided among income groups according to 2001 GNI per capita, calculated using the World Bank atlas method:

Low income: $ 745 or less
Lower middle income: $ 746–$ 2975
Upper middle income: $ 2976–$ 9206
High-income: $ 9206 or more.

Appendix 2

List of countries included in the analysis

Austria	1	Haiti	4	Panama	3
Bangladesh	4	Honduras	4	Papua New Guinea	3
Belgium	1	Hong Kong	1	Paraguay	3
Belize	3	Hungary	2	Peru	3
Benin	4	Iceland	1	Philippines	3
Bolivia	3	India	4	Rwanda	4
Brazil	2	Indonesia	4	Senegal	4
Burkina Faso	4	Ireland	1	Sierra Leone	4
Burundi	4	Israel	1	Singapore	1
Cameroon	4	Italy	1	Spain	1
Canada	1	Jamaica	2	Sri Lanka	3
Central African	4	Japan	1	Sweden	1
Chad	4	Kenya	4	Switzerland	1
Chile	2	Korea, Rep.	2	Syria	3
China	4	Lesotho	4	Thailand	3
Colombia	3	Luxembourg	1	Togo	4
Congo, Rep.	4	Madagascar	4	Trinidad and Tobap.	2
Costa Rica	3	Malawi	4	Tunisia	3
Cote d'Ivoire	4	Malaysia	2	Turkey	2
Denmark	1	Mali	4	United Kingdom	1
Dominican Rep.	3	Mauritania	4	United States	1
Ecuador	3	Mexico	2	Uruguay	2
Egypt	3	Morocco	3	Venezuela	2
El Salvador	3	Nepal	4	Zambia	4
Finland	1	Netherlands	1	Zimbabwe	4
France	1	New Zealand	1		
Gambia	4	Niger	4		

Note: 1 = high-income country; 2 = upper middle-income country; 3 = lower middle-income country and 4 = lower income country.

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 5

MATHEMATICAL CHARACTERIZATION OF BEHAVIORAL MARKET DYNAMICS: FROM STYLIZED AGENTS TO AGGREGATE PRICE PROCESSES

M. Bagella, R. Ciciretti and G. Susinno

Abstract

Ex-post analysis of the evolution of financial indexes tends to match, in a causal relationship, exogenous events contained in an information stream, and observed market reactions. It is obvious that market players may react to globally available information, the effects of a self-reinforced endogenous mechanism. This may destabilize the global system even in the absence of an external perturbation, and may play a major role in explaining some observed stylized facts. Indeed, empirical analysis of financial markets has shown a number of these stylized facts, such as heavy tails or volatility "bursts," which are difficult to explain in terms of the evolution of fundamental economic variables. Indeed, the non-Gaussian, non-stable character of empirical distributions, such as excess demand or stock returns, demonstrates the weakness of any "independent agent" approach to modeling the real market. Starting with existing literature on the characterization of the behavior of random economies with many interacting agents, we identify a set of micro-economic interaction rules, which could help to explain the macro-economic observed market behavior. Following the work of Bornholdt and extending that of Brock and Durlauf, we will consider interacting agents whose payoff exhibit both a strategic complementarity with their nearest-neighbors actions and an eventual global substitutability with the global market state. In this setup we reconstruct a price process related to the imbalance between buyers and sellers. Finally, we investigate how the frustration resulting from the tendency of local imitation, with an additional coupling with the average

state of the system, reproduces the main observed stylized facts of real financial markets. We show how in this framework even the largest crash may emerge as a natural intrinsic metastable dynamic of the system induced by a collective phenomena, such as crowd effects or "herd" behavior.

1. INTRODUCTION

A number of stylized facts have characterized the dynamics of the new Millennium financial markets. This is a change which cannot only be ascribed to the resorption of the "dot.com" speculative bubble since March 2000 or to the complex geo-political context following the 9/11 terrorist attack, but it seems that the way investors reacted to the news caused a drastic change. Following a recent research on the U.S. market by the Economics Department of the University of Rome "Tor Vergata" (Bagella *et al.*), the news impact on investors' behavior underwent a change after March 2000. This seems to have been reinforced by the 9/11 terrorist attack. Before the end of the speculative bubble, the evolution of the implicit risk premium on equity prices was closely related to the analyst views. The impact of positive (negative) news, both aggregate- and firm-specific, was to reduce (augment) the implicit risk premium by acting on the confidence of investors. Such a reaction was reflected by the offer/demand trade off on prices. This causal relation, between news and prices, was also confirmed by the joint evolution between the Consumer Prices Index (CPI) (University of Michigan) and implicit risk premiums. Also, the pre-2000 markets were over-reacting to good news, probably induced by a joint bullish bias of financial analysts and investors. All this seems to have been strongly modified by the end of the last speculative bubble via a reduced impact of news on implicit risk premiums. This indicates a decoupling between investors' feelings and financial analysts' valuations, an effect that persists and is reinforced after September 11, 2001. It is observed by Bagella *et al.* that after the terrorist attack there is no significant impact of positive news and CPI on implicit risk premiums, while there is little persistence of the impact of negative news. Observing markets' evolution for the last 20 years, there are some stylized facts that cannot be explained by only invoking the influence of geopolitical and economic news on the strategies and appetites of investors (Cont and Bouhaud). It is difficult to understand why 9/11 has produced a market movement more than five times smaller than the black Monday crash. On October 19, 1987, all major market indexes experienced a drop of more than 30%, a date that subsequently became known as "Black Monday." The Dow Jones Industrial Average plummeted 508 points, losing 22.6% of its total value. The S&P500 dropped by 20.4%, falling from 282.7 to 225.06.

This was the greatest loss Wall Street had ever experienced in a single day, and the informational causes of such an event are still an issue for many economists and market practitioners. There are schools of thought, invoking the effect of derivatives, portfolio insurance, and program trading but there is not a universally accepted statement. Standard economic theory supposes an efficient market mechanism where prices react instantaneously and in an unbiased manner to new information. It can hardly explain the difference in the dynamics of such extremes. Indeed, it seems that significant price fluctuations are not necessarily related to the arrival of information or variation in fundamental economic variables. Therefore, modelers are led to conclude that market movements may also be caused by the intrinsic metastable dynamics of the system induced by a collective phenomena, such as crowd effects or "herd" behavior (Cont and Bouchaud). In the light of former considerations, it appears as a major issue to modelers, both academicians and investment strategists, to investigate the dynamical evolution of systems where investors are allowed to influence each other while all being influenced by a global publicly available information. That is introducing a behavioral component on Economics and Financial modeling. The complexity of such an approach may discourage standard modelers, since it imposes the determination of how micro-economics dynamics, and idiosyncratic investors' behavior, may propagate and contaminate the entire financial system, creating fashion rump-up phases eventually destroyed by sudden panic crises. However, the close collaboration between economists and physicists may help to shed some light by a careful merger between Economics constraints, and methods and tools akin to natural sciences (Brock and Durlauf).

Recently, the literature on this subject has witnessed increasing contributions from physicists. Some of these works look like high-tech versions of dead-end models already investigated by economists, but some of it may be worth reading. Analytical and methodological tools from statistical physics, applied to understand interacting particles dynamics, can be transposed to understand a complex dynamical system such as financial markets. In 2003, Feigenbaum wrote an instructive review, accessible to both economists and physicists, to some of the recent theoretical approaches employed by physicists in this literature, and compared them with methods used by pure economists. In this paper, presented at the XIII International Conference on Banking and Finance held at the University of Rome "Tor Vergata," we tackle the problem of understanding how exogenous news and analysts' views may influence random economies with many interacting agents. The standard approach of micro-economic theory is to consider preferences of individual economic agents as fixed initial data. In other words, as stated by Koopmans (1957), an agent is not allowed "to indulge in a certain randomness in his responses to given circumstances." But what if this stringent condition is relaxed? It is still possible to lay down theoretical results bearing some usefulness for the understanding of market dynamics? A careful analysis of the economic literature shows some essays characterizing the behavior of random economies with many interacting agents. Hans Föllmer's work in 1973 is, with evidence, one of the first to tackle this problem by allowing random preferences of economic agents (Hildenbrand, 1974) and assumes the probability of law governing that randomness dependent on the agent's environment. Such systems allow for bull/bear market phases induced by local imitation, which tend to organize the investors as in a rump-up fashion process. The higher the aggregation process the higher the fear for a potential trend inversion, bringing the system progressively toward an extremely instable configuration, which may lead to a crash or a local dip. This follows a fundamental result of statistical mechanics that states that the statistical distribution of the average dynamics of interacting complex systems is independent of the detailed specification of agent–agent interactions, as long as they interact. As a consequence, even if it is almost impossible to determine the overall evolution of a particular agent, the macroscopic behavior of the system can be recovered, given the geometry of agents' network from random interactions drawn from adequate probability distributions. There is no need for an exact specification of the microscopic interactions.

This chapter is divided into seven sections, starting with the introduction, Section 1. Section 2 introduces the framework of a binary choice (buy/sell) model, as in Brock and Durlauf. Since in the case of global rationality the model can be solved, we will discuss how single/multiple equilibria may appear from a random economy with many interacting agents. In Section 3 we extend the model to allow local reinforcement by neighbors imitation and a feedback effect in terms of information about the global system status. In Section 4 we define a price process as a function of the imbalance between buyers and sellers. In Section 5 we introduce the effect of exogenous information (bad/good news) on agent's private utilities and show how a flow of positive news may allow the system to reach a higher level of macroscopic organization. Finally, in Section 6 we discuss the results and compare them with market observations. Section 7 concludes.

2. RANDOM ECONOMIES

In the following we will resume the Brock and Durlauf model framework, since it will constitute the basis of our analysis. The fundamental idea is to consider that there are observables accessible both by the agent and the modeler, while "other" observables are accessible only by the agent. Therefore, following Brock and Durlauf, we first consider a population of N agents:

The action s_i of each agent $i = 1, \ldots, N$ belongs to a binary choice set, i.e. $s_i \in \{-1; +1\}$:

The agent's choice is made to maximize a payoff V.

The characteristics available for the choice are: – [–] observables to the modeler and agent $i : O_i$.

[–] Unobservables to the modeler but observables to the agent ε_i, the ε_i are logistically distributed random shocks derived from an idiosyncratic decision process of the agent i, given his available information set.

The ε_i are considered as extreme values distributed, such as:

$$P(\varepsilon_i(1) - \varepsilon_i(-1) \leq x) = \frac{1}{1 + \mathrm{e}xp(-\beta_i x)}$$

Setting aside the econometric practical desire to obtain the random utility term logistically distribution, this assumption can be considered as strong and universal. Indeed, it means that the probability density function of agents' actions $s_i = \{-1; +1\}$ are distributed according to a Maxwell-Boltzmann distribution function and the β term can be interpreted as a *market temperature* describing the degree of randomness in the behavior of agents. In complex physical systems, the Maxwell-Boltzmann distribution is independent of the detailed specification of agent–agent interactions, as long as they exist. Therefore this fact endows the Maxwell-Boltzmann distribution with universality. As a consequence, even if it is almost impossible to determine the overall evolution of a particular agent, the macroscopic behavior of the system can be described from random interactions drawn from adequate probability distributions, even without the exact specification of the microscopic interactions.

The individual decision process is defined by:

$$min_{\{s_i\}} - V(s_i, O_i, p_i(s_{-i}), \varepsilon_i)$$

where s_{-i} denotes the vector of all choices except i. Therefore $p_i(s_{-i})$ denotes the individual's beliefs concerning the choices of other agents, which is assumed independent of the realization of any ε_i.

Assume that the payoff function can be decomposed as:

$$V(s_i, O_i, p_i(s_{-i}), \varepsilon_i) = U(s_i, O_i, p_i(s_{-i})) + S(s_i, O_i, p_i(s_{-i})) + \varepsilon_i$$

Then assume:

$$S(s_i, O_i, p_i(s_{-i})) = -E_i\left\{\sum_{i\neq j}\frac{J_{ij}}{2}(s_i - s_j)^2\right\}$$
$$= \sum_{i\neq j} J_{ij}(s_i E_i\{s_j\} - 1) \tag{1}$$

as a social utility term.

And assume:

$$U(s_i, O_i, p_i(s_{-i})) = h(s_i, O_i, p_i(s_{-i}))\cdot s_i + k_i = h_i \cdot s_i + k_i$$

as a personal utility term, given the macroscopic behavior of the system seen by agent i.

In the case when agents all possess rational expectations, subjective expectations are replaced by their mathematical counterparts, such as:

$$E_i(s_j) = E(s_j)$$

then we have:

$$E(s_i) = tanh\left(2\beta_i h_i + 2\sum_{i\neq j}\beta_i J_{ij} E(s_j)\right) \tag{2}$$

and it is easy to verify that it admits at least one solution.

This is a well-known result in statistical mechanics because it corresponds to the Bragg and Williams (1934) approximation of the two-dimensional Ising Model (Ising, 1925). It accounts for the fact that if only long-range interactions arising from local imitation are considered, then the model admits analytical solution. It must be noted that in economics, if the axiom of global rationality holds, Equation 2 represent a continuous mapping of $[-1, 1]^N$ to $[-1, 1]^N$ (see Brock and Durlauf) and, by the Brouwer's fixed point theorem, there is at least one self-consistent expectation for this binary choice model. Moreover, for a homogeneous system, such as N identical agents with identical observable characteristics $h_i = h$, $\beta_i = \beta$, $J_{ij} = J$, any M is a self-consistent solution for the average choice if it solves:

$$M = tanh(2\beta h + 2\beta JNM)$$

Let $\xi = 2\beta JN$, then we have the following configurations: 2. [1.] $h = 0$: For $0 < \xi < 1$, the fundamental equilibrium $M^0 = 0$ is unique and stable.

For $\xi > 1$, the fundamental equilibrium $M_0 = 0$ becomes unstable, and two new equilibria appear, the bull market equilibrium M^+ and the bear market equilibrium M^-, which are both stable. At the bull (resp. bear) market equilibrium, more than half of the agents have the status "+1" (resp. "−1"). [2.] $h \neq 0$: For $h > 0$ and $0 < \xi < 1$, there is a unique equilibrium for the system, which is shifted to the bull market phase. By contrast, as $h < 0$, the system shifts to the bear market phase. Moreover, the equilibrium is stable.

For $\xi > 1$, the system has two stable equilibria, m^+, m^- and one unstable, m^* if $|\beta h| < \beta h^* = H_c$, where βh^* is determined by:

$$cosh^2\left(H_c \pm \sqrt{\xi(\xi - 1)}\right) = \xi$$

Finally, if $|\beta h| > H_c$ only one stable equilibrium remains.

3. LOCAL COMPLEMENTARITY AND GLOBAL SUBSTITUTABILITY

Assume that the system admits a structure of nearest-neighbor interacting agents, that is, denote by n_i the number of neighbors of agent $i, \forall i \in$ n-dimensional periodic lattice, and:

$$J_{ij} = \begin{cases} J_{ij} \neq 0 & if\ j \in n_i \\ 0 & otherwise. \end{cases}$$

This term will account for the social interaction $S(s_i, O_i, p_i(s_{-i}))$, and $J_{ij} \geq 0$ measures the strategic complementarity between individual choices and the expected choices of his neighbors.

Moreover we assume that, with respect to the average macroscopic state of the system, the personal utility of each agent may exhibit either a strategic complementarity (if they belong to the minority group) or a strategic substitutability (if they belong to the majority group). In such cases, local interactions will tend to align the expectations of each agent while the interactions with the expected average global state of the system will push agents in the minority to join the majority, and agents in the majority to join the minority. Therefore, let us rewrite the personal utility as (Boenholdt):

$$\begin{aligned} U(s_i, O_i, p_i(s_{-i})) = h_i s_i + k_i &= -\alpha_{ij} C(s_i) s_i \frac{1}{N} \sum_{j=1}^{N} E(s_j) + k_i \\ &= -\alpha_{ij} C(s_i) E(\overline{M}) \cdot s_i + k_i; \end{aligned}$$

with $\alpha_{ij} \geq 0$. Moreover, we choose $C(s_i)$ such that:

$$C(s_i) = \begin{cases} +1 & if\ s_i\ belongs\ to\ the\ majority\ i.e.: sign(s_i) = sign(\overline{M}) \\ -1 & Otherwise; \end{cases}$$

In this case, the payoff V admits both a strategic complementarity [ ] between individual choices and the expected choices of others:

$$\frac{\partial^2 V(s_i, O_i, p_i(s_{-i}), \varepsilon_i)}{\partial s_i \partial E_i(s_j)} = J_{ij} \geq 0$$

and a term in the private utility, which can exhibit either strategic complementarity or substitutability (Cooper and John, 1988) between individual choices and the expected average status of the system $E(\overline{M})$:

$$\frac{\partial^2 V(s_i, O_i, p_i(s_{-i}), \varepsilon_i)}{\partial s_i \partial E(\overline{M})} = -\alpha_{ij} C(s_i)$$

Finally, we assume that the best expectation of agent i about the state of his neighbor j, is the observed state s_j, i.e.:

$$E_i(s_j) \equiv s_j$$

and:

$$\overline{M} = \frac{1}{N} \sum_{j=1}^{N} s_j$$

Given the assumed distribution of $\varepsilon(s_i)$, we deduce that:

$$\begin{aligned} P(s_i = +1) &= \frac{1}{1 + exp(-2\beta \cdot [\sum_{<ij>} J_{ij} s_j - \alpha_{ij} C(s_i) \overline{M}])} \\ P(s_i = -1) &= 1 - P(s_i = +1); \end{aligned} \tag{3}$$

Lets consider a network of N agents. Each agent i, located in a node of a network, is modeled by a simple state or spin. We will start from a simple two state model, such as $s_i = \pm 1$, $s_i = 1$ representing a buyer and $s_i = -1$ a seller [ ]. Eventually, as in Iori, we may introduce a “0” state to identify locally inactive traders. We impose on the network a periodicity condition, for example, agents on the boundaries of the lattice are connected north–south and east–west. This corresponds to a torus.

Each trader interacts with his nearest neighbors with a constant interaction energy $J_{ij} = J$, if sites i, j are directly connected and $J_{ij} = 0$ is not. Also, each agent has access to the global average state of the system. Therefore, in this model there are two kind of actors, those who tend to mimic the majority ($C_i = -1$) and those who try to take advantage, being in the minority ($C_j = +1$). There is always the probability (Equation 3) of a switch from majority to minority players but this can happen only at a cost. The transition rule is that a trader who is in the majority group will try to switch to the minority one in order to take maximum advantage of a future fashion movement (buy low, sell high). On the other hand, a minority agent can be dissatisfied with his returns and may eventually decide to join the crowd. In that case, as exposed in [ ], a majority agent i will always act with a strategy spin $C_i = +1$, while a minority agent j will act with a strategy spin $C_j = -1$. If this applies continuously, the payoff of agent i can be written as:

$$V(s_i, O_i, p_i(s_{-i}), \varepsilon_i) = -\alpha_{ij}|E(\overline{M})|s_i + \sum_{<ij>} J_{ij} s_i E_i(s_j) + \varepsilon(s_i) + k_i \tag{4}$$

Here, $\alpha_{i,j}$ can be seen as the average coupling impact of the expected global state $\overline{M}$ on agent i.

This simple model gives rise to non-trivial dynamics. A high level of imbalance between buyers (+1 state) and sellers (−1 state) generates an metastable system with global structures created by the local herding. Those structures may be followed by a sudden order disruption and rapid rearrangement typical of over-critical systems. The characteristics that are stylized in this toy market model are: "Local reinforcement by neighbors imitation."

- *Global feedback effects in terms of information about the global system status.* Minority/Majority players provide a non-vanishing probability to jump from one group to the other. This is an important aspect since a trend follower's desire will be to exit (resp. enter) the market before the crowd. Therefore a successful trader will be in the minority when changing his position and in the majority during inactivity.
- *A idiosyncratic behavior in the decision process seen as a thermal noise.* This noise is introduced to take into account the bounded rationality of an agent. A real market player cannot act following a fully rational expectations otherwise, for an agent to have rational beliefs, he has to have an idea of what all the other agents are going to do. They in turn have to know about what he is going to do, which means he also has to have beliefs about their beliefs

about what he is going to do! And they need to have beliefs about these beliefs and so on, leading to an infinitely complicated problem that no economist, let alone the average investor, can solve (Feigenbaum, 2003). Therefore, in the decision process, a market actor must introduce a subjective estimate on the way to act, such as a noise with respect to the global system and his neighbors' state.

The dependence of the payoff with the average state of the system produces a feedback mechanism that may introduce Self Organized Critical behavior (Bak *et al.*) and intermittency in the dynamics. This construction tends to mimic, in a simple and stylized manner, the influence of an aggregate price process on individual investors, which in turn take their positions according both to the feeling of their nearest neighbors and their own idiosyncratic beliefs on the macroscopic evolution of the system.

The effect of the macroscopic price process on the individuals may act in different ways according to the agent's strategy and behavior. The agent may be in a state of noisy trader aligning his/her actions to those of the majority. Conversely, the agent may decide on a contrary position, deciding to act as the minority does. Moreover, the role of the heat-bath dynamics is to model the relative strength in the decision process between the idiosyncratic feeling of an agent (which can be seen as pure thermal noise) and the local/global state of the system.

This model specification, assuming also a lattice structure, will allow us to simulate the dynamics of the system. It remains to define how prices may be created in such a toy market.

3.1. The price process

As proposed by Kaizoji *et al.* (2002), we may think of two groups of market players and a clearing system mechanism. Fundamentalists will produce an order size X^F proportional to the misalignment of the price $p(t)$ with respect to a "fundamental" price $p^*(t)$, therefore:

$$X^F(t) \propto \ln(p^*(t)) - \ln(p(t))$$

On the other hand, noise trader orders X^N are proportional to the average imbalance measured as $M(t)$, where N is the number of interactive traders, therefore:

$$X^N(t) \propto M(t)$$

The clearing mechanism imposes that:

$$X^N(t) + X^F(t) = \ln(p^*(t)) - \ln(p(t)) + \lambda M(t) \equiv 0$$

Assuming, for simplicity, $p^*(t) = p = 1$, then one obtains:

$$r(t-1,t) = \ln\left(\frac{p(t)}{p(t-1)}\right) = \lambda[M(t) - M(t-1)]$$

Therefore, as a first approximation we assume the increments of the global average imbalance as a proxy for the prices log-returns in the artificial market.

So it is worth noting the recent proposition made by Cross *et al.* In their recent preprint they propose a similar price mechanism based on the average imbalance between buyers and sellers, i.e.:

$$\begin{aligned}
&\sigma(t) = M(t) \\
&p(t) = p(t-1) \cdot exp(\sqrt{\tau}\Delta W(t) + k\Delta\sigma(t))
\end{aligned}$$

where $W(t)$ represents the creation of new, uncorrelated and globally available information over a time period τ. The variable $\Delta\sigma(t) = \sigma(t) - \sigma(t-1)$ is the most recent change in market sentiment and the constant $k > 0$ determines the average effect that a single agent has on the market price. The larger the value of k, the more the market price is influenced by internal market dynamics as opposed to the generation of new market information. This approach is slightly weaker since there is no general global mechanism turning the macroscopic state into a microscopic contribution to the agent's action flip mechanism.

4. EXOGENOUS INFORMATION

The model allows us to introduce an additional source of randomness, such as an exogenous source of global information I_ε:

$$V(s_i, O_i, p_i(s_{-i}), \varepsilon_i) = -\alpha_{ij} s_i(|E(\overline{M}) \cdot | + I_\varepsilon) + \sum_{<ij>} J_{ij} s_i E_i(s_j) + \varepsilon(s_i) + k_i \quad (5)$$

In this case, we allow for two kind of information:

- **Bad news**: $I_\varepsilon(t) < 0$ will force players to pay more attention to the global evolution eventually decreasing the characteristic correlation length of the system. This tends to destabilize the system. However, if the system is in a stable configuration (no imbalance: $|M(t)| \simeq 0$), bad news can be absorbed by the system without crashing.
- **Good news**: $I_\varepsilon(t) > 0$ will make players more self-confident in listening to their neighbors. Enthusiasm driven by good news may eventually lead to a higher level of organization. The higher limit of metastable under-critical configuration, produced by good news, can eventually amplify the reversion process induced by the spontaneous order breaking—a bigger crash.

The previous observations can intuitively be derived from the equilibrium analysis we made by inspecting Equation 2. Indeed a large imbalance in the system will produce a higher value of $|E(\overline{M})|$ and eventually this could push the

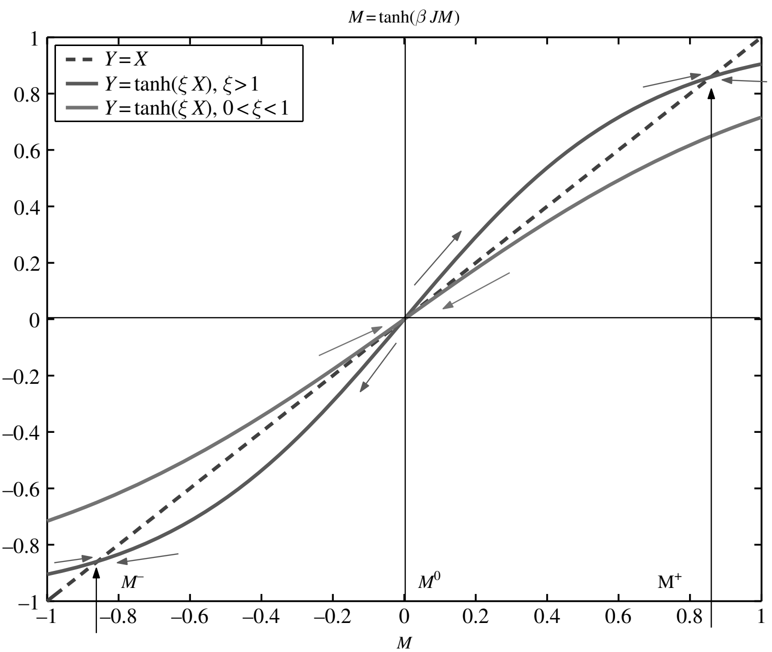


Fig. 5.1. A graphical solution for the number of equilibria in a binary choice model, with interactions assuming personal indifference, is $U(-1) = U(+1)$

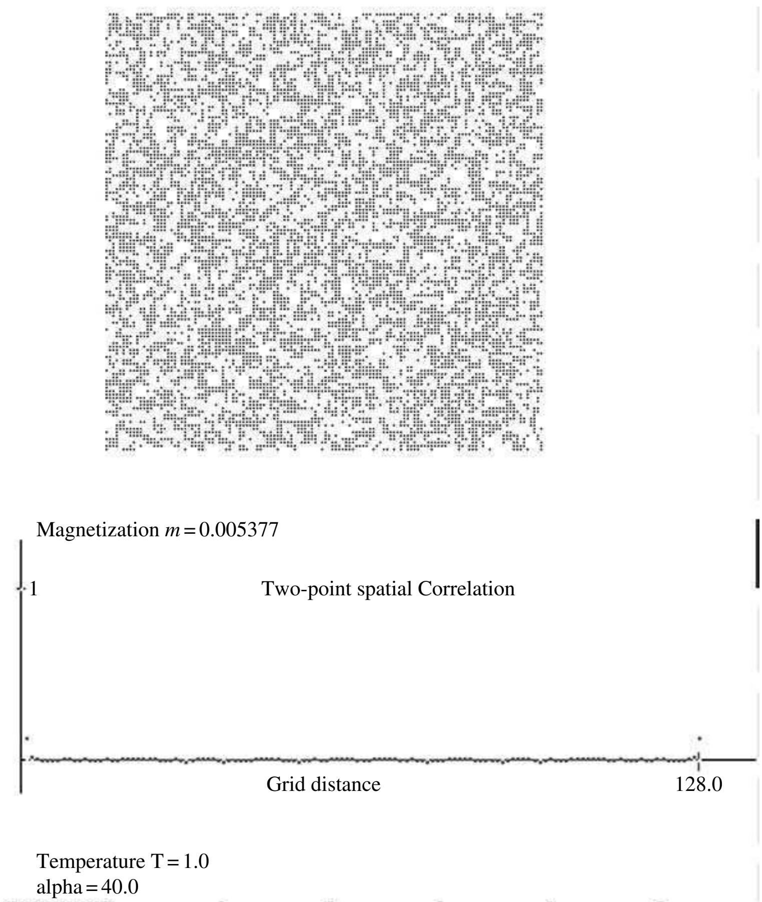


Fig. 5.2. T = 1: Initial State

personal utility term above the critical value H_c, disrupting the multiple equilibria setup. Conversely, a flow of good information will moderate the influence of $|E(\overline{M})|$, allowing the system to reach a higher level of macroscopic organization. This can be seen in the numerical tests obtained by applying the simulation method described in the next section. The system in its initial status is shown in Figure 5.1 and, after N time steps, the system is in an ordered phase (Figure 5.2). An increase in the order of the system at $N + 1$ steps may destroy the internal organization, as seen in Figure 5.3.

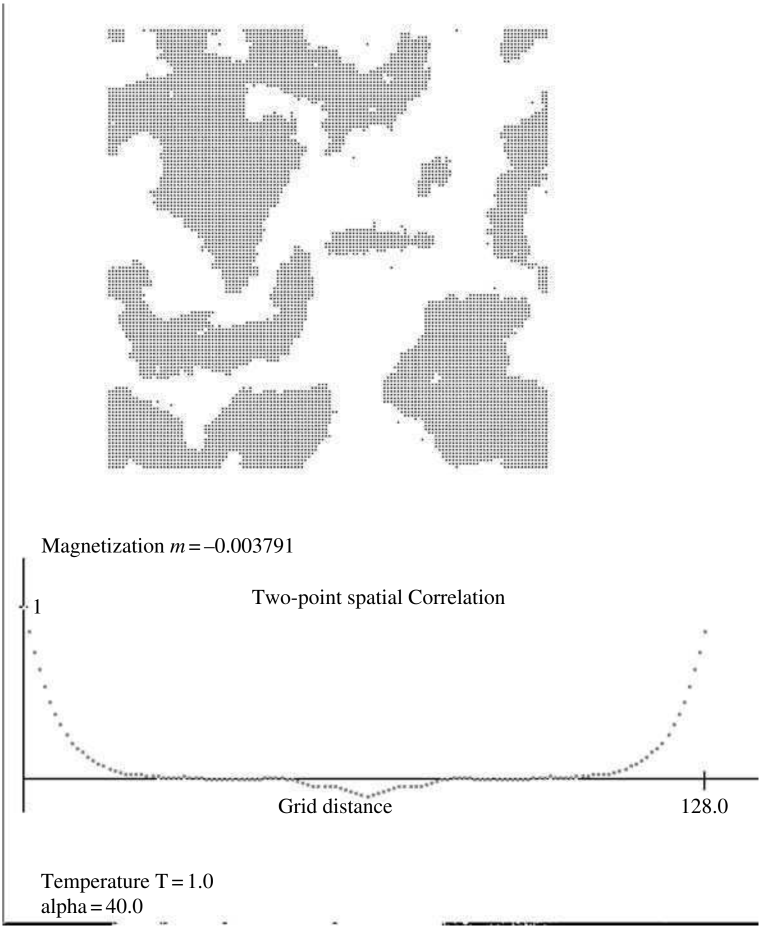


Fig. 5.3. T = N: Ordered phase

5. DISCUSSION OF THE RESULTS

First, we observe in the real-Time Metropolis simulations a number of interesting features, which are shared with real market phenomenology. In Figure 5.4, the evolution of the daily prices for the S&P500 is reported. The time windows spans from 1982 to 2004.

It is interesting to note that the crash induced by the September 11, 2001 terrorist attack has a relative amplitude that is five times smaller than the jump of Black

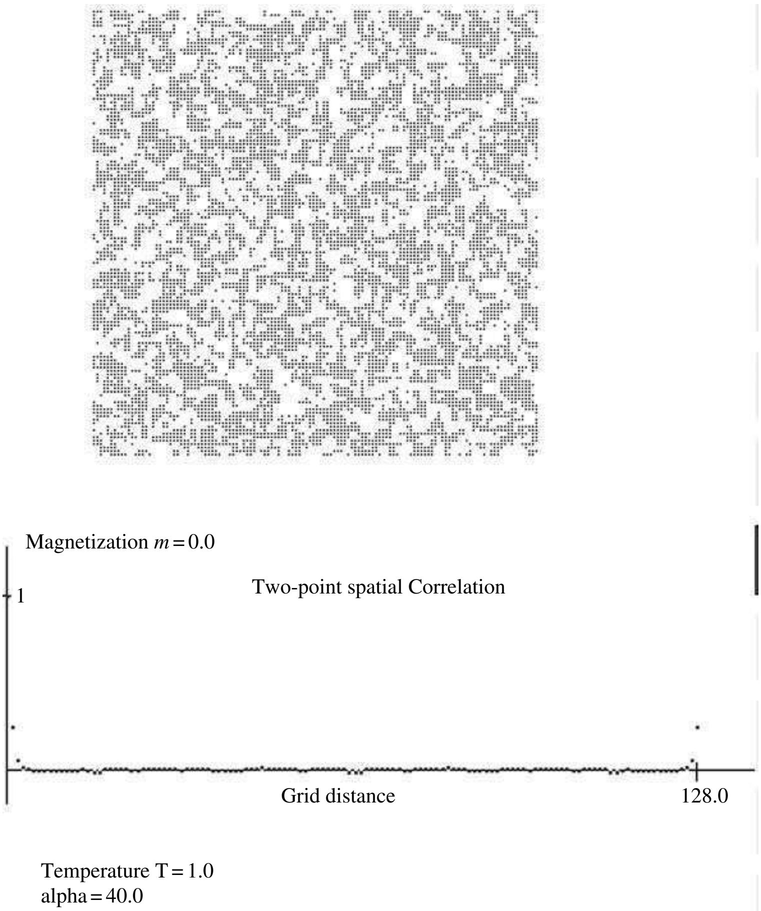
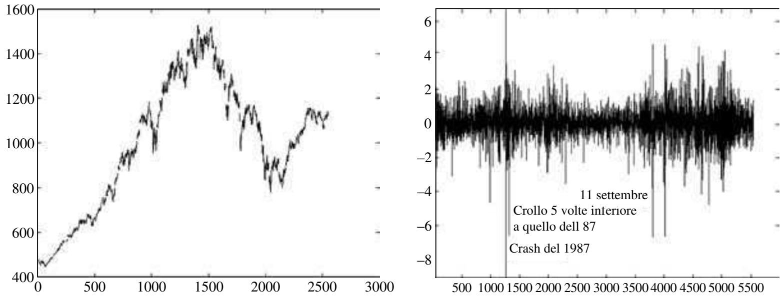


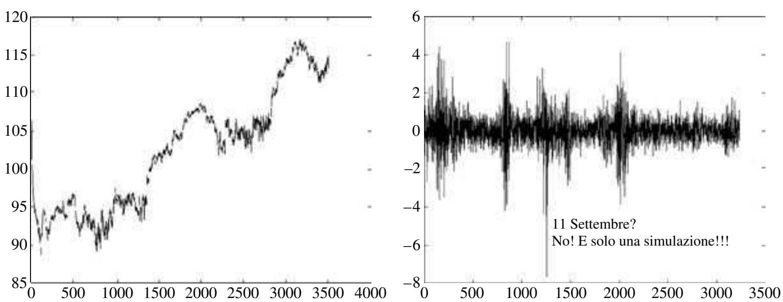
Fig. 5.4. T $= N + 1$: Ordered phase destroyed

Monday 1987, where the market had a drawback of almost 20% (Figure 5.5). Indeed, the terrorist attack took place in a period when the inflationary bubble of the late 1990s already started to reabsorb (Bagella, *et al.*). In the light of the model presented here, we may imagine that the maximum metastable state has been reached by the year 2000, when the dot.com bubble started to collapse, and at the time of the 9/11 attack, the system had already reached a state stable enough to absorb the shock. It seems that significant price fluctuations may not necessarily be related to the arrival of information or variations in fundamental



(a) (b)

Fig. 5.5. (a) s&P500 evolution from 1982 to 2004; (b) Normalized log returns



(a) (b)

Fig. 5.6. (a) Simulated price evolution; (b) Normalized log returns

economic variables as reported in Figure 5.6. As a comparison, in Figure 5.7 we show the results of a simulation. We took a grid of 128×128 agents with $\alpha = 40$ and a idiosyncratic noise equivalent to a temperature $T = 1°K < T_c$. We chose a temperature T smaller than the critical temperature T_c, since we wanted a herding process to take place. For temperatures higher than T_c, the idiosyncratic noise would be high enough to dominate the dynamics. We want agents to listen to their neighbors, not only to their subjective decision process. Listening to neighbors moves the system towards a high level of imbalance,

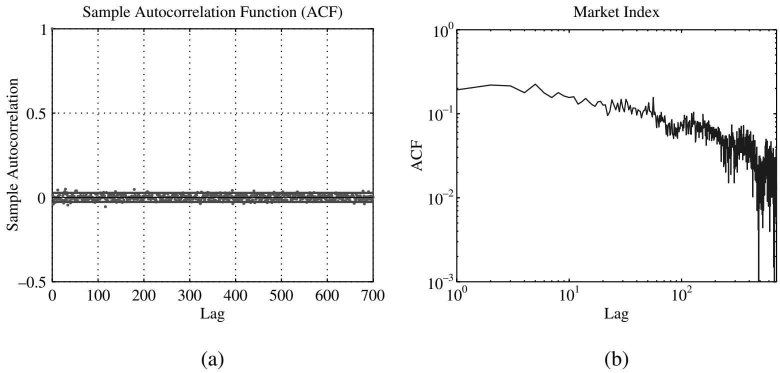


Fig. 5.7. (a) S&P500 returns autocorrelation; (b) Autocorrelation of the absolute returns

making it riskier to remain with the majority of the crowd. As the order of the system increases, people start to be increasingly influenced by the global information, waiting for the decision to jump out of the majority. The interesting observation is that such a system may naturally evolve to a metastable correlated configuration, where the smallest perturbation may trigger a massive migration from the ordered metastable state to a less ordered but stable one. The system is also able to reproduce the persistency observed in the autocorrelation of the absolute returns (Figure 5.8) while, as in the market signal shown in Figure 5.9, there is no autocorrelation for the returns. We observe also that there is an anal-

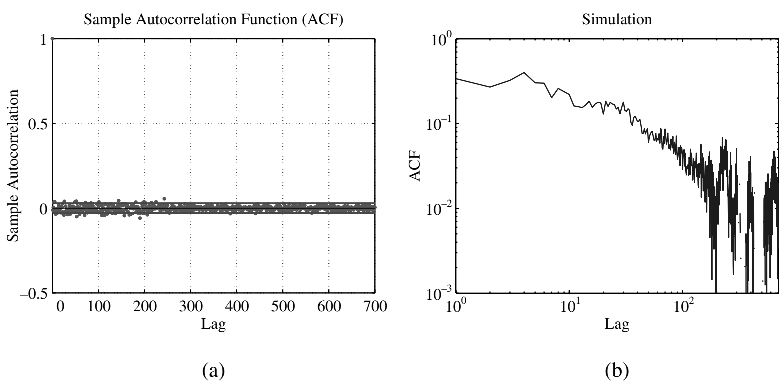


Fig. 5.8. (a) Model returns autocorrelation; (b) Autocorrelation of the absolute returns

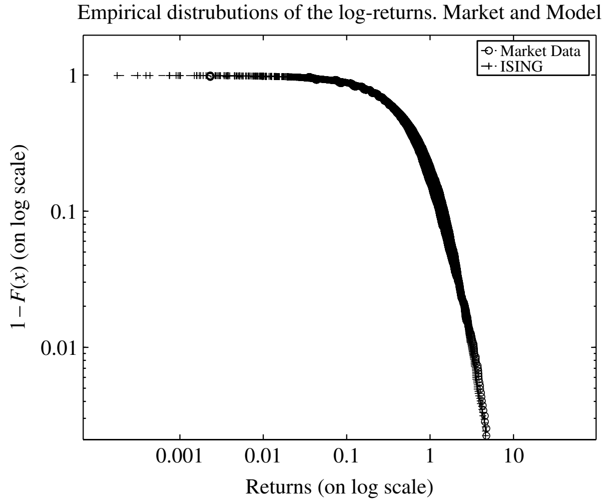


Fig. 5.9. Empirical distribution

ogy between the absolute value of the imbalance $|M(t)|$ and the volumes in the market, and market instability should coincide with high volumes exchanges.

Finally, we observe that such a simplified model also gives rise to a distribution of returns with fat tails. Given a high threshold u, the distribution of excess values of $|r(t)|$ over threshold u is defined by:

$$F_u(y) = Pr(|r(t)| - u \leq y || r(t)| \geq u) = \frac{F(y+u) - F(u)}{1 - F(u)}$$

which represents the probability that the value of $|r(t)|$ exceeds the threshold u by at most an amount y, given that $|r(t)|$ exceeds the threshold u. Following a theorem of Belkema and de Haan (1974), for a sufficiently large threshold u, the distribution function of the excess may be approximated by the generalized Pareto distribution (GPD) such that, as the threshold increases, the excess distribution $F_u(y)$ converges to the GPD, which is:

$$G(x) = \begin{cases} 1 - \left(1 - \xi \dfrac{x}{\beta}\right)^{-1/\xi}, & if \ \xi \neq 0 \\ 1 - e^{-x/\beta}, & if \ \xi = 0 \end{cases}$$

The parameter $1/\xi$ is called the tail index. For $\xi > 0$, the distribution is heavy-tailed and $E[x^t]$ is infinite for $t \geq 1/\xi$. In most market time series, ξ lies between 0.25 and 0.5.

In our example, with a cut-off of 1.05 for the financial and synthetic Ising time series, we observe: $\xi = 0.24 \pm 0.06$ and $\xi = 0.33 \pm 0.06$, respectively.

6. ARE TRADERS MAD?

If, on the one hand, we live in a infinitely rational world, we may be tempted to reject the approach presented here. After all, how can a rational human being act in such a way to contribute to burn a fortune? Indeed, it can happen merely because infinite rationality, when thinking in a multi-period, multi-agents economy, cannot be applied efficiently to each agent. Each decision is always taken, subjectively conditioned to the available information. This margin of subjectivity opens up the probability of being a better/worse player than one's neighbors. To win the game, each player must be on the minority side just before the crowd is attracted by this side, and then with the majority during the fashion rump-up. As the game proceeds, the neighbors imitation tends to become more dangerous, since all the players are looking for the right moment to jump out from the majority to join the minority. Think about gain consolidation or stop loss. During the herding phase, all agents have a tendency to imitate each other, and the more the imitation orders the system, the higher the risk of a rapid inversion. At a given stage, each majority market player wants to satisfy his appetite for future gains but has also to deal with his increasing fear. In critical conditions, the system is so stressed that even the smallest perturbation may trigger a catastrophic reaction. In that way each agent can behave rationally, acting upon available information, but the aggregate behavior can produce a catastrophic event giving rise to a irrational but natural crash.

6.1. Empirical distribution

Therefore, it is tempting to think that extreme market movements may be an intrinsic metastable dynamics of the system, induced by a collective phenomena such as crowd effects or "herd" behavior.

7. CONCLUSIONS

In a driftless market, with no definite impact of analysts' view on implicit risk premiums, stylized behavioral models help to shed some light on the news/volumes/books/prices dynamical relationship in the view of an operational characterization of market dynamics (Figure 5.10). It appears that, from a purely speculative approach, the mathematical investigation of behavioral group dynamics allows for a deeper understanding and efficiency enhancement of investment strategies, such as Equity Market Neutral or Statistical Arbitrage. These are strategies where investment decisions are helped by the identification of market inefficiencies through the application of specific mathematical models. From a regulatory point of view, this approach can help to design stability indicators and gives a useful tool to investigate news' impact on the financial system. This in turn can help design more efficient information policies. As quoted by Feigenbaum (2003), financial managers only concerned with determining an optimal portfolio of investments given the current economic climate, may neglect the behavior of other market players only if their behavior remains fixed. If this is not the case, as recent studies indicate, they should be wary of behaviorist models. This is because they are vulnerable to the Lucas (1976) critique "if the behavior of agents depends on the economic environment, one cannot correctly assess the response of changes to environmental parameters with a model that holds the behavior fixed." In this context, Statistical Mechanics may provide useful tools for the investigation.

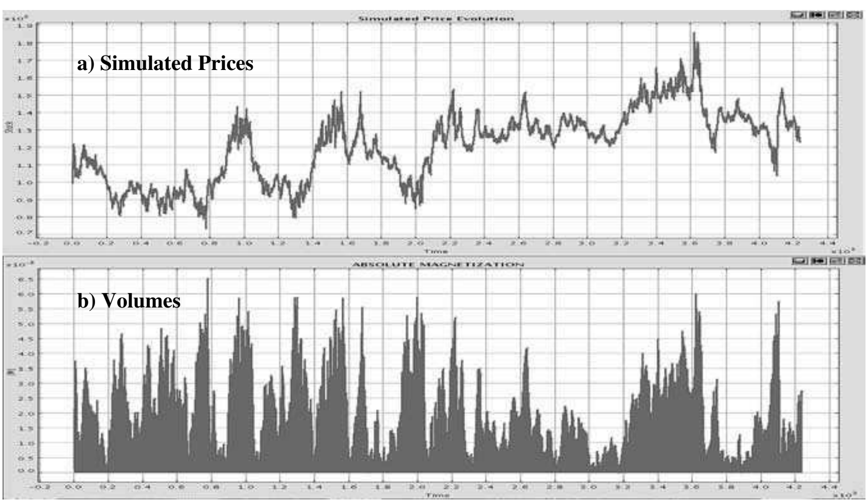


Fig. 5.10. Prices and Volumes joint evolution

7.1. Prices and Volumes joint evolution

The Ising Model is one of the simplest fundamental models of statistical mechanics. It is used to describe phenomena such as magnets; liquid/gas co-existence; alloys of two metals and, probably, bearish/bullish market players. These systems are described by local spins with values $s_i = \pm 1$ corresponding to a binary choice set. The choices can be interpreted as up/down spins, atom A or atom B in an alloy, buy/sell, etc. These variables generally describe the status of an element in a site i of a lattice. The Ising model, even in its simplest form, is extremely powerful in describing the *Order/Disorder* phase transitions.

The macroscopic behavior of a system depends on its lattice structure but only a limited set of elementary configurations admit an analytical solution. However, even the most stylized description of the system may be useful to the investigation. The frequent use of two-dimensional lattices in Ising models, as a first approximation, is often the first step toward the understanding of the origins of observed macroscopic dynamics.

Models akin to statistical mechanics have only recently been introduced in micro-economic theory by Föllmer (1974) and the Economic literature on this subject is still scarce. As in Brock and Durlauf, we show how under the assumption of perfect, global rationality the macroscopic equilibria deriving from local strategic complementarity can explain a fashion rump-up at the origin of a collective behavior responsible for bull/bear market phases, and how information can generate a bull $\leftrightarrow$ bear transition. The physical interpretation of global rationality hypothesis is that it assumes that "there is no short-range order apart from that which follows from long-range order" corresponding to the well-known Bragg and Williams (1934) approximation in statistical mechanics.

By considering a two-dimensional n-periodic square lattice, we perturb the theoretical construct of Brock and Durlauf by allowing both local imitation, by strategic complementarity between agents, and global substitutability between agents and aggregated market status. In such a case we obtain a system that admits intermittency in dynamics without reaching a defined equilibrium. If we consider the organization induced by local imitation (e.g. a market index, an aggregated price process, etc.), as an additional source of information available to the agent, a metastable dynamics is recovered, which admits the same stylized facts observed from empirical analysis of market evolution.

In our configuration, each agent sits on a node of a grid with four other neighbors. So there is no hierarchy or cluster structure on the connections of each player. By changing the network configuration, moving toward more complex connections, the possibility to compute analytically the characteristic parameters of the system (such as the critical temperature T_c or β_c) is almost lost. Therefore one has to rely on numerical simulations in order to investigate the dynamical properties of the system.

In 1998, Cont and Bouchaud proposed a simple model where a random communication structure between agents may give rise to the stylized facts observed in empirical studies of high-frequency market data. From an empirical viewpoint, Bonanno *et al.* (2003) analyzed the topological characterization of the minimal spanning tree (National Institute of Standards and Technology) obtained by considering the price returns correlations in stock markets. In this case, they find that the empirical tree has features of a complex network that cannot be reproduced, even as a first approximation, by a random market model and by the one-factor model. It is tempting to consider a tantamount relationship between the connection structure they obtain, comparing large correlation matrices of stocks returns, and the distribution of the connections between agents. In this case we could try to map the minimal connected tree obtained, by analyzing the correlated evolution of market returns of a large number of assets into an agents' network of "spin agents" and compare the model's behavior with observed data. We plan to tackle this problem in the near future.

ACKNOWLEDGEMENT

Economics Department University of Rome "Tor Vergata" Computational and Quantitative Finance Finance-and-Physics Economics Department University of Rome "Tor Vergata."

BIBLIOGRAPHY

Bagella, M., Becchetti, L. and Ciciretti, R. *Market Analysts Reaction: The Effect of Aggregate and Firm Specific News, mimeo*.

Bak, P., Tang, C. and Weisenfeld, K. "Self-organized criticality." *Phys. Rev. Lett.* **59**: 381.

Bannerjee, A. (1992) "A simple model of herd behavior." *Quarterly Journal of Economics* **107**: 797–818.

Belkema, A.A. and de Haan (1974) "Residual lifetime at great age." *Annals of Probability* **2**: 792–804.

Berlevy, G. and Veronesi, P. (2003) "Rational Panics and Stock Market Crashes." *Journal of Economic Theory* **110(2)**: 234–63.

Bonanno, G. *et al.* (2003) "Topology of Correlation-Based minimal spanning trees in real and model markets." *Phys. Rev.* E **68**.

Bornholdt, S. *Expectation bubbles in a spin model of markets: Intermittency from frustration across scales*, cond-mat/0105224.

Bragg, W.L. and Williams, E.J. (1934) *Proceeding of the Society of London Series A* **145**, 699.

Brock, W.A. and Durlauf, S.N. *Interaction-based models*, Handbook of econometrics vol. **5** Elsevier.

Cont, R. and Bouchaud, J.-P. *Herd Behavior and Aggregate Fluctuations in Financial Markets*, condmat/9712318.

Cooper, R. and John, A. (1988) "Coordinating Coordination Failures in Keynesian Models." *The Quarterly Journal of Economics*, **103(3)**: 441–63.

Cross, R. *et al. A Threshold Model of Investor Psychology*, preprint http://math.gmu.edu/~harbir/cgls092004b.pdf.

Cutler, D.M. *et al.* "What Moves Stock Prices?" *J. Port. Manag.*, Spring, 4–12.

Feigenbaum, J. (2003) "Financial Physics," *Rep. Prog. Phys.* **66**: 1611–49.

Föllmer, H. (1974) "Random Economies with Many Interacting Agents." *J. Math. Econ.* **1**: 63–6.

Hildenbrand, W. (1974) "Random Preferences and Equilibrium Analysis." *Journal of Economic Theory* **3**: 414–29.

Iori, G. *A microsimulation of traders activity in the stock market: The role of heterogeneity, agents interactions and trade frictions*, http://netec.mcc.ac.uk/WoPEc/data/Papers/wpawuwpfi9905005.html.

Ising, E. (1925) *Z. Phys.* **31**: 235.

Kaizoji, T. *et al.* (2002) "Dynamics of price and trading volume in a spin model of stock markets with heterogeneous agents." *Physica A* **316**: 441–52.

Koopmans, T.C. (1957) *Three Essays on the State of Economic Science*, New York, Toronto, London: McGraw-Hill.

Lucas, R.E. (1976) "Econometric policy evaluation: A critique." In: Brunner, K. and Maltzer, A.H. (eds) The Phillips Curve and Labor Markets. North-Holland Publishing Company.

Metropolis, N., Rosembluth, A., Rosembluth, M. and Teller, A. (1953) "Equations of state calculations by fast computing machines." *J. Chem. Phys.* **21**: 1087.

National Institute of Standards and Technology, NIST, http://www.nist.gov/dads/.

Orlané, A. (1995) "Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion." *Journal of Economic Behavior and Organization*, **28**: 257–74.

Osanger, L. (1944) *Phys. Rev.* **65**: 117.

Shiller, R. (1989) *Market Volatility*. Cambridge, MA: The MIT Press.

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 6

CONDITIONAL ASSET PRICING MODEL: AN APPLICATION TO THE KOREAN STOCK MARKET

Sungho Choi, Rifat Gorener and John R. Norsworthy*

Abstract

This chapter investigates the risk-return relationship using the 4-state model on the Korean Stock Market. The 4-state model tests some elements of Kahneman and Tversky's (1979) prospect theory. We obtained data from the Korea Stock Exchange (KSE), which comprised daily stock returns of 40 randomly chosen companies from KOSPI 200 for the period of 1991 to 2001. We find that the 4-state model gives a higher explanatory power, implying strong reference dependence. We also find that rotation of the coordinate axes further improves the explanatory power. Finally, we find asymmetric valuation of gains and losses in the data for all the companies.

Keywords: prospect theory, asset pricing, reference dependence, 4-state model

JEL Classification Codes: D8, G1.

1. INTRODUCTION

The capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) predicts that equilibrium prices will be set such that expected returns in excess of the risk-free rate will be proportional to the covariance with aggregate risk that is measured by the return on the market portfolio. Since these pioneering

*Corresponding author

studies, much further research has been done on the subject of asset pricing modeling. Even though it has two major attractions, simplicity and implications, empirical evidence does raise serious questions about the CAPM (and the market model). Ross (1976) offered a model of security pricing known as arbitrage pricing theory (APT) that allows for multiple sources of risk. Although the APT is subject to less criticism, the number of factors and the nature of the risk factor are not defined by the APT theory itself. Also important, as discussed in this chapter, the frequency of observations is not directly addressed.

Zin (2002) argues that the recent interest in behavioral finance models of asset pricing is motivated partly by a desire to have models that appear realistic in light of experimental evidence, and partly by their success in moment-matching exercises. Kahneman and Tversky (1979) pioneered the application of prospect theory in behavioral economics. In developing their prospect theory, they argue that people are not as rational in their calculations as economic models assume. Instead, they argue that people repeatedly make errors in judgment that can be predicted and categorized. Based on Kahneman and Tversky's (1979) prospect theory, Norsworthy *et al.* (2003) examined the risk-return relationship in the U.S. market by applying the 4-state model, which uses a four-way partition of the observations based on expected values of individual asset and market returns. The 4-state model captures the most important element of Kahneman and Tversky's (1979) prospect theory – dependence of expected returns on the current frame of reference. They also show the asymmetric valuation of gains and losses of equal size and also diminishing marginal sensitivity to gains and losses as described in Tversky and Kahneman (1991).

Norsworthy *et al.* (2003) measure the descriptive power of the 4-state model and compare it with the market model. They achieve this by analyzing daily returns data for 100 individual stocks during 1984–98: the 30 Dow-Jones industrials, and samples of 30 and 40 stocks from the S&P mid-cap and S&P small-cap index lists. They report that the four-way partition doubles the explanatory power of the conventional asset pricing model (APM), the market model for large cap stocks, and improves that of medium and small cap stocks by more than five times on average. The explanatory power is increased further when the partitioning procedure is adjusted by rotation of the coordinate axes to reflect expected asset and market returns. Because the partitioned model increases measured systematic risk, diversifiable risk is lower than that which that conventional models predict.

In this chapter, we examine the risk-return relationship of the Korean stock market using that 4-state model. Since they have showed the robustness of 4-state model against the market model for the U.S. market, we extend the test of the 4-state model to an international market to further test its validity. We expect that this model will yield a significant improvement in explanatory power over the market model. We also expect that the rotation of the coordinate axes will improve the explanatory power greatly, and finally expect that there will be asymmetric valuation of gains and losses.

We obtained data from the Korea Stock Exchange (KSE), our data set comprising daily stock returns of 40 randomly chosen companies from KOSPI 200 for the period of 1991 to 2001. We find that the 4-state model gives higher explanatory power, thus implying a strong dependence in a short-term reference frame. We also find that the rotation of axis improves the explanatory power further. The result appears both in the tabulated results and in the visual evidence of plots of the observations and fitted values. Finally we find asymmetric valuation of gains and losses in 1991–2001 data for all 40 companies. The symmetry hypothesis, concerning valuation of positive and negative returns of equal magnitude, is rejected.

The rest of the chapter is organized as follows: Section 2 presents a brief review of the literature related to prospect theory in finance. Section 3 presents a history of the South Korean market system and summarizes the relevant regulations. Section 4 describes the data and methodology and present summary statistics. Section 5 presents the results of the analysis on the Korean market using the 4-state model. Concluding remarks are presented in Section 6.

2. LITERATURE REVIEW

Kahneman and Tversky (1979) have formalized loss-averse behavior and certain other apparent anomalies as behavioral elements in their "prospect theory." In prospect theory, the utility function is replaced by a valuation function that evaluates changes in expected income from the current level based on a current reference frame that conditions expectations. While increases in income are weighted by a small marginal utility, decreases in income are weighted by a larger marginal utility. Tversky and Kahneman's (1991) value function – which replaces the expected utility value function in the investor's decision – has three characteristics:

1. *reference dependence* (gains and losses are defined relative to a reference point);
2. *loss aversion* (the value function is asymmetric – steeper in the negative than in the positive domain); and
3. *diminishing sensitivity* (the marginal value of both gains and losses decreases with their respective sizes).

Shiller (1998) provides an excellent review of prospect theory in finance, and notes that expected utility theory (and its rational expectations derivative) is still the dominant paradigm for investor decisions in finance and for economic decisions generally. The full range of asset pricing models is surveyed and assessed in Cochrane (2000). Many of those models have been used to identify anomalies in investor behavior. Cochrane (1999) also discusses the broad category of alternative APMs, referred to as multifactor models.

A recent paper by Barberis *et al.* (1999) discusses prospect theory in the context of the national income accounts, incorporating consumption and wealth – changes in wealth – as part of the reference frame. They also provide an excellent guide to the recent financial literature addressing prospect theory and loss aversion, which, if paralleled here, would be redundant. A difficulty with the national accounts context is that the data are infrequently measured (quarterly) and sometimes substantially revised in annual and lower frequency incorporations of new data. Thus wealth, investment, income and consumption estimates often change considerably. The approach here is to omit these admittedly important variables from the model specification, partly in the interest of a much larger data space for analysis and testing models of asset valuation. Another strain in the literature follows the practice of the first Kahneman and Tversky paper, and obtains responses by surveying. This approach seems less satisfactory as an alternative to working with a large panel of historical data based on actual (rather than hypothetical) decisions. As Kahneman and Tversky themselves remark, the responses may be influenced by the phrasing of the questions. Our results have the virtue of reproducibility at some expense in the scope that can be claimed for the results.

Recently, prospect theory has been entering the literature relating to investor behavior. In many cases, however, it carries baggage from its origins in psychology. Many sensible and plausible ideas from that field do not come into finance in forms that are readily testable, particularly with high frequency data. This matter is well illustrated by Shefrin (2000), who identifies and discusses several practices associated with reference dependence, including mental accounting, hedonic editing, regret minimization and (what is more commonly called) isolation. These concepts are perhaps best explored in hypothetical decision situations, as in the original Kahneman and Tversky paper. The scope for analyzing them in historical financial market data seems small.

Allais (1953) asserts that valuation of future outcomes according to mathematical expectations is unlikely, and is subject to the Bernoulli Paradox. He asserts that risk aversion and diminishing marginal sensitivity to large gains and losses is more consistent with human psychology. Friedman and Savage (1948) explain the apparent inconsistency of simultaneously buying insurance and lottery tickets with a utility function that has several inflection points, including one (a kink) at the origin. Their formulation is consistent with loss aversion and diminishing marginal sensitivity. Soros (1994) observes that investor expectations for future asset performance influence future asset prices, and so the exogeneity assumed in the EMH and mathematical expectations theories do not hold. He uses the terms "participant bias" and "reflexivity" to describe the interaction between expectations of investors and realization of market valuation in the future. Pettengill *et al.* (1995) show that expected returns depend on the state of the market. In a bear market, expected returns are lower, but are higher in a bull market. They do not address prospect theory directly. In a similar vein, Jaganathan and Wang (1996) view expected returns in the CAPM context, including the return on human capital and other wealth. They find that higher conditional βs characterize periods of expected positive economic growth, and lower conditional βs define periods of expected decline.

Asymmetric valuation of gains and losses may interact with reference dependence, and affect the responses of both buyers and sellers to changes of price or profit relative to the reference level. The response to change is expected to be more intense when the changes are unfavorable, below expectations, than when the changes are favorable. In this vein, Putler (1988) estimated separate demand elasticities for increases and for decreases in the retail price of shell eggs, relative to a reference price estimated from the series of earlier prices. The estimated elasticities were -1.10 for price increases and -0.45 for price decreases, indicating that price increases have a significantly greater impact on consumer decisions than decreases.

This paper measures the K-T reference dependence in terms of four states, which are defined by combining the directions of movements of the asset price and the market. Benartzi and Thaler (1993) use myopic loss aversion theory to explain the equity risk premium puzzle. Investors are assumed to be loss-averse, meaning that they are more sensitive to losses than to gains. They evaluate their portfolios frequently as if they were operating with a time horizon of about one year, although they have long-term investment goals. In simulations, Benartzi and Thaler find that the size of the equity premiums is consistent with previously estimated parameters of "prospect theory" if investors evaluate their portfolios annually. Odean (1998) finds a *disposition effect*, whereby investors are reluctant to realize losses, and too eager to realize gains, in the records of 10,000 investors at a large brokerage house. This effect contradicts expected utility decision-making. This lock-in effect on the downside is not the same as loss aversion, although it may contribute to, or reflect, the diminishing marginal sensitivity predicted by prospect theory.

3. OVERVIEW OF THE KOREAN MARKET

An historical overview of the Korean stock market is necessary and much of the historical information was adapted from the Bank of Korea (2002). The Korean financial markets have experienced great changes due to active financial liberalization and financial market opening since the early 1990s. The currency crisis at the end of 1997 accelerated the process of change in the Korean financial system. These changes have resulted in an increase of the participation of foreign capital in the domestic financial industry and markets.

The KSE was established in March 1956 and the Korean capital market became an organized market. The legal basis for its operations was provided by the Securities and Exchange Act, which came into effect in January 1962. In addition to the established KSE, an over-the-counter (OTC) market for stocks was established by the Korea Securities Dealers Association in April 1987. The OTC market was reorganized as the Korea Securities Dealers Association Automated Quotation (KOSDAQ) market, a Korean edition of NASDAQ, in January 1997. The Korea Over the Counter Bulletin Board (the third stock market) was launched in March 2000. One of the most interesting features of the Korean stock market is the existence of daily price change limits on each individual stock. The daily return of individual stock cannot exceed 15% in absolute value for the current period, whereas previously the limit was 8% in absolute value.

The government first allowed direct foreign investment in stocks in 1981. In January 1992, the government permitted foreigners to make direct investments in listed stocks within an overall limit for all foreigners of 10% of the total stocks for each individual issue and a ceiling of 3% per issue for a single foreign investor. Finally, the government expanded the investment ceilings on foreign investment in stocks several times, completely lifting them in May 1998, with the exception of investment in public corporations. The initial opening of the bond market took place in July 1994, relatively later than the stock market, with foreign investment being allowed in convertible bonds issued by small and medium enterprises. Then range of bonds permitted to foreign investors was expanded step by step to include non-guaranteed medium- and long-term bonds issued by small and medium enterprises and non-guaranteed convertible bonds issued by large enterprises. All restrictions on foreign investment in listed bonds were, however, abolished in December 1997, immediately after the currency crisis.

The KSE market is a systematic and competitive market operated by the KSE for transactions of securities (stocks, bonds, and mutual funds, etc.). The Korea Composite Stock Price Index (KOSPI), an indicator of prices of all stocks listed on the KSE, represents the stock price level at a comparative point of time, assuming that the level of stock prices at a base point of time is 100. At present, the KSE computes and announces KOSPI based on the total market value, taking the date of January 4, 1980 as the base point of time. The number of listed companies at KSE was 776 at the end of 1997, but the number dropped to 693 by the end of 2001, due to the process of corporate restructuring after 1998. The KOSDAQ market is a secondary market operated by Kosdaq Stock Market, Inc. It has no exchange floor and transactions are made through a computer network system. The KOSDAQ Composite Index, the stock price index of the KOSDAQ market, has been announced since July 3, 1997, setting the date of July 1, 1996 as its base point of time. The index calculation formula is the same as that of KOSPI. The number of companies listed on the KOSDAQ market decreased from 359 at the end of 1997 to 331 at the end of 1998. However, it soared remarkably to 616 at the end of 2001, due to rapid growth in the information technology industry. The Korea Over The Counter Bulletin Board was brought into operation to facilitate transactions in the stocks of companies unable to obtain a listing on the KSE or KOSDAQ or companies whose stocks were delisted from KSE or KOSDAQ.

The major investors in the primary bond market are institutional investors such as banks, investment trust management companies, and insurance companies, and matters related to bond issuance between issuers and investors are mainly handled by securities companies. The outstanding amount of bond issuance at the end of the 1990 was only 59 trillion Won but it has expanded rapidly since 1998, reaching 427 trillion Won at the end of June 2001. This is attributable to the substantially increased issue of Treasury bonds to finance the fiscal deficit, Deposit Insurance Fund Bonds, Non-Performing Loans Management Fund Bonds issued to support the restructuring of financial institutions, and Monetary Stabilization Bonds for absorbing the considerable increase in liquidity brought about by the current account surplus and support for financial restructuring.

In Korea, the KSE and Korea Futures Exchange (KOFEX) make up the institutionalized market. In the KSE, KOSPI 200 futures, KOSPI 200 options, and equity options are available. In the KOFEX, which was opened in April 1999, KOSDAQ 50 index futures and options, US dollar futures and options, CD interest rate futures, and Korea Treasury Bonds futures are traded. In the OTC financial derivatives market, futures exchange, swaps, options, forward rate agreements and other derivative products are traded. A KOSPI 200 futures market was established in the KSE in May 1996. The KOSPI 200 index is a spot index of total market price that was developed by the KSE for futures, and options transactions and this index comprises 200 stocks selected from the listed stocks in the KSE market, by considering such properties as industrial representation and liquidity. The total market price of KOSPI 200 index constituent stocks was approximately 70% of the entire market value of all stocks in KSE market at the time when the index's constituent stocks were selected. A KOSPI 200 options market opened in the KSE in July 1997. There are four contract months which are made up of three consecutive months including the spot month and one additional near-term month from the quarterly cycle (March, June, September, and December). And the last trading day is the second Thursday of each contract month. An equity options market opened in the KSE in January 2002. The listed stocks are Samsung Electronics, KT, SK Telecom, Korea Electric Power Corp., POSCO, Kookmin Bank, and Hyundai Motor Co. The last trading day is the second Thursday of each contract month. There is no price change limit.

Korea's financial supervisory system was reorganized as an integrated financial supervisory system under the Act on the Establishment of Financial Supervisory Organizations, which went into effect in April 1998. Under the system, the Financial Supervisory Commission (FSC) and its executive arm, the Financial Supervisory Service (FSS), are in charge of the supervisory business of almost all financial institutions, including banks and non-bank financial institutions. Major matters that the FSC deliberates and resolves are the formulation and amendment of regulations relevant to the supervision of financial institutions, permission of establishment, merger, conversion or assignment or assumption of the business of financial institutions, permissions relevant to the operation of financial institutions, examination of and sanction against financial institutions and administration, supervision and surveillance of the securities and futures markets. The Securities and Futures Commission (SFC) was established under the FSC to investigate unfair transactions in the securities and futures markets, oversee financial accounting standards and audits, and conduct the duties delegated to it by the FSC of the administration, supervision, and surveillance of the securities and futures markets. Related to this, the SFC instructs and supervises the FSS.

4. DATA, METHODOLOGY, AND HYPOTHESES

The analysis of the 4-state model on Korean stock market is carried out for 5 largest companies plus 35 randomly selected companies from January 1991 through to December 2001, using daily stock returns. The data set is obtained from the KSE and contains date, securities code, opening/high/low/closing price, trading volume, trading value, company name, standard price, number of listed shares, P-E ratio, dividend yield, market value, industry code. We carefully drew a sample from the KOSPI 200 companies. Among KOSPI 200 companies, 139 companies have been a component of the index continuously for the sample period of 1991 through to 2001.

The five largest companies in KOSPI 200, as of December 2001, were Samsung Electronics, SK Telecom, Korea Electric Power, POSCO and Hyundai Motor and the sample includes all of these. The large cap stocks are a random sample of ten drawn from the KOSPI 200, with market cap values between $500 billion and $5 trillion Won on December 31, 2001. The mid cap stocks are a random sample of ten stocks from the KOSPI 200, based on values between \$100 billion and \$500 billion Won at the same date. The small cap stocks are a random sample of 15 stocks from the KOSPI 200 based on values less than \$100 billion Won for the same date. The KOSPI Index serves as the indicator of general market performance for the market models, and for the 4-state variants of the market model. Table 6.1 summarizes the sample description in our sample. In addition to analyze entire sample period, we broke up the sample period into two periods based on the major event in Korean market. The three break points are June 1996, December 1997, and November 1998. The analysis is carried out for the 11-year period for stocks and the hypothesis tests are repeated for the each sub-period.

As explained by Norsworthy *et al.*, the four-way partition of the model is associated with different conditional decision rules consistent with four reference states that may present in the market. These reference states determine the formation of investor expectations, according to current movements in asset and market returns. The partitioning embodies the assumption that investors perceive and expect regularities not only at the aggregate level (across partitions) represented by the simple APMs, but also in the four different states of the market defined by the partition. The 4-state model implicitly assumes that investor expectations are symmetric about the x- and y-axes. That is, that expectations are different according to the asset and current market returns as they lie above or below zero. However, if the investor expects non-zero returns for the asset and market returns, then deviations should be measured above and below these expected values. Therefore, we employ two different rotations: mean rotation and optimum rotation. It should be clearly understood that the rotation of the axes is for purposes of classifying the observations into the four states or reference frames. The *observed* values of r_M and r_A, not the *rotated* values, are the basis for estimating the 4-state model.

We employ the same hypothesis testing as Norsworthy *et al.*, since we want to verify the robustness of the 4-state model for the Korean stock market. The following hypotheses are formally tested by their paper and we test the hypotheses for each stock in all periods using as F-test based on the respective R-squares of the alternative models.

Panel A. The 4-state Conditional Asset Pricing Model

As in conventional asset pricing models, asset and market returns are defined as

$$r_A = \ln(1 + R_A) \ and \ r_M = \ln(1 + \ln(R_M) \tag{1}$$

Table 6.1 Sample Description – Company Names and Ticker Symbols in the Study

Ticker	Old Ticker	DataStreamCode	NAME	Symbol
700203000	50510	314529	ASIA CEMENT MNFG.	ACM
700721000	52000	502636	BYUCKSAN	BYS
701260000	62850	314547	CHUNG HO COMNET	CHC
700104000	18520	756962	CHEIL JEDANG	CJ
700099000	38500	777453	DONGBU HANNONG CHEMICAL	DHC
700021000	75060	756963	DAELIM INDUSTRIAL	DLI
700015000	22500	777413	DOOSAN CORPORATION	DS
700660000	40720	502668	DONGSHIN PHARM.	DSP
700354000	88000	756965	DAISHIN SECURITIES	DSS
701404000	64290	314658	GPS CORP.	GPS
700104000	22510	502644	HITE BREWERY	HB
700983000	37010	756968	HANWHA CHEMICAL	HC
700544000	80050	314679	HYUNDAI DEPT. STORE	HDS
700024000	45500	777883	HANKOOK TIRE	HKT
700538000	67510	756971	HYUNDAI MOTOR	HM
700230000	35010	777882	HANKUK PAPER MNFG.	HPM
700345000	88030	777402	HYUNDAI SECURITIES	HS
701468000	37090	315258	HANSOL CHEMIENCE	HSC
700088000	43000	777395	HANWHA CORPORATION	HW
700639000	50530	777460	HYUNDAI CEMENT	HYC
701051000	66520	314663	ILJIN CORP.	ILJIN
700595000	37080	314655	ISU CHEMICAL	ISC
700312000	40680	777463	IL SUNG PHARM.	ISP
700757000	40550	777464	ILYANG PHARM.	IYP
701576000	74000	314713	KOREA ELECTRIC POWER	KEP
700225000	40540	502686	KEUNWHA PHARM.	KP
700157000	37049	314739	KUM YANG	KY
700626000	66020	777450	LG CABLE & MCH.	LGCM
700261000	64010	755743	LG ELECTRONICS INV.	LGEI
700305000	66100	314577	NEXANS KOREA	NXK
700549000	53040	501936	POSCO	POSCO
700581000	55040	314808	POONGSAN	PS
700341000	50520	756979	SSANGYONG CEMENT	SC
700612000	28570	756980	SK CHEMICALS CO.	SKC
701767000	84300	314710	SK TELECOM	SKT
700593000	64050	772091	SAMSUNG ELECTRONICS	SSE
700915000	64530	314841	SAMSUNG ELTO.MECH.	SSEM
701636000	88230	314622	SAMSUNG SECURITIES	SSS
700640000	64520	314839	SAMSUNG SDI	SSSDI
701245000	69500	502719	SAMSUNG TECHWIN	SSTW

and R_A and R_M are the dividend-adjusted daily asset and market (S&P500) returns from CRSP (Center for Research in Asset Prices, University of Chicago). The time subscript below is suppressed except where ambiguity would result.

The conventional equation for asset returns, is the market model,

$$r_A = \alpha + \beta r_M + \varepsilon_A \tag{2}$$

where r_M is the return on a market portfolio, conventionally represented by the Standard & Poor's Index of 500 large capitalization stocks, denoted hereafter the S&P500.

The 4-state model is estimated in the form

$$r_A = \sum_{i=1}^{4} p_i(\alpha_i + \beta_i r_M) + \varepsilon_A \tag{3}$$

where p_i, $i = 1 \ldots 4$ is a dummy variable denoting partition i as defined as shown below, and α_i and β_i are the respective intercepts and slopes in partition i.

The inclusion of $\{p_i\}$ on the right hand side of (1.6) introduces an element of endogeneity into the equation. As explained in the papers referenced in the text, the equation is estimated in implicit form, so that there is no "right hand" side. This procedure removes the parameter bias that may result from endogeneity. Intuitively, one may conceive the investor as making the buying or selling decision at (say) 30 minutes before the close of the market, so that even the direction of asset and market returns are at that time (largely) exogenous. It is an empirical question as to how often the directions of asset and market returns change in the last minutes of the trading day. Our preliminary analysis shows that 91 percent of the time, the sign of the asset return based on the prior day's closing price is the same at 3:30 as at the close. It is well-known that a considerable portion of trading takes place quite near the end of the trading day.

Partition of the Investor's Reference Frame by Signs
of Asset & Market Returns

Partition or State	**Sign of Asset Return**	**Sign of Market Return**
$p = 1$	$r_A \geq 0$	$r_M \geq 0$
$p = 2$	$r_A < 0$	$r_M \geq 0$
$p = 3$	$r_A < 0$	$r_M < 0$
$p = 4$	$r_A \geq 0$	$r_M < 0$

Behavior represented in (1.6) allows the investor to have different valuation decision rules, depending on the current market conditions, as defined by the directions of movement of r_A and r_M during the day of observation.[1] Moreover, the 4-state model allows for *loss aversion*, and more generally, asymmetric responses to positive and negative movements in r_A and r_M. The explanatory power of the 4-state model, as well as its behavioral representativeness, is enhanced by rotation of the reference axes so that the base axis passes through the historic means of r_A and r_M. The rotation is *only for defining the investor's reference frame and the associated partitions*.[2] As explained in Norsworthy *et al*. (2003) and Gorener (2003), this step incorporates the historical means of r_A and r_M as expectations for future asset and market returns, and hence asset valuation. Rotation also adjusts the expected trade-off between asset and market returns to its historical value. The estimation of the model is carried out with the unadjusted returns data.

Hypothesis 1: The explanatory power of the market model is not improved by partitioning.

The hypothesis is tested for whole time period, 1991–2001. It constitutes a direct test of the K-T proposition that the context of market-asset determines expected returns.

Hypothesis 2: The rotation of the classification axes from (0,0) to pass through (μ_A, μ_M) *does not improve the description of historical asset pricing by the 4-state model.*

This is a further elaboration of the K-T context hypothesis: that not only do the current states of asset and market returns matter, but that these states are determined relative to expected values of $\mathrm{r_A}$ and $\mathrm{r_M}$. An approximate F-test compares the rotated and unrotated models for each asset. In addition to the test of the 4-state model, we also test on the market model whether the mean rotation adds the explanatory power. Finally we test whether the optimal by rotated 4-state model gives more explanatory power than the mean rotated model.

5. EMPIRICAL RESULTS*

We start by estimating the market model using the Matlab program by ordinary least squares (OLS) and by using the heteroskedasticity adjustment to the coefficient standard errors (White, 1980). Then the 4-state model is estimated using the same method. Table 6.2 presents all coefficients for each model for the

* In interpreting the empirical results, it should be noted that stock prices were generally falling during the period under study.

Table 6.2 Coefficients of MM and 4-state Models, Mean and Optimal Rotations

Panel A: Optimally Rotated Market Model						
	α	Test Statistics	P-Val	β	Test Statistics	P-Val
ACM	−5.028E-05	−0.1360175	0.8918164	0.6142059	28.194545	0
BYS	−0.0001694	−0.397844	0.6907728	0.692088	32.238984	0
CHC	−6.2E-05	−0.13541	0.892299	0.8768	7.998043	0
CJ	0.000139	0.400782	0.688608	0.637046	28.83976	0
DHC	−0.00014	−0.36534	0.714885	0.661937	29.54557	0
DLI	−7.8E-05	−0.22359	0.823095	0.736871	37.628	0
DS	−1.3E-05	−0.03105	0.975229	0.593481	26.64573	0
DSP	−9.8E-06	−0.02253	0.98203	0.76472	7.719894	0
DSS	−1.4E-05	−0.04167	0.966768	0.804111	46.79855	0
GPS	−0.00031	−0.77819	0.436518	0.701933	34.50019	0
HB	0.000273	0.717737	0.472974	0.604646	26.75374	0
HC	−0.00023	−0.64632	0.51812	0.75712	40.07721	0
HDS	9.42E-05	0.245775	0.805872	0.68563	32.89918	0
HKT	−0.00023	−0.57844	0.563007	0.786394	4.016215	6.06E-05
HM	5.13E-05	0.155387	0.876527	0.719742	39.84238	0
HPM	−1.6E-05	−0.04292	0.965771	0.597311	28.17402	0
HS	−9E-05	−0.27154	0.785996	0.814897	44.6756	0
HSC	−0.0003	−0.75873	0.448073	0.662217	34.50263	0
HW	−0.00032	−0.86089	0.389367	0.697431	35.2351	0
HYC	−0.00017	−0.45425	0.649681	0.656923	30.20854	0
ILJIN	−0.00033	−0.82906	0.407135	0.734441	4.746663	2.16E-06
ISC	1.12E-05	0.030828	0.975409	0.685016	33.91526	0
ISP	7.84E-05	0.2019	0.840009	0.698077	35.66622	0
IYP	−0.00023	−0.59349	0.552895	0.677539	31.9244	0
KEP	7.99E-05	0.274135	0.783999	0.719532	40.11293	0
KP	5.78E-05	0.12502	0.900516	0.771956	7.615151	0
KY	−0.00024	−0.55988	0.575603	0.82249	5.794904	7.5E-09
LGCM	−3.9E-05	−0.12074	0.903907	0.69519	38.31499	0
LGEI	8.44E-05	0.279728	0.779705	0.795272	51.67517	0
NXK	−0.00021	−0.49451	0.620984	0.809875	4.970031	7.06E-07
POSCO	0.000449	1.449805	0.147215	0.691877	33.04573	0
PS	−5.8E-05	−0.16715	0.867262	0.695855	34.32987	0
SC	−0.00049	−1.2695	0.20436	0.712387	21.55672	0
SKC	−0.00018	−0.50527	0.613408	0.698379	34.19794	0
SKT	0.000223	0.455704	0.648635	0.796159	3.637718	0.00028
SSE	0.000493	1.638296	0.101462	0.792249	44.68689	0
SSEM	0.000113	0.358888	0.719703	0.749648	44.57831	0
SSS	0.000199	0.653553	0.513449	0.804001	46.54304	0
SSSDI	0.000156	0.491123	0.623374	0.696426	38.00489	0
SSTW	−0.00045	−1.32711	0.184569	0.725583	40.03773	0

Table 6.2 (*Continued*)

Panel B: Mean Rotated Market Model						
	α	Test Statistics	P-Val	β	Test Statistics	P-Val
ACM	2.35E-05	0.0726	0.9422	−0.3460	−20.36	0.0000
BYS	6.34E-05	0.1882	0.8507	−0.1736	−13.24	0.0000
CHC	2.82E-05	0.0828	0.9340	−0.1080	−3.49	0.0005
CJ	6.34E-05	0.2159	0.8291	−0.3108	−14.44	0.0000
DHC	6.08E-05	0.1890	0.8501	−0.2425	−16.91	0.0000
DLI	4.30E-05	0.1493	0.8813	−0.3737	−28.61	0.0000
DS	7.67E-06	0.0182	0.9855	−0.6167	−32.74	0.0000
DSP	8.85E-06	0.0158	0.9874	−1.0413	−15.79	0.0000
DSS	1.22E-05	0.0326	0.9740	−0.9705	−45.41	0.0000
GPS	0.000136	0.4312	0.6664	−0.2340	−18.13	0.0000
HB	0.000107	0.3378	0.7355	−0.2449	−16.19	0.0000
HC	0.00012	0.4125	0.6800	−0.2898	−22.52	0.0000
HDS	3.97E-05	0.1292	0.8972	−0.2275	−15.59	0.0000
HKT	8.90E-05	0.2710	0.7864	−0.1254	−1.99	0.0472
HM	2.58E-05	0.0967	0.9230	−0.3080	−22.03	0.0000
HPM	9.59E-06	0.0245	0.9804	−0.6324	−29.91	0.0000
HS	5.39E-05	0.2018	0.8401	−0.3657	−27.25	0.0000
HSC	0.000127	0.3948	0.6930	−0.2397	−17.40	0.0000
HW	0.000151	0.5000	0.6171	−0.2852	−19.35	0.0000
HYC	7.59E-05	0.2452	0.8064	−0.2874	−18.03	0.0000
ILJIN	0.000121	0.3708	0.7108	−0.1412	−2.17	0.0303
ISC	3.26E-06	0.0113	0.9910	−0.1110	−7.80	0.0000
ISP	3.18E-05	0.1025	0.9183	−0.2009	−15.58	0.0000
IYP	9.94E-05	0.3134	0.7540	−0.2395	−16.19	0.0000
KEP	4.62E-05	0.1811	0.8563	−0.4395	−32.42	0.0000
KP	1.18E-05	0.0336	0.9732	−0.0462	−1.79	0.0736
KY	9.38E-05	0.2750	0.7833	−0.1040	−3.70	0.0002
LGCM	2.40E-05	0.0796	0.9365	−0.5365	−28.05	0.0000
LGEI	5.00E-05	0.2062	0.8366	−0.3838	−34.15	0.0000
NXK	7.21E-05	0.2122	0.8319	−0.0878	−1.70	0.0894
POSCO	0.000259	0.9550	0.3397	−0.4687	−29.28	0.0000
PS	3.17E-05	0.1051	0.9163	−0.4052	−26.06	0.0000
SC	0.00022	0.7121	0.4765	−0.2421	−14.30	0.0000
SKC	8.93E-05	0.2977	0.7659	−0.3022	−19.78	0.0000
SKT	8.86E-05	0.2530	0.8003	−0.1291	−1.58	0.1142
SSE	0.000305	1.2310	0.2184	−0.4350	−24.59	0.0000
SSEM	6.36E-05	0.2477	0.8044	−0.3811	−27.30	0.0000
SSS	0.000116	0.4724	0.6367	−0.3513	−30.44	0.0000
SSSDI	8.52E-05	0.3189	0.7498	−0.4063	−26.36	0.0000
SSTW	0.000239	0.8544	0.3930	−0.3598	−24.11	0.0000

(*Continued*)

Table 6.2 (*Continued*)

Panel C: Base Market Model						
	α	Test Statistics	P-Val	β	Test Statistics	P-Val
ACM	−7.256E-05	−0.142979	0.8863161	0.7462507	20.889671	0
BYS	−0.0003701	−0.5279839	0.5975485	0.724882	13.438133	0
CHC	−0.00027	−0.22242	0.824	1.146664	4.453013	8.77E-06
CJ	0.000198	0.433865	0.664417	0.82531	23.63646	0
DHC	−0.00026	−0.42153	0.673395	0.793212	19.20339	0
DLI	−0.00012	−0.23431	0.814761	1.089803	30.25403	0
DS	−1.5E-05	−0.02455	0.980416	0.609258	16.3123	0
DSP	−1E-05	−0.01208	0.990365	0.754269	14.35335	0
DSS	−2E-05	−0.03739	0.970177	1.339766	33.57016	0
GPS	−0.00059	−0.90862	0.363622	0.906117	21.58135	0
HB	0.000432	0.821762	0.411276	0.715165	18.53329	0
HC	−0.00042	−0.72264	0.469959	1.145044	29.03688	0
HDS	0.000169	0.294036	0.76875	0.905403	24.18964	0
HKT	−0.00071	−0.80468	0.421065	0.857333	14.73315	0
HM	7.67E-05	0.169185	0.865662	1.026498	30.33066	0
HPM	−1.8E-05	−0.03874	0.969104	0.728408	22.62954	0
HS	−0.00016	−0.28982	0.771972	1.398952	36.91778	0
HSC	−0.00053	−0.88432	0.376593	0.80618	21.47602	0
HW	−0.00053	−0.95704	0.338621	0.951805	25.56748	0
HYC	−0.00027	−0.50246	0.615381	0.839565	21.69343	0
ILJIN	−0.00087	−1.10201	0.270541	0.755518	9.113745	0
ISC	2.24E-05	0.043515	0.965294	0.932241	24.82777	0
ISP	0.000152	0.251657	0.801323	0.911225	21.92475	0
IYP	−0.00042	−0.68844	0.491229	0.839567	20.68391	0
KEP	9.88E-05	0.285075	0.775606	0.939271	37.34514	0
KP	0.000251	0.265234	0.790847	0.511573	9.651655	0
KY	−0.00091	−0.90387	0.366133	0.897827	3.389281	0.00071
LGCM	−5E-05	−0.12159	0.903231	0.949491	30.36253	0
LGEI	0.000122	0.294643	0.768287	1.210675	43.51538	0
NXK	−0.00083	−0.81578	0.414686	0.714384	8.195482	0
POSCO	0.000547	1.485446	0.137528	0.889058	31.69564	0
PS	−8.4E-05	−0.17506	0.861041	0.96702	27.14256	0
SC	−0.00091	−1.47225	0.141054	0.954936	20.92093	0
SKC	−0.0003	−0.55491	0.578997	0.960397	26.56103	0
SKT	0.00068	0.7753	0.438222	1.005543	16.33172	0
SSE	0.000694	1.693351	0.09049	1.193666	46.1893	0
SSEM	0.00016	0.376252	0.706755	1.084496	39.87084	0
SSS	0.000321	0.698248	0.485075	1.312951	41.65487	0
SSSDI	0.000203	0.507604	0.611768	0.932907	33.94172	0
SSTW	−0.00067	−1.42342	0.154715	1.052144	30.91068	0

Table 6.2 (*Continued*)

Panel D: Optimally Rotated 4-State Market Model												
	α_1	α_2	α_3	α_4	β_1	P-Val	β_2	P-Val	β_3	P-Val	β_4	P-Val
ACM	0.027135	0.002288	−0.02445	−0.00261	0.64107	0	0.294033	2.49E-08	0.683242	0	0.285196	2.67E-08
BYS	0.036382	0.000605	−0.03108	−6.90E-05	0.369811	0.000104	0.192014	0.017791	0.612083	0	0.168769	0.00421
CHC	0.036829	−0.00401	−0.05362	0.005339	0.811339	0.042234	1.123821	0	4.57702	0.142282	0.83363	0
CJ	0.024453	0.004838	−0.0206	−0.0036	0.776372	0	0.642696	0	0.857452	0	0.484424	0
DHC	0.031819	0.001121	−0.02693	−0.0032	0.557115	0	0.258291	0.003508	0.770929	0	0.315644	1.18E-08
DLI	0.029371	0.00176	−0.02422	−0.00144	1.156011	0	0.933032	0	1.109413	0	0.854348	0
DS	0.032245	0.00298	−0.02868	−0.00319	0.356888	3.55E-07	0.333102	4.00E-10	0.565054	0	0.281363	0
DSP	0.038169	0.001771	−0.03249	−0.00265	0.351613	1.25E-05	0.250197	0.005084	0.594655	0	0.229325	0.000401
DSS	0.029161	0.001608	−0.02275	−0.00245	1.311074	0	1.041875	0	1.490202	0	1.039076	0
GPS	0.034122	0.004681	−0.0309	−0.00127	0.755493	6.50E-09	1.04945	0	0.887066	0	0.733596	0
HB	0.028701	0.001356	−0.02356	−0.00356	0.580968	0	0.295832	9.70E-07	0.793698	0	0.426634	0
HC	0.033465	0.00406	−0.02693	−0.00287	0.96769	0	1.106511	0	1.025183	0	0.819351	0
HDS	0.030547	0.002217	−0.02496	−0.00363	0.6092	0	0.466864	4.10E-09	0.815041	0	0.542203	0
HKT	0.026634	0.003253	−0.02233	−0.00396	0.731994	0	0.452899	0	0.993919	3.05E-08	0.436946	0
HM	0.023295	0.003046	−0.02195	−0.00494	1.127552	0	0.771149	0	0.959308	0	0.867871	0
HPM	0.02729	0.002732	−0.02297	−0.00351	0.509194	0	0.310141	8.00E-10	0.667196	0	0.324457	0
HS	0.028603	0.001697	−0.02308	−0.00267	1.449221	0	1.06732	0	1.505728	0	1.018946	0
HSC	0.032692	0.003982	−0.02964	−0.00396	0.485946	0	0.43923	0	0.697596	0	0.357036	2.36E-08
HW	0.031674	0.004532	−0.02596	−0.00197	0.762442	0	0.896073	0	0.759156	0	0.665279	0
HYC	0.028509	0.003242	−0.0254	−0.00491	0.678983	0	0.476042	0	0.726909	0	0.524258	0

(*Continued*)

Table 6.2 (*Continued*)

Panel D: Optimally Rotated 4-State Market Model												
	α_1	α_2	α_3	α_4	β_1	P-Val	β_2	P-Val	β_3	P-Val	β_4	P-Val
ILJIN	0.032474	0.003107	−0.02981	−0.0046	0.458144	5.53E-07	0.485603	0	0.46522	0.149719	0.453786	0
ISC	0.029701	0.00174	−0.02284	−0.00398	0.497476	6.00E-09	0.466408	0	0.876326	0	0.565212	0
ISP	0.033055	0.002344	−0.02719	−0.00305	0.612235	0	0.469858	3.25E-08	0.890302	0	0.518457	0
IYP	0.034326	0.004918	−0.029	−0.0019	0.554608	7.84E-06	0.945189	0	0.971003	0	0.681777	0
KEP	0.018062	0.002801	−0.01521	−0.00407	0.923688	0	0.564841	0	0.92409	0	0.634992	0
KP	0.042225	−0.00141	−0.03394	−0.00095	0.175619	0.070671	−0.01871	0.801477	0.585967	0	0.01136	0.864888
KY	0.037489	−0.00224	−0.04167	0.006249	0.427724	0.003649	0.937244	0	2.919803	0.135427	0.526358	0
LGCM	0.02205	0.002802	−0.01893	−0.00295	0.86603	0	0.518184	0	0.849074	0	0.508183	0
LGEI	0.02181	0.003233	−0.01846	−0.00405	1.250909	0	1.009629	0	1.306935	0	0.985472	0
NXK	0.037157	0.004835	−0.03514	−0.00214	0.584088	2.45E-07	1.008986	0	0.286423	0.607896	0.673279	0
POSCO	0.017815	0.002172	−0.01457	−0.0037	0.965949	0	0.484857	0	0.998489	0	0.576381	0
PS	0.027054	0.002492	−0.02052	−0.00299	0.722921	0	0.46997	2.12E-07	0.986773	0	0.487772	0
SC	0.03346	0.0032	−0.02718	−0.0022	0.780111	1.40E-09	0.886619	0	1.092976	0	0.699703	0
SKC	0.030906	0.003768	−0.02584	−0.00274	0.734191	8.00E-10	0.921782	0	0.987832	0	0.77433	0
SKT	0.02502	0.004125	−0.0231	−0.00458	1.09978	0	0.853937	0	1.572292	0.000104	0.812515	0
SSE	0.020304	0.00503	−0.01818	−0.00597	1.28616	0	1.012379	0	1.221321	0	1.009736	0
SSEM	0.021825	0.004564	−0.01953	−0.00562	1.151124	0	0.859263	0	0.980491	0	0.8481	0
SSS	0.02487	0.002897	−0.02118	−0.00391	1.520093	0	1.038455	0	1.357389	0	1.096906	0
SSSDI	0.020936	0.004372	−0.0176	−0.00472	0.966126	0	0.687414	0	0.955817	0	0.633157	0
SSTW	0.024981	0.004009	−0.02165	−0.00559	1.041074	0	0.796195	0	0.992054	0	0.820519	0

Table 6.2 (*Continued*)

Panel E: Mean Rotated 4-State Market Model												
	α_1	α_2	α_3	α_4	β_1	P-Val	β_2	P-Val	β_3	P-Val	β_4	P-Val
ACM	−0.00814	0.023052	0.007085	−0.01808	−0.27396	0.000478	0.41204	1.00E-09	−0.32531	0.000312	0.657726	0
BYS	−0.01439	0.023337	0.013234	−0.02063	−0.39772	0.000256	0.65985	0	−0.68175	2.84E-06	0.845792	0
CHC	−0.00965	0.026166	0.011906	−0.00191	−0.34064	0.031738	0.40061	0.002793	−0.40854	0.00377	2.015164	0.060582
CJ	0.017062	−0.00668	−0.01239	0.010263	−0.31446	0.150466	0.917402	0	−0.42003	0.14344	0.671133	0
DHC	−0.01512	0.023793	0.013843	−0.01671	−0.10998	0.124216	0.486762	0	−0.48493	2.00E-05	0.902577	0
DLI	−0.00995	0.022011	0.007885	−0.015	−0.33968	0.000324	0.658277	0	−0.46396	0.000644	0.930497	0
DS	−0.00399	0.031854	0.003393	−0.02709	0.07799	0.2249	0.241249	0.000362	0.119177	0.035242	0.562335	0
DSP	−0.00439	0.038718	0.004146	−0.03339	0.534543	0	0.411177	2.30E-05	0.668177	0	0.653969	2.00E-10
DSS	−0.00352	0.022086	−0.00103	−0.01787	0.193576	0.002192	1.050005	0	−0.24994	0.021144	1.092367	0
GPS	−0.01426	0.025114	0.014842	−0.01682	−0.31034	0.001898	0.466366	0	−0.46445	0.000381	1.009778	0
HB	0.019094	−0.01181	−0.01759	0.01442	−0.31752	0.002691	0.746473	0	−0.09867	0.286982	0.612961	0
HC	−0.01134	0.023809	0.009376	−0.0157	−0.41585	0.002072	0.700689	0	−0.77982	1.04E-08	1.07713	0
HDS	0.025278	−0.0109	−0.02089	0.013914	−0.1282	0.42994	0.962869	0	−0.16514	0.114597	0.774841	0
HKT	−0.01025	0.018552	0.011303	−0.01526	−0.09662	0.18812	0.606642	0	−0.20488	0.034165	0.805962	0
HM	0.016976	−0.00689	−0.01816	0.007661	−0.14029	0.19688	0.972336	0	−0.07606	0.786292	1.041298	0
HPM	−0.00481	0.025804	0.003851	−0.02176	0.062896	0.221148	0.392882	0	0.019672	0.83847	0.56312	0
HS	−0.00931	0.018805	0.010406	−0.01442	−0.37974	0.02228	1.022499	0	−0.37345	0.004405	1.16463	0
HSC	−0.014	0.024628	0.015452	−0.02092	−0.12506	0.170116	0.470974	0	−0.23659	0.041908	0.804575	0
HW	−0.01084	0.022879	0.010474	−0.01532	−0.55693	0.000257	0.538478	0	−0.69175	8.14E-06	0.87969	0
HYC	−0.00974	0.019934	0.012117	−0.01671	−0.38404	0.012936	0.614471	0	−0.32824	0.001621	0.774584	0

(*Continued*)

Table 6.2 (*Continued*)

Panel E: Mean Rotated 4-State Market Model												
	α_1	α_2	α_3	α_4	β_1	P-Val	β_2	P-Val	β_3	P-Val	β_4	P-Val
ILJIN	−0.00792	0.023429	0.01309	−0.01738	−0.97386	0.288208	0.511515	0	−0.57398	0.001372	0.830887	0
ISC	0.031773	−0.00654	−0.02622	0.007179	0.277476	0.106934	1.001673	0	0.571379	0.000159	0.801744	0
ISP	0.028572	−0.01302	−0.02529	0.016822	−0.07975	0.634607	0.983567	0	0.148231	0.104683	0.64387	0
IYP	−0.01397	0.023809	0.013154	−0.01677	−0.21439	0.017089	0.439538	0	−0.62007	3.76E-05	0.948889	0
KEP	0.015032	−0.00619	−0.01595	0.007353	−0.01202	0.849067	0.848674	0	0.153289	0.136445	0.856535	0
KP	0.036177	−0.01549	−0.02694	0.017955	−0.30994	0.124821	0.824732	0	−0.08993	0.31112	0.684218	0
KY	−0.01441	0.026385	0.018802	−0.00433	−0.24581	0.002356	0.248779	0.000422	−0.32314	0.013813	1.96023	0.081654
LGCM	−0.00474	0.017465	0.005001	−0.01493	−0.27437	0.001663	0.664187	0	−0.28604	0.028902	0.728773	0
LGEI	0.017318	−0.00653	−0.01622	0.007653	−0.18137	0.219482	1.115232	0	0.057779	0.352822	1.125505	0
NXK	−0.01257	0.025484	0.017037	−0.01945	−0.85877	0.315922	0.483365	0	−0.40632	0.000616	0.871489	0
POSCO	0.012183	−0.00633	−0.0126	0.008075	−0.13093	0.195927	0.806794	0	0.030607	0.556012	0.823169	0
PS	−0.01008	0.021149	0.008203	−0.01453	−0.04888	0.434558	0.551083	0	−0.47512	2.78E-05	0.903151	0
SC	−0.01245	0.019768	0.009374	−0.01421	−0.30524	0.004337	0.694429	0	−0.74543	1.11E-07	0.997555	0
SKC	−0.01023	0.018988	0.011079	−0.01597	−0.25939	0.106792	0.622084	0	−0.62402	0.000186	0.930817	0
SKT	0.016238	−0.00931	−0.01521	0.011444	−0.09674	0.323359	1.00374	0	0.063876	0.463404	0.844748	0
SSE	0.013296	−0.00612	−0.01309	0.007661	−0.12936	0.279187	1.081446	0	−0.05091	0.542862	1.148605	0
SSEM	0.017273	−0.00838	−0.01656	0.007831	−0.05472	0.639072	0.932108	0	0.02432	0.834573	1.037911	0
SSS	0.020682	−0.00909	−0.01597	0.00952	0.242314	0.01923	1.119305	0	−0.17551	0.121643	1.218851	0
SSSDI	0.015264	−0.00658	−0.01442	0.007367	−0.13942	0.283158	0.869777	0	0.060389	0.550914	0.89047	0
SSTW	−0.00872	0.014345	0.011825	−0.01188	−0.40902	0.018703	0.840382	0	−0.17275	0.199075	1.008546	0

Table 6.2 (*Continued*)

Panel F: Base 4-State Model												
	α_1	α_2	α_3	α_4	β_1	**P-Val**	β_2	**P-Val**	β_3	**P-Val**	β_4	**P-Val**
ACM	0.014853	0.011797	−0.01496	−0.01793	0.553581	0	−0.16719	0.068426	0.757486	0	−0.20065	0.051896
BYS	0.016001	0.015337	−0.01952	−0.02282	0.77306	0	−0.59923	3.40E-05	0.88489	0	−0.348	0.005644
CHC	0.018625	0.013317	−0.00011	−0.01664	0.517248	9.22E-06	−0.33711	0.017016	2.074932	0.057492	−0.38324	0.075774
CJ	0.011368	0.010329	−0.01081	−0.01144	0.634841	0	−0.22967	0.161588	0.883148	0	−0.47626	0.102014
DHC	0.018337	0.015538	−0.01584	−0.02046	0.535968	0	−0.42513	0.000134	0.923035	0	−0.13101	0.125243
DLI	0.015329	0.012428	−0.01367	−0.01666	0.812728	0	−0.25154	0.076205	0.977876	0	−0.29255	0.005433
DS	0.017578	0.016441	−0.02031	−0.02101	0.382829	0	−0.18377	0.096131	0.602164	0	−0.20264	0.02971
DSP	0.021757	0.021422	−0.02145	−0.02097	0.591048	0	−0.51388	0.000529	0.905826	0	−0.37349	0.001137
DSS	0.010958	0.011956	−0.01166	−0.0159	1.217907	0	−0.32009	0.002848	1.260249	0	−0.24377	0.10411
GPS	0.018713	0.016173	−0.01631	−0.02115	0.58891	0	−0.39948	0.002496	1.027306	0	−0.24393	0.02683
HB	0.014974	0.011913	−0.01656	−0.01728	0.594722	0	−0.37915	8.18E-05	0.683279	0	−0.1157	0.221321
HC	0.017342	0.010927	−0.01547	−0.01635	0.812836	0	−0.70187	5.72E-07	1.085003	0	−0.59885	6.80E-05
HDS	0.015486	0.014887	−0.01564	−0.02	0.722483	0	−0.3585	0.014925	0.869695	0	−0.21163	0.051313
HKT	0.01252	0.01157	−0.01348	−0.01784	0.69129	0	−0.19063	0.047923	0.86866	0	0.003459	0.980956
HM	0.009329	0.009583	−0.01104	−0.01393	0.997242	0	−0.24078	0.007002	0.912273	0	−0.34858	0.187946
HPM	0.015596	0.013073	−0.01581	−0.01935	0.502441	0	−0.18108	0.14666	0.6527	0	−0.06364	0.383524
HS	0.011945	0.014313	−0.01365	−0.01679	1.196058	0	−0.19037	0.14163	1.192239	0	−0.31856	0.139658
HSC	0.016477	0.015798	−0.01988	−0.02275	0.554823	0	−0.21701	0.060019	0.838178	0	−0.12952	0.287123
HW	0.015876	0.011545	−0.01492	−0.01762	0.69013	0	−0.63092	4.75E-05	0.89329	0	−0.56905	0.001239
HYC	0.0131	0.014392	−0.01631	−0.01529	0.731451	0	−0.21523	0.05053	0.788307	0	−0.42941	0.02249

(*Continued*)

Table 6.2 *(Continued)*

Panel F: Base 4-State Model												
	α_1	α_2	α_3	α_4	β_1	P-Val	β_2	P-Val	β_3	P-Val	β_4	P-Val
ILJIN	0.01729	0.014046	−0.01705	−0.01025	0.575697	0	−0.52397	0.003887	0.841788	0	−1.4395	0.274918
ISC	0.017366	0.012713	−0.01623	−0.01795	0.547083	0	−0.48528	0.000538	0.816649	0	−0.23192	0.080904
ISP	0.019257	0.019405	−0.01868	−0.02249	0.562669	0	−0.17746	0.226328	0.899789	0	0.003089	0.971101
IYP	0.017758	0.01454	−0.01659	−0.01862	0.534532	0	−0.54395	0.000286	0.954844	0	−0.26644	0.008253
KEP	0.007789	0.008466	−0.00977	−0.01392	0.844876	0	−0.07336	0.189977	0.791997	0	0.001146	0.988008
KP	0.020493	0.023521	−0.02276	−0.02449	0.597356	0	−0.28553	0.083876	0.853565	0	−0.19224	0.025959
KY	0.020052	0.019808	−0.0038	−0.01958	0.354741	1.11E-07	−0.27415	0.03862	1.977587	0.081435	−0.23723	0.008659
LGCM	0.009339	0.00971	−0.01151	−0.01659	0.83489	0	−0.14496	0.273936	0.839085	0	−0.19294	0.06606
LGEI	0.008727	0.007592	−0.01011	−0.01514	1.089833	0	−0.42551	0.00089	1.030318	0	−0.02234	0.724529
NXK	0.019471	0.017232	−0.0189	−0.01798	0.559679	0	−0.39727	0.000756	0.890029	0	−1.02689	0.35384
POSCO	0.008266	0.008114	−0.00862	−0.012	0.816553	0	−0.13025	0.139598	0.776768	0	−0.00344	0.946319
PS	0.013214	0.013063	−0.01303	−0.01789	0.700354	0	−0.31697	0.004163	0.940972	0	−0.03676	0.678512
SC	0.014772	0.010233	−0.01407	−0.01693	0.767102	0	−0.70339	7.66E-07	1.002347	0	−0.39029	0.001068
SKC	0.013591	0.013068	−0.01494	−0.01779	0.700669	0	−0.51877	0.001807	0.965903	0	−0.26444	0.212047
SKT	0.011595	0.012148	−0.01252	−0.01499	0.839674	0	−0.10772	0.208927	0.967717	0	0.049788	0.570493
SSE	0.007953	0.008845	−0.00866	−0.01277	1.138607	0	−0.16323	0.13265	1.026592	0	−0.07333	0.381465
SSEM	0.008697	0.011264	−0.01209	−0.01473	1.0095	0	−0.10616	0.295879	0.855689	0	−0.1112	0.329796
SSS	0.01024	0.013689	−0.01241	−0.01498	1.194723	0	0.136886	0.104467	1.064915	0	−0.24579	0.032482
SSSDI	0.008138	0.009414	−0.01006	−0.01274	0.86491	0	−0.14919	0.151295	0.817624	0	−0.055	0.546046
SSTW	0.010012	0.01204	−0.01177	−0.0127	0.92432	0	−0.15949	0.23411	1.012116	0	−0.4558	0.017931

entire sample period. We first estimate each model without rotation of the axes. Then, we rotate the axes to pass through the mean point defined by the historical average asset and market returns, and estimate each model. Finally, the optimum rotation of the axes is calculated and we estimate the two models based on the optimum rotation. Gorener (2003) describes the meaning and effects of asset rotation in the 4-state model.

As it is shown in Table 6.2, the coefficients of β for market models are statistically significant, except for β in the mean rotation model for SKT. This result contradicts Fama and French (1992). They reported that β was not statistically significant and had a negative sign. In our analysis, we find all coefficients (β) are significant in both models. In the market model, their signs are positive in the base and optimally rotated case but are negative in the mean rotated case. None of constant terms (α) are significant at any level in the market model case. For the 4-state model case, all constant terms are significant for all models. However, the coefficients of the 4-state model give a mixed result. For the base 4-state model, β_1 and β_3 are all statistically significant and the signs of β_1 and β_3 are positive. On the other hand, β_2 and β_4 give a different result. There are 13 out of 40 β_2 coefficients are not significant and all coefficients have negative signs, while there are 21 insignificant coefficients for β_4 having negative signs. The result of the mean rotated 4-state case is somewhat similar to the base case but in an opposite way. In this case, β_2 and β_4 are all statistically significant and the signs of β_2 and β_4 are positive. On the other hand, there are 22 β_2 coefficients that are not significant and all coefficients have negative signs, while there are 16 insignificant coefficients for β_3 having all negative signs. The optimally rotated 4-state case gives a different result compared to the other cases. In this case, β_1 coefficients are all statistically significant and the signs of them are all positive. For β_2 coefficients, only one coefficient is not significant and all signs of coefficients are positive except one. Three β_3 coefficients are not statistically significant and all have positive signs. Lastly, all β_4 coefficients have a positive sign and only one coefficient is not significant. The analyses of the sub-groups give a similar result, even though they are not reported in the paper.

Table 6.3 presents the adjusted R-squared values for both the market model and the 4-state model with no rotation, mean rotation, and optimal rotation cases for the whole sample period. As this table shows, the explanatory power of the 4-state model far exceeds that of the market model for almost all stocks. The adjusted R-square value of the market model without rotation ranges form 3.5% to 53%, averaging 26.5%. Doing a 4-way partition only (4-state base model) improves the average greatly to 55%, more than double the adjusted explanatory power. This result also shows that expectations of future asset returns are strongly affected by average historical performance. Even though mean rotation of axis

Table 6.3 Adjusted R-Squared, 4 State and Market Models, 1991–2001

	4ST			MM		
Symbol	**Mrot**	**Optrot**	**Base**	**Mrot**	**Optrot**	**Base**
ACM	0.5872	0.6605	0.5552	0.269	0.377	0.2008
BYS	0.5432	0.5953	0.5328	0.133	0.4784	0.1147
CHC	0.2099	0.2673	0.2046	0.1479	0.7696	0.0962
CJ	0.5607	0.6603	0.5675	0.2544	0.4053	0.278
DHC	0.5904	0.6641	0.5739	0.1996	0.4377	0.1749
DLI	0.6445	0.7103	0.6267	0.4023	0.5428	0.3495
DS	0.6558	0.6687	0.5546	0.351	0.3521	0.1187
DSP	0.4646	0.4662	0.4175	0.5309	0.5859	0.0847
DSS	0.7016	0.7332	0.6615	0.6317	0.646	0.4418
GPS	0.5929	0.6548	0.5812	0.2153	0.4929	0.203
HB	0.5647	0.6605	0.5709	0.1718	0.3651	0.1832
HC	0.636	0.6974	0.6187	0.3335	0.5719	0.3149
HDS	0.5904	0.6598	0.5867	0.1965	0.4701	0.2346
HKT	0.2252	0.2629	0.218	0.1127	0.6193	0.1001
HM	0.6409	0.727	0.6476	0.3189	0.5186	0.3892
HPM	0.6776	0.7027	0.603	0.3541	0.3565	0.2124
HS	0.66	0.7399	0.6488	0.4985	0.6655	0.4496
HSC	0.6136	0.706	0.5935	0.1964	0.4375	0.1847
HW	0.6104	0.6729	0.598	0.2733	0.4865	0.261
HYC	0.5876	0.6554	0.5715	0.2455	0.4318	0.2237
ILJIN	0.3887	0.4137	0.3857	0.11	0.5387	0.1014
ISC	0.6494	0.7063	0.621	0.0598	0.469	0.2799
ISP	0.6043	0.6881	0.6017	0.1704	0.4866	0.215
IYP	0.5862	0.658	0.5727	0.2054	0.458	0.1897
KEP	0.6813	0.7772	0.6897	0.4347	0.5185	0.4688
KP	0.3637	0.4233	0.3565	0.016	0.5964	0.0349
KY	0.3064	0.3631	0.3019	0.098	0.6778	0.0871
LGCM	0.6689	0.7302	0.6404	0.4549	0.4825	0.3796
LGEI	0.7101	0.7932	0.7113	0.4845	0.6315	0.526
NXK	0.2806	0.3083	0.2786	0.0672	0.6565	0.0573
POSCO	0.6389	0.7483	0.646	0.4205	0.4776	0.4263
PS	0.6438	0.7137	0.6166	0.384	0.4832	0.3193
SC	0.5334	0.5958	0.5226	0.233	0.5076	0.225
SKC	0.6038	0.6662	0.5896	0.2929	0.4878	0.2714
SKT	0.2367	0.2904	0.2403	0.124	0.6348	0.1382
SSE	0.6992	0.7947	0.703	0.5238	0.6287	0.5305
SSEM	0.6742	0.7754	0.6775	0.4231	0.5611	0.4541
SSS	0.6705	0.7705	0.6763	0.4669	0.6475	0.4869
SSSDI	0.6303	0.7449	0.6356	0.3857	0.4843	0.4055
SSTW	0.6349	0.7039	0.6258	0.3775	0.5266	0.3688

improves the explanatory power slightly without an exception for both the market model and the 4-state model, the optimum rotation of the 4-state model gives the greatest power for most cases, except for three stocks: CHC, NXK and SKT. Even with these exceptions, the explanatory power of the optimum rotation of the market model exceeds that of the 4-state model. On average, the adjusted R-square of the optimal rotation of axis on the 4-state model is about 63%, while the market model is about 52%.

Table 6.4 summaries the average adjusted R-square value and also their ranges for the whole period as well as the sub-periods. As this table indicates, the

Table 6.4 Average Adjusted R-Squared and Their Ranges, 4 State and Market Models, for Various Sample Periods

		OPTIMAL ROTATION		**MEAN ROTATION**		**NO ROTATION**	
Period	**AdjRSQ**	**MM**	**4-ST**	**MM**	**4-ST**	**MM**	**4-ST**
1991–96	MEAN	0.474939	0.74841	0.24335	0.711023	0.234786	0.631443
	MAX	0.643343	0.824195	0.639387	0.802216	0.520273	0.726433
	MIN	0.242788	0.64921	−0.00061	0.59858	0.022669	0.527999
1991–97	MEAN	0.500562	0.734922	0.211425	0.705242	0.279725	0.635786
	MAX	0.724216	0.813493	0.619688	0.804469	0.512335	0.724257
	MIN	0.354311	0.308224	−1.8E-06	0.305008	0.041994	0.255505
1991–98	MEAN	0.513957	0.694749	0.201538	0.675051	0.287206	0.610533
	MAX	0.70944	0.820692	0.660563	0.812725	0.550945	0.740979
	MIN	0.383874	0.211623	−5.8E-06	0.200031	0.031602	0.172291
1996–2001	MEAN	0.534699	0.648873	0.210749	0.611044	0.278317	0.556427
	MAX	0.804132	0.851277	0.686855	0.799653	0.615715	0.780862
	MIN	0.331931	0.236613	0.004042	0.214574	0.022467	0.193988
1997–2001	MEAN	0.527818	0.654888	0.257123	0.620381	0.265076	0.548256
	MAX	0.831111	0.861029	0.67178	0.809785	0.634455	0.78985
	MIN	0.30291	0.205736	−0.00085	0.180498	0.03114	0.161638
1998–2001	MEAN	0.511064	0.653475	0.240001	0.614983	0.254382	0.540281
	MAX	0.871345	0.850053	0.752869	0.797406	0.616229	0.778589
	MIN	0.266637	0.204305	−0.00115	0.188218	0.02875	0.16704
1991–2001	MEAN	0.52413	0.63076	0.289215	0.564063	0.264535	0.550625
	MAX	0.7696	0.7947	0.6317	0.7101	0.5305	0.7113
	MIN	0.3521	0.2629	0.016	0.2099	0.0349	0.2046

average adjusted R-square value is somewhat stable across the different sample period for the market model. The difference is within the 5% range. However, the average R-square value of the period 1991 to 1996 varies the most against those values of the whole time period. Again, even though the mean rotation improves the explanatory power, the optimal rotation of axis greatly adds the explanatory power to the market model. On the other hand, the average adjusted R-squared of the 4-state model gives a different result from the market model for the different sample period. As Table 6.4 clearly shows, there is a large difference in the average adjusted R-squared value among different groups. For the optimal rotation and mean rotation cases, the average adjusted R-squares value is different from the whole time period value, sometime varying by more than 15%. In addition, the average adjusted R-squared value of the period 1991 to 1996 and the period 1991 to 1997 are especially high, compared to other sub-periods as well as the whole period. For example, the average adjusted R-square of the 1991–96 value is 74% while the 1996–2001 value is about 64%. Therefore, we can conclude that there is a structural change in those time periods in the Korean Market.

Table 6.5 presents F-statistics for tests of the hypotheses. The first column of the table represents the F-statistic values of the first hypothesis, showing that partitioning the asset pricing model – without rotation – does not improve the explanatory power for all 40 stocks. As shown in the F-test statistic values, they are large numbers, indicating rejection of the null hypothesis. Moreover, the coefficients of the APM models differ considerably among partitions. This evidence strongly supports the idea that investor expectations are heavily influenced by the frame of reference that includes current market conditions, as described by movements in asset and market returns. The F-statistics are all significant, so that the null hypothesis is rejected in all testable cases. Hypothesis 1 is rejected for all 40 stocks. Therefore, we can conclude that the 4-state model is better than the market model in explaining expected asset returns for the Korean stock market.

As we already saw in the adjusted R-squared values, rotating the axes in both the market and the 4-state model increases the explanatory power further. In Table 6.5, we formally test the rotation of axes in both the market model and the 4-state model as well. As mentioned in Norsworthy *et al.* (2005), in the second hypothesis test, rotation adds nothing to partitioning and cannot be conducted as a nested test, because formally the number of estimated parameters is the same for the rotated and unrotated models. However, the calculation of the rotated data incorporates the means of the asset and market returns, so that it is not unreasonable to “penalize” the rotated model by two degrees of freedom. The explanatory power of the rotated model is considerably greater in any

Table 6.5 F-tests of Hypotheses for 4 State and Market Models

Ticker	BaseMM vs Base4ST	BaseMM vs MrotMM	BaseMM vs OptMM	Base4ST vs Mrot4ST	Base4ST vs Opt4ST	Mrot4ST vs Opt4ST
ACM	410.20	2480.88	909.94	120.33	318.90	664.10
BYS	460.65	5458.57	1832.89	35.68	159.12	396.80
CHC	70.98	19340.03	6485.72	11.19	88.71	241.99
CJ	344.71	2310.96	743.75	−22.78	280.90	901.59
DHC	481.81	3925.75	1330.94	62.73	276.20	675.61
DLI	382.33	3487.89	1231.09	77.88	296.85	699.39
DS	503.56	1508.19	1089.88	452.93	353.88	120.53
DSP	294.42	2356.61	2928.55	136.03	94.41	10.35
DSS	334.26	1590.89	1572.12	208.04	276.56	364.40
GPS	464.77	4650.27	1562.42	45.03	219.57	552.48
HB	464.87	2787.15	918.17	−20.73	271.73	868.36
HC	410.30	4813.38	1621.52	73.81	267.37	624.54
HDS	438.48	3938.35	1280.57	14.95	221.15	627.47
HKT	78.37	9710.26	3249.48	15.27	63.47	158.31
HM	377.54	2860.64	878.29	−27.74	298.78	969.29
HPM	506.24	989.51	768.21	356.80	344.50	259.79
HS	292.33	4960.30	1724.87	51.60	359.63	944.02
HSC	517.33	3841.35	1289.50	81.26	393.50	966.89
HW	431.48	3766.26	1269.04	50.15	235.80	587.94
HYC	417.83	3242.28	1104.71	61.05	250.57	605.58
ILJIN	238.64	7089.07	2367.89	8.57	49.96	132.01
ISC	463.05	3704.37	1080.19	125.80	298.56	595.40
ISP	499.60	4511.09	1466.59	11.07	284.90	826.60
IYP	461.29	4134.37	1390.49	51.28	256.76	646.17
KEP	366.43	1533.96	454.78	−39.14	403.90	1323.68
KP	257.64	9941.93	3305.60	18.38	119.68	318.47
KY	159.05	12652.36	4230.65	10.83	99.38	274.57
LGCM	373.44	1643.59	705.07	133.30	341.80	698.32
LGEI	330.47	2933.32	916.67	−4.88	407.43	1235.97
NXK	158.59	12097.54	4041.98	5.09	45.01	124.43
POSCO	319.78	1348.25	437.84	−29.28	417.30	1335.90
PS	399.14	2698.70	988.65	118.58	348.87	751.54
SC	321.22	4676.60	1567.43	36.41	186.42	475.20
SKC	399.17	3623.84	1231.79	56.09	236.37	575.95
SKT	70.05	9738.14	3240.58	−6.38	73.27	233.53
SSE	299.22	2631.02	867.95	−18.20	459.08	1430.57
SSEM	356.65	2577.23	815.85	−14.50	447.71	1385.20
SSS	301.46	3991.47	1306.87	−26.38	421.85	1341.31
SSSDI	325.21	1854.02	588.27	−20.91	440.16	1381.08
SSTW	353.71	3048.97	1028.35	39.25	271.08	716.48

event. Under these circumstances, Hypothesis 2 – that rotation adds nothing to partitioning – is rejected uniformly in the panel of 40 stocks. For both the market model and the 4-state model, all F-statistics are significant at any reasonable significance level. As can be seen in Table 6.5, the mean rotated model is better than the base model in both the market model and the 4-state model and the optimum rotation gives more explanatory power than the mean rotation in the 4-state model. One interesting point in F-statistics is that several test statistics of the base vs. optimal rotation are smaller than the statistics of the base vs. mean rotation. This is because we heavily penalize the optimum rotation than the mean rotation so that the degrees of freedom are smaller in the optimal rotation case.

Table 6.6 presents the angles and kappa for both the market model and the 4-state model that are used in rotations of axis. The interesting point in this table is that most of kappas are negative, and only 15 stocks have a positive kappa but they are generally large companies in these cases. Therefore, the negative rotation is common for the Korean stocks for the entire sample period. Although it is not discussed here, for the entire sample period the KOSPI return was actually negative. The positive kappa for the relatively large companies could be explained by the Korean market specific situations. Those large companies are owned by a conglomerate called *Chebol* (similar to the *Keiretsu* in Japan) and those companies are the favorite among investors for the bear market. Therefore, their returns are not heavily affected by the general market condition. The five large companies in the Korean market had positive returns for the same sample period, even though the market was down.

In Appendix 1, Figures A6.1 through to A6.10 show the fitted values of the market model, the optimally rotated market model, the 4-state model, and the optimally rotated 4-state model in panels A through to D, respectively. We select 10 companies out of the sample of 40 companies and plot the fitted values for both unrotated and optimally rotated models. As the figures show, the market model has a single straight line, passing close to the origin. On the other hand, the 4-state model has four different lines explaining four different reference frames. The rotated and unrotated 4-state models for quadrants 1 and 3 usually show the characteristic kink expected from the prospect theory – the Friedman and Savage (1948) and Allais (1953) kink. This holds for 10 out of the 10 companies. It is also notable that the β coefficients of the rotated 4-state models have positive slopes, while in the unrotated models, the second and fourth quadrants typically have perverse slopes. We conclude that the rotation operation brings the 4-state model into greater agreement with investor behavior. In addition, Figures A6.1 to A6.10 show the evidence of asymmetry of gains and losses. As you can see in these figures, kinks at the origin in the plots of fitted values for quadrants 1 and

Table 6.6 Rotation Angles (°) and Kappa (Radians), 4-state Model

	Market Model				4-State Model			
	OPTIMAL ROTATION		MEAN ROTATION		OPTIMAL ROTATION		MEAN ROTATION	
Ticker	Angle	Kappa	Angle	Kappa	Angle	Kappa	Angle	Kappa
ACM	−24.93	−0.44	−84.44	−1.47	−42.30	−0.74	−84.44	−1.47
BYS	−33.93	−0.59	−88.97	−1.55	−45.18	−0.79	−88.97	−1.55
CHC	−39.87	−0.70	−88.56	−1.55	−79.02	−1.38	−88.56	−1.55
CJ	20.61	0.36	88.15	1.54	46.44	0.81	88.15	1.54
DHC	−29.25	−0.51	−88.49	−1.54	−43.92	−0.77	−88.49	−1.54
DLI	−23.85	−0.42	−86.72	−1.51	−60.48	−1.06	−86.72	−1.51
DS	−30.06	−0.52	−58.03	−1.01	−43.02	−0.75	−58.03	−1.01
DSP	−37.62	−0.66	−39.13	−0.68	−44.28	−0.77	−39.13	−0.68
DSS	−24.39	−0.43	−58.50	−1.02	−61.38	−1.07	−58.50	−1.02
GPS	−29.61	−0.52	−89.35	−1.56	−60.66	−1.06	−89.35	−1.56
HB	25.65	0.45	89.14	1.56	46.62	0.81	89.14	1.56
HC	−27.00	−0.47	−89.09	−1.55	−58.86	−1.03	−89.09	−1.55
HDS	27.00	0.47	87.84	1.53	47.88	0.84	87.84	1.53
HKT	−37.44	−0.65	−89.47	−1.56	−45.18	−0.79	−89.47	−1.56
HM	19.89	0.35	85.48	1.49	54.72	0.96	85.48	1.49
HPM	−22.86	−0.40	−64.23	−1.12	−41.94	−0.73	−64.23	−1.12
HS	−25.11	−0.44	−87.44	−1.53	−60.84	−1.06	−87.44	−1.53
HSC	−28.62	−0.50	−89.29	−1.56	−43.38	−0.76	−89.29	−1.56
HW	−26.19	−0.46	−89.28	−1.56	−54.18	−0.95	−89.28	−1.56
HYC	−26.01	−0.45	−88.57	−1.55	−46.08	−0.80	−88.57	−1.55
ILJIN	−35.91	−0.63	−89.56	−1.56	−45.72	−0.80	−89.56	−1.56
ISC	24.21	0.42	77.02	1.34	47.52	0.83	77.02	1.34
ISP	28.71	0.50	87.61	1.53	49.68	0.87	87.61	1.53
IYP	−29.07	−0.51	−89.09	−1.55	−58.86	−1.03	−89.09	−1.55
KEP	12.60	0.22	86.41	1.51	45.18	0.79	86.41	1.51
KP	40.50	0.71	88.52	1.54	43.92	0.77	88.52	1.54
KY	−38.88	−0.68	−89.58	−1.56	−72.18	−1.26	−89.58	−1.56
LGCM	−17.91	−0.31	−81.51	−1.42	−45.72	−0.80	−81.51	−1.42
LGEI	18.18	0.32	87.10	1.52	57.60	1.01	87.10	1.52
NXK	−39.87	−0.70	−89.54	−1.56	−61.02	−1.06	−89.54	−1.56
POSCO	12.78	0.22	89.32	1.56	45.90	0.80	89.32	1.56
PS	−22.41	−0.39	−85.18	−1.49	−46.98	−0.82	−85.18	−1.49
SC	−28.98	−0.51	−89.58	−1.56	−58.32	−1.02	−89.58	−1.56
SKC	−25.65	−0.45	−88.72	−1.55	−58.32	−1.02	−88.72	−1.55
SKT	36.18	0.63	89.45	1.56	57.60	1.01	89.45	1.56
SSE	17.64	0.31	89.46	1.56	55.44	0.97	89.46	1.56
SSEM	18.09	0.32	87.74	1.53	51.84	0.90	87.74	1.53
SSS	22.05	0.38	88.86	1.55	60.66	1.06	88.86	1.55
SSSDI	15.75	0.27	88.20	1.54	48.06	0.84	88.20	1.54
SSTW	−21.78	−0.38	−89.43	−1.56	−52.56	−0.92	−89.43	−1.56
MEAN	−9.68	−0.17	−19.04	−0.33	−14.28	−0.25	−19.04	−0.33

3 are the visual evidence of the asymmetry of gains and losses. All 10 selected companies have kinks at the origin and this supports the asymmetry.

6. CONCLUSIONS

We investigate the risk-return relationship using a four-state model on the Korean Stock Market. The 4-state model has been developed by Norsworthy *et al.* and was tested on the U.S. market. Based on Kahneman and Tversky's (1979) prospect theory, they tested the dependence of expected returns on the current frame of reference, asymmetric valuation of gains and losses of equal size and diminishing marginal sensitivity to gains and losses in the context of a 4-state model.

We find that the 4-state model gives a higher explanatory power, implying strong reference dependence, i.e. dependence on current market conditions, which determine the partitions where the observations are classified. This added explanatory power resides in the relationship between risk and expected or required return, which is discontinuous and state- (or reference frame-) dependent, so that the market model miss-specifies the true relationship between risk and return.

We also find that the rotation of axis to include expectations based on historical returns improves the explanatory power further. The useful insight is that the rotation has psychological meaning. The psychological import is asserted by the clearly greater explanatory power of the rotated model: the observations are translated into expectations space. This result appears both in the tabulated results and in the visual evidence of the plots of the observations and fitted values.

Finally we find asymmetric valuation of gains and losses in 1991–2001 data for all 40 companies. The symmetry hypothesis concerning valuation of positive and negative returns of equal magnitude is rejected. The rotated and unrotated 4-state models for quadrants 1 and 3 usually show the characteristic kink expected from prospect theory – the kink predicted by Friedman and Savage (1948) and Allais (1953). These kinks give a visual evidence of asymmetry of gains and losses.

BIBLIOGRAPHY

Allais, Maurice (1953) "Le Comportment de l'Homme Rationnel devant le Risque, Critique des Postulats et Axiomes de l'Ecole Americaine." *Econometrica* **21**: 503–46.

Bank of Korea (2002) *Financial System in Korea*. Seoul: The Bank of Korea.

Barberis, N., Huang, M. and Santos, T. (1999) "Prospect Theory and Asset Prices." Working Paper 7220, National Bureau of Economic Research, Cambridge MA.

Benartzi, S. and Thaler, R. (1993) "Myopic Loss Aversion and the Equity Premium Puzzle." Working Paper No. 4369, National Bureau of Economic Research, Cambridge MA. (May).

Cochrane, J. (1999) "New Facts in Finance." Working Paper, National Bureau of Economic Research, Cambridge MA.

——— (2000) *Asset Pricing*. Princeton NJ: Princeton University Press.

Fama, E. and French, K. (1992) "The Cross-Section of Expected Stock Returns." *Journal of Finance* **47**: 427–65.

Friedman, M. and Savage, L. (1948) "The Utility Analysis of Choices Involving Risk." *Journal of Political Economy* **56**: 279–304.

Gorener, Rifat (2003) "Investor Behavior and Rotated Asset Pricing Models: Empirical Evidence." PhD dissertation, Rensselaer Polytechnic Institute, 2003.

Kahneman, D. and Tversky, A. (1979) "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* March: 263–91.

Knetsch, Jack and Sinden, J.A (1984) "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measure of Value." *Quarterly Journal of Economics* August: 507–21.

Lintner, J. (1965) "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets." *Review of Economics and Statistics* **47**: 13–37.

Norsworthy, J.R., Schuler, R.E., Gorener, R., Morgan, I.W. and Li, D. (2003) "Expected Utility, Prospect Theory and Asset Prices." Working paper, Center for Financial Studies, Rensselaer Polytechnic Institute and Financial Economic Network, SSRN, Working Paper Series, posted Nov 22, 2003.

Pettengill, G., Sundaram, S. and Mathur, I. (1995) "The Conditional Relationship between Beta and Returns." *Journal of Financial and Quantitative Analysis* March: 101–16.

Putler, Daniel (1988) "Reference Price Effects and Consumer Behavior." unpublished, Economic Research Service, U.S. Department of Agriculture, Washington, DC.

Ross, S.A. (1976) "The arbitrage theory of capital pricing theory." *Journal of Economic Theory*, December.

Sharpe, W. (1964) "Capital asset prices: a theory of market equilibrium under conditions of risk." *Journal of Finance* **19**: 425–42.

Shefrin, H. (2000) *Beyond Greed and Fear*. Boston, MA: Harvard Business School Press.

Shiller, R.J. (1998) "Human Behavior and the Efficiency of the Financial System." Working Paper 6375, National Bureau of Economic Research, January.

Soros, G. (1994) "The Alchemy of Finance: Reading the Mind of the Market." New York: John Wiley & Sons, Inc.

Thaler, R.H. (1997) "The Equity Premium Puzzle." *Journal of Economic Perspectives*, June.

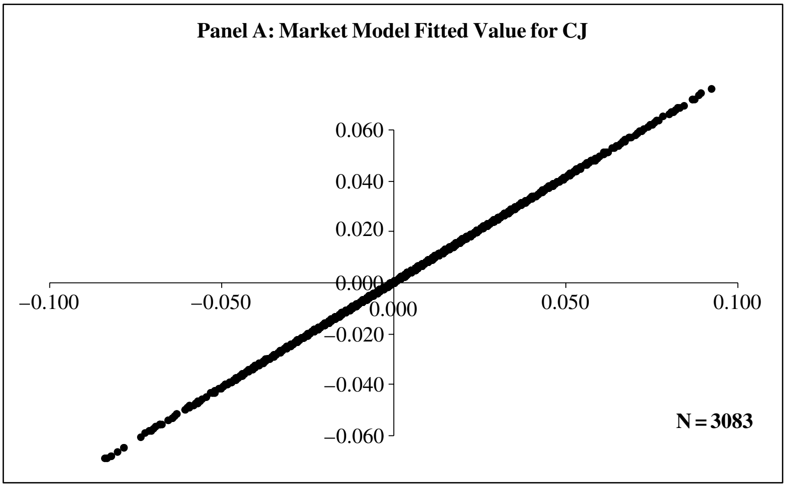
Tversky, A. and Kahneman, D. (1991) "Loss Aversion in Riskless Choice: a Reference Dependent Model." *Quarterly Journal of Economics* 1039–61.

White, H. (1980) "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity." *Econometrica* **48**: 721–46.

Zin, S. (2002) "Are behavioral asset-pricing models structural?" *Journal of Monetary Economics* **49**: 215–28.

Appendix 1

Fitted Value for 10 Selected Companies



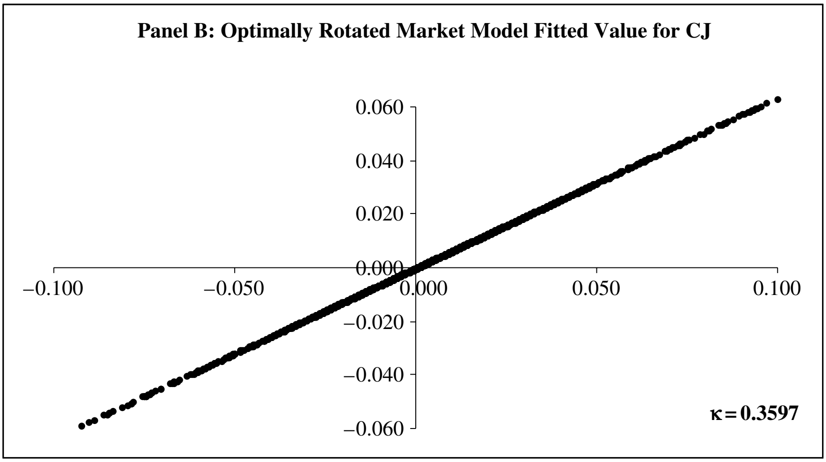


Fig. A6.1. CJ

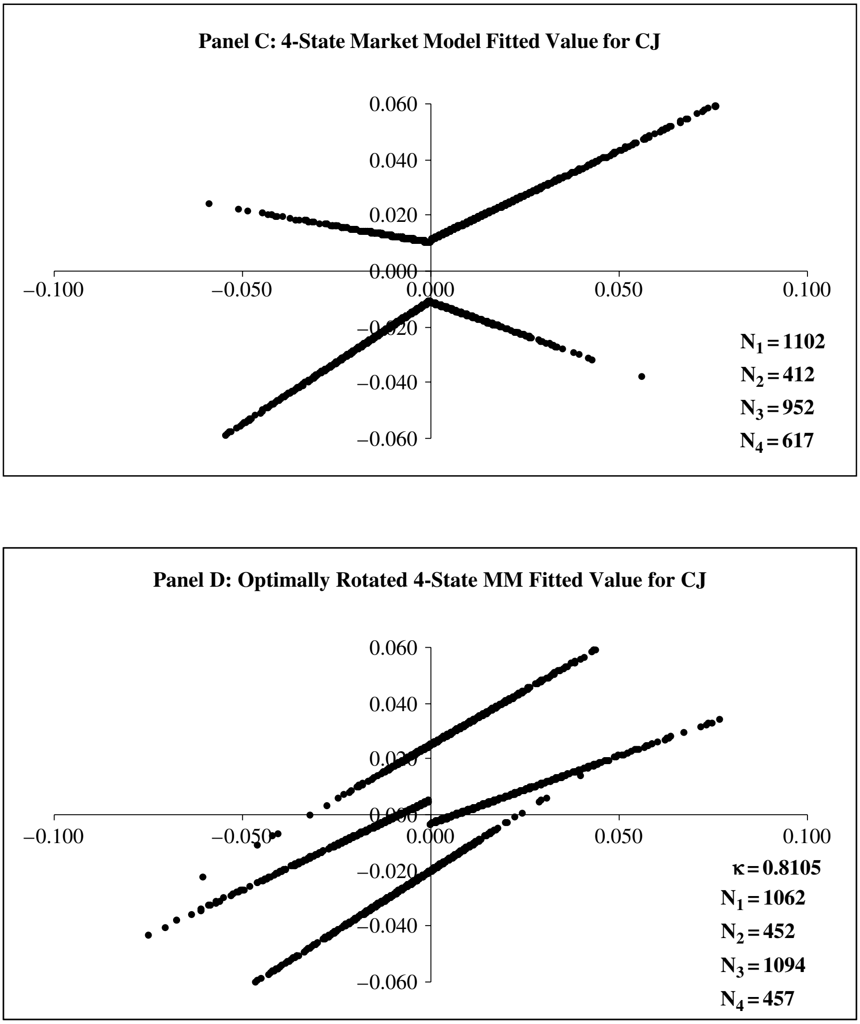
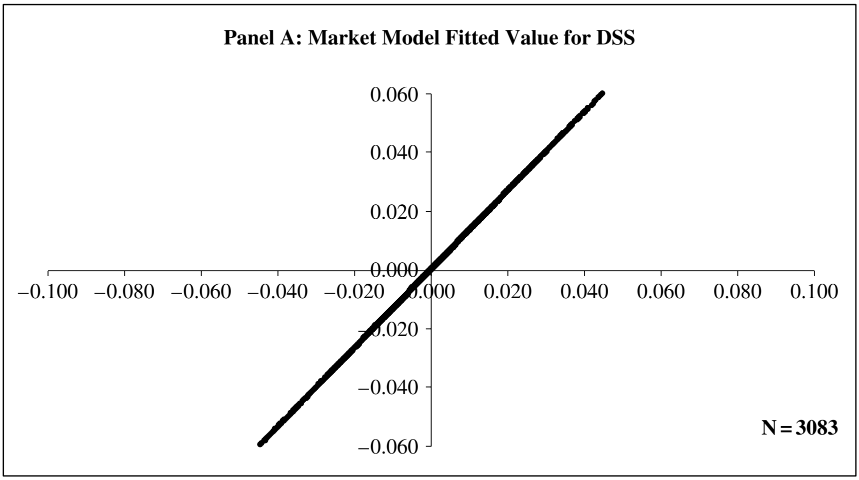


Fig. A6.1. (*Continued*)


Panel A: Market Model Fitted Value for DSS
0.060
0.040
0.020
0.000
−0.100 −0.080 −0.060 −0.040 −0.020 0.000 0.020 0.040 0.060 0.080 0.100
−0.020
−0.040
−0.060
N = 3083

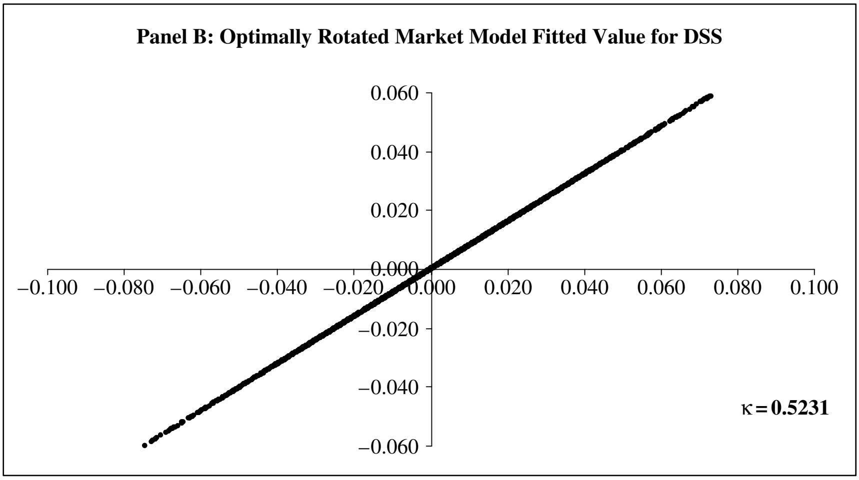
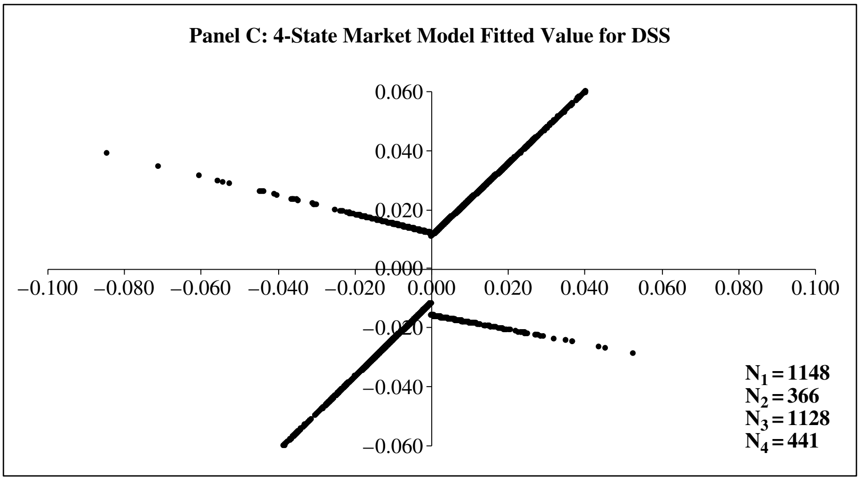

Panel B: Optimally Rotated Market Model Fitted Value for DSS
0.060
0.040
0.020
0.000
−0.100 −0.080 −0.060 −0.040 −0.020 0.000 0.020 0.040 0.060 0.080 0.100
−0.020
−0.040
−0.060
κ = 0.5231

Fig. A6.2. DSS



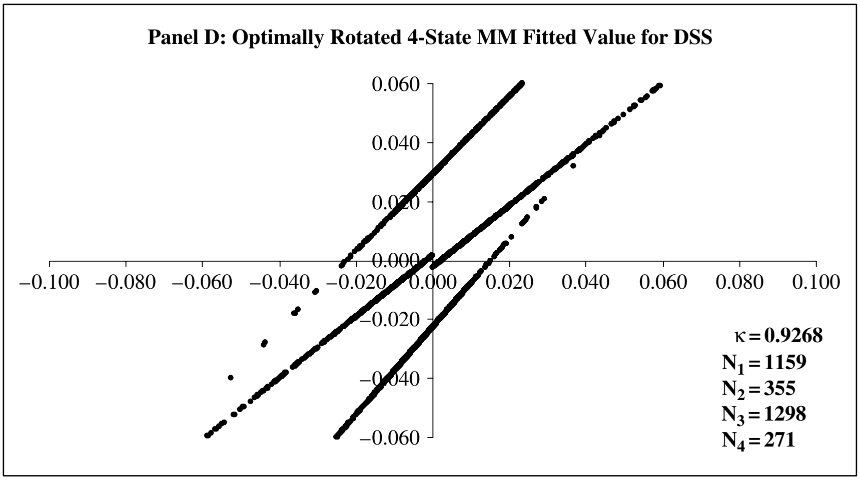
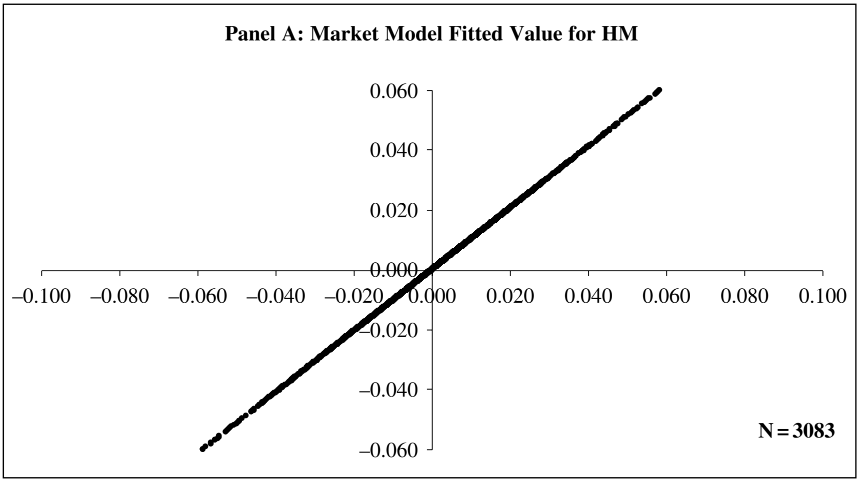


Fig. A6.2. (*Continued*)


Panel A: Market Model Fitted Value for HM
0.060
0.040
0.020
0.000
–0.020
–0.040
–0.060
–0.100 –0.080 –0.060 –0.040 –0.020 0.000 0.020 0.040 0.060 0.080 0.100
N = 3083

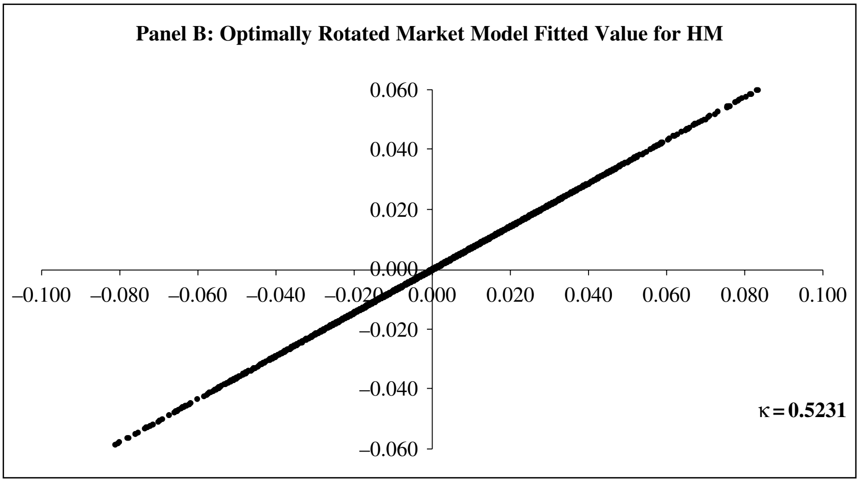
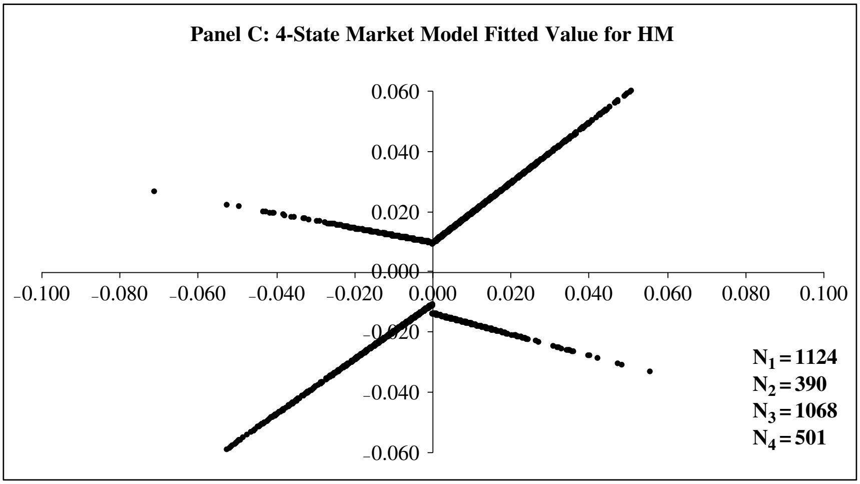

Panel B: Optimally Rotated Market Model Fitted Value for HM
0.060
0.040
0.020
0.000
–0.020
–0.040
–0.060
–0.100 –0.080 –0.060 –0.040 –0.020 0.000 0.020 0.040 0.060 0.080 0.100
κ = 0.5231

Fig. A6.3. HM



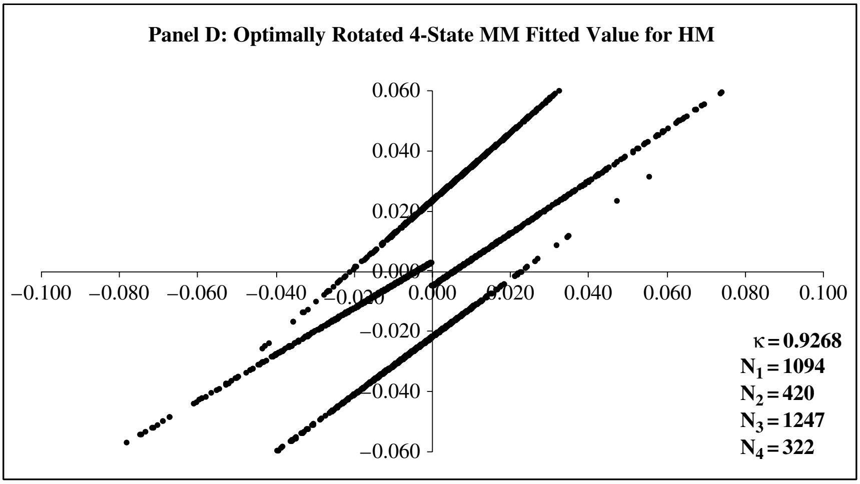
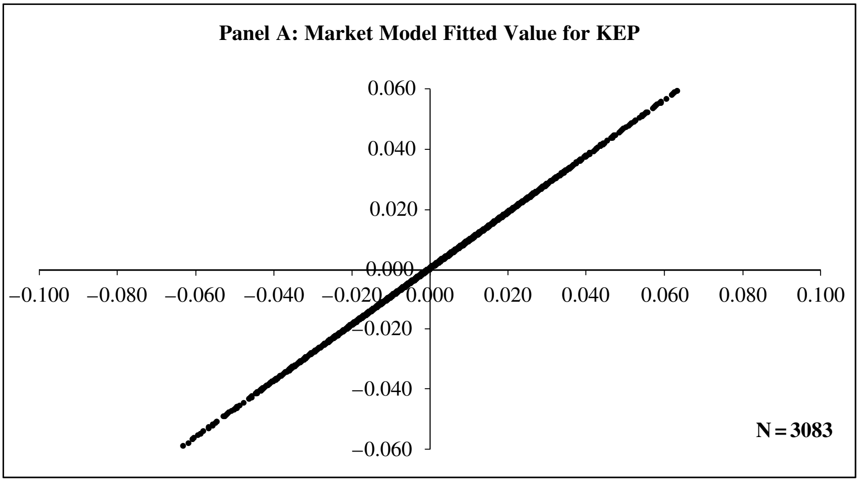


Fig. A6.3. (*Continued*)


Panel A: Market Model Fitted Value for KEP
0.060
0.040
0.020
0.000
–0.100 –0.080 –0.060 –0.040 –0.020 0.000 0.020 0.040 0.060 0.080 0.100
–0.020
–0.040
–0.060
N = 3083

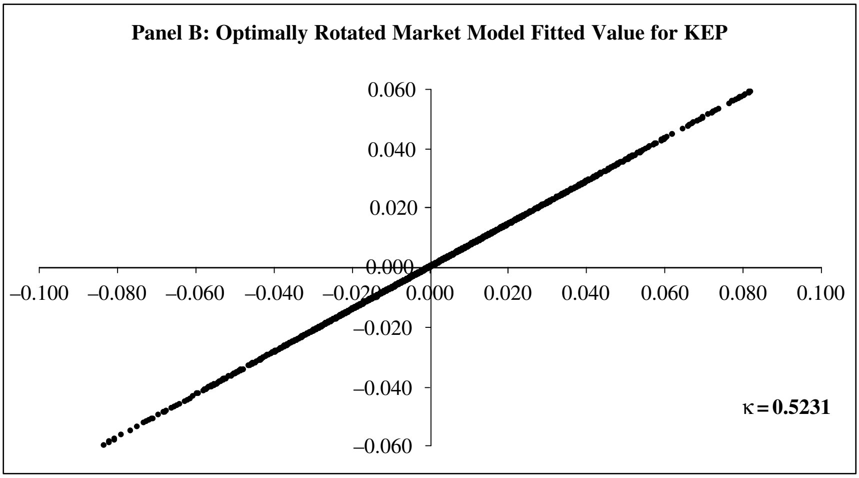
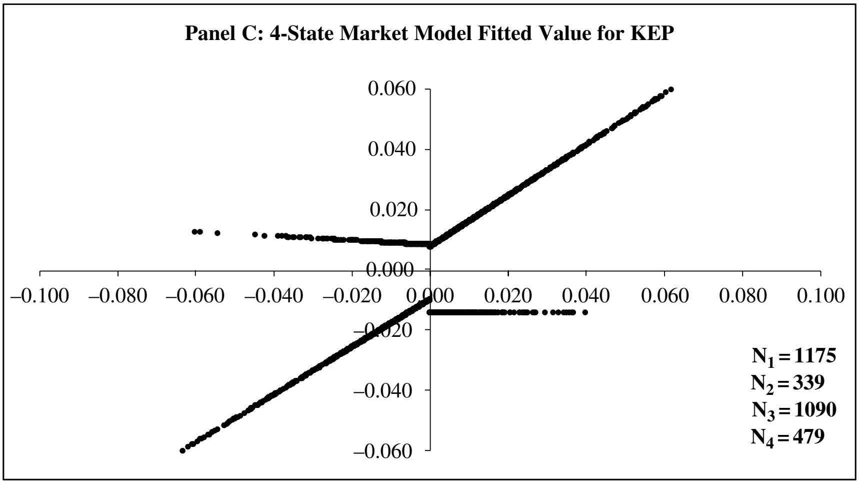

Panel B: Optimally Rotated Market Model Fitted Value for KEP
0.060
0.040
0.020
0.000
–0.100 –0.080 –0.060 –0.040 –0.020 0.000 0.020 0.040 0.060 0.080 0.100
–0.020
–0.040
–0.060
κ = 0.5231

Fig. A6.4. KEP



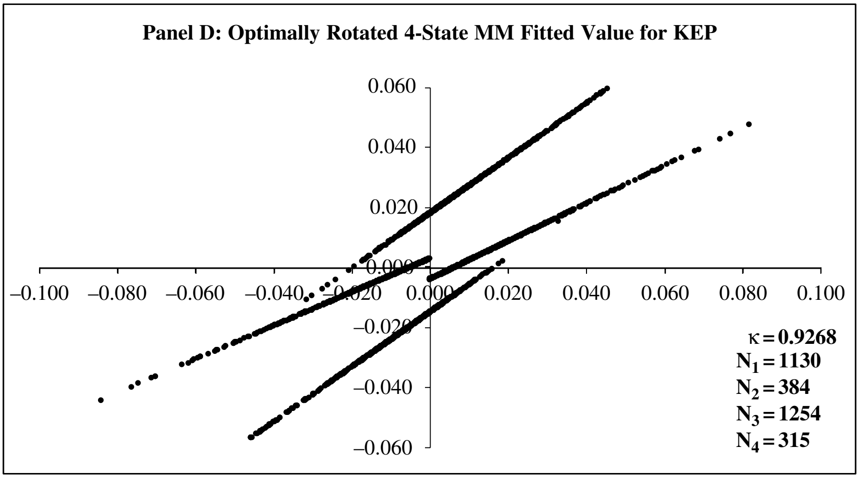


Fig. A6.4. (*Continued*)

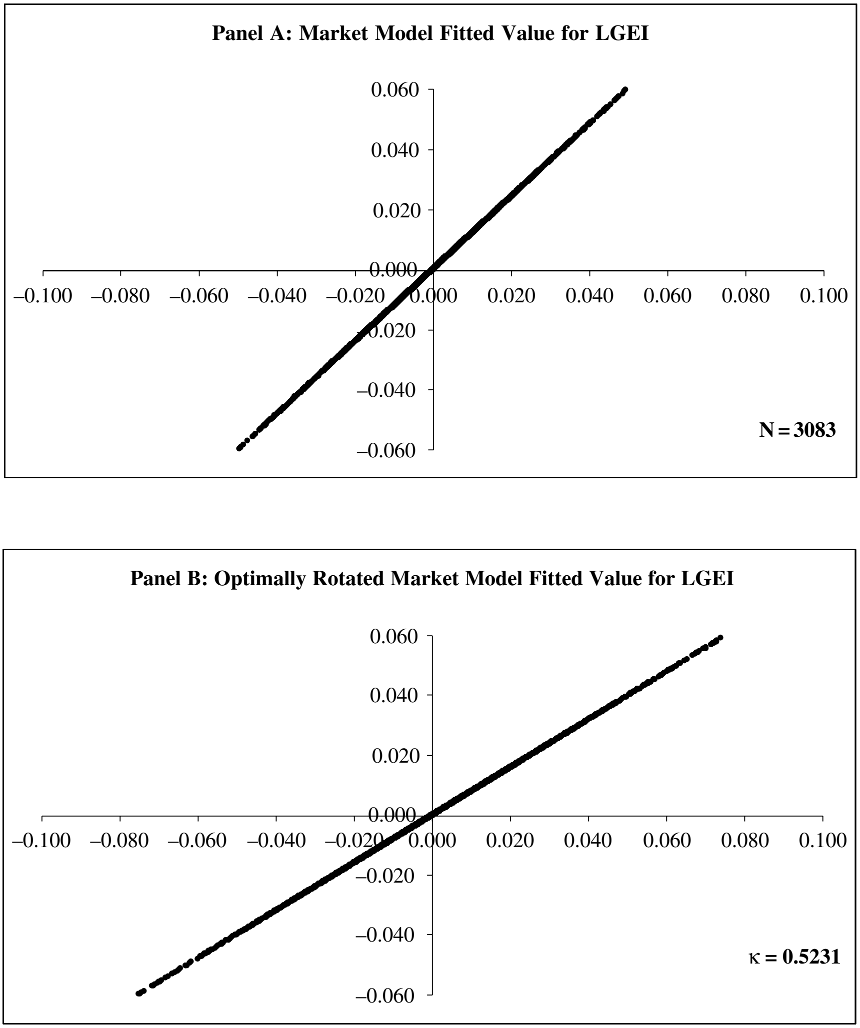
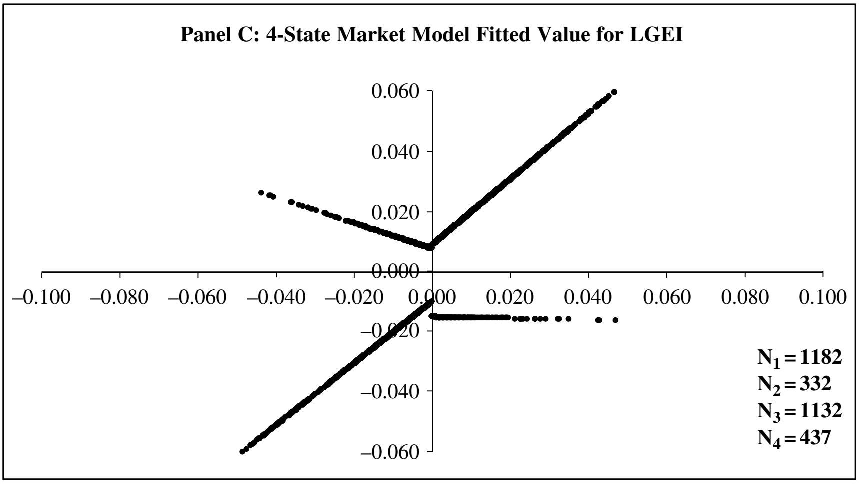

Panel A: Market Model Fitted Value for LGEI
0.060
0.040
0.020
0.000
–0.020
–0.040
–0.060
–0.100
–0.080
–0.060
–0.040
–0.020
0.000
0.020
0.040
0.060
0.080
0.100
N = 3083
Panel B: Optimally Rotated Market Model Fitted Value for LGEI
κ = 0.5231

Fig. A6.5. LGEI



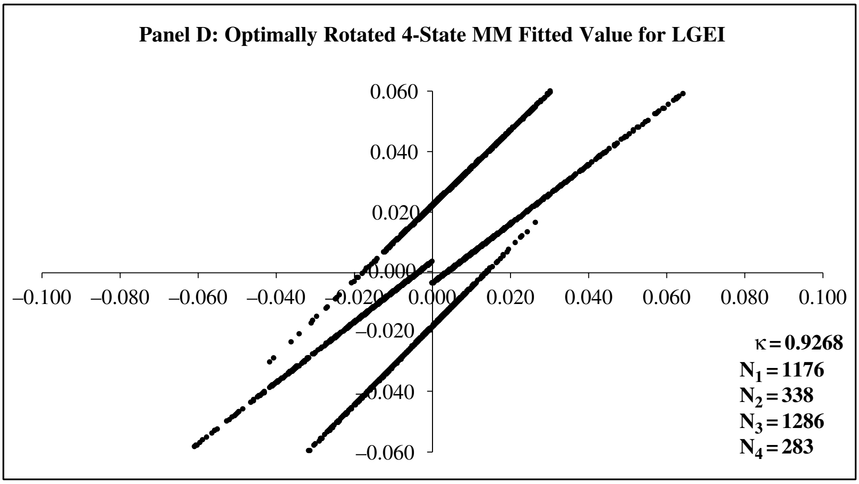
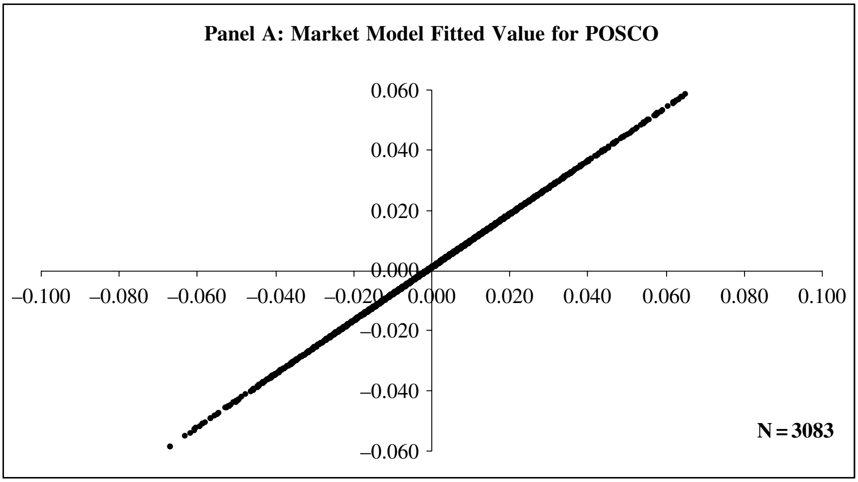


Fig. A6.5. (*Continued*)



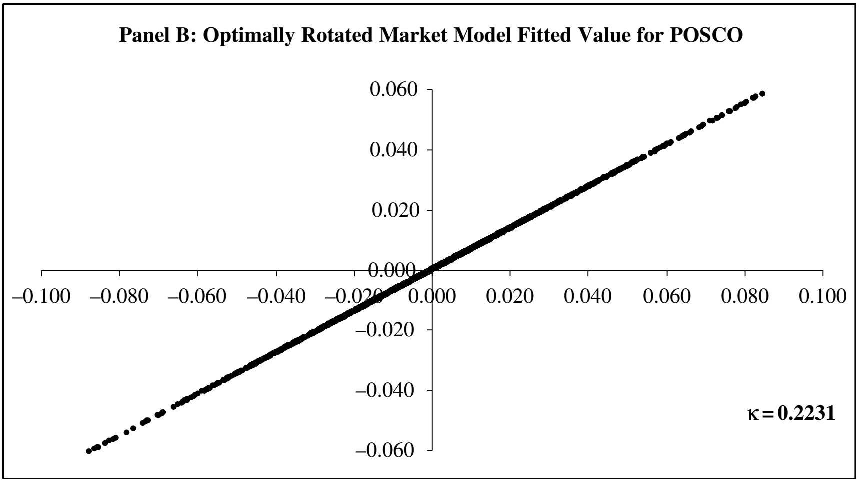
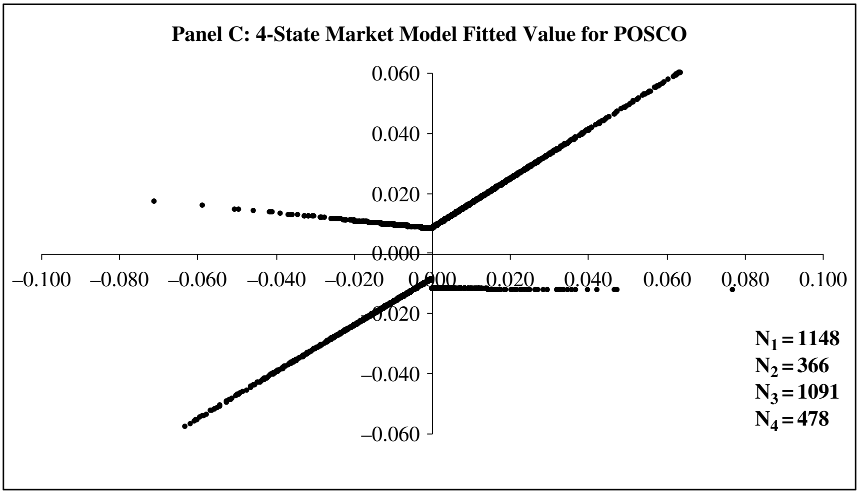


Fig. A6.6. POSCO



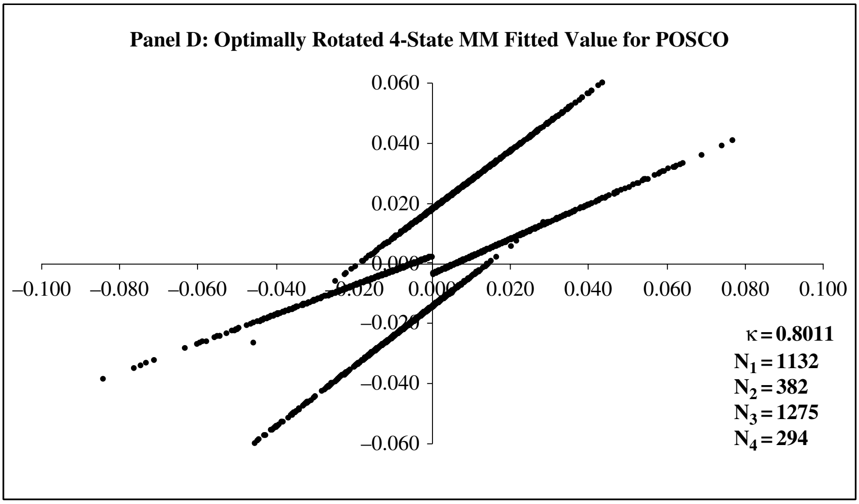
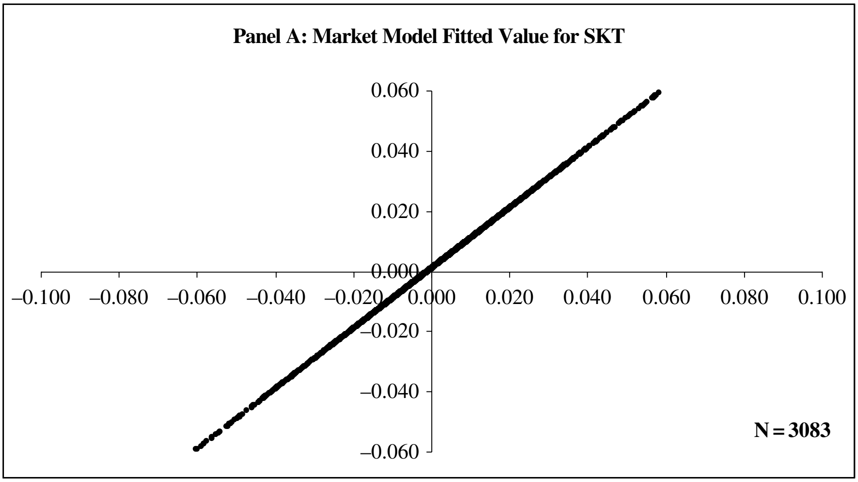


Fig. A6.6. (*Continued*)


Panel A: Market Model Fitted Value for SKT
0.060
0.040
0.020
0.000
–0.020
–0.040
–0.060
–0.100 –0.080 –0.060 –0.040 –0.020 0.000 0.020 0.040 0.060 0.080 0.100
N = 3083

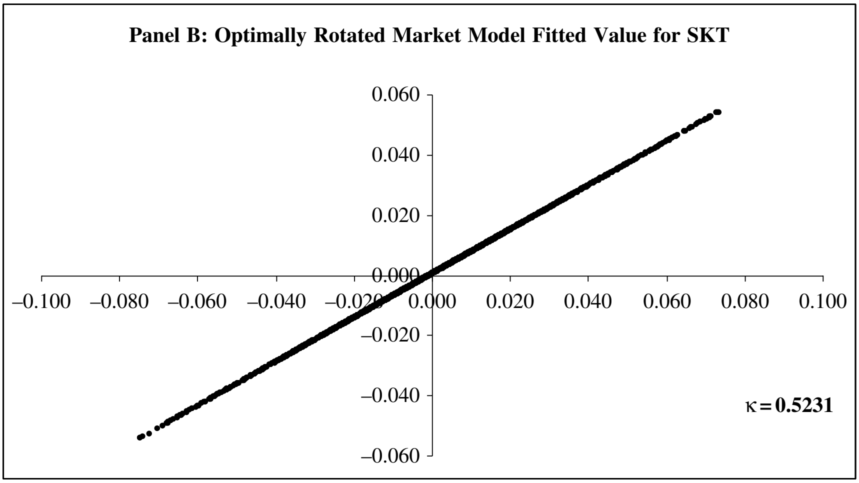
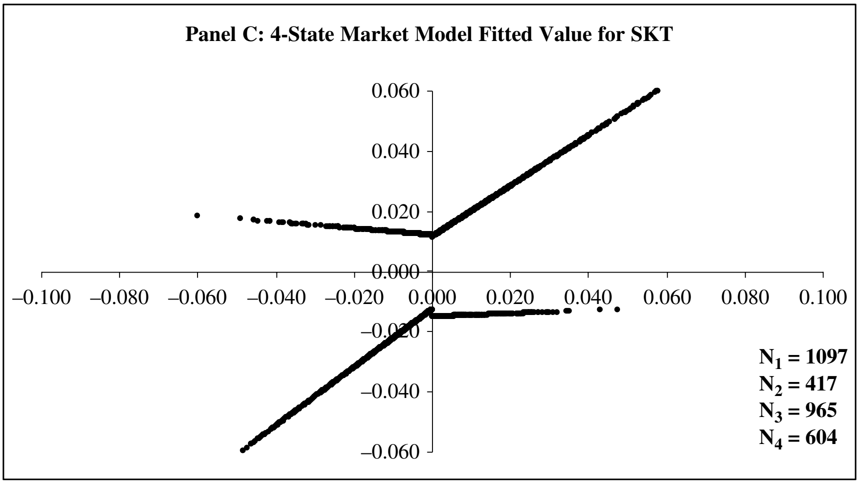

Panel B: Optimally Rotated Market Model Fitted Value for SKT
0.060
0.040
0.020
0.000
–0.020
–0.040
–0.060
–0.100 –0.080 –0.060 –0.040 –0.020 0.000 0.020 0.040 0.060 0.080 0.100
κ = 0.5231

Fig. A6.7. SKT



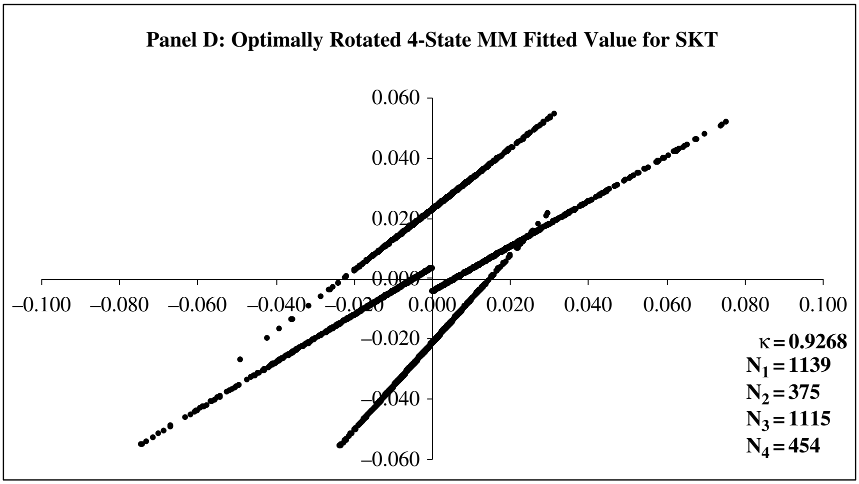


Fig. A6.7. (*Continued*)

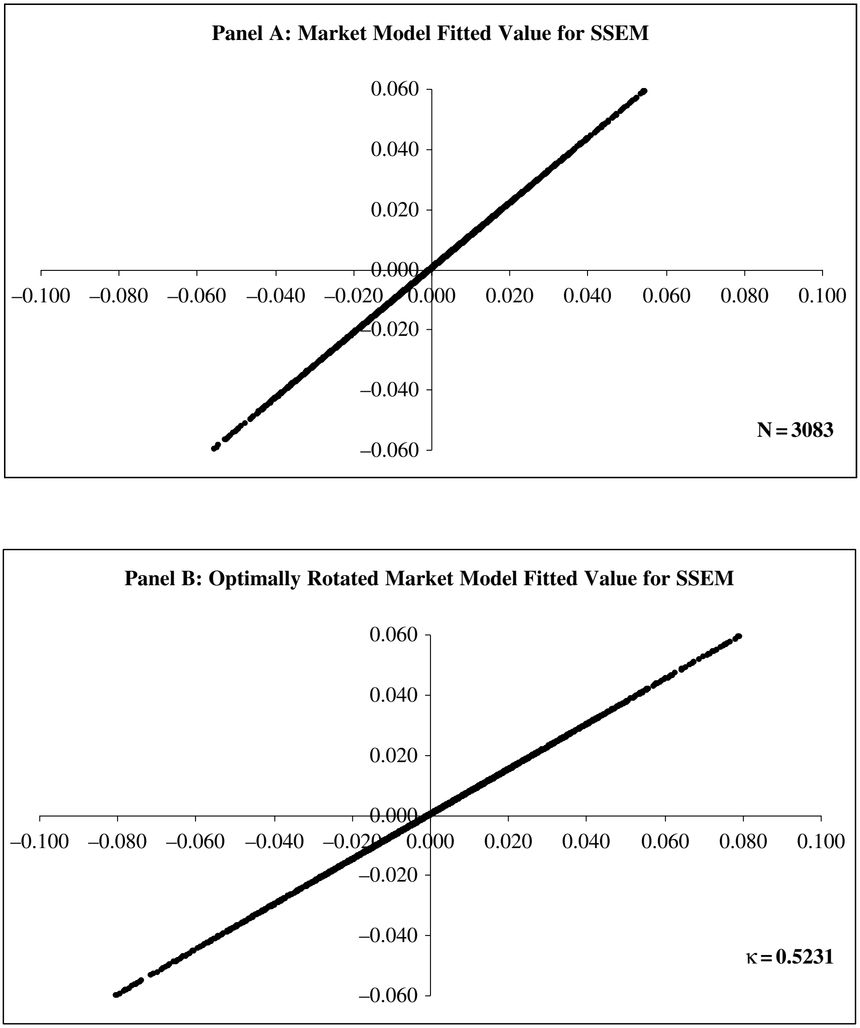
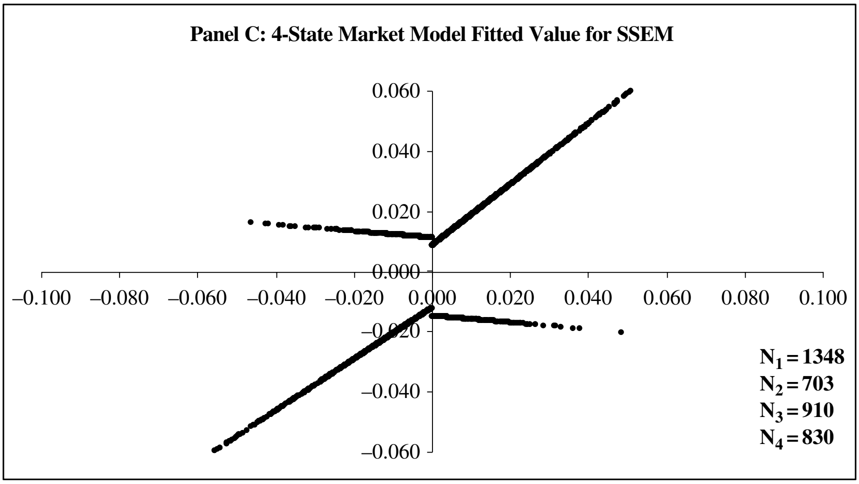

Panel A: Market Model Fitted Value for SSEM
0.060
0.040
0.020
0.000
–0.020
–0.040
–0.060
–0.100 –0.080 –0.060 –0.040 –0.020 0.000 0.020 0.040 0.060 0.080 0.100
N = 3083
Panel B: Optimally Rotated Market Model Fitted Value for SSEM
0.060
0.040
0.020
0.000
–0.020
–0.040
–0.060
–0.100 –0.080 –0.060 –0.040 –0.020 0.000 0.020 0.040 0.060 0.080 0.100
κ = 0.5231

Fig. A6.8. SSEM



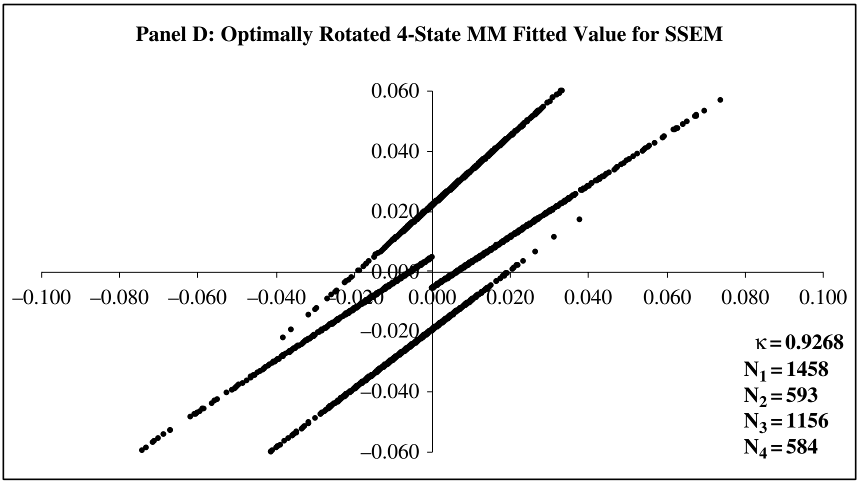
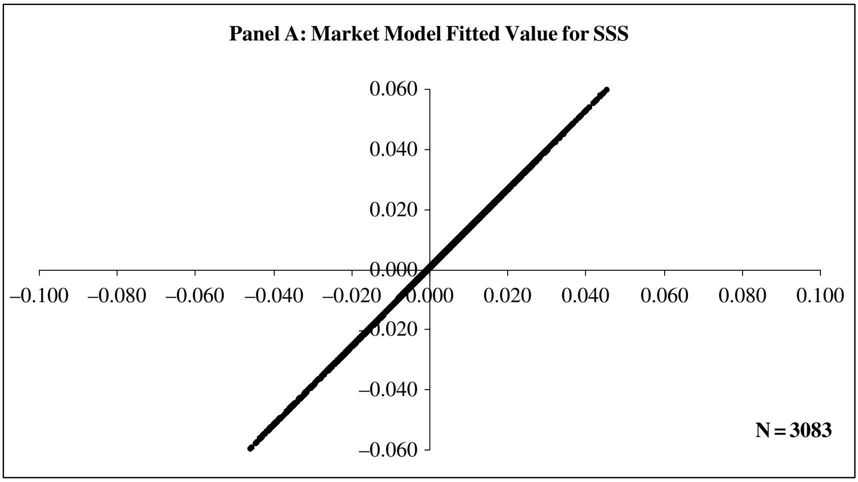


Fig. A6.8. (*Continued*)


Panel A: Market Model Fitted Value for SSS
0.060
0.040
0.020
0.000
–0.020
–0.040
–0.060
–0.100 –0.080 –0.060 –0.040 –0.020 0.000 0.020 0.040 0.060 0.080 0.100
N = 3083

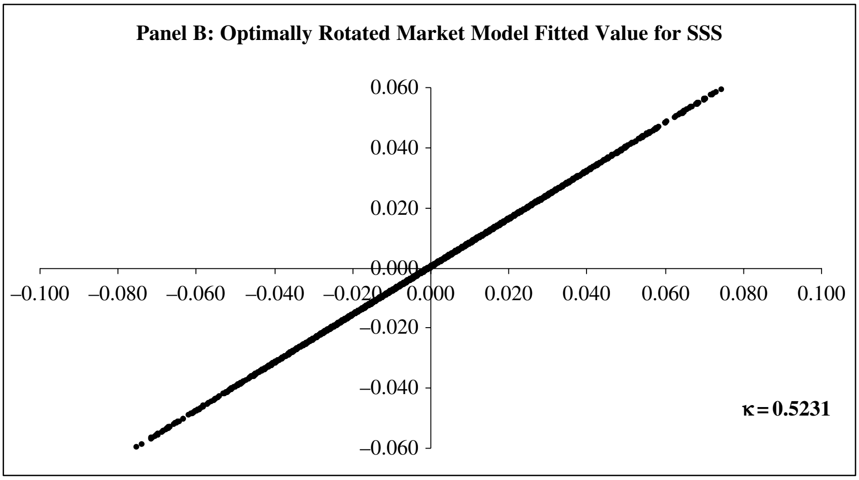
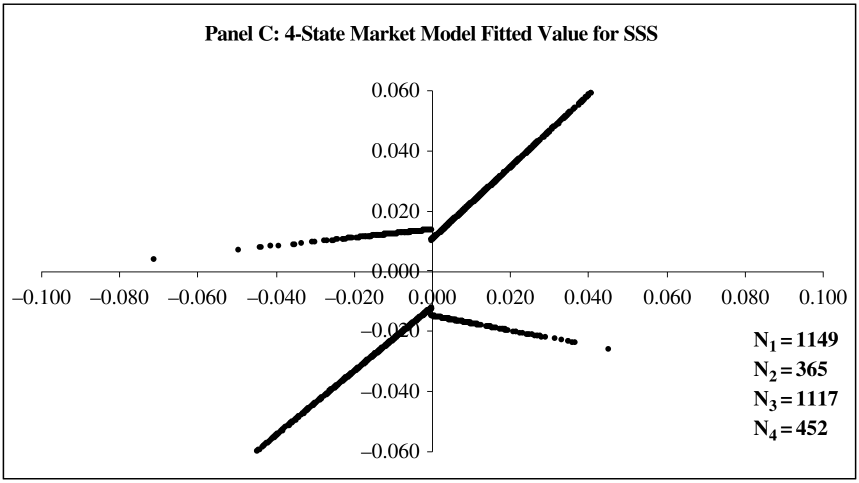

Panel B: Optimally Rotated Market Model Fitted Value for SSS
0.060
0.040
0.020
0.000
–0.020
–0.040
–0.060
–0.100 –0.080 –0.060 –0.040 –0.020 0.000 0.020 0.040 0.060 0.080 0.100
κ = 0.5231

Fig. A6.9. SSS



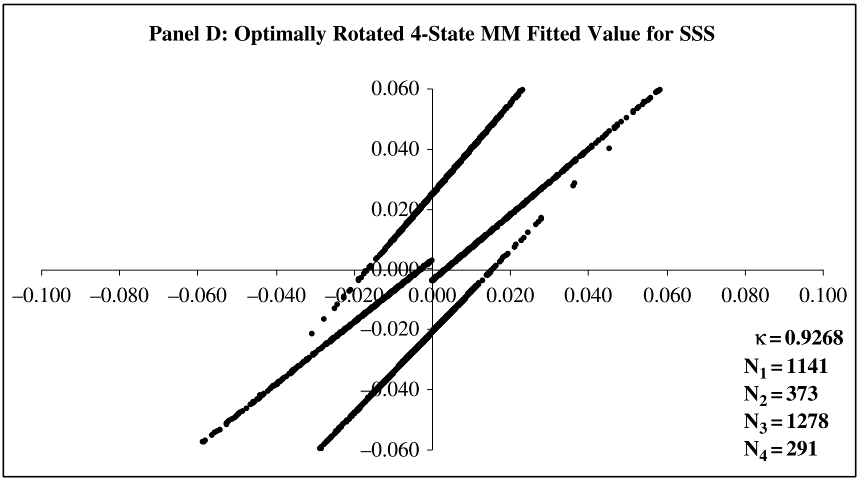
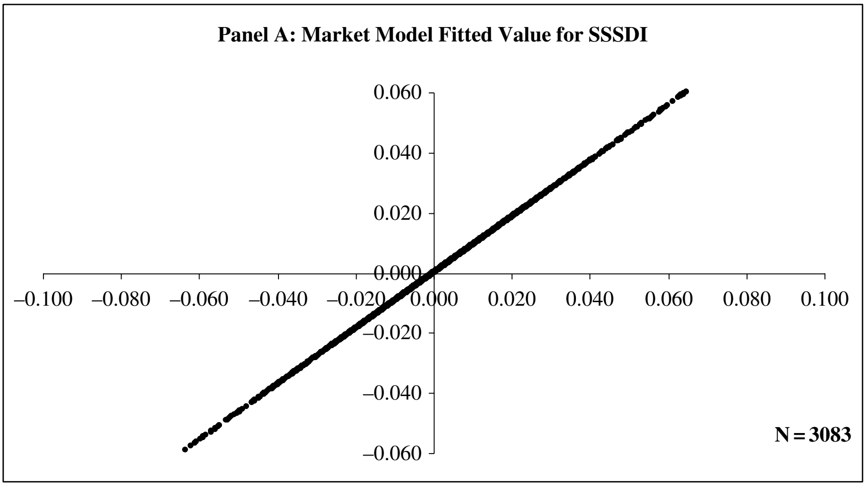


Fig. A6.9. (*Continued*)



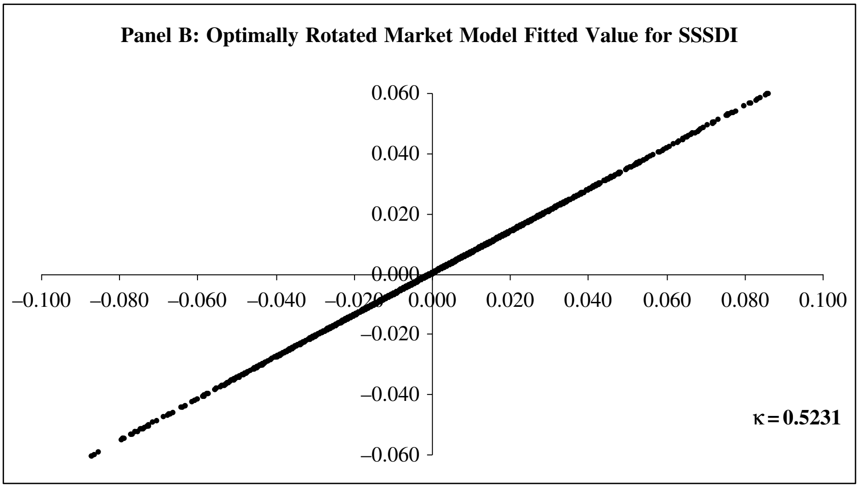
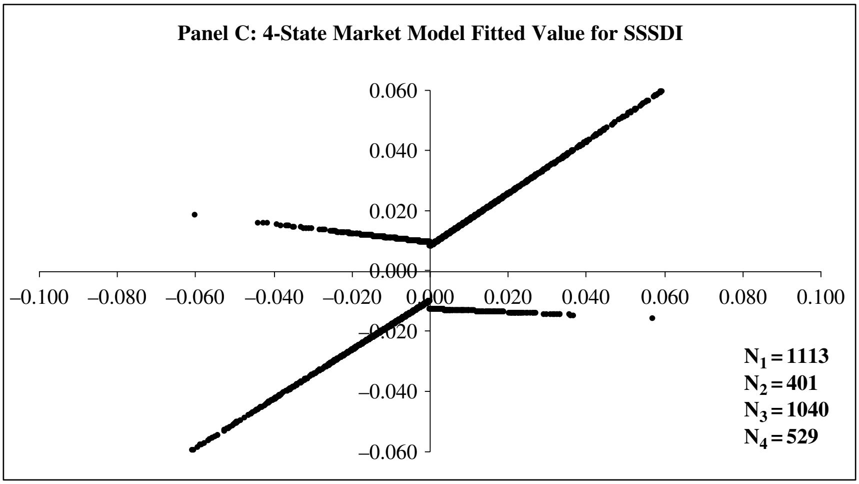


Fig. A6.10. SSSDI



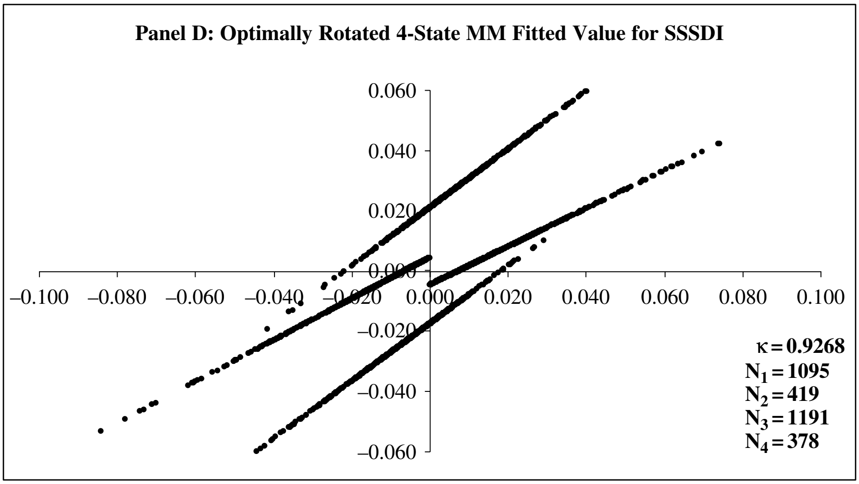


Fig. A6.10. (*Continued*)

This Page is Intentionally Left Blank

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 7

DIVIDEND POLICY OF BANK INITIAL PUBLIC OFFERINGS

Wolfgang Bessler*, James P. Murtagh and Dona D. Siregar

Abstract

This chapter investigates the short-term valuation effects and the long-run performance of bank initial public offerings (IPOs) in the United States for the period from 1970 to 1997. Overall, the empirical results provide significant evidence that the dividend policy of bank IPOs differs from that of non-banks. The dividend policy of banks has a significant impact on the long-run performance. Most importantly, banks that were acquired later on outperform the benchmark significantly but banks that continue to operate independently as well as banks that failed underperform. Moreover, the timing of the dividend initiation is an important characteristic that separates the outperformers from the underperformers.

1. INTRODUCTION

The dividend policy of firms has been one of the most important research topics in the finance literature since the publication of the seminal paper on the irrelevance of dividend policy by Miller and Modigliani (1961). In a recent paper, Fama and French (2001) provide empirical evidence that on average the relative number of dividend paying firms has been decreasing over the last decades. Especially start-up firms and initial public offerings (IPOs), i.e. firms listed on NASDAQ, have developed a tendency to avoid initiating dividend payments. However, the relative increase of the IPO group in relation to all listed firms accounts mainly for the decline of the average number of dividend paying firms. The fact that non-financial or industrial firms do not start paying dividends immediately after

*Corresponding author

going public can easily be explained with the investment opportunities and the cash flow needs of these firms. Nevertheless, of those firms that continued to be traded after going public for an extended period of time (alive firms), i.e. after accounting for those IPOs that merged or delisted (failed) after going public, about half of those IPOs eventually start paying dividends. Thus, even for IPOs, dividend policy seems to be an important signal during the first periods after listing.

Moreover, the empirical evidence indicates that the dividend policy of banks is unique in that dividend changes are used to signal the quality of a bank in an environment that is best characterized by significant information asymmetries (Bessler and Nohel 1996, Slovin *et al.* 1999, Bessler and Nohel 2000). Thus, we observe a distinctly different behavior between banks and industrial firms (non-banks) with respect to dividend policy. In addition, the capital market reactions following dividend announcements are much stronger for banks. Because most of the previous studies suggest that the dividend policy of IPOs is different from that of established firms and because the stock market reaction to dividend changes by banks is different from that of industrial firms, it is interesting to investigate the dividend policy of banks after they went public (IPOs). This specific question has not been addressed in the literature so far. Thus, we add to the empirical evidence on bank behavior, initial public offerings, and dividend policy, by investigating the valuation effects of the dividend policy of bank IPOs.

The focus of this study is on investigating both the long-run performance (BHARs) and the short-term valuation effects (CARs) following dividend initiation announcements by banks that went public in the United States during the period from 1970 to 1997. For this we first analyze the long-run performance of dividend initiations especially with respect to the timing of the first payment. Second, we test whether there are significant differences in the performance among different categories of bank IPOs, i.e. banks that merged later on, were dropped, or continued to operate independently (alive) at least up to the end of the sample period. Finally, we examine the capital market reaction around the time of the dividend initiation announcements. Thus, we are interested in the stock market reaction around the dividend initiation date (short-term valuation effects) as well as in the long-run performance following a dividend initiation.

The rest of this chapter is organized as follows. The next section reviews the literature with respect to dividend policy, initial public offerings, bank initial public offerings, and bank dividend policy. Section III provides a description of the data and methodology employed in this study. The results are presented and discussed in Section IV. Section V concludes the paper.

2. REVIEW OF THE LITERATURE

The literature that is relevant to our research question comes from various research areas of corporate finance and banking. First, agency problems and dividend policy need to be addressed in that dividend announcements provide information to shareholders about the future performance of the firm. Thus, dividend changes are an important signal that usually results in significant short-term and long-run valuation effects. This aspect is addressed in the next section. Moreover, the literature on dividend policy of IPOs is important in this context and is discussed in the second section. With respect to the banking literature, the empirical findings on the performance of bank IPOs and the valuation effects of dividend changes by banks are interesting issues that are reviewed in Section III and Section IV, respectively.

2.1. Dividends as an Information Signal

In a world of symmetric information, all economic agents have the same information with respect to the valuation of a firm. However, this assumption does not hold any longer under more realistic assumptions, for example, when one of the agents is better informed about the firm's prospects than the other agents. In such an environment it is reasonable to assume that managers possess an information advantage about their own firm. Therefore, financial decisions may signal a change in the quality of the firm to the market. One of these actions that management can employ to reliably signal information to shareholders is a change in dividend policy (Williams 1992, Allen and Michaely 1995, Bessler and Ellermann 2004). Consequently, dividend changes, especially dividend initiations and dividend increases may be used as an active strategy to convey positive information to the market (Michaely *et al.* 1995). In contrast, dividend decreases and dividend omissions are usually associated with negative information. In this case, although this information revelation is not actively pursued and intended by management, it cannot be avoided due to the weak financial situation of the firm.

There are two main hypotheses that are helpful in explaining what information is contained in a dividend announcement: the earnings hypothesis (cash flow) and the free cash flow hypothesis. The earnings hypothesis proposes that by paying out cash to the shareholder, the management signals to the market that the firm has good investment projects and is able to generate positive cash flows in the long term. An increase in the level of dividends is viewed as a positive signal by the financial market because firms committed to paying dividends indicate that they are capable of generating positive cash flows in the long-run. A decrease in dividends is viewed as a negative signal and may suggest up-coming long-term financial problems. Consequently, investors should lower their valuation of these firms. Studies by Lintner (1956), Fama and Babiak (1968), Battacharya (1979), John and Williams (1985), and Miller and Rock (1985) find evidence for this hypothesis. In addition, Asquith and Mullins (1983), Healy and Palepu (1988) and Venkatesh (1989) report on average positive stock price reactions to the announcement of dividend initiations under the assumption of the earnings signaling hypothesis.

The free cash flow theory hypothesizes that a firm with substantial free cash flows will have a tendency to overinvest by accepting marginal investment projects with negative net present values (Jensen 1986). If managers are overinvesting, an increase in dividend payments will decrease the available cash flows and therefore limit overinvesting. Hence, the market value of the firm should increase. In contrast, a decrease in dividends may facilitate overinvesting. As a consequence the stock price should decrease.

The valuation effects of dividend increases and dividend decreases, however, should be viewed and interpreted carefully. Some argue that the utilization of dividends as a signal depends on the availability of other signals to the firm. Larger firms have more ways to signal their quality at reasonable costs. They may utilize analyst reports as an effective and less costly practice to signal the quality of their projects. In contrast, the signaling opportunities are different for smaller firms. With limited alternatives available, dividends are a reasonable signaling mechanism. Thus, the relative valuation effects of dividend changes may be a function of firm size. Ellermann (2003) provides supporting empirical evidence for this view for Germany in that the valuation effects of dividend decreases are more severe for IPOs than for established firms.

Ambarish *et al.* (1987) as well as John and Lang (1991) propose that dividends may only be one measure among others in evaluating a firm's quality. For example, they are interpreted by the market in the context of the investment opportunities that are available to the firms. Thus, the optimal signals used are determined by the nature of the firms' investment opportunities. Established firms, conceptually characterized by having valuable assets in place and limited opportunities to invest, often use a large pay-out ratio as their primary signal. Growth firms, characterized by few assets in place and valuable opportunities to invest, do not often employ dividends but instead use investments as their main signal. Again, these models predict that the announcement of a dividend increase results in larger stock price increases for established firms compared to growth firms.

In addition, John and Lang (1991) investigate insider trading prior to the announcement of dividend changes. In their study they show that the announcement effect of dividends is influenced by the nature of a firm's investment opportunities and by the productivity of its current capital investments. However, not all dividend increases are viewed by the market as good news. In some cases, an increase in dividends is a signal that the firm may not have sufficient outstanding investment opportunities. The authors suggest that the interpretation of an increase in dividend has to be based on insider trading activity immediately prior to the announcement.

In a similar study of the relation between dividend policy and investment opportunities, Lang and Litzenberger (1989) examined the announcement effect of large dividend changes and linked it to investment opportunities available to the firm by utilizing Tobin's Q measurement. They find that large dividend changes are significantly affected by investment opportunities. The average abnormal returns at the dividend announcement date are more than three times larger for firms with average Qs of less than one than for firms with average Qs of greater than one. Dividend increase and decrease announcements result in similar effects when each event is analyzed separately.

Dyl and Weigand (1998) hypothesized that the initiation of cash dividends coincides with a reduction in the risk of a firm's earnings and cash flows. Using a sample of 240 firms (NYSE and AMEX) that initiated dividend payments during the period from January 1972 to December 1993, they show that the variance of daily returns as well as the average beta decreases in the year following the dividend initiation. Thus, it seems fair to conclude that management is in a position to use dividend changes to signal the quality of the firm. The important question to investigate, however, is whether all firms can employ dividend changes in the same manner, or whether the magnitude of the impact depends on the maturity of the firm (e.g. IPO) as well as on the industry (e.g. banking) in which the firm is operating.

2.2. Initial Public Offerings and Dividend Policy

Lipson *et al.* (1998) compare the performance of IPOs that initiate dividends with those that do not. The analysis is carried out by building two groups of matching firms. One group consists of firms that do not pay dividends. This group is matched with dividend-initiating firms, controlling for the time of the going public and the industry. Another group of firms is matched with the dividend-initiating firms by size and industry (size matched) but these are already paying dividends. They argue that a firm should engage in signaling activities, especially to differentiate itself from other firms that the market perceives as having similar prospects. By grouping the samples, the authors examined comparable IPOs in terms of life cycle and future growth opportunities. They found that raw and industry adjusted earnings increase for the initiating firms in the first year after the dividend initiation, but not in the second year. Earnings surprises for initiating firms are more favorable than for non-initiating firms by the second year following the dividend initiation. However, the earnings surprises of the initiating firms are not significantly different from the size-matched samples or industry averages. Thus, the study suggests that if dividend initiations signal future earnings prospects, the signal must differentiate a newly public firm from other newly public firms but not from established firms in the industry. Thus, there is a strong size effect instead of an industry effect.

Similar to the work of DeAngelo *et al.* (1996), Lipson *et al.* (1998) also found that changes in dividend levels can be a valid signal but only if a significant commitment of cash is used. The dividend commitments of initiating firms represent about 5% of earnings. If non-initiating firms had matched the dividend yield, dividend-to-sales ratio, or dividend-to-assets ratio of the initiating firms, the dividend commitments would have been about 8.5% of earnings. They reported that the difference in dividend commitments between initiating and non-initiating firms is significant. This suggests that dividend commitments may be sufficiently large to support a signaling equilibrium in the context of dividend initiations. In addition, Lipson *et al.* (1998) provide evidence that dividend-initiating firms are usually larger and more profitable than the non-initiating firms that went public at the same time.

The dividend policy of initial public offerings may also depend on a number of other factors, such as hot issue markets, accounting standards and the banking system. In hot issue markets, for example, the Neuer Markt in Germany during the period from 1998 to 2000 (Bessler and Kurth 2005), the dividend policy of IPOs may not be that important to the investor because of the extraordinary return opportunities from stock price increases compared to the only marginal return contribution of divided payments. Consequently, these IPOs did not initiate dividend payments or even stopped paying dividends. In contrast, in a stock market environment with a more normal valuation level but with severe information asymmetries, dividend payments may still be an important signaling device even for IPOs. Bessler (2001) and Ellermann (2003) report that due to the relatively low accounting and reporting standards in Germany, most of the IPOs during the period 1980–95 did pay dividends before and at the time of the IPO but decreased and cut their dividends in the years after going public. This is in sharp contrast to the dividend behavior of non-bank IPOs in the United States. Thus, in an environment of information asymmetries, dividend policy may still be an important means to convey information reliably. However, various IPO studies also provide empirical evidence that the going public process is characterized by severe agency problems and that other means are available to signal the quality of the firm, such as venture capital involvement, underwriter reputation, and extended lock-up periods (Bessler and Kurth 2006).

2.3. Bank Initial Public Offerings

Houge and Loughran (1999) investigate the long-run performance of banks that went public, as measured by the five-year post-IPO returns. They find empirical evidence that the bank IPOs do not experience underperformance until two or three years after the offering. However, they find significant underperformance with respect to several market benchmarks over a five-year holding period. According to Houge and Loughran, the reason for this result is that the banks maintained initially a relatively constant proportion of loaned assets throughout the pre-IPO period and did not experience a dramatic shift in profitability after the offering. Compared to the industry average, the banks in the sample reported lower levels of loan loss provisions during the pre-IPO years. Following the offering, however, the banks increased their loan loss allowances up to the aggregate industry level. Banks usually use these provisions for loan losses to adjust for higher current and future levels of loan write-offs.

The increase in post-offering loan charges is consistent with the banks adopting a marginally riskier loan strategy to grow their asset base. Banks with more aggressive loan growth around the public offering have a significantly higher proportion of post-IPO loan loss provisions than banks with more conservative growth rates. The poor long-run performance of the banks is directly attributed to the high growth institutions, while the low growth banks outperformed the benchmarks. This result is interesting and important in that it is in contrast to the findings for IPOs of non-financial firms. IPOs usually underperform the benchmark and firms with high growth potential seem to have a relatively better performance. Moreover, the performance of banks also seems to be related to firm size. In fact, size is found to be an important explanatory variable of post offering returns. Larger banks in the sample lagged the non-IPO bank index by −20.2%, while smaller banks matched the benchmark over the five-year holding period. The more negative valuation effects of larger banks are consistent with the stock market reaction of dividend cuts and omissions by commercial banks as reviewed in the next section.

2.4. Valuation Effects of Bank Dividend Announcements

There exists sufficient empirical evidence that the dividend policy of banks is special and is significantly different from that of non-banks (Bessler and Nohel

1996, 2000). The multidimensional aspect of the asymmetric information problems faced by banks and bank customers, shareholders, and examiners is an important aspect in arguing that banks are different. Quarterly dividend payments and annual dividend increases have been common for banks in the United States. Shareholders may expect regular dividend payments from those financial institutions that are viable and that currently are not faced with severe financial difficulties. In addition to their shareholders' anxiety, banks have to consider the assurance needs and confidence aspects of their customers. Quarterly announcements of stable or growing dividends may therefore be utilized by banks as a means for providing positive information about the bank's solvency to investors, customers, and regulators alike. Hence, dividends provide some positive information about the bank's current success and about the future viability of the bank. In contrast, dividend cuts lead to strong negative valuation effects for banks of about −8% for a two-day period and up to −12% for a two-week period (Bessler and Nohel 1996). In a world with information asymmetries, banks that go public (bank IPOs) may consider to start paying dividends early on in order to signal their quality and viability to shareholders. Important research questions are whether the timing of the dividend initiation is an important signal and whether the weaker banks can duplicate this signal and fool the market about their quality. In addition, due the information asymmetries, banks are especially exposed to contagion effects from dividend decreases of other banks (Bessler and Nohel 2000).

In sum, this literature review suggests that compared to industrial firms, banks are special with respect to both the long-run performance of initial public offerings as well as to the dividend announcement effects. Thus, it is interesting to investigate in greater detail the dividend policy of bank initial public offerings, especially the short-term and long-run performance as well as its dependency on the future status and success of the bank.

3. DATA AND METHODOLOGY

3.1. Data

The original sample includes all banks that went public in the United States between 1970 and 1997. The list of these bank IPOs was obtained from the SDC Platinum database. As in other studies we include only bank IPOs with an offering price equal to or greater than $5 per share and total proceeds raised equal to or greater than $1.5 million. Moreover, the following IPOs were excluded from the sample: spin-offs, unit issues, reverse LBOs, and ADRs. For every bank in the sample we obtain daily and monthly stock return data and S&P500 returns from the Center for Research in Security Prices (CRSP). The same source is used for the dividend announcement date and dividend payment date, dividend amount, and dividend codes. The final sample includes 431 bank IPOs that are listed both in the SDC and CRSP.

The status of a bank IPO is obtained from the CRSP data coding schemes called "delisting codes". The coding scheme categorizes firms in five main groups: active, mergers, exchanges, liquidations, and dropped. This study focuses on active, mergers, and dropped bank IPOs. "Active" means a bank was still operating from the time it went public until the end of the period covered in this study, December 31, 2000. "Mergers" are banks that were acquired. Banks that were permanently delisted from trading at their current exchange are classified as "dropped". In Table 7.1, the 431 banks are grouped based on the four digits of the Standard Industrial Classification (SIC) code system and the delisting codes.

In the empirical analysis of the long-run and short-term performance of bank IPOs, the number of banks in the sample is sometimes lower than the original 431 IPOs. This is due to several reasons related to the objective of our analysis as well as to the availability of stock price data. First, 11 banks were excluded because they were not in the category of Alive, Merged or Dropped. Second, 13 banks were omitted due to the lack of stock return data especially at the beginning of the listing period. Third, only those banks that started paying dividends within 12 quarters after listing were included, thus excluding 60 banks that begun dividend payments later on. Forth, 95 banks that never paid dividends during the period examined in this study had to be omitted. The final sample for our long-run and short-term performance study consists of 252 bank initial public offerings.

Table 7.1 Summary of Bank Initial Public Offerings Sample

Bank Classification	SIC	Original	Not Alive, Merged, Dropped
	6000	2	–
National Commercial Banks	6021	22	–
State Commercial Banks	6022	68	1
Commercial Banks, NEC	6029	2	–
Savings Institutions, Federally Chartered	6035	200	6
Savings Institutions, Not Federally Chartered	6036	129	4
Branches and Agencies of Foreign Banks	6081	1	–
Functions Related to Deposit Banking, NEC	6099	4	–
Offices of Bank Holding Companies	6712	3	–
Remaining Bank IPOs		431	420

3.2. Methodology

In this study we investigate both the long-run performance as well as the short-term valuation effects of bank IPOs with respect to dividend initiations. For measuring the long-run performance we employ BHARs and for the short-term valuation effect we use CARs as described in the next two sections.

3.2.1. Measuring Long-run Performance

In the empirical investigation the standard buy and hold abnormal returns (BHAR) approach is used to measure the long-run performance of bank initial public offerings relative to the market index. Calculating buy and hold abnormal returns is the usual method to investigate the long-run performance of IPOs (Ritter 1991), although other methods have been discussed in the literature (Bessler and Kurth 2006). BHARs are calculated as the geometric return of the monthly stock returns of bank IPOs minus the geometric return of the monthly market returns (S&P500) over various investment periods ranging from 1 month to 36 months. Thus, the buy and hold abnormal return for an individual stock is calculated as follows:

$$\text{Buy and Hold Abnormal Return} = \prod_{t=1}^{T}(1 + R_{t(IPO)}) - \prod_{t=1}^{T}(1 + R_{t(S\&P500)})$$

where $R_{t(IPO)}$ is the monthly return of a bank IPO and $R_{t(S\&P500)}$ is the monthly return of the S&P500.

3.2.2. Measuring Short-term Valuation Effects

The standard event study methodology is applied to analyze the market effects of the first dividend announcement of a bank IPO, with the announcement date of the first dividend payment as the particular event date. Throughout this chapter, the terms "first dividend payment" and "dividend initiation" are used interchangeably. As an event window we employ the usual 21-day period from -10 to $+10$ around the announcement date. The market model is employed to model the expected return of the bank IPO over the event period. The parameters of the market model are estimated over a period from day -100 (or less) to day -11 prior to the dividend announcement.

Abnormal returns during the event window are calculated as the daily returns during the event window minus the expected returns. For measuring the significance of a short-term valuation effect of a dividend announcement we use the standard approach. For a single event, the H_o hypothesis is that $C\hat{A}R_i \sim N(0, \sigma_i^2)$, where $C\hat{A}R_i$, is the cumulative daily abnormal return of the dividend initiation announcement event of bank *i*. The significance test of H_o is constructed by using the standardized cumulative abnormal returns that are calculated as follows:

$$SC\hat{A}R_i = \frac{C\hat{A}R_i}{\hat{\sigma}_i}$$

where $\hat{\sigma}_i$ is replaced with the $\hat{\sigma}_{\varepsilon i}$ from the estimation of the market model. Under the null hypothesis, the distribution of the standardized cumulative abnormal returns follows a Student t distribution with L – 2 degrees of freedom, where L is the length of the estimation window.

The returns from grouping several events are assumed to be $C\overline{A}R_i \sim N(0, \bar{\sigma}_i^2)$, with

$$C\overline{A}R = \frac{1}{N}\sum_{i=1}^{N} C\hat{A}R_i$$

$$\bar{\sigma}_i^2 = \frac{1}{N^2}\sum_{i=1}^{N} \sigma_i^2$$

where N is the number of banks and $\hat{\sigma}_i$ is a consistent estimator of $\bar{\sigma}_i$ so that

$$\hat{\bar{\sigma}}_i^2 = \frac{1}{N^2}\sum_{i=1}^{N} \hat{\sigma}_i^2$$

can replace $\bar{\sigma}_i^2$.

For testing the significance of the abnormal returns we employ the procedures suggested in Campbell *et al.* (1997). The method is applied to test for the significance of an individual bank as well as for a group of banks. The banks are grouped by the delisting code and the time when the first dividend payment of a bank IPO is announced. The significance of the null hypothesis is tested using the J_1 and J_2 procedures, as described in Campbell *et al.* (1997). They have the following form:

$$J_1 = \frac{C\overline{A}R}{\hat{\bar{\sigma}}_i} \sim N(0,1) \qquad J_2 = \left(\frac{N(L-4)}{L-2}\right)^{\frac{1}{2}} S\overline{CA}R \sim N(0,1)$$

where $SC\overline{A}R = \frac{1}{N}\sum_{i=1}^{N} SC\hat{A}R_i$.

4. EMPIRICAL RESULTS

In this section we present our findings from our empirical analysis of the impact of dividend initiations on the performance of bank IPOs in the United States. We first analyze whether the delisting codes, i.e. the future status of the bank, has a significant impact on the bank behavior with respect to dividend policy. We then investigate the long-run performance as well as the short-term valuation effects of the dividend policy of bank IPOs.

4.1. Importance of the Delisting Codes

Over the period from 1970 to 1997, the number of banks that went public each year varies greatly as is shown in Figure 7.1a. It becomes immediately evident that a higher number of banks went public in the years between 1983 and 1988 than in the two periods before and after the 1980s. Thus, we observe three different time periods that at a first glance could be due to a hot issue market for banks during the 1980s. However, this period is usually not classified as a hot issue market for IPOs. The main reason for this observation appears to be bank deregulation in the early 1980s.

A more detailed analysis indicates that saving institutions and commercial banks are the major categories of financial institutions that contribute to the dramatic increase of bank IPOs during that period (Figure 7.1b). Of the 431 banks that went public in this period, 54.3% were eventually acquired (merged), 23%

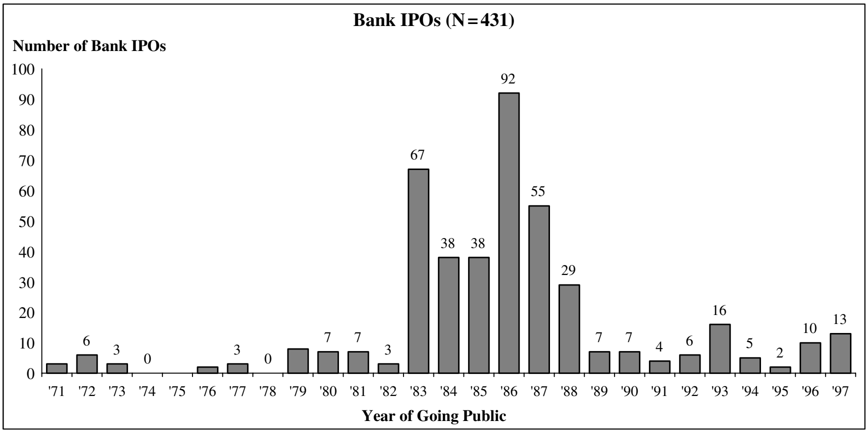


Fig. 7.1a. Number of Bank IPOs, 1971–1997

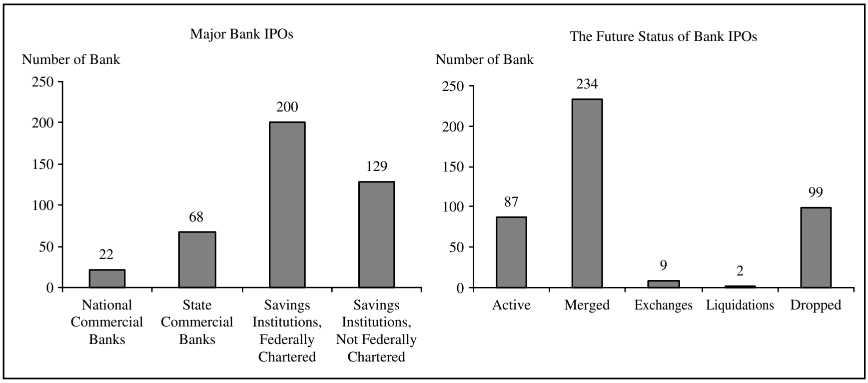


Fig. 7.1b. Bank IPO Classification and the Future Status of Bank IPOs

were delisted (dropped), and 20.2% continued to operate independently (alive). The remaining 2.5% of banks are in the group of exchanges and delistings. In the period of high bank IPO activity (1983–88) about 60% of the banks that went public eventually merged, suggesting that these IPOs could have been part of an exit strategy either of the owners, of bank management or of both. Of these banks, 25% were delisted (dropped) and only 14% were still active at the end of the period.

In order to show the importance of the dividend policy for banks, we compare the status of the bank IPOs to the status of non-bank IPOs over time. Not surprisingly, few non-banks started paying dividends in the first year after going public. This proportion rises to slightly more than 10% over the next two years and then remains at that level for the rest of the 10-year period considered in this study (Figures 7.2a and 7.2b). By the end of this time period, half (50%) of the non-bank IPOs have either merged or were dropped. In comparison, nearly 70% of the bank IPOs have merged or have been dropped in the first 10 years, as is shown in Figure 7.3a. However, the proportion of banks that pay dividends is considerably higher, reaching 30% in the first year and exceeding 40% in the second and third year. As already mentioned, by the tenth year after going public, only about 30% of the bank IPOs are still active. However, two-thirds of these banks are paying dividends as shown in Figure 7.3b. For reasons given below we use quarterly data. The analysis clearly reveals that the dividend policy of bank IPOs is different from that of non-bank IPOs. Obviously, this is in sharp contrast to the findings and conclusions of Fama and French (2001) for non-banks but underscores the notion that banks are special and need to be investigated separately (Bessler and Nohel 1996, Bessler and Nohel 2000).

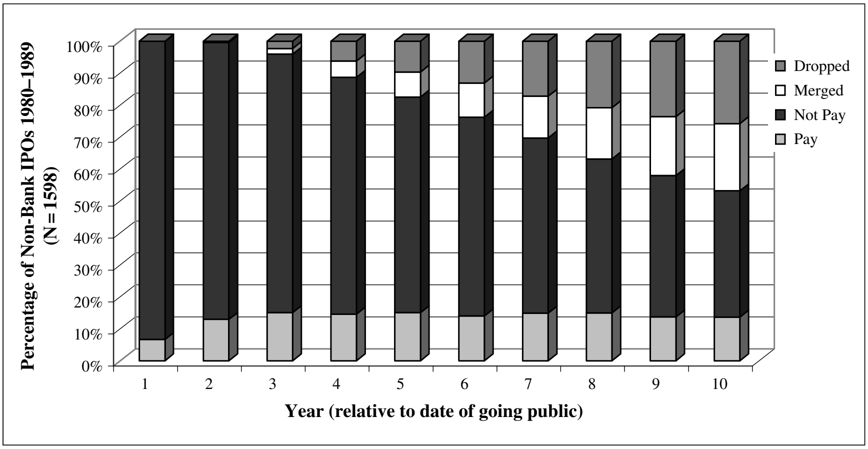


Fig. 7.2a. Non-bank IPOs (1980–1989) and Their Delisting Codes Over Time

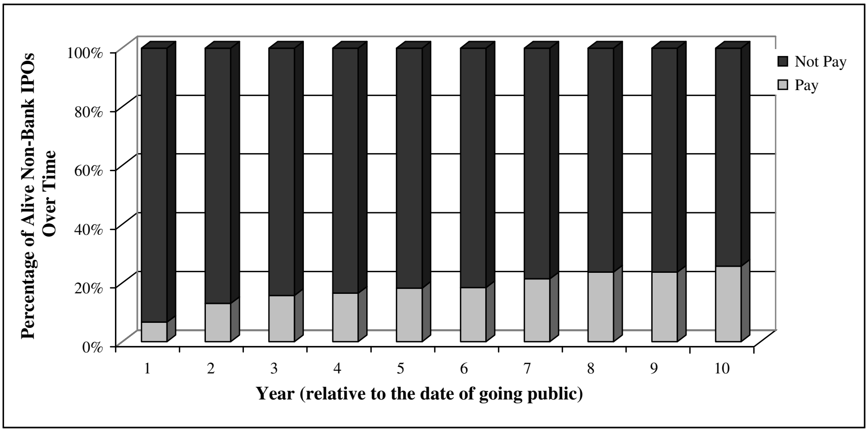


Fig. 7.2b. Non-bank IPOs (1979–1989): Proportion of Paying vs Non-Dividend Paying Firms

On the one hand, a higher proportion higher proportion of banks appear to pay dividends much earlier and, on the other hand, banks continue to pay dividends for the first 10 years after going public. Thus, it seems fair to conclude that dividends are an important mechanism for banks to signal their quality, viability, and possibly solvency to shareholders and regulators alike.

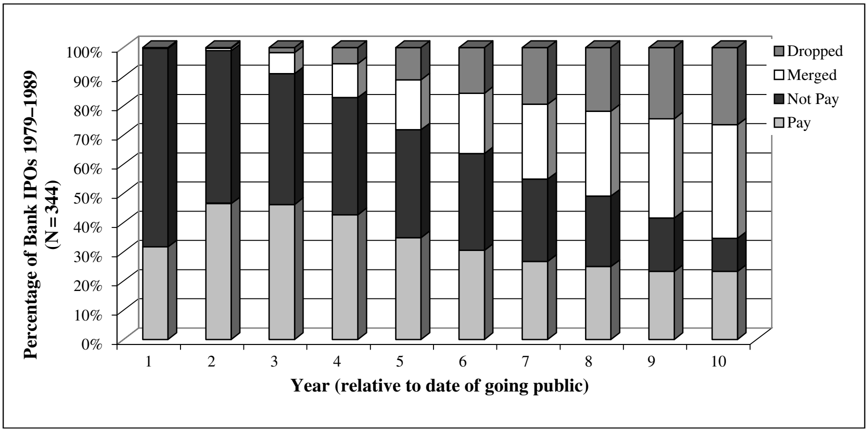


Fig. 7.3a. Bank IPOs (1979–1989) and Their Delisting Codes Over Time

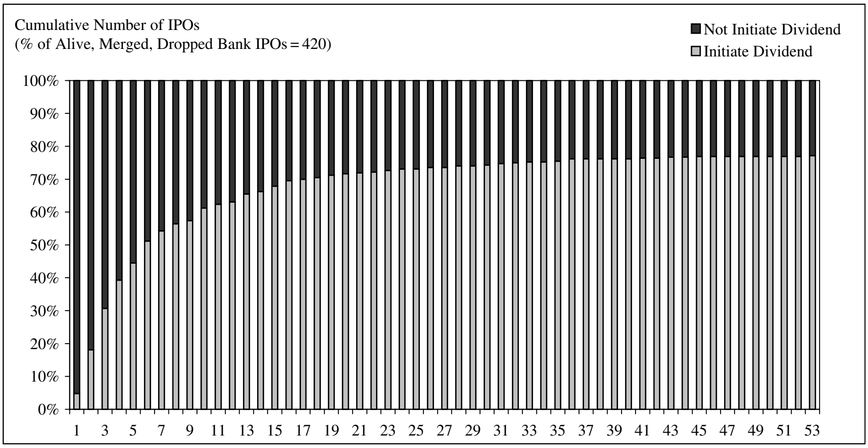


Fig. 7.3b. Bank IPOs (1979–1989): Proportion of Paying vs Non-Dividend Paying Banks (Quarterly)

The first dividend payment of a bank IPO is defined as the time when a bank pays its first regular dividend. The timing decision of bank dividend payments is measured in quarters or years relative to its date of going public. For example, a bank making its first dividend payment in quarter 1 means that the bank makes the first dividend payment within the first three months after it went public. One year is equal to twelve months relative to the going public date. The use

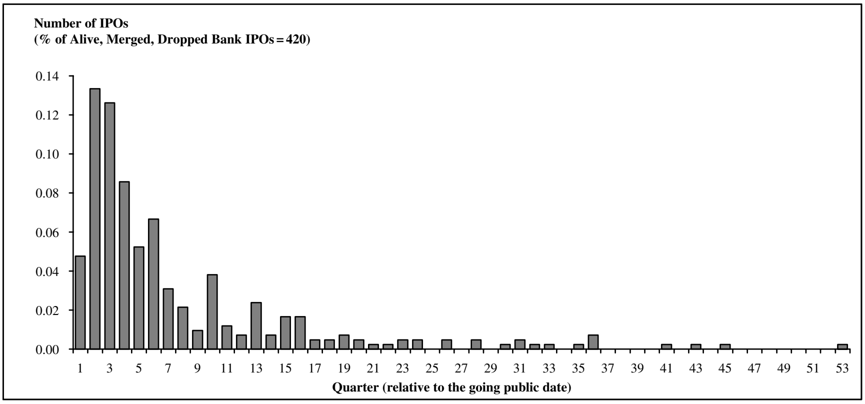


Fig. 7.4. The Timing of First Dividend Payment of Bank IPOs

of the calendar year would be inappropriate because the dividend payment of an IPO that went public at the end of a year and that started paying dividends immediately after going public, i.e. at the end of the first quarter, might otherwise be classified as a non-payer in the year of the IPO. Figure 7.4 shows the timing of the first dividend payment of bank IPOs. The graph reveals an interesting pattern of the timing decisions. Most bank IPOs paid the first dividend within the first year after going public. A smaller number of IPOs started paying in the second year, and even a smaller number began dividend payments in the third year. The rest of the banks initiated dividends in later years.

4.2. Long-run Performance

The long-run performance of bank IPOs is measured by the 36-months market-adjusted buy and hold abnormal returns (BHAR). From Figure 7.5a it is evident that the average BHAR for the entire sample of bank IPOs ($N = 420$) is positive. Returns increase modestly in the first year and more rapidly to nearly 20% in the next 18 months. The returns decline in the last 6 months bringing the 3-year performance to slightly less than 10%. This result is in contrast to the usual findings of a negative long-run performance of IPOs (Ritter 1991, Bessler and Thies 2005). This result for the entire sample may change when the IPOs are separated by certain criteria. When the full sample of bank IPOs is further categorized by the delisting code, a different pattern emerges. Banks that eventually merged have positive BHAR of about 30%, while banks that were eventually dropped just break even after three years. Banks that stayed

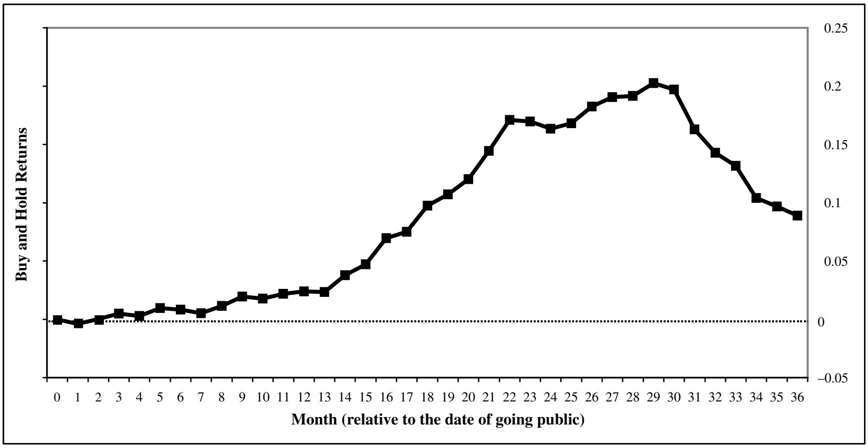


Fig. 7.5a. Long-Run Performance (BHAR) of Bank IPOs

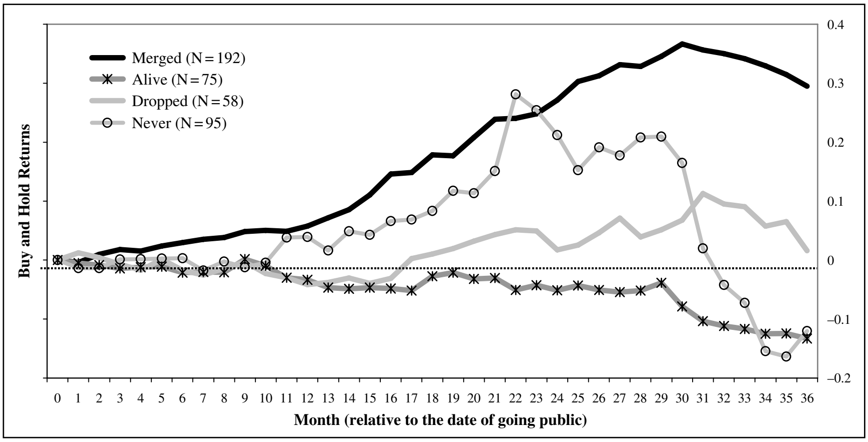


Fig. 7.5b. Long-Run Performance (BHAR) by Delisting Code

alive and either paid a dividend in the first three years or never paid a dividend have a negative long-run performance in excess of −10%. Thus, to come to any meaningful and final conclusions about the long-run performance of bank IPOs, it is important to identify those characteristics that may explain, determine and impact the performance of banks. Usually, early dividend initiations and regular dividend payments are an important indicator that may signal the quality and the future performance of a bank and in particular of a bank IPO. However, the future status of a bank separates the better performing IPOs from the weaker performing IPOs. This is a new insight that needs to be explored in greater detail.

Analyzing only those banks that paid dividends within the first three years of the IPO, a similar pattern emerges. In Table 7.2a and in Figure 7.6a, it is shown that the 36-months returns of the group of merged banks are positive while the returns of the groups of alive and dropped banks are negative. Thus, we are finding a similar pattern. To further investigate these insights, the same results are presented in Table 7.3 and in Figure 7.6b, but in a different way. This time the banks are grouped by delisting codes and the time of dividend initiation.

Table 7.2a BHAR from Day of Going Public, by Timing of First Dividend Payment

Average Buy and Hold Returns (%)			
	Alive	***Merged***	***Dropped***
1st year	0.5436	17.1998	−5.2845
2nd year	−6.1639	16.0158	−4.9720
3rd year	−14.050	10.8661	−5.6213
T-test significance from zero			
1st year	0.9179	8.7983*	−5.3111*
2nd year	−5.7918*	7.7308*	−1.3871
3rd year	−5.8702*	9.4208*	−3.1419*
Two Population T-test for First Dividend Timing			
	Merged-Alive	***Merged-Dropped***	***Alive-Dropped***
1st year	8.1543*	10.2502*	5.0333*
2nd year	9.5231*	5.0694*	−0.3188
3rd year	9.3779*	7.7452*	−2.8207*

Table 7.2b Two Population T-Test for 36 Months BHAR from Time of Going Public, by Delisting Codes

	1st *Yr* – 2nd *Yr*	**1st *Yr* – 3rd *Yr***	**2nd *Yr* – 3rd *Yr***
Alive	5.5074*	5.9188*	3.0107*
Merged	0.4156	2.7904*	2.1719**
Dropped	−0.0840	0.1645	0.1621

* Significance at the 1%-level.
** Significance at the 5%-level.

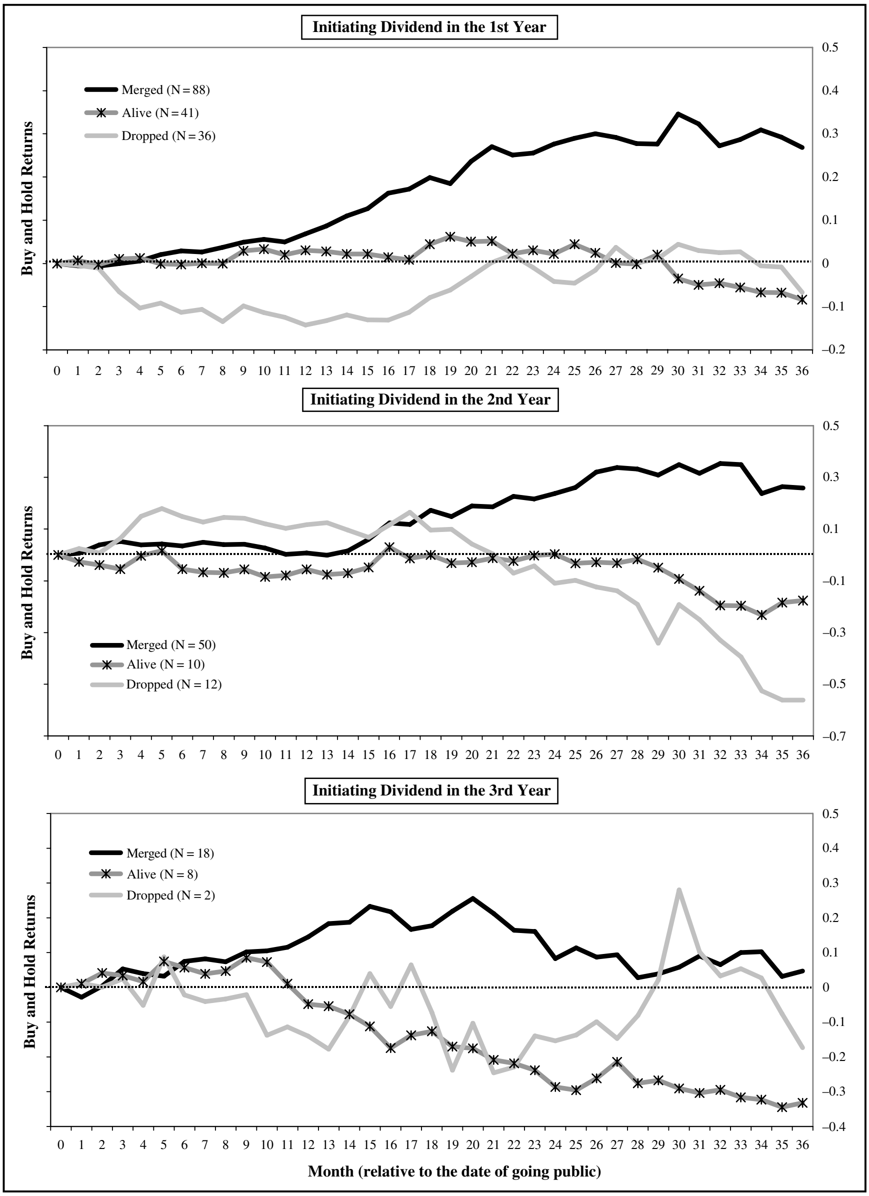


Fig. 7.6a. BHAR from Day of Going Public, by Timing of First Dividend Payment

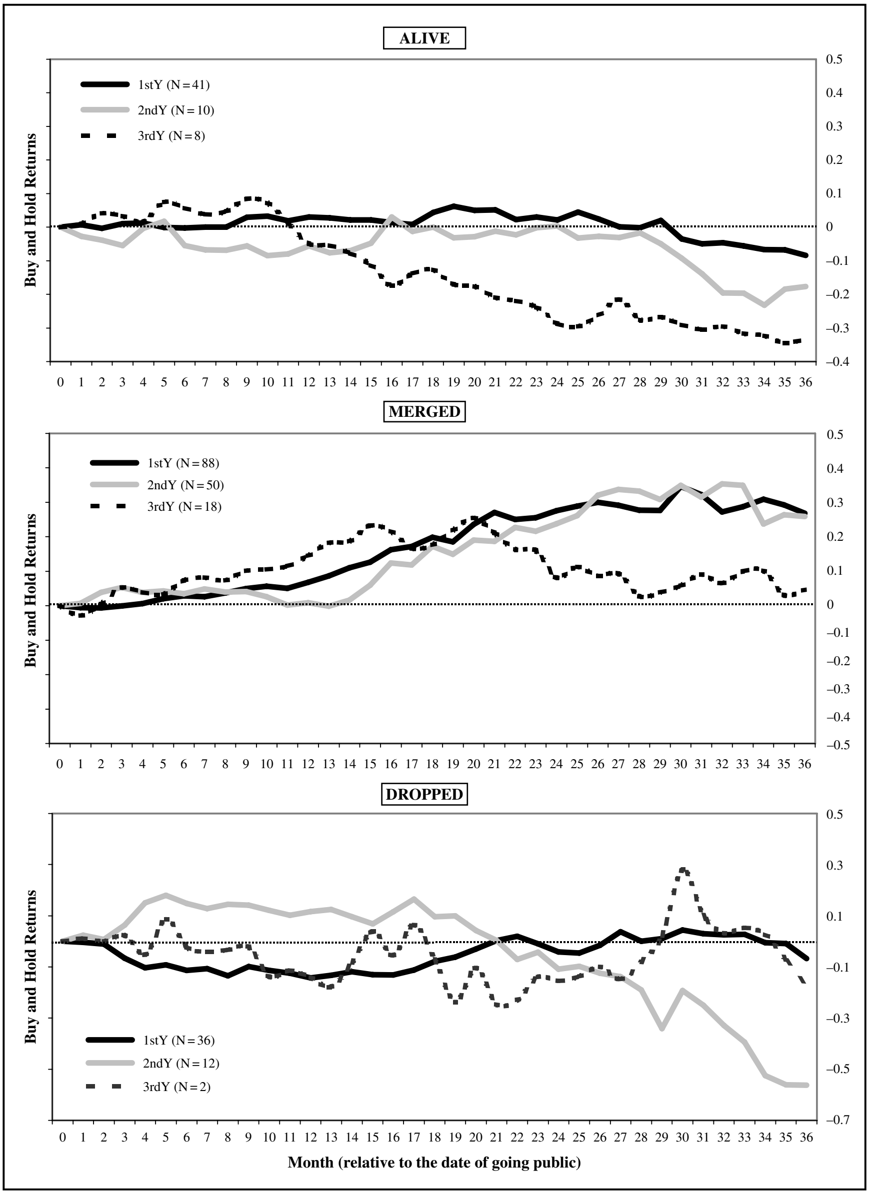


Fig. 7.6b. BHAR from Day of Going Public, by Delisting Codes

Table 7.3a BHAR Pre and Post First Dividend Payment, by Timing of First Dividend Payment

Average Buy and Hold Returns (%)			
	Alive	***Merged***	***Dropped***
1st year	0.0349	18.9322	−3.5873
2nd year	−12.1206	14.9884	−27.9249
3rd year	−29.5185	3.6621	−40.550
T-test significance from zero			
1st year	0.0673	8.5983*	−5.5046*
2nd year	−6.4419*	8.5102*	−4.8468*
3rd year	−10.6693*	2.6550*	−6.9909*
Two Population T-test for First Dividend Timing			
	Merged-Alive	***Merged-Dropped***	***Alive-Dropped***
1st year	8.3536*	9.807*	4.3482*
2nd year	10.5187*	7.1229*	2.6076*
3rd year	10.7331*	7.4155*	1.7172

Table 7.3b Two Population T-Test BHAR Pre and Post First Dividend Payment, by Delisting Codes

	1st *Yr* – 2nd Yr	**1st *Yr* – 3rd *Yr***	**2nd *Yr* – 3rd *Yr***
Alive	6.2280*	10.4989*	−10.6693*
Merged	1.3987	5.8772*	5.0630*
Dropped	−4.1974*	6.3327*	1.5448

Additional insights into the importance of the dividend policy of bank IPOs can be expected form analyzing the performance after the first dividend payment. The results are reported in Tables 7.4 and 7.5. The graphs in Figures 7.7a and 7.7b show the long-run performance of BHAR from the date of the dividend initiation. Analyzing the findings from this perspective reveals a slightly different pattern. The results from the two population t-test indicate that, in paired comparisons, the BHARs are significantly different from each other in each time period. The long-run performance of those bank IPOs that continued operations (alive) is

Table 7.4 Event Study Test for All Announcements by Delisting Codes

	CAR (%)	***J1***	***J2***	***N***
Alive	1.0961	3.0745*	4.4379*	55
Merged	1.2935	6.1964*	9.8658*	150
Dropped	2.2378	6.0593*	6.6989*	47

Table 7.5a Cumulative Average Abnormal Returns (CARs) of Dividend Initiation Announcement by Timing of First Dividend Payment

1st Year	***CAR (%)***	***J1***	***J2***	***N***
Alive	0.1406	0.3237	2.0635**	37
Merged	1.9091	7.1430*	12.7162*	84
Dropped	2.1497	4.9512*	6.0515*	33
2nd Year	***CAR (%)***	***J1***	***J2***	***N***
Alive	4.5394	4.9507*	4.2364*	10
Merged	0.1463	0.3834	−1.0980	49
Dropped	0.5373	0.7397	0.2788	12
3rd Year	***CAR (%)***	***J1***	***J2***	***N***
Alive	1.2110	1.4952	2.4622**	8
Merged	1.5585	2.3550**	2.9034*	17
Dropped	13.8944	6.2079*	7.2102*	2

Table 7.5b Cumulative Average Abnormal Returns of Dividend Initiation Announcement by Timing of first Dividend Payment and Delisting Codes

Alive	***CAR (%)***	***J1***	***J2***	***N***
1st year	0.1406	0.3237	2.0635**	37
2nd year	4.5394	4.9507*	4.2364*	10
3rd year	1.2110	1.4952	2.4622**	8
Merged	***CAR (%)***	***J1***	***J2***	***N***
1st year	1.9091	7.1430*	12.7162*	84
2nd year	0.1463	0.3834	−1.0980	49
3rd year	1.5585	2.3550**	2.9034*	17
Dropped	***CAR (%)***	***J1***	***J2***	***N***
1st year	2.1497	4.9512*	6.0515*	33
2nd year	0.5373	0.7397	0.2788	12
3rd year	13.8944	6.2079*	7.2102*	2

consistently negative. In addition, the performance is more negative the longer the bank delays to initiate a dividend payment. The results from the t-test indicate that the individual yearly returns are each significantly different from each other. A similar pattern also emerges for the banks that eventually delisted (dropped). In this case the 1–2 year and 1–3 year returns are significantly different from each other. For the group of banks that eventually merged, we find that those bank IPOs that initiated the dividend payments in the first year after going public have the strongest outperformance. If the acquisition (merger) was planned or

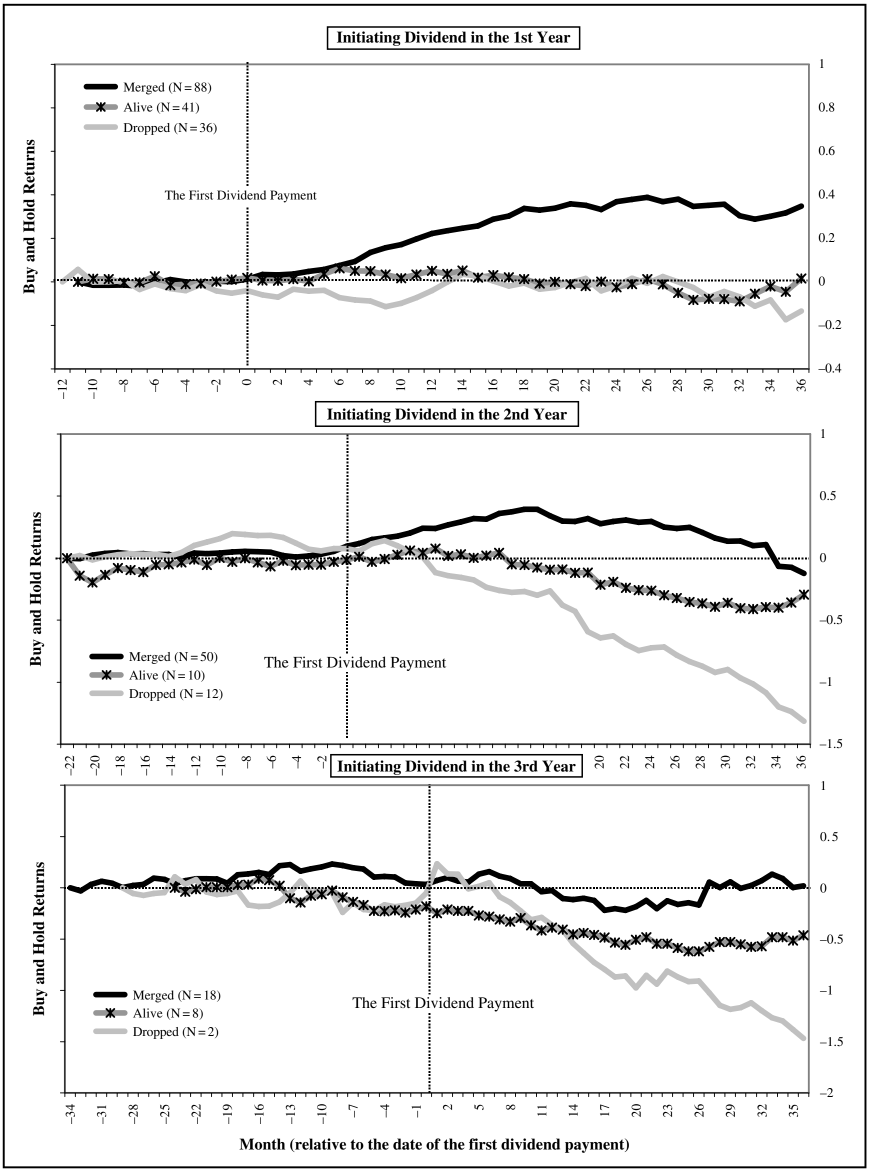


Fig. 7.7a. BHAR Pre and Post First Dividend Payment, by Timing of First Dividend Payment

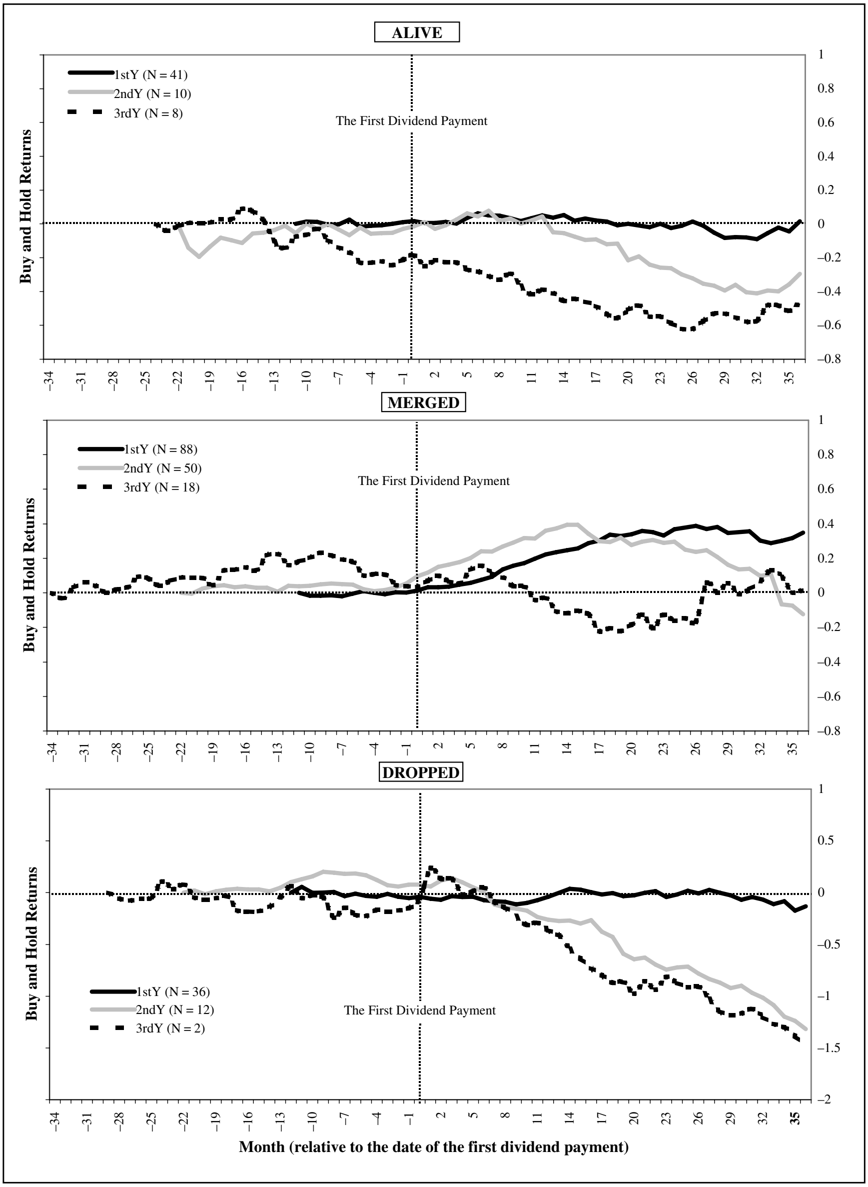


Fig. 7.7b. BHAR Pre and Post First Dividend Payment, by Delisting Codes

expected by its bank management at the time of the going public decision, then this suggests that the early dividend initiation was a means to signal quality and to help increase the market value of the bank. However, for this group, the 1–3 year and 2–3 year returns are different from each other.

4.3. Short-term Valuation Effects

In addition to analyzing the long-run performance, we expect additional insights into the relevance of the dividend policy of banks in general and dividend initiations of bank IPOs in particular, from analyzing the stock returns around the announcement date in greater detail. In Figures 7.8a and 7.8b we report the average abnormal returns (AR) and average cumulative abnormal returns (CAR) in the 21-day event window surrounding the dividend initiation announcement. The results and test statistics are reported in Table 7.4. On average, the stock market reaction to the dividend initiation announcement was positive, regardless of the eventual delisting code. Interestingly, the banks that eventually were delisted (dropped) have the greatest positive abnormal returns of 2.24%. This finding appears to be surprising, but it is possible and sensible, however, that the market initially interpreted this dividend initiation as a more positive signal for those banks that were considered as weaker or problem financial institutions at that time. Again, this result supports the notion that signals are only reliable if they are costly and cannot be duplicated by weaker institutions. In fact, such a dividend initiation by weaker banks must have been a costly cash outflow for

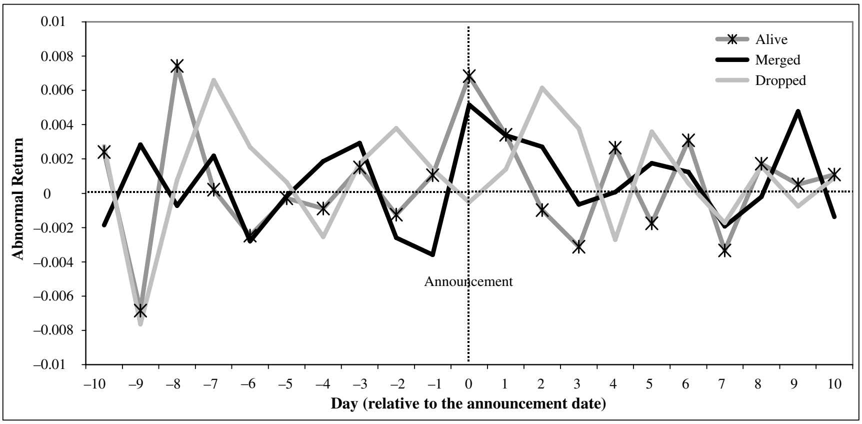


Fig. 7.8a. Average Abnormal Returns of Dividend Initiation Announcements

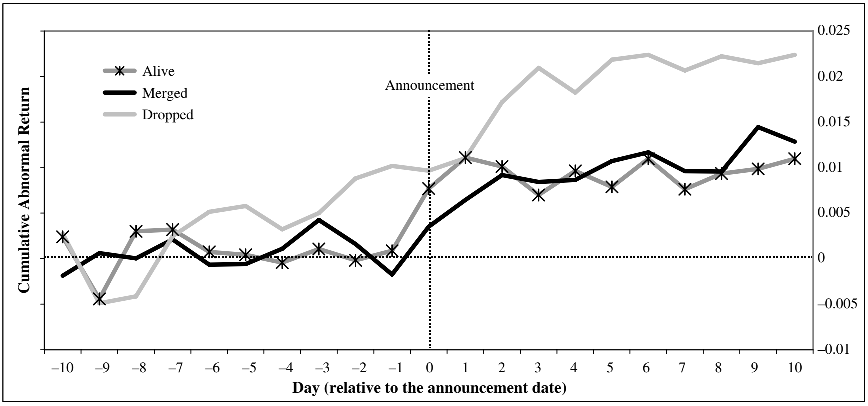


Fig. 7.8b. Cumulative Abnormal Returns (CARs) of Dividend Initiation Announcements

the bank. Thus, it is not surprising that this initial positive valuation effect does translate into a negative long-run performance.

We finally analyze the dividend initiation announcement by delisting codes. The results and test statistics are reported in Tables 7.5a and 7.5b. In Figures 7.9a–7.9d the average abnormal returns and average cumulative abnormal returns of the dividend initiation announcements are presented. The graphs

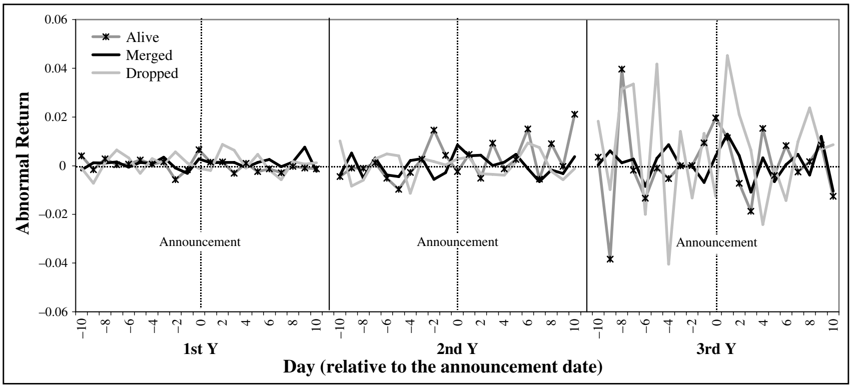


Fig. 7.9a. Abnormal Returns of Dividend Initiation Announcement by Timing of First Dividend Payment

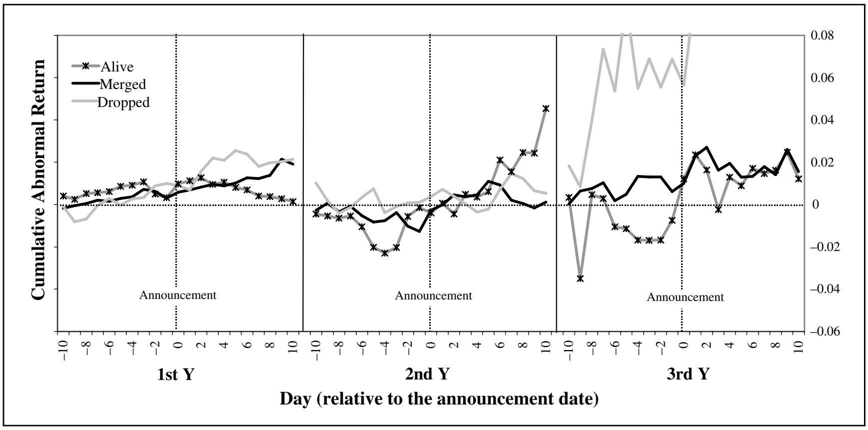


Fig. 7.9b. Cumulative Abnormal Returns (CARs) of Dividend Initiation Announcement by Timing of First Dividend Payment

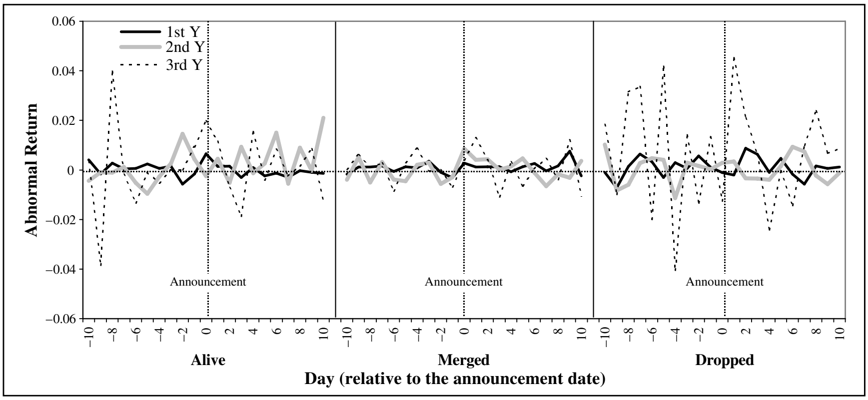


Fig. 7.9c. Abnormal Returns of Dividend Initiation Announcements by Timing of First Dividend Payment and Delisting Codes

are separated by delisting codes and the timing of the dividend payment. All CARs are positive and most are statistically significant. Only the 2nd year merged and 2nd year dropped CARs are not significant although they are positive. The strongest positive returns are found in the group of the 3rd year dropped banks (+13.89%, N = 2) and the group of the 2nd year alive banks (+4.54%, N = 10). However, there does not appear to be a consistent pattern

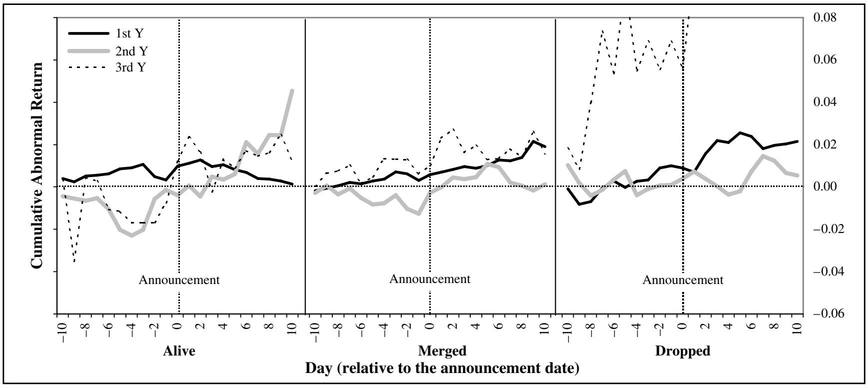


Fig. 7.9d. Cumulative Abnormal Returns of Dividend Initiation Announcements by Timing of First Dividend Payment and Delisting Codes

in the returns between delisting groups. For the group of banks that continued operations (alive), an initiation of a dividend payment in the 2nd year leads to the most positive stock market reaction. The group of merged banks shows the highest returns when dividend payments are initiated in the 1st year. The opposite is true for the group of dropped banks, where the highest returns are seen in the two banks that initiated dividends in the 3rd year.

5. CONCLUSIONS

Dividend policy and the role that dividend announcements play in communicating manager's private information to shareholders have attracted a considerable amount of research since the seminal paper of Miller and Modigliani (1961). So far there is significant empirical evidence which suggest that management can use dividend changes to signal the quality of the firm in that dividend increases result in positive stock price reactions and dividend decreases lead to negative stock price reactions in the short- as well as in the long-run. The important question to investigate, however, is whether all firms can employ dividend changes in the same manner or whether the magnitude of the valuation effects depends on the maturity of the firm (e.g. IPO) as well as on the industry (e.g. banking) in which the firm is operating. Most of the empirical research has focused on established firms instead of IPOs as well as on industrial firms instead of banks. The objective of this study was to investigate the short-term valuation effects as well as the long-run performance of IPOs for banks in the United

States over the period from 1970 to 1997. The empirical results suggest that, on average, bank IPOs outperform the market over the first 36 months after going public. This result is in contrast to the immense empirical evidence for industrial firms (non-banks) where most studies find negative long-run valuation effects, suggesting again that banks are special and different from industrial firms.

However, by separating the sample of bank IPOs with respect to the future status of the bank, the results change. First of all, only those banks that were acquired later on outperformed the benchmark. Second, the group of banks that continued to operate independently over an extended period of time underperformed the benchmark. An underperformance was also found for the group of bank IPOs that eventually failed. Moreover, the initiation of a dividend payment is an important characteristic that separates the outperformers from the underperformers. Thus, in an environment that is characterized by information asymmetries, dividend initiations are an important signaling mechanism for banks that just went public to convey reliably positive information about their quality and solvency to the market and regulators alike.

Overall, the empirical results provide significant evidence that the dividend policy of banks is quite different from that of non-banks and that the dividend policy of bank IPOs has a significant impact on the long-run performance of banks.

BIBLIOGRAPHY

Allen, F. and R. Michaely (1995) “Dividend policy,” in: Jarrow, R. (ed.), *Handbooks in Operations Research & Management Science* **9**: 793–837.

Ambarish, R., K. John and J. Williams (1987) “Efficient signaling with dividends and investments.” *Journal of Finance* **42**: 321–44.

Asquith, P. and D. Mullins, Jr. (1983) “The impact of initiating dividend payments on shareholder’s wealth.” *Journal of Business* **56**: 77–96.

Battacharya, S. (1979) “Imperfect information, dividend policy and the ‘bird in the hand’ fallacy.” *Bell Journal of Economics and Management Science* **10**: 259–70.

Bessler, W. (2001) *Dividendenpolitik von Wachstumsunternehmen,* in: *Achleitner and Bassen* (eds), Investor Relations am Neuen Markt, Schäffer-Poeschel. pp. 291–317.

Bessler, W. and H. Ellermann (2004) “Theoretische Ansätze zur Dividendenpolitik, in: Achleitner and Thoma” (eds) *Handbuch Corporate Finance, Cologne*, Chapter 2.1.6. pp. 1–26.

Bessler, W. and A. Kurth (2005) “Exit strategies of venture capitalists in hot issue markets: evidence from Germany.” *Journal of Entrepreneurial Finance and Business Ventures* **10**: pp. 17–51.

Bessler, W. and A. Kurth (2006) "Agency problems and the performance of venture-backed IPOs in Germany: exit strategies, lock-up periods, and bank ownership." *European Journal of Finance*, Forthcoming.

Bessler, W. and T. Nohel (1996) "The stock-market reaction to dividend cuts and omissions by commercial banks." *Journal of Banking and Finance* **20**: 1485–508.

Bessler, W. and T. Nohel (2000) "Asymmetric information, dividend reductions, and contagion effect in bank stock returns." *Journal of Banking and Finance* **24**: 1831–48.

Bessler, W. and S. Thies (2005) "The long-run performance of initial public offerings in Germany." *Managerial Finance*, forthcoming.

Campbell, J.Y., A.W. Lo and A.C. MacKinlay(1997) *The Econometrics of Financial Markets*. Princeton, New Jersey: Princeton University Press. pp. 149–80.

DeAngelo, H., L. DeAngelo and D. Skinner (1996) "Reversal of fortune: dividend signaling and the disappearance of sustained earnings growth." *Journal of Financial Economics* **40**: 341–71.

Dyl, E. and R. Weigand (1998) "The information content of dividend initiations: additional evidences." *Financial Management* **27**: 27–35.

Ellermann, H. (2003) *Dividendenpolitik und Long-run Performance in Deutschland*. Wiesbaden: Deutscher Universitäts Verlag.

Fama, E. and H. Babiak (1968) "Dividend policy: an empirical analysis." *Journal of American Statistical Association* **63**: 113–61.

Fama, E.F. and K.R. French (2001) "Disappearing dividends: changing firm characteristics or lower propensity to pay?" *Journal of Financial Economics* **60**:.

Healy, P. and K. Palepu (1988) "Earning information conveyed by dividends initiations and omissions." *Journal of Financial Economics* **21**: 14976.

Houge, T. and T. Loughran (1999) "Growth fixation and the performance of bank initial public offerings, 1983–1991." *Journal of Banking and Finance* **23**: 1277–301.

Jensen, M.C. (1986) "Agency costs of free cash flow, corporate finance, and takeover." *American Economic Review* **76**: 323–9.

John, K. and L. Lang (1991) "Insider trading around dividend announcements: theory and evidence." *Journal of Finance* **46**: 1361–89.

John, K. and J. Williams (1985) "Dividends, dilution and taxes: a signaling equilibrium." *Journal of Finance*, **40**: 1053–70.

Lang, L. and R. Litzenberger (1989) "Dividend announcements: cash flow signaling vs. free cash flow hypothesis." *Journal of Financial Economics* **24**: 181–91.

Lintner, J. (1956) "The distribution of incomes of corporating among dividend, retained earnings and taxes." *American Economic Review* **46**: 7–113.

Lipson, M., C. Maquieira and W. Megginson (1998) "Dividend initiations and earnings surprises." *Financial Management* **27**: 36–45.

Michaely, R., R.H. Thaler and K.L. Womack (1995) "Price reactions to dividend initiations and omissions: overreaction or drift?" *Journal of Finance* **50**: 573–608.

Miller, M.H. and F. Modigliani (1961) "Dividend policy, growth and the valuation of shares." *Journal of Business* **34**: 411–33.

Miller, M.H. and K. Rock (1985) "Dividend policy under asymmetric information." *Journal of Finance* **40**: 1031–51.

Ritter, J.R. (1991) "The long-run perfomance of initial public offerings." *Journal of Finance* **46**: 3–27.

Slovin, M.B., M.E. Sushka and J.A. Polonchek (1999) "An analysis of contagion and competitive effects at commercial banks." *Journal of Financial Economics* **54**: 197–225.

Venkatesh, P.C. (1989) "The impact of dividend initiation on the information content of earnings announcements and returns volatility." *Journal of Business* **62**: 175–97.

Williams, J.T. (1992) "Signaling with dividends," in: Newman, P., Murray, M., Eatwell, J. (eds), *The New Palgrave Dictionary of Money and Finance*, Vol. I. London, pp. 458–61.

This Page is Intentionally Left Blank

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 8

FINANCIAL SECTOR DEVELOPMENT AND GROWTH: RE-EXAMINING THE NEXUS

George Mavrotas* and Sang-Ik Son

Abstract

The chapter tries to delve deeper into the relationship between financial sector development, broadly defined to go beyond financial deepening, and economic growth by using a new database including 65 countries (both industrial and developing ones) over the period 1960–99 and by also exploring new routes regarding the measurement of financial sector development. Empirical results obtained from the estimation of dynamic panel data models using various GMM estimators seem to suggest that financial sector development contributes to economic growth although the magnitude of the impact varies depending *inter alia* on the level of development (industrial vis-à-vis developing countries).

Key words: Financial sector development, growth, industrial countries, developing countries, dynamic panel data models, GMM estimators.

JEL Classification No.: E44, O16.

1. INTRODUCTION

The overall nexus between finance and economic growth has been the subject of a rather voluminous literature, both theoretical and empirical, which goes back to the seminal contribution by Goldsmith (1969) as well as the money-growth literature of the 1960s, in particular, Gurley and Shaw (1960), Tobin (1965) and

*Corresponding author

Patrick (1966). A further interest in the above relationship appeared in the early 1970s following the influential works of McKinnon (1973) and Shaw (1973) and the associated with financial repression literature.[1]

The 1990s have witnessed a revived interest in the above relationship, although the focus of much of the recent literature on the subject has been on the interaction between financial sector development (broadly defined to go beyond financial liberalization) and economic growth. The turning point was the study by King and Levine (1993a), which relaunched the finance and growth literature, by placing a new emphasis on financial depth as an important determinant of the overall growth process. The study by Wachtel and Rousseau (1995) reported important evidence on the above relationship from long time series for several countries. Since then a series of studies were published on the finance-growth nexus (Wachtel 2004, provides an excellent discussion). Yet a number of issues remain unresolved and call for further research.

On the empirical front, much of the empirical literature has used cross-section analysis to examine the macroeconomic association between the development of the financial sector of the economy and the long-term growth rate. The cross-section approach has been used by many studies, including Wallich (1969), Fry (1980), Khatkhate (1988), Barro (1991), Roubini and Sala-i-Martin (1992a), Atje and Jovanovic (1993), Quah (1993), King and Levine (1992, 1993a) and Pill (1997) among others. However, despite its popularity, the cross-section approach adopted in much of the above literature has certain limitations and shortcomings. In terms of measurement problems, country officials sometimes define, collect and measure variables inconsistently across countries (Levine and Zervos, 1996). In addition, the above approach regresses the average data of sampled countries over a certain period, and thus, can only reveal the "average effect" of a variable across countries. However, it is reasonable to expect the effects to be rather different across countries. Furthermore, it is not only likely that the long-run causality may vary across countries but it is also possible that the long-run relationships themselves will exhibit substantial variation (Arestis and Demetriades, 1997).

Another issue of crucial importance, as well as of relevance to the above relationship, is the measurement of financial sector development. Since there is no concrete definition of financial development, measuring financial

---

[1] It is clearly beyond the scope of the present chapter to review the vast literature on finance and growth. The interested reader should refer to Fry (1988), Hermes and Lensink (1996), Arestis and Demetriades (1997), Levine (1997), World Bank Policy Report on *Finance for Growth: Policy Choices in a Volatile World* (2001), Demirguc-Kunt and Levine (2001), Mavrotas and Kelly (2001), Green and Kirkpatrick (2002) and Wachtel (2004) among others.

sector development is not an easy procedure. As argued rightly by Bandiera *et al*. (2000), an ideal index of financial sector development should attempt to measure both the various aspects of the deregulatory and the institution-building process in financial sector development. However, measuring the above aspects is a difficult if not an impossible task. Various measures of financial sector development have been used in empirical work in the recent past. Common measures of financial development used in the literature have been financial depth or selected financial indicators. Financial depth in particular has been used extensively in much of the early as well as recent literature as a measure of financial sector development. However, it could be well argued that when we consider the likely channels through which a more developed financial system helps promote growth, it becomes evident that, though useful and readily available, banking depth (usually measured as M2/GDP or M3/GDP) is unlikely to be a wholly reliable indicator of financial sector development (see Honohan, 2004, for an excellent recent discussion on this issue).

In view of the above discussion, the main purpose of this chapter is to try to contribute to the empirical literature on the relationship between financial sector development and economic growth by adopting a different approach compared to most of the previous studies on the subject. Our contribution has three main elements: Firstly, we use in our empirical analysis a new panel data set composed of 2,535 observations from the adjusted data for 65 countries (both developed and developing ones) over the 1960–99 period, which, to the best of our knowledge, is a larger data set than most of the previous studies on the topic. Our database, which is constructed on the basis of the World Bank database (described in detail in Beck *et al*., 1999), is a fresh attempt to extend and develop the database on financial development and structure. Second, the chapter is a clear departure from much of the empirical literature on the subject, since it employs different measures of financial sector development instead of using only the standard (but at the same time problematic) financial depth indicator. More precisely, we constructed a financial sector development index, using the method of principal components, which was subsequently used in the econometric analysis. Finally, in the econometric analysis we employed relatively recently developed econometric estimators related to the estimation of dynamic panel data models, such as the GMM two-step estimator and the GMM system-estimator. Furthermore, we conducted sensitivity analysis to test the robustness of the empirical results obtained.

The rest of the chapter is organized as follows: Section 2 discusses issues related to the measurement of financial sector development, of crucial importance in the present chapter, as well as data issues before we proceed with econometric methodology issues and the estimation of the econometric models in Section 3. Section 4 concludes the chapter.

2. DATA ISSUES

2.1. Measuring Financial Sector Development: Exploring New Routes

As already mentioned, a number of studies have studied the relationship between financial development and economic growth. Nevertheless, the term "financial development" has not yet received a concrete definition. This is mainly due to the fact that the financial structure is not only complicated in an economy, but also has evolved differently in the development process of different countries. Goldsmith (1969) pointed out that "financial development is a change in financial structure; hence, the study of financial development essentially requires information on changes in the financial structure over shorter or longer periods of time. Financial development can be studied either by information on the flows of financial transactions over continuous periods of time or by the comparison of financial structure at different points of time."

More recently Beck, Demirguc-Kunt, and Levine (1999) presented a comprehensive assessment of the development, structure and performance of the financial sector, and introduced the sources of statistics on the size, activity and efficiency of various financial intermediaries and markets across a broad spectrum of countries and through time. This chapter employs some measures of financial sector development suggested by Beck, but in the context of our new dataset.

To capture the measure of size of financial intermediaries we use, in line with Beck, the ratio of deposit money bank domestic assets to deposit money bank domestic assets plus central bank domestic assets (hereafter, Commercial-Central Bank, or CMB). This indicator measures the relative importance of deposit money banks relative to central banks. This indicator is persuasive in as much as central banks lose relative importance as we move from low- to high-income countries, and the other financial intermediaries gain relative importance. Thus, a measure of the relative size of financial intermediaries is a useful indicator of development.

As another measure of the size of financial intermediaries Beck *et al.* (1999) proposed the ratio of liquid liabilities to gross domestic product (GDP). In the present paper, Liquid Liabilities (LQ) equals currency plus demand and interest-bearing liabilities of banks and other financial intermediaries divided by GDP. LQ has been a typical measure of financial depth, which is the broadest available indicator of financial intermediation, including all financial sectors of central bank assets, deposit money banks assets, and other financial institutions assets.

In order to measure the activity of financial intermediaries, following Beck *et al.* (1999), we employ the ratio of private credit by deposit money banks and other financial institutions to GDP (hereafter, Private Credit or PCR). This indicator isolates credit issued to the private sector as opposed to credit issued to governments and public enterprises; thus it measures the mobilized savings that are channeled to private firms.

These financial variables can capture different aspects of the financial sector development process as compared to a simple financial depth indicator. Therefore they are more appropriate to study the finance-growth relationship. However, we still need an eclectic indicator to capture in a comprehensive way all kinds of changes in financial sector in terms of activity, structure and size, rather than separate variables dealing with single aspects, respectively. In view of this, in this chapter we constructed, by using principal component analysis, a Financial Sector Development Index (FSDI), which is the linear combination of the financial indicators PCR, CMB and LQ:

$$Z1_{it} = a_{1i\cdot}PCR_{it} + a_{2i\cdot}CMB_{it} + a_{3i\cdot}LQ_{it} = FSDI_{it} \tag{1}$$

where $Z1_{it}$ is the first principal component and coefficient vector $(a_{1i\cdot},\ a_{2i\cdot},\ a_{3i\cdot})$ calculated from the time-series data for each country. Hence, FSDI is our main financial sector development indicator to encompass the three financial indicators previously discussed.[2]

2.2. Financial and Other Variables

All raw data for the variables used in the empirical analysis have been obtained from the electronic version 2001 of the IMF's *International Financial Statistics* and the electronic version 2001 of World Bank's *World Development Indicators*, except Ethiopia's GDP data, which was obtained from UN's *Yearbook of National Accounts*. The raw data set covers 65 countries over the period 1960–99 (40 years), but the time span of data employed after adjustment is 1961–99 (39 years) for 65 countries.[3] The raw data can be distinguished into two main groups: stock variables and flow variables. Whereas stock variables are measured at the end of a period, flow variables are defined relative to a period. This presents problems in measuring both in terms of correct timing and in terms of deflating correctly. To address the above problems a data adjustment process is required.

[2] The method of principal components involves transforming the sub-variables into a new set of variables which will be pairwise uncorrelated and of which the first will have the maximum possible variance, the second the maximum possible variance among those uncorrelated with the first, and so forth. This approach has also been used by Demetriades and Luintel (1996), Bandiera *et al*. (2000) and Kelly and Mavrotas (2003) although not in the context of panel data analysis.

[3] See Appendix for a list of countries included in the sample.

Regarding data adjustment, we used the method proposed by Beck *et al.* (1999) and Beck Levine and Loayza (1999). More precisely, we deflated the end-of-year financial balance sheet items (f) by the end-of-year consumer price indices (CPI) and also deflated the GDP series by the annual CPI. Then, we computed the average of the real financial balance sheet item in year t and $t-1$ and divided the average by real GDP measured in year t. Accordingly, Private Credit (PCR) is calculated using IFS data and the following formula:

$$PCR_{it} = \{(0.5)^*[f_{it}/CPI(e)_{it} + f_{i,t-1}/CPI(e)_{i,t-1}]\}/[GDP_{it}/CPI(a)_{it}] \tag{2}$$

where, f stands for credit by deposit money banks and other financial institutions to the private sector (IFS lines 22d + 42d), GDP is from IFS (line 99b), $CPI(e)$ is end-of-period CPI (IFS line 64) and $CPI(a)$ is the average annual CPI. The f and end-of-period CPI are either the value for December or, where not available, the value for the last quarter. In case the end-of-period CPI in 1960 and 1961 is not available, the average annual CPI is used. In addition, some data on CPI were estimated using the average annual increase rate of the following 3 years,[4] where CPI data in the early 1960s are missing or not available. It is useful to note that the data from 1999 in Euro-zone countries are reported in Euro currency, so the data were converted to the equivalent values in national currency.

CMB, which is the ratio of commercial bank domestic assets divided by commercial bank plus central bank domestic assets, is calculated using IFS data and the following formula:

$$CMB_{it} = DB_{it}/[DB_{it} + CB_{it}] \tag{3}$$

where *DB* is assets of deposit money banks (IFS lines 22a–d) and *CB* is central bank assets (IFS lines 12a–d).

The data on *LQ* is obtained from "liquid liabilities (M3) as percent of GDP" in the World Development Indicators 2001 of the World Bank. If the data from the World Bank were not fully available for the period of 1961–99, we used money and quasi-money (M2), which is calculated using IFS data and the following formula:

$$LQ_{it} = \{(0.5)^*[m_{it}/CPI(e)_{it} + m_{i,t-1}/CPI(e)_{i,t-1}]\}/[GDP_{it}/CPI(a)_{it}] \tag{4}$$

where m is money (IFS line 34) plus quasi-money (IFS line 35), GDP (IFS line 99b), $CPI(e)$ is end-of-period CPI (IFS line 64), and $CPI(a)$ is the average annual CPI.

---

[4] The employed method of estimation is $\mathrm{CPI(t)} = \mathrm{CPI(t+1)}/[\mathrm{CPI(t+4)}/\mathrm{CPI(t+1)}]^{1/3}$.

As already discussed, the *FSDI* is calculated as the linear combination of the financial indicators *PCR, CMB* and *LQ* by using principal component analysis. Under the assumption of heterogeneity across countries, we estimated coefficients of the principal components for each country in our sample.[5]

2.2.1. The Set of Conditioning Variables

To explore the link between financial sector development and the growth variables, we also use a set of conditioning variables containing the other explanatory variables in the growth model. Under the open economy assumption, the conditioning information set includes the basic input variables, control and policy variables as well as open economy variables.

The basic input variable is related to scale effects, i.e. that an expansion of the aggregate labour force, L, raises the per capita growth rate for the economy in the endogenous growth model. In particular, under the assumptions of learning-by-doing and knowledge spillovers, the per capita growth rate would increase over time as the labour force grows over time. We consider a simple neoclassical production function with labour-augmenting technology for firm i; $Y_i = F(K_i, A_iL_i)$, where A_i is the index of knowledge available to the firm. Under the assumptions of learning-by-doing and knowledge spillovers,[6] the change in each firm's technology term, A_i, corresponds to the economy's overall learning and is proportional to the change in the aggregate capital stock, K. Thus, we can replace A_i by K in the above equation, and if the production function takes the Cobb-Douglas form, then output for firm i is given by:

$$Y_i = A \cdot (K_i)^{\alpha} \cdot (KL_i)^{1-\alpha}. \tag{5}$$

The private marginal product of capital can be obtained by differentiating with respect to K_i, and assuming $k_i = k$;

$$\partial Y_i / \partial K_i = A\alpha L^{1-\alpha}. \tag{6}$$

[5] Thus, for instance, the financial sector development index of Australia is calculated as:

$$FSDI_{AUS\ t} = 0.5297PCR_{AUS\ t} + 0.5020CMB_{AUS\ t} + 0.5927LQ_{AUS\ t}$$

where the coefficient vector of the first principal component is calculated from the time-series data of Australia.

[6] First, learning-by-doing works through each firm's investment. Specifically, an increase in a firm's capital stock leads to a parallel increase in its stock of knowledge. Second, each firm's knowledge is a public good that any other firm can access at zero cost. In other words, once discovered, a piece of knowledge spills over instantly across the whole economy. This assumption allows us to replace A_i by K.

A firm's profit can be written as:

$$L_i \cdot [f(k_i, K) - (r+\delta) \cdot k_i - w] \tag{7}$$

where $f(\cdot)$ is the intensive form of the production function (5), δ is the depreciation rate, $r+\delta$ is the rental price of capital and w is wage rate. Profit maximization and zero-profit condition imply:

$$\begin{aligned} \partial y_i/\partial k_i = f(k, K) = r + \delta, \text{ or } r = \partial Y_i/\partial K_i - \delta \\ \partial Y_i/\partial L_i = f(k, K) - k \cdot f(k, K) = w. \end{aligned} \tag{8}$$

Substituting (8) into the condition for optimization, $\gamma_c = (1/\theta) \cdot (r - \rho)$, then from the Keynes-Ramsey rule:

$$\gamma_c = (1/\theta) \cdot (A\alpha L^{1-\alpha} - \delta - \rho), \tag{9}$$

where γ_c equals growth rate, θ is the elasticity of marginal utility and ρ is the rate of time preference. Therefore, this result reflects the positive effect of L on the private marginal product of capital by satisfying the condition of $f'(L) > 0$, and an expansion of labour force raises the per capita growth rate (Barro and Sala-i-Martin, 1995). Data on the variable representing Scale Effects (SE) are obtained from "Labour force, total" in the World Development Indicators 2001.

2.2.2. Control and Open Economy Variables

The control variables employed in the empirical analysis are the two policy variables, i.e. the inflation rate (INFL) and the ratio of government expenditure to GDP (GEXP) as indicators of macroeconomic stability in the growth equation (although the latter could also be viewed as a measure of private sector activity). The data source for both variables is the World Development Indicators. Under the assumption of an open economy, our conditioning information set includes two open economy variables: Openness to Trade (OTR) and Foreign Direct Investment (FDI). The variable *OTR* is the sum of exports and imports as share of GDP. Data on Trade Openness are obtained from IFS (IFS lines 90c + 98c).

The theoretical foundation, regarding the effects of FDI on growth, derives from either neo-classical models or endogenous growth models. In neoclassical models of growth, FDI increases the volume of investment and its efficiency, and leads to long-term level effects and medium-term, transitional increases in growth. Endogenous growth models, on the other hand, consider long-run growth as a function of technological progress, and provide a framework in which FDI can permanently increase the rate of growth in the host economy through technology transfer, diffusion, and spillover effects. The data on *FDI* are obtained from "Foreign direct investment, net inflows (% of GDP)" in the World Development Indicators 2001 of the World Bank.

3. ECONOMETRIC METHODOLOGY AND RESULTS

3.1. Empirical Model

Static panel data models analyze the impact of financial development on growth at a certain time period. However, it might seem more persuasive to argue that financial development affects economic growth over a number of periods, and growth responds to financial development with a time lag. This is quite plausible in our analysis with regard to financial and monetary variables. We consider the following simple distributed-lag model:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \in_t \tag{10}$$

where Y denotes economic growth and X represents a set of financial variables and the other independent variables. For the sake of simplicity we can introduce the Koyck approach that the $\beta' s$ are all of the same sign and decline geometrically as $\beta_i = \beta_0 \delta^i (i = 0, 1, 2 \ldots)$, where $0 < \delta < 1$. The result implies that current and recent financial developments are expected to affect the current growth rate more heavily than ones in the distant past. In line with the Koyck transformation, we take:

$$Y_t = \alpha(1-\delta) + \delta Y_{t-1} + \beta_0 X_t + v_t \tag{11}$$

where $v_t = u_t - \delta u_{t-1}$. For our analysis with panel data we rewrite equation (11) to specify an autoregressive panel data model as follows:

$$\begin{aligned} y_{it} &= y_{i,t-1}\delta + X_{it}\beta'_{ik} + \in_{it}, \\ \in_{it} &= \mu_i + v_{it}, \\ & k = 2, \ldots, K; i = 1, \ldots, N; t = 1, \ldots, T \end{aligned} \tag{12}$$

where y represents per capita real GDP growth, X is a $(K-1) \times 1$ row vector of the "independent" variables, which includes *FSDI, SE, GEXP, INFL, OTR, FDI*, δ is a scalar, β is the $(K-1) \times 1$ column vector of the slope parameters, μ_i is an unobserved country-specific time-invariant effect which allows for heterogeneity, and v_{it} is the disturbance term.

Economic relationships are dynamic in nature, and one of the advantages of dynamic panel data models is that they allow the researcher to delve deeper into the dynamics of adjustment. The dynamic relationship is characterized by the presence of lagged dependent variables among the regressors. However, the inclusion of a lagged dependent variable among the regressors causes autocorrelation problems, since the lagged dependent variable is correlated with the error term. This renders the OLS estimator biased and inconsistent even if the error terms are not serially correlated. For the fixed effects estimator, the within groups transformation wipes the individual effects, but $(y_{i,t-1} - y_{i\cdot-1})$ will still be correlated with $(v_{i,t} - v_{i\cdot})$ even if the $v_{i,t}$ are not serially correlated. This is because $y_{i,t-1}$ is correlated with $v_{i\cdot}$ by construction. The same problem occurs with the random effects GLS estimator. In view of the above problems and in order to obtain consistent and efficient estimators for dynamic panel data models, in the present paper we use the econometric methodology developed by Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995) and Blundell and Bond (1998).

3.2. Two-step GMM Difference Estimator

Arellano and Bond (1991) have proposed a methodology for obtaining more efficient estimators once the model has been differenced, by using all the orthogonality conditions that exist between the lagged values of y_{it} and the disturbances v_{it}. They suggest that, if the X_{it} are strictly exogenous in the model with exogenous variables, the moment conditions are:

$$E(y_{i,t-s}\Delta v_{it}) = 0;\ \text{for } t = 3, \ldots, T \text{ and } 2 \leq s \leq t-1 \tag{13}$$

$$E(X_{is}\Delta v_{it}) = 0;\ \text{for } t = 3, \ldots, T \text{ and } 1 \leq s \leq T. \tag{14}$$

and the valid instruments are $Z_i = diag\ [y_{i1}, \ldots, y_{is}\ \ X_{i1}', \ldots, X_{iT}'], (s = 1, \ldots, T-2)$. That is,

$$Z_i = \begin{bmatrix} [y_{i1}X_{i1}'X_{i2}'] & & 0 \\ & [y_{i1}y_{i2}X_{i1}'X_{i2}'X_{i3}'] & \\ & & \cdots \\ 0 & & [y_{i1}\ldots y_{iT-2}X_{i1}'\ldots X_{iT}'] \end{bmatrix}$$

Table 8.1 GMM Estimation Results (Full Sample)

Dependent variable y_{it}

Regressors	GMM-diff		System
	GMM1	GMM2	
$y_{i,t-1}$	0.392325	0.516243	0.471444
	(0.0361)	(0.0000)	(0.0000)
FSDI	0.041133	0.042388	0.210102
	(0.2417)	(0.0000)	(0.0000)
lnSE	0.045193	0.041977	0.165219
	(0.8869)	(0.7206)	(0.0000)
GEXP	−0.926085	−1.088344	−0.069733
	(0.0810)	(0.0001)	(0.0932)
INFL	−0.001866	−0.001827	−0.000166
	(0.2043)	(0.0038)	(0.0018)
OTR	−0.014773	−0.005337	0.126457
	(0.8889)	(0.8846)	(0.0000)
FDI	0.000299	0.000687	0.003675
	(0.8590)	(0.2765)	(0.0005)
Sargan	45.03[734]	35.23[734]	37.44[734]
Dif-Sargan	–	–	37.44[37]
Serial	–	−0.2258[(a)]	0.1834[(a)]
Correlation	–	–	–
	–	–	–
	–	0.0001[(b)]	0.5821[(b)]
RMSE	–	0.8302	0.4212
R^2	0.0164	0.0218	0.7817
no. of Obs.	2535	2535	2535

Notes:

(i) Parentheses report p-values of t-statistics.

(ii) Sargan test reports the χ^2-statistic with degrees of freedom in [ ].

(iii) Serial correlation tests: (a) m_2 statistic; the null hypothesis is that the errors in the first-differenced regression exhibit no second-order serial correlation; their signs indicate the sign of the estimated autocorrelation coefficients in the residuals; (b) p-value of t-statistics of coefficient for v_{t-2} in regressing v_t on v_{t-1}, v_{t-2}, and X.

(iv) RMSE: root mean squared error.

(v) The estimates were derived using RATS.

Using these instruments we obtained the empirical results based on the one-step GMM estimator reported in Table 8.1. The results show that most one-step estimates are insignificant.

On the other hand, when we relax the assumption of strict exogeneity of the explanatory variables, and adopt instead the assumption that all the explanatory variables are weakly exogenous, the valid instruments set is now:

$$Z_i = diag\ [y_{i1}, \ldots, y_{is} X_{i1}', \ldots, X_{i,s+1}'], (s = 1, \ldots, T-2).$$

In this case the moment conditions are:

$$E(y_{i,t-s} \Delta v_{it}) = 0; \text{ for } t = 3, \ldots, T \text{ and } 2 \leq s \leq t-1 \qquad \text{(13 again)}$$

$$E(X_{i,t-s} \Delta v_{it}) = 0; \text{ for } t = 3, \ldots, T \text{ and } 1 \leq s \leq t-1. \qquad (15)$$

Using this set of instruments we performed the GMM estimation and obtained the two-step GMM estimator results reported in Table 8.1. The two-step estimates for Y_{t-1}, *FSDI, GEXP, INFL* become significant at the 1% level. The improvement (as compared to the one-step results) is due to minimizing the asymptotic variance, resulting in more efficient GMM estimators.[7]

3.3. GMM System Estimator

Blundell and Bond (1998) have argued that, when the lagged dependent and the explanatory variables are persistent over time, lagged levels of these variables are weak instruments for the regression equation in differences. The instruments' weakness has repercussions on both the asymptotic and small-sample performance of the difference estimator. As the variables' persistence increases, the asymptotic variance of the coefficients obtained with the difference estimator rises, so that the asymptotic precision of this estimator deteriorates.[8] Furthermore, according to Griliches and Hausman (1986), differencing may exacerbate the bias due to errors in variables by decreasing the signal-to-noise ratio, so that the simple difference estimator may be affected by measurement errors (Levine, Loayza and Beck 2000). In order to deal with these concerns, Blundell and Bond (1998) suggest the use of Arellano and Bover's (1995) system estimator, which reduces the potential biases and imprecision associated with the usual difference estimator.

In view of this, in what follows we employ the GMM system estimator for the estimation of the model using again the same dataset. Arellano and Bover

---

[7] It has been shown that the asymptotic standard errors associated with the two-step estimates are generally around 30% lower than those associated with one-step estimates (Arellano and Bond, 1991).

[8] They show the result of Monte Carlo experiments, namely, that the weakness of the instruments produces biased coefficients in small samples. This bias is exacerbated with the variables' over time persistence, the importance of the specific-effect, and the smallness of the time-series dimension.

(1995) present an estimator that combines the regression in differences with the regression in levels. The instruments for the regression in differences are the same as above, and the moment conditions in Equations (13) and (14) apply to the first part of the system, i.e. the regression in differences. The instruments for the regression in levels are the lagged differences of the corresponding variables. The additional moment conditions in the second part of the system, i.e. the regression in levels, are given as follows:

$$E[(\Delta y_{i,t-s} \cdot \in_{it}] = 0; \text{ for } s = 1, \tag{16}$$

$$E[(\Delta X_{i,t-s} \cdot \in_{it}] = 0; \text{ for } s = 1, \tag{17}$$

where $\in_{it} = \mu_i + v_{it}$. Thus, the additional valid instruments $Z_{yi} = diag\ [\Delta y_{i2}, \ldots, \Delta y_{iT-1}]$ are available for y_{t-1}. Under the assumption of strict exogeneity of the explanatory variables, $Z_{xi} = [\underline{\underline{X}}_{i1}', \ldots, \underline{\underline{X}}_{iT}']$, where $\underline{\underline{X}}_{it} = X_{it} - \underline{X}_{i\cdot}$ and $\underline{X}_{i\cdot} = X_{it}/T$, are additional valid instruments for the second equation of the transformed system. Therefore, the range of choices for valid instruments for the explanatory variables are $Z_{xi} = [X_{i1}', \ldots, X_{iT}', \underline{\underline{X}}_{i1}', \ldots, \underline{\underline{X}}_{iT}']$, which is a Breusch, Mizon and Schmidt (BMS) – type estimator.

The moment conditions in Equations (13), (14), (16) and (17) can be expressed more compactly as:

$$E(\mathbf{Z}_{si}'\mathbf{q}_i) = 0, \tag{18}$$

where:

$$\mathbf{Z}_{si}' = \begin{bmatrix} Z_{di} & 0 \\ 0 & Z_{li} \end{bmatrix}, \mathbf{q}_i = \begin{bmatrix} \Delta\in_i \\ \in_i \end{bmatrix}$$

with $Z_{di} = diag\ [y_{i1}, \ldots, y_{is}, X_{i1}', \ldots, X_{iT}',\ \underline{\underline{X}}_{i1}', \ldots, \underline{\underline{X}}_{iT}']$, (s = 1, … ,$T-2$) and $Z_{li} = diag\ [\Delta y_{i2}, \ldots, \Delta y_{iT-1}]$. Using these moment conditions with the GMM procedure, we can obtain the system estimator. The system GMM estimator is a combination of the GMM differenced estimator and a GMM levels estimator. This combination is linear for the system GMM estimators, which are given by:

$$\begin{pmatrix} \delta \\ \beta \end{pmatrix} = (\mathbf{q}_{-1}'\mathbf{Z}_s(\mathbf{Z}_s'\mathbf{Z}_s)^{-1}\mathbf{Z}_s'\mathbf{q}_{-1}')^{-1}(\mathbf{q}_{-1}'\mathbf{Z}_s(\mathbf{Z}_s'\mathbf{Z}_s)^{-1}\mathbf{Z}_s'\mathbf{q}) \tag{19}$$

In this case we use the instrument set of $Z_{di} = diag\ [y_{i1}, \ldots, y_{i,T-2}\ \underline{\underline{X}}_{i1}', \ldots, \underline{\underline{X}}_{iT-2}']$ and $Z_{li} = diag\ [\Delta y_{i2}, \ldots, \Delta y_{iT-1}]$ to obtain the empirical results shown in Table 8.1.

Overall, the reported results are satisfactory. All the coefficients for independent variables are statistically significant at the 1% level, except *GEXP*, which is significant at the 10% level. The coefficient for the financial variable *FSDI* is positive, and all the other estimates have the expected signs. The value of R^2 is also very high.

3.4. Robustness Checks (Sensitivity Analysis)

The consistency of the GMM estimator depends on whether lagged values of Y and X are valid instruments in the growth regression. To address this issue, we consider a Sargan test of overidentifying restrictions which tests the overall validity of the instruments by analyzing the sample analog of the moment conditions used in the estimation process. The relevant test statistic is given by:

$$S = \Delta v' Z \left[\sum_{i=1}^{N} Z_i'(\Delta v_i)(\Delta v_i)' Z_i \right]^{-1} Z' \Delta v \; asy \; \sim \chi^2(p-k) \qquad (20)$$

where p refers to the number of moment conditions and k is the number of parameters to be estimated. In general, under the null that the moment conditions are valid, S is asymptotically chi-squared distributed with $p-k$ degrees of freedom.

Ahn and Schmidt (1995) show that the maximum number of orthogonality conditions for GMM estimation is $T(T-1)/2+(T-2)$, which represents all the moment conditions implied by the assumptions that the v_{it} are uncorrelated among themselves and with μ_i and y_{i0}. The $T(T-1)/2$ moment conditions are the orthogonality conditions of $E(y_{i_s}\Delta v_{it}) = 0$ in the first-differenced equation, while $(T-2)$ is the orthogonality conditions of $E(v_{it}\Delta v_{it}) = 0$ in the level equation.[9] The 1% critical value for the chi-squared distribution even with 100 df is 135.81. Hence, the null of hypothesis that the instruments are valid is not rejected at any level of significance.

For the system estimators, the Difference Sargan (DS) tests are also used to test the validity of the level moment conditions that are utilized by the system estimators. The *DS* statistic is obtained as the difference between the S-statistic in the system model and that in the differenced model. *DS* is asymptotically

[9] When the homoskedasticity restriction is available, the number of moment conditions is $(T-2)$. However, we allowed the error term to be heterogeneous in our specification, so that, there are $T(T-1)/2$ restrictions in the first-differenced equation and $T-2$ in the level equation. Hence, in our case of $T = 39$, the number of df for the two-step GMM difference estimators is $p-k = (39 \times 38)/2 - 7 = 734$, and that for the system estimators is $p-k = [(39 \times 38)/2 + 37] - 7 = 771$.

chi-squared distributed with $(p_s - k) - (p_d - k)$ degrees of freedom under the null that the level moment conditions are valid. In our case the relevant statistic $DS = 37.44$ with 37 df (p-value $= 0.4488$), thus it does not reject the null hypothesis at any level.

Arellano and Bond (1991) have proposed a test for the hypothesis that there is no second-order serial correlation for the disturbances of the first-differenced equation. The test is important because the consistency of the GMM estimator relies on the assumption that $E[\Delta v_t \Delta v_{t-2}] = 0$. The m_2 test statistic of Arellano and Bond (1991) takes the form:

$$m_2 = [v^\wedge{}_{-2}{}' v^\wedge *]/v^{\wedge 1/2} \sim N(0, 1), \tag{21}$$

where v is a vector of residuals, v_{-2} is the vector of residuals lagged twice, and $v^\wedge *$ is the vector of trimmed v to match v_{-2}. This test statistic is the standardized second-order residual autocovariances (Bond, 2002). In the present paper we calculate the value of the statistic expressed as $m_2 = \mathrm{cov}(v^\wedge{}_{-2},\ v^\wedge *)/\mathrm{var}(v^\wedge)$. The test results do not reject the null of no second-order serial correlation in all cases.[10] In addition, we test the null hypothesis by using the t-statistic of coefficient for v_{t-2} in regression v_t on $v_{t-1},\ v_{t-2}$ and X. The result shows that the coefficient for v_{t-2} is statistically insignificant in the relevant regression for the GMM system estimator, supporting the fact that v_t is uncorrelated with v_{t-2}.

Finally, we test the predictive accuracy of the system estimator and the two-step estimator by using the RMSE (the root mean squared error):

$$\mathrm{RMSE} = \sqrt{\frac{1}{T^0} \Sigma_i (y_i - y^\wedge{}_i)^2} \tag{22}$$

where T^0 is the number of periods being forecasted. As shown in Table 8.1, the RMSE for the system estimator is 0.4212, while that of two-step estimator is 0.8302. Thus, the GMM system estimator performs more precisely than the two-step estimator.

3.5. Economic Prediction from the Dynamic Model

The above results seem to suggest a statistically significant impact of the financial sector development indicator on per capita GDP in a dynamic panel data setting.

[10] In the case of GMM system estimator we obtained $m_2 = 0.1838$ with the significance level 0.4270 $(= 0.8541/2)$ in the normal distribution i.e. 0.4270 exceeds the significance level of $\alpha = 0.10$, and the null of no second order serial correlation is not rejected.

Let us interpret the economic meanings of the above results. The interpretation is based on the GMM system estimator results.

Our central variable, namely financial sector development (*FSDI*) is positively associated with economic growth. Recall that in the dynamic model (12), the coefficient of the lagged dependent variable δ postulates the speed of adjustment, which represents lag effects. The financial development in past periods affect growth rate in the current period with geometrically declining influences as $\beta_k 0.4714^i$ $(i = 0, 1, 2 \ldots)$. In other words, if we consider polynomials in the lag operators $D(L), A(L), B(L)$ as:

$$D(L) = A(L)^{-1}B(L) = d_0 + d_1L + d_2L^2 + d_3L^3 + \cdots , \tag{23}$$

then $A(L) = 1 - 0.4714L,\ B(L) = \beta_0$. The impulse response functions will be $d_0 = \beta_0,\ d_1 = 0.4714\beta_0,\ d_2 = 0.4714d_1,\ d_3 = 0.4714d_2$ and so on. Thus, the proceeding traces through the effects on growth of a one-time innovation in "independent" variables.

In the case of *FSDI*, $\beta_0 = 0.2101$. This suggests that 1 unit increase of *FSDI* will affect the growth rate by 0.2101% increase in the current period, then one period later it will cause a 0.099% increase, two periods later 0.0466%, three periods later 0.022%, four periods later 0.0104%, and so on. These effects of the financial "shocks" on the real output for the full sample of 65 countries are plotted in Figure 8.1. From this dynamic property, we can see that a change in financial sector development does not affect real growth rate with one-shot effect, but exerts persistent sizable impacts on growth within the context of a distributed-lag pattern.

3.6. Does the Level of Development Matter?

It would be reasonable to assume different impacts of financial sector development on growth between developed and developing countries. To test this hypothesis we divided the full sample into two sub-groups of 24 industrial countries and 41 developing countries and obtained the regression results reported in Table 8.2 using the GMM system estimator.

In view of these results, the impulse response functions of y on the financial indicator for the two country groups are plotted in Figure 8.2.

As shown in Figure 8.2, in the developed country group *FSDI*'s influence on Y will peter out quickly, whereas in the developing country group, distant past values of *FSDI* will have considerable impacts on Y. The mean lag is 1.098 for the developed country group and 2.177 for the developing country

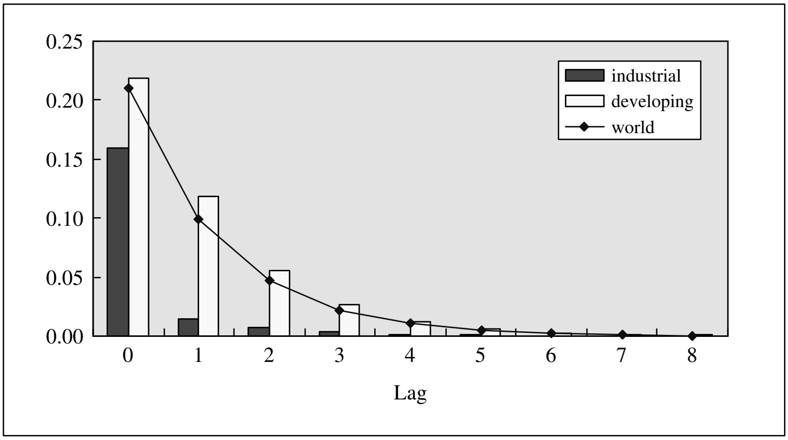


Fig. 8.1. Impulse Response Functions of *Y* on *FSDI* (based on GMM system estimator results)

Table 8.2 Estimation Results for Industrial and Developing Countries (GMM System Estimator)

Dependent variable y_{it}

Regressors	GMM System Estimator	
	Industrial	Developing
$y_{i,t-1}$	0.0895	0.5407
	(0.6212)	(0.0000)
FSDI	0.1599	0.2184
	(0.0005)	(0.0000)
ln*SE*	1.0672	0.1054
	(0.0000)	(0.0000)
GEXP	0.5376	−0.1365
	(0.0989)	(0.0006)
INFL	−0.0011	−0.0001
	(0.0924)	(0.0238)
OTR	0.4636	0.0560
	(0.0000)	(0.0048)
FDI	0.0017	0.0020
	(0.3976)	(0.1062)
R^2	0.74	0.76

Notes:
(i) Parentheses report p-values of t-statistics.
(ii) The estimates were derived using RATS.

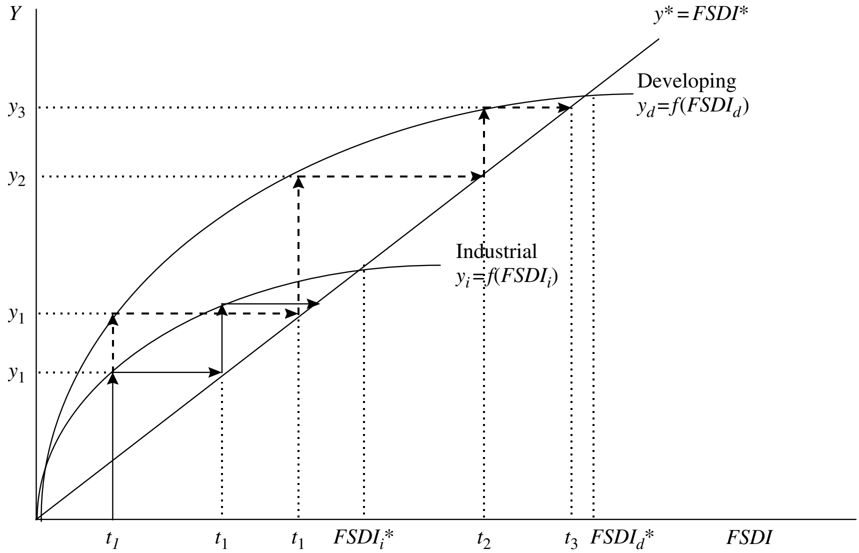


Fig. 8.2. Dynamics of adjustment in the relationship between *y* and *FSDI* (based on GMM system estimator results)

group. In addition, the magnitude of the coefficient for developing countries 0.2184 is larger as compared to 0.1599 for industrial countries. Hence, the results seem to suggest that the effect of financial sector development in developing countries is more persistent and larger than those in industrial countries. In other words, if a gap occurs between finance and economic growth, in industrial countries the gap will be filled more quickly and the equilibrium will be restored within a relatively shorter period of time as compared to the case of developing countries.

Apart from the financial variables (of central importance to the study), we can derive some tentative economic implications on the basis of the results obtained in connection with the conditioning information set of variables. As far as scale effects are concerned, the estimated coefficient of *SE* seems to suggest the existence of scale effects, that a 1% increase in the effective labour force leads on average to about a 0.1652% increase in the real output in the current period, and then affects it with geometrically declining effects, namely, 0.077% one period later, 0.0368% two periods later, 0.0173% three periods later, and so on. Thus, the finding suggests that per capita growth has a positive linkage with the scale effect, and responds to a change in the labour force with time lags. Endogenous growth theories imply some benefits from larger scale. In particular, if there are significant setup costs at the country level for inventing or adapting new products or production techniques, then larger economies would perform better. The variable *SE* is logarithmic, thus its coefficient indicates the output elasticity of labour force. When we run regressions by splitting the full sample into two country groups, the estimated coefficient for *SE* of the industrial country group is 1.0672, which is larger than 0.1054 of the developing country group. This seems to suggest that the output elasticity of labour force in industrial countries is higher than in developing countries.

Government expenditure plays an important role in the growth process and it could affect economic growth positively or negatively. Our results show that the full sample country group is associated with a negative estimate for government expenditure in the dynamic models. However, it is interesting that, when we split the full sample into two country groups, the industrial country group reveals a positive effect, whereas the developing country group appears with a negative effect concerning government expenditure.

The relationship between inflation and economic growth is more complex because inflation affects economic growth indirectly through real money balances in saving or investment functions, rather than directly. Our empirical results seem to suggest that for the full sample the estimated coefficient on *INFL* is negative, as shown in Table 8.1. The finding supports the argument that inflation has a negative effect on growth, even if the magnitude of the impact is small. This seems also to be the case when we split the sample into the two country groups of developed and developing countries.

Turning to the open economy variables used in the study, the estimated coefficient of *OTR* is significantly positive: 0.1246 in the case of the full sample. When we divide the full sample into two sub-groups to capture potentially different effects related to different levels of development, the estimated coefficients of *OTR* are 0.4636 and 0.0560 for industrial and developing countries, respectively. It is notable that the magnitude of coefficient for industrial countries is much larger than that of developing countries. This result tells us that foreign trade affects GDP much more in industrial countries than in developing countries. Thus, it is closely related to economic growth. In the case of foreign direct investment, the estimated coefficient of *FDI* is significantly positive: 0.0036 for the full sample. It means that foreign direct investment influences positively real per capita growth in the dynamic process. However, for the sub-groups of developing and industrial countries, the reported results are rather inconclusive since *FDI* enters insignificantly in both cases, even if the coefficient turns to be positive as expected.

4. CONCLUDING REMARKS

The paper tried to delve deeper into the relationship between financial sector development, broadly defined to go beyond financial deepening, and economic growth by using a new database including 65 countries (both industrial and developing ones) over the period 1960–99 and by also exploring new routes regarding the measurement of financial sector development. Empirical results obtained from the estimation of dynamic panel data models using various GMM estimators (including the GMM system estimator) seem to suggest that financial sector development contributes to economic growth although the magnitude of the impact varies depending *inter alia* on the level of development (industrial vis-à-vis developing countries).

Our results seem also to indicate that the effect of financial sector development in developing countries is more persistent and larger than those in industrial countries. In other words, if a gap occurs between finance and economic growth, in industrial countries the gap will be filled up more quickly and the equilibrium will be restored within a relatively shorter period of time as compared to the case of developing countries. We also found that per capita GDP growth has a positive linkage with scale effects and responds to a change in the labour force with time lags. Empirical results in line with *a priori* expectations were also derived regarding the impact of open economy variables used in the study and inflation. Our findings seem also to be robust in view of the sensitivity analysis we carried out.

Needless to say, the reported findings are far from conclusive regarding the above relationship at the world global level since an even larger database would be essential covering most countries in the world before we ended up with robust policy conclusions. Other factors that may also be inserted into the picture, such as institutions, may improve our overall understanding on how financial sector development really works in industrial and developing countries. Furthermore, the available empirical evidence seems to provide policymakers with inadequate advice regarding the sequencing of financial sector developments (Wachtel, 2004). Experimenting with industry and firm data may be also rewarding regarding the above relationship (a route already taken by some recent studies in this area).[11] These are important challenges for future research on the finance-growth nexus.

Finally, an important issue calling for further research is related to the overall finance-growth-poverty reduction relationship of relevance to many developing countries undertaking a series of reforms in their financial sector, in par-

[11] See Wachtel (2004) for an insightful discussion.

ticular, illuminating the channels through which financial sector development can contribute to the overall development process and poverty-reducing growth (Green *et al*., 2003).

ACKNOWLEDGEMENT

George Mavrotas is grateful to Leonardo Becchetti, Iftekhar Hasan, Alexander Karmann, Robert Lensink, Marco Mazzoli and conference participants for very helpful comments and suggestions. The usual disclaimer applies. The views expressed in the chapter are those of the authors and should not be necessarily attributed to UNU-WIDER.

BIBLIOGRAPHY

Ahn, S.C. and P. Schmidt (1995) "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics* **68**: 5–27.

Arellano, M. and S. Bond (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies* **58**: 277–97.

Arellano, M. and O. Bover (1995) "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics* **68**: 29–51.

Arestis, P. and P. Demetriades (1997) "Financial Development and Economic Growth: Assessing the Evidence," *Economic Journal* **107**(May) 783–99.

Atje, R. and B. Jovanovic (1993) "Stock Markets and Development," *European Economic Review* **37**: 632–40.

Bandiera, O., G. Caprio, P. Honohan, and F. Schiantarelli (2000) "Does Financial Reform Raise or Reduce Saving?" *The Review of Economics and Statistics* **82(2)**: 239–63.

Barro, R.J. (1991) "Economic Growth in a Cross-Section of Countries," *The Quarterly Journal of Economics* **56**: 407–43.

Barro, R.J. and X. Sala-i-Martin (1995) *Economic Growth*. McGraw-Hill.

Beck, T., A. Demirguc-Kunt and R. Levine (1999) "A New Database on the Financial Development and Structure," Policy Research Paper, No.2146, World Bank.

Beck, T., R. Levine and N. Loayza (1999) "Finance and the Sources of Growth," Policy Review Working Paper, No.2057, World Bank.

Benhabib, J. and M.M. Spiegel (2000) "The Role of Financial Development in Growth and Investment," *Journal of Economic Growth* **5**(December): 341–60.

Blundell, R. and S. Bond (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Models," *Journal of Econometrics* **87**: 115–43.

Bond, S. (2002) "Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice," *CEMMAP Working Paper*, cwp09/02, The Institute for Fiscal Studies.

Demetriades, P. and Luintel, K. (1996) "Banking Sector Policies and Financial Development in Nepal," *Oxford Bulletin of Economics and Statistics* **58(2)**: 355–72.

Demirguc-Kunt, A. and R. Levine (eds) (2001) *Financial Structure and Economic Growth*. Cambridge, Mass: MIT Press.

Fry, M.J. (1988) *Money, Interest, and Banking in Economic Development*. Baltimore and London: The Johns Hopkins University Press.

Goldsmith, W. (1969) *Financial Structure and Development*. New Haven: Yale University Press.

Green, C.J. and C. Kirkpatrick (2002) "Finance and Development: An Overview of the Issues," *Journal of International Development* **14(2)**: 207–10.

Green, C.J., C. Kirkpatrick and V. Murinde (2003) "How Does Finance Contribute to the Development Process and Poverty Reduction?" mimeo, Birmingham Business School, University of Birmingham, May.

Gurley, G.J. and Shaw, S.E. (1960) *Money in a Theory of Finance*. Washington, DC: Brookings Institution.

Hausman, J. (1978) "Specification Tests in Econometrics," *Econometrica* **46**: 1251–71.

Hermes, N. and Lensink, R. (eds) (1996) *Financial Development and Economic Growth: Theory and Experience from Developing Countries*. Routledge Studies in Development Economics.

Honohan, P. (2004) "Financial Development, Growth and Poverty: How Close are the Links?" *World Bank Policy Research Working Paper No. 3203*, February, World Bank, Washington DC.

Kelly, R. and G. Mavrotas (2003) "Financial Sector Development: Futile or Fruitful? An Examination of the Determinants of Saving in Sri Lanka," *WIDER Discussion Paper No. 2003/14*, World Institute for Development Economics Research, United Nations University, Helsinki, Finland.

Khatkhate, D.R. (1988) "Assessing the Impact of Interest Rates in Less Developed Countries," *World Development* **16**: 577–88.

King, R.G. and R. Levine (1992) "Financial Indicators and Growth in a Cross Section of Countries," *World Bank Working Paper*, WPS 819, World Bank.

King, R.G. and R. Levine (1993a) "Finance and Growth: Schumpeter Might be Right," *Quarterly Journal of Economics* **108**: 717–38.

King, R.G. and R. Levine (1993b) "Finance, Entrepreneurship, and Growth: Theory and Evidence," *Journal of Monetary Economics* **32**: 513–42.

Levine, R. (1992) "Financial Structures and Economic Development," *World Bank Working Papers*, WPS 849, World Bank.

Levine, R. (1997) "Financial Development and Economic Growth: Views and Agenda," *Journal of Economic Literature*. **35**: 688–726.

Levine, R. and S. Zervos (1996) "Stock Market Development and Long-Run Growth," *The World Bank Economic Review* **10(2)**: 323–39.

Mavrotas, G. and R. Kelly (2001) "Savings Mobilisation and Financial Sector Development: The Nexus," *Savings and Development* **XXV(1)**: 33–66.

McKinnon, R.I. (1973) *Money and Capital in Economic Development*. Washington, D.C.: Brookings Institution.

Patrick, H.T. (1966) "Financial Development and Economic Growth in Underdeveloped Countries," *Economic Development and Cultural Change* **14**: 74–89.

Pill, H. (1997) "Real Interest Rates and Growth: Improving on Some Deflating Experiences," *Journal of Developing Studies* **34(1)**: 85–110.

Quah, D. (1993) "Empirical Cross-section Dynamics in Economic Growth," *European Economic Review* **37**: 426–34.

Roubini, N. and X. Sala-i-Martin (1992) "Financial Repression and Economic Growth," *Journal of Development Economics* **39**: 5–30.

Shaw, E.S. (1973) *Financial Deepening in Economic Development*. New York: Oxford University Press.

Tobin, J. (1965) "Money and Economic Growth," *Econometrica* **33**: 671–84.

Wachtel, P. (2004) "How Much Do We Really Know About Growth and Finance?," *Research in Banking and Finance* **4**: 91–113.

Wachtel, P. and P. Rousseau (1995) "Financial Intermediation and Economic Growth: A Historical Comparison of the US, UK and Canada," in M. Bordo and R.S. Irwin (eds), *Anglo-American Finance*.

Wallich, H.C. (1969) "Money and Growth: A Cross-section Analysis," *Journal of Money, Credit and Banking* **1**: 281–302.

World Bank (2001) *Finance for Growth: Policy Choices in a Volatile World*. New York: Oxford University Press.

Appendix

Countries included in the sample (65)

Industrial Countries (24)

1. Australia
2. Austria
3. Belgium
4. Canada
5. Cyprus
6. Denmark
7. Finland
8. France
9. Germany
10. Greece
11. Iceland
12. Italy
13. Ireland
14. Japan
15. Luxembourg
16. Malta
17. Netherlands
18. Norway
19. New Zealand
20. Portugal
21. Sweden
22. Switzerland
23. United Kingdom
24. United States

Developing Countries (41)

Africa (15)

25. Burundi
26. Cameroon
27. Cote d'Ivoire
28. Ethiopia
29. Gabon
30. Ghana
31. Kenya
32. Morocco
33. Niger
34. Nigeria
35. Rwanda
36. Senegal
37. Sierra Leone
38. South Africa
39. Tanzania

Middle East (2)

40. Egypt
41. Iran

Asia and Pacific (10)

42. Fiji
43. India
44. Indonesia
45. S. Korea
46. Malaysia
47. Nepal
48. Pakistan
49. The Philippines
50. Sri Lanka
51. Thailand

South America (14)

52. Colombia
53. Costa Rica
54. Dominican Rep.
55. Ecuador
56. El Salvador
57. Guatemala
58. Haiti
59. Honduras
60. Jamaica
61. Mexico
62. Panama
63. Paraguay
64. Trinidad and Tobago
65. Venezuela

This Page is Intentionally Left Blank

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 9

THE (CORPORATE) EQUITY RISK PREMIUM/(CORPORATE) BOND RISK PREMIUM NEXUS IN THE U.S. MARKET

Leonardo Becchetti, Andrea Carpentieri, Rocco Ciciretti, Nazaria Solferino

Abstract

We analyze the relationship between the risk of holding a stock and the risk of holding a corporate bond on a large sample of option adjusted delta credit spreads (DCSs), disaggregated by rating, industry and maturity. We find that the implied equity risk premium (IERP) significantly and positively affects the IG corporate bond risk premium (proxied by DCSs). The nexus reveals that a common risk tolerance (perception of risk) component affects both bond and equity markets. This risk tolerance component is highly sensitive to economic (financial scandals and crises) and political (September 11, Iraqi and Afghani wars) shocks and its impact on DCSs is not captured by traditional DCSs determinants such as stock market performance and volatility, institutional investors and consumers confidence and interest rate levels.
Implied Equity Risk Premium – Bond Spreads

JEL Classification No.: G11–G12

1. INTRODUCTION

One of the most important issues that have recently captured the attention of large part of the empirical literature on credit risk is the identification of factors affecting corporate bond credit spreads. The origins of this literature trace back to the pioneering work of Merton (1974), who introduced structural models of default, by modeling the lower re-organization boundary and the allocation of residual values upon liquidation exogenously. Structural models of default specify a particular firm value process, and assume that default is triggered when firm value falls below some explicit threshold. They therefore require the knowledge of the underlying firm value and of its volatility. This approach has given origin to a series of models, either removing or adding some hypotheses to the original framework. Longstaff and Schwartz (1995) introduce stochastic interest rates. Leland and Toft (1996), make default endogenous and relate bond pricing to the structure of the firm value. Zhou (1997) adds the possibility of jumps in asset values; and Duffie and Lando (2001) develop a model with imperfect information, assuming that only a noisy process, but not the forcing variable, that is, the value of an unlevered firm, may be observed.

Another strand of literature on corporate debt pricing, that of reduced form models, has instead focused its attention on exogenous specifications of default outcomes and recovering rates based on arbitrage-free valuation. This class of models, investigated by Jarrow and Turnbull (1995), Jarrow, Lando and Turnbull (1997), Duffie and Singleton (1999), and others, assigns probabilities of default and recovery rates exogenously, but derives pricing formulas that can be calibrated to data.[1] Although reduced form models have been useful in fitting the observed credit spreads, their abstraction from the underlying firm value process makes them less useful in suggesting determinants of credit spread changes. This is the reason why our empirical analysis is more related to that of structural models, given that the interpretation provided by structural models is useful for predicting the determinants of credit spread changes. In these models, equity possesses characteristics similar to a call option, while the debt claim exhibits features analogous to those of a portfolio that has a claim on firm value, being short on a call option. For example, consider a simple structural model of default (as in Collin-Dufresne *et al.*, 2001) with firm value following the risk-neutral process:

$$\frac{dV}{V} = (r-\delta)dt + \sigma dz^{Q} + \lambda(dq^{Q} - pdt) \qquad (1)$$

where, V is firm value, r the spot rate, δ the firm payout rate, σ the firm's volatility, λ the size of a firm-value jump, and p the risk-neutral probability,

[1] More specifically, also in the model of Duffie and Lando (2001), for not perfectly observable assets, structural models have an intensity of default and therefore they become reduced form models. This paper represents an attempt to unify these two approaches.

or intensity, of such a jump (with dq^Q being the risk neutral transition density of the jump process). Default is assumed to occur the first time the firm value reaches a threshold K. In this model it is implicitly assumed that K is the amount of debt outstanding. This structural framework suggests that factors affecting changes in credit spreads are changes in: i) the spot rate r, ii) the slope of yield curve, iii) leverage $\frac{K}{V}$; iv) the volatility σ; v) the probability or magnitude of a downward jump $\{p, \lambda\}$[2]; vi) the business climate.[3] Most of these variables have been used in the first empirical works on the determinants of bond returns. In these works, Fama and French (1989, 1993) evaluate the relationship between aggregate stock and bond returns. Cornell and Green (1991) and Kwan (1996) analyze the relationship between the two markets at aggregate index and firm level. The main result of these first studies is that low-grade bond returns are relatively more correlated with stock returns, while high-grade bond returns are relatively more correlated with government bond returns.

More recent empirical analyzes have shifted the focus from bond yields to changes in the yield differential between corporate and government returns at the same maturity. This variable, defined as delta credit spread, is considered a more accurate proxy of corporate credit risk, since it measures the excess return required by investors for the additional risk involved in holding corporate instead of government bonds. In their main findings, sources of risk affecting credit spreads have been shown to be at least four:

1) changes in risk perception by market investors;
2) default risk of bond issuers;
3) uncertainty about the timing of default; and
4) uncertainty about the recovery value, or the value reimbursed to bondholders at maturity.

Business cycle is obviously an important driver for many of these risks. For example, Fama and French (1995) explain that up to 30% of the variability of DCSs is not accounted for by default risk. Pedrosa and Roll (1998) provide a first descriptive analysis of DCSs for fixed bond indexes classified for grade, industry and maturity. Collin-Dufresne *et al.* (2001) explain around 25% of the variability of DCSs of Lehman Brothers U.S. bond investment grade indexes.

---

[2] Implied volatility in observed option prices suggest that markets do account for the probability of large negative jumps in firm value. Increases in either the probability or the magnitude of a negative jump should widen credit spreads.

[3] Even if the probability of default remains constant for a firm, changes in credit spreads can occur due to changes in the expected recovery rate. The expected recovery rate, in turn, should be a function of the overall business climate.

The authors find that regression residuals are highly cross-correlated and their principal component analysis shows that DCS variability is largely explained by a principal component, which they interpret as generated by local demand and supply shocks. Elton *et al.*, (2001) explain corporate bond returns mainly in terms of systematic non-diversifiable risk premium and, after that, default risk premium and fiscal components. Using the typical Fama and French three-factor risk model, the authors aim to explain the determinants of risk premium for risk associated with corporate bonds (and their importance), and find that a substantial part is due to the systematic risk, as it happens for stock returns. Huang and Kong (2003) use nine Merrill Lynch investment grade and high-yield U.S. indexes and explain up to 30% of investment grade and 60% of high-yield credit spreads.

Our work contributes to this literature by introducing a measure of risk tolerance among variables affecting credit spreads. For this purpose we build from asset pricing dividend cash flow models a measure of implied equity risk premium, which we interpret as a time varying common risk tolerance factor affecting also the investor perception of risk in holding corporate bonds. Buying a stock and a bond is not the same thing. The first is a bet on firm profits, and the second is a bet on its capacity to repay the debt. Nonetheless, our aim is to test whether there are common risk components associated to the two bets.

The chapter is divided into six sections (including introduction and conclusions). In the second section we introduce and discuss our measure of implied equity risk premium. In the third section we describe our database and comment our descriptive findings. In the fourth section we present our empirical finding and in the fifth section we discuss our econometric results. Section 6 concludes.

2. THE EXTRACTION OF THE IMPLIED EQUITY RISK PREMIUM

Elton *et al*. (2001) use Fama and French (1993) risk factors to explain spread components influenced by systematic risk. These are the excess return on the market, the return on a portfolio of small stocks minus the return on a portfolio of a large stocks (the SMB factor), and the return on a portfolio of high stocks, minus the return on a portfolio of low book-to-market stocks (the HML factor).

We focus in particular on the first of these factors, but unlike these authors, our variable is not the same as in Fama and French, where it is a measure of equilibrium derived from the CAPM mode.[4]

Our implied equity risk premium is calculated as the average implied risk premium extracted from a Discounted Dividend (DD) asset-pricing model on individual S&P constituents. The same approach is followed by Claus and Thomas (2001), who estimate the post-1985 equity premium for the U.S. market, using I/B/E/S earning forecasts and a risk-free rate, proxied by long-term government bond returns. The authors find the risk premium to be of the order of 3%, much less than the historical difference between secular average yearly returns on stocks and bonds in the U.S. market and less than the 8% average equity premium observed since 1926. This result documents the extreme confidence in the future of financial investors in the post Berlin-wall era before the stock market bubble burst in March 2000. Similar results are obtained by the same authors on equity markets of Canada, France, Germany, Japan, and the United Kingdom and by Jagannathan *et al.* (2001) who estimate negative implied risk premia for limited periods in the last two decades. The advantage of the DD approach is that we exactly measure financial investors' current availability to pay for holding a stock, after controlling for the impact of consensus forecasts on future earnings, exposition to systematic non-diversifiable risk and risk-free rate. The DD approach also avoids three limits of the CAPM factors. First, the correspondence between excess return and additional risk is valid only in equilibrium when all CAPM hypotheses hold. Second, the commonly used stock index return is just a proxy of the overall stock market return, which should be the proper variable to consider in CAPM models. Third, stock market excess return is also a measure of stock market performance and, as such, it is expected to reduce and not to increase DCSs.

---

[4] In Fama and French (1993) the risk premium (*Rm–Rf*) is estimated as follow. *Rf* is the one-month Treasury bill rate, observed at the beginning of the month; *Rm* is the value-weighted monthly percent return on the stocks in the 25 size-BE/ME (the ratio of the book value of a firm's common stock, BE, to its market value, ME) portfolios, plus the negative-BE stocks excluded from the portfolios. In its turn, the 25 size-BE/ME stock portfolios are formed as follows. Each year, *t* from 1963 to 1991, NYSE quintile breakpoints for size are used to allocate NYSE, Amex and NASDAQ stocks five size quintiles. Similarly, NYSE quintile breakpoints for BE/ME are used to allocate NYSE, Amex and NASDAQ stocks to five book-to-market equity quintiles. The 25 size-BE/ME stock portfolios are formed as the intersections of the five size and the five BE/ME groups.

Following Adriani *et al.* (2004) and Bagella *et al.* (2004), our implied risk premium is extracted from the following asset pricing formula:

$$P_{obs} = \sum_{t=0}^{5} \frac{DPS_0(1+E[g_u])^t}{(1+r_{CAPM})^t} + \frac{DPS_0(1+E[g_u])^6}{(1+r_{CAPM})^6(R-G)} \tag{2}$$

where P_{obs} is the daily observed closing price; DPS is the dividend per share; $E[g_u]$ is the I/B/E/S expected earning per share growth rate.[5] $r_{CAPM} = r_f + \beta * PR$ (risk free rate plus beta times risk premium). The r_f used in our formula is the 3-month U.S. T-bill issued by the U.S. Government. β is a non-diversifiable risk component and is calculated by regressing weekly individual stock returns on weekly returns of the stock market index (S&P 500 COMPOSITE INDEX return) on a time window that includes the two preceding years. $(R-G)$ is the difference between the discount rate of a stock that behaves like the rest of the economy $(R = r_f + PR)$ and the nominal perpetual rate of growth of the economy. We arbitrary fix the threshold for the two-stage approach at the end of the fifth year. In the second stage (from the sixth year on) the stock behaves like the rest of the economy, that is, as a perpetual asset yielding the perpetual nominal growth rate of the economy, G, and being discounted at the rate R where $R = r_f + PR$. The unknown variable in Equation (1) is exactly the implied equity risk premium (PR), which is extracted on the basis of all other parameter values with a simple computational algorithm. As is well known, this formula is highly sensitive to the perpetual rate of growth G. We therefore choose a nominal rate of growth of 3% and perform a sensitivity analysis around this value checking whether our findings are robust when the nominal rate of growth is set at 2.5 or 3.5%. What we observe in our work is just a shift of the implied risk premium, but no significant changes in its law of variation. Another apparently critical point is the choice of the threshold in the two-stage growth formula. We choose the fifth year and we perform a robustness check also in this case. Shifts (forward) in the threshold produce two compensating effects. On the one hand, an additional year of (first-stage) high earnings growth, for a given stock price, raises the implied risk premium. On the other hand, the starting value of the terminal rate of growth formula is anticipated one year and therefore less heavily discounted, thereby reducing the implied risk premium. Once again, the robustness check is shown to affect mainly the absolute value of the implied risk premium and much less its law of variation. Econometric findings described in the following section are robust to this change as well.

---

[5] Since data on dividend growth forecasts are not available, the usual approach is to proxy this variable with I/B/E/S earning growth forecasts, under the assumption of a constant dividend/earning ratio.

3. DATA SOURCE AND SAMPLE DESCRIPTION

3.1. Descriptive Statistics on the Dependent Variable

We use as alternative dependent variables, a set of IG corporate bonds indexes[6] on option-adjusted spreads (OASs), provided by Merrill Lynch. The use of OASs insulates our target variable (changes in credit risk) from changes in the value of options attached to corporate bonds. The problem is common because many corporate bonds are callable, or have call options that allow issuers to repurchase them at a convenient time in order to re-finance their investment at lower interest rates.

Our dataset includes 207 IG indexes, from AAA to BBB rating, with 83 monthly observations for an observation period ranging from January 1997 to November 2003. Maturities are classified according to the following buckets 1–3, 3–5, 5–7, 7–10, 10–15 and 15+ years and data are available also for different ratings and macro-industries (Financials, Industrials, and for the U.S. Utilities). Provided by Merrill Lynch, an investment bank that sells its information (and its fixed-income indexes) on the market, our dataset benefits from a strong producer's incentive to release high-quality data.[7] Moreover, our dataset does not include "matrix prices" or matrix interpolation with available data for missing observations as occurs in the widely used Lehman Brothers Fixed Income Database used in most empirical analyzes (Sarig and Warga 1989; Collin-Dufresne *et al.*, 2001; Elton *et al.*, 2001).

In Table 9.1 we report descriptive statistics for a selected number of fixed-bond indexes in our database, ranked by maturity, ratings and industry classifications. The table shows that DCS volatility is monotonically increasing in credit risk, with B rating indexes having a volatility that is three times higher than that of AAA indexes. Auto and Telecommunications indexes are those with the highest DCSs volatility among industries. In the same table skewness, kurtosis and Jarque-Bera tests clearly show that our DCS series have a non-normal distribution and excess kurtosis, exactly as stock return series. First-order autocorrelation co-efficients and Box-Pierce diagnostics reveal the presence of significant autocorrelation in these series.

---

[6] All indexes are rebalanced in the last day of the month to account for entries, exits or transition of individual bonds to different investment grade or maturity classes. To avoid these changes affecting our dependent variables, DCSs are calculated as differences between the first day of the month (in which the rebalancing has already occurred) and the day before the revision which follows.

[7] This incentive, which is not ensured when data are provided by market makers given the illiquidity of corporate bond markets, is taken into account by Collin-Dufresne *et al*. (2001)., Elton *et al.* (2001) and Duffee and Singleton (1999).

Table 9.1 Descriptive statistics of DCSs for most representative corporate bond indexes and of their main determinants

Mean	S.dev	Min	p10	p50	p90	Max	Sk	Kurt	J-B	B-P	ac(1)	AV	numB
US_{DCS}	.35	11.80	−44	−9	−1	15	47	.32	7.25	63.82	6.98	.1392	3789
US_{AAA}	−.28	5.71	−16	−6	−1	6	21	.60	5.33	23.71	9.68	.0630	117
US_{AA}	−.25	7.01	−16	−8	0	8	24	.73	4.62	16.42	12.66	.1527	454
US_{A}	.23	9.87	−37	−10	0	13	29	.19	5.77	27.09	6.23	.0643	1734
US_{BBB}	.86	17.37	−64	−12	−1	18	79	.77	9.17	139.86	11.03	.1861	1484
US_{1-3Yrs}	.92	13.93	−62	−9	−1	17	53	−.12	9.06	127.01	13.38	.0945	653
US_{3-5Yrs}	.29	11.80	−44	−9	−1	15	47	.32	7.25	63.82	6.98	.1392	3789
US_{5-7Yrs}	−.05	12.10	−45	−11	−1	14	47	.40	7.25	64.58	10.07	.1725	505
$US_{7-10Yrs}$	.58	12.18	−39	−10	−1	15	46	.45	6.39	42.53	8.87	.1370	697
$US_{10-15Yrs}$	.07	10.94	−28	−9	−1	14	53	1.34	8.96	147.79	11.50	.1350	194
US_{15+Yrs}	−.34	11.91	−37	−10	−2	16	39	.04	5.28	17.95	7.44	.1193	1077
$US_{FinCorp}$	−.24	12.05	−53	−12	−1	15	34	−.43	7.38	69.03	4.35	.0728	1222
$US_{Banking}$	−.88	8.21	−29	−11	−1	9	20	−.11	4.32	6.24	8.15	.1235	567
$US_{Brokerage}$	−1.22	10.77	−34	−12	−2	11	36	−.52	5.26	21.35	7.18	.1035	153
$US_{Fin\&Invest}$	.43	19.39	−105	−13	0	18	72	−1.15	14.55	479.84	5.12	−.0622	375
$US_{Insurance}$	.86	15.99	−19	−11	−1	14	115	4.60	32.63	3329.01	12.81	.2887	127
$US_{IndCorp}$	.41	12.76	−37	−11	−2	15	51	.85	6.58	54.43	7.46	.1563	2130
$US_{AutoGroup}$	.04	26.30	−121	−20	−1	29	78	−.94	9.18	144.29	3.52	−.0278	136
$US_{BasicInd}$	.01	10.25	−28	−8	−1	14	26	.21	4.30	6.41	9.64	.1616	239

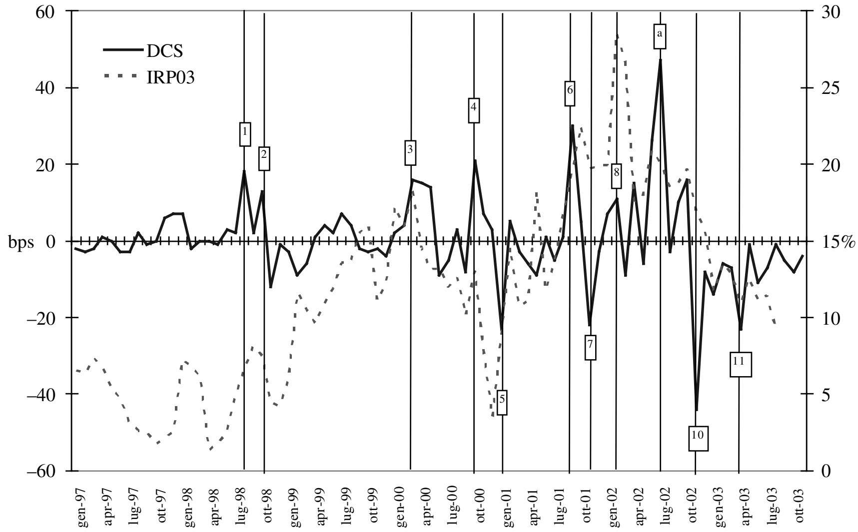
Table 9.1 (*Continued*)

Mean	S.dev	Min	p10	p50	p90	Max	Sk	Kurt	J-B	B-P	ac(1)	AV	numB
$US_{ConsCycl}$	−.31	15.24	−66	−13	−1	18	44	−.71	7.25	69.28	7.61	−.0106	213
$US_{CapGoods}$	.72	10.17	−25	−9	−1	14	33	.45	4.25	8.23	12.78	−.0739	188
$US_{ConsNoCycl}$	−.63	7.40	−29	−7	−1	8	20	−.24	5.98	31.48	12.32	.2309	279
US_{Energy}	.64	12.75	−30	−10	−1	14	74	2.37	15.01	576.67	18.37	.2511	344
US_{Media}	−1.64	22.52	−59	−21	−3	15	126	2.17	15.70	623.08	11.52	.1632	154
$US_{RealEstate}$	−.98	11.25	−31	−13	−2	12	48	1.04	6.97	69.42	26.63	.3320	95
$US_{ServiceCycl}$	.86	19.15	−50	−14	−1	24	98	2.02	11.16	287.21	9.50	−.1052	222
$US_{ServNoCycl}$	.69	19.47	−30	−17	−1	17	116	2.90	16.89	784.05	7.60	.1068	50
$US_{Telecom}$	1.54	29.34	−76	−21	−1	20	130	1.91	10.83	262.30	25.60	.2017	249
$US_{Tech\&Elect}$	2.27	19.11	−48	−15	−1	20	100	1.82	10.92	262.63	9.13	.0061	97
$US_{UtilityCorp}$	1.90	20.37	−61	−13	0	17	123	2.85	19.13	1011.94	19.82	−.0488	335
$IRP03$	.11	.06	.01	.03	.12	.20	.28						
$Y5y$	−3.61	31.79	−70.4	−45.8	−2.85	33.9	81.1						
$Russell2$	.67	6.49	−19.49	−6.68	.97	8.43	16.42						
$CCCB$	117.82	21.75	61.4	82.6	126.4	139.1	144.7						
$Beta$	.93	.27	.03	.59	.91	1.27	2.42						
$Swsp5$	.21	8.77	−19.25	−9	−.5	13	34						
Vol	7.35	54.77	−.93	−.42	−.01	.79	417.19						

The table shows, for each series of monthly delta credit spreads, some position and dispersions measures like mean, standard deviation (S.dev), minimum and maximum values (Min, Max) and some percentiles (p10, p50, p90). Skewness (Sk) and kurtosis (Kurt) have been inserted to check normality of historical series, both summarized by the Jarque-Bera (J-B) statistical test. The Box-Pierce (B-P) statistic, with a lag of ten periods, and first-order autocorrelation co-efficient (ac(1)), test the hypothesis of no autocorrelations. The last column shows the average number of bonds to control their constituents.

3.2. Stylized Facts on IERP and DCS Co-Movements

Figure 9.1 plots the patterns of both credit spread changes and the equity implied risk premium over the entire sample period. We can recognize a break in DCSs volatility, which seems on the rise after the burst of the stock market bubble in March 2000. Another relevant finding is the rise in the implied equity risk premium after 9/11 to high levels and its decline only after the beginning of the Afghanistan and of the Iraqi war. These descriptive findings seem to confirm that the beginning of the "war on terrorism" era had significant effects on the perception of risk of financial investors. Invasions of Iraq and Afghanistan were perceived, at least at the moment of their announcement, as potential solutions to the international crisis and therefore became factors that increased investors' confidence in the future. When interpreting these findings, we must not forget that we observe a specific geographical market and therefore we are investigating the perception of risk in holding not generic assets but, specifically, U.S. stocks and bonds. Changes in our risk measure may also be driven by substitution effects and movement of investors from one geographical market to another. By observing Figure 9.1 more in detail and looking at any sharp DCS change, we identify nine relevant episodes. Economic crises (Russia, Brazil and Argentina, corresponding to points 1, 2 and 8 on Figure 9.1, respectively) and the stock market bubble burst (March 2000, point 3), economic scandals (as Enron, Worldcom and Vivendi, corresponding to points 8 and 9), terrorist attacks (U.S. embassy in Kenya and Tanzania, U.S. Navy in Yemen and 9/11, corresponding to points 1, 4 and 6, respectively) are all related to the highest positive DCS peaks. All these periods seem characterized by a parallel decrease in investors' confidence and by a rally of the implied equity risk premium. The opposite happens in points 5, 7, 10 and 11 (Presidential elections, Afghanistan War, Microsoft judgement, and Iraq War and two reductions of FED rates) where we have reductions in IRP and downward peaks of DCSs.



Legend: List of events characterizing the main DCS'peaks versus bottom in the Figure (Right scale for IRP03, left scale for DCS)
1 (August 98): Russian Crisis and terrorism attack on US Embassy in Kenya and Tanzania;
2 (November 98): Brazilian Crisis, military action in Iraq and Clinton impeachment;
3 (March 00): Asset Price Bubble;
4 (October 00): Terrorism attack on US Navy in Yemen;
5 (Nov-Dec 00; Jan 01): Bush election and Fed interest rate cut;
6 (September 01): Twin Tower;
7 (November 01): Afghanistan War and Fed interest rate cut;
8 (February 02): Argentina Crisis and Enron Scandal;
9 (July 02): Worldcom and Vivendi Scandals;
10 (November 02): Microsoft Antitrust Judgement
11 (March 03): Iraq War.
Source: Own elaboration on Thomson Financial (for IRP03) and Merrill Lynch (for DCS) data

Fig. 9.1. Monthly Patterns of Implied Risk Premium and Delta Credit Spreads Curve

It is interesting to observe that the highest peak in DCSs (corresponding to the most relevant change in perception of risk of holding corporate bonds) is in correspondence of financial scandals (July 2002) and not of September 11, while the lowest level of confidence in the future of equity investors (highest implied risk premium in the stock market) is in correspondence of the Enron scandal. These findings seem to show that reduction of confidence (or perceived higher risk) in bond and stock market are a cumulative process affected by aggregate political shocks and by more specific financial market shocks, where the latter seem to generate the highest marginal contribution (highest DCSs change), at least in bond markets.

4. EMPIRICAL FINDINGS

We test the impact on DCSs of the implied equity risk premium described in Section 5, net of the following selected set of control factors:

i) a measure of interest rate levels, represented by the return of a 5-year benchmark;
ii) a measure of stock returns which is widely representative of both large and small capitalization stocks. This why we chose the Russell 2000 stock index. As is well known, the effect of stock market variables on credit spreads is based on structural models (Merton, 1974), which illustrate how positive stock returns increase the value of firm asset and reduce the probability of failure;
iii) the Consumer Confidence Conference Board Indicator as a proxy of the business cycle. Since credit spreads measure excess risk of corporate with respect to government bonds, they are obviously expected to be negatively correlated with the business cycle (see, among others, Van Horne, 2001; Duffie and Singleton, 2003);
iv) the total volume of institutional investors sales and purchases of HY US corporate bonds, recorded by Lehman Brothers since 1998, as a measure of the signaling effects of institutional investors trades. This variable tests whether institutional investors' information not captured by other controls (i.e. stock market performance) affects DCSs;
v) the exposition to non-diversifiable risk measured by the average beta of S&P constituents, calculated as explained in Section 2;
vi) the beta of each individual stock reflects its sensitivity to aggregate shocks. The average beta is the sensitivity to aggregate shocks of constituents of our sample stocks (those present in the S&P500 from the beginning to the end of the sample period). Since this variable is a measure of levered beta, it is positively affected by changes in leverage. It therefore keeps track of the average indebtedness of stock index constituents and can also be considered a proxy for the average default risk;
vii) the spread between the 5-year swap rate and the 5-year government bond under the assumption, common in this literature, that a reduced liquidity in the swap market necessarily implies a parallel and amplified effect in the corporate bond market (Collin *et al.*, 2001).

In order to evaluate the significance of the above-mentioned variables, we perform several estimates, each of them using as a dependent variable one of the database corporate bond indexes, classified by investment grade, maturity and industry. Given the autocorrelation and non-normality of DCSs, evidenced by descriptive statistics presented in Table 9.1, we use heteroskedasticity and autocorrelation robust Newey-West (1987) standard errors. The optimal truncation lag is obtained by following the automatic method of selection suggested by Newey-West (1994). More specifically, we compare the baseline specification:

$$\begin{aligned} DCS_{i,t+1} = {} & \alpha_0 + \alpha_1 \times dY5y_t + \alpha_2 \times dCCCB_t + \alpha_3 \times dVol_t \\ & + \alpha_4 \times Russell2_t + \alpha_5 \times dSwsp5t \end{aligned} \tag{3}$$

with the modified specification in which equity implied risk premium and its associated beta are considered as additional regressors:

$$\begin{aligned} DCS_{i,t+1} = {} & \alpha_0 + \alpha_1 \times dY5y_t + \alpha_2 \times dCCCB_t + \alpha_3 \times dVol_t \\ & + \alpha_4 \times Russell2_t + \alpha_5 \times dSwsp5t \\ & + \alpha_7 \times IRP03_t + \alpha_8 \times Beta_t \end{aligned} \tag{4}$$

where $DCS_{i,t+1}$ is the monthly change in the option adjusted spread of the index bond i from period t to period $t+1$, $Y5y$ is the 5-year U.S. treasury yield, *Russell*2 is the monthly return of the *Russell*2000*index*.[8] *CCCB* is the level of the Consumer Confidence Indicator, *Vol* is the volume of professional insiders trades of High Yield U.S. bonds with Lehman Brothers, *Swsp*5 is the rate of change of the 5-year swap rate, *IRP*03 is the equity implied risk premium calculated according to the DD formula in Equation (2) under the assumption of a 3% perpetual nominal growth rate in the economy.[9]

5. OUR RESULTS

Econometric results presented in Table 9.2 document an R^2 of 47% in the baseline equation of U.S. IG corporate bonds, which does not include our implied equity risk premium. This confirms that the set of controls included in the

[8] *d* before each variable indicates that the associated variable is in first difference.

[9] Our robustness checks show that results of the paper do not change when we perform additional estimates by using alternatively 2.5% and 3.5% perpetual nominal growth rates. As expected, changes in this parameter only determine a shift of the implied risk premium without affecting its law of variation. Results are omitted for reasons of space and are available from the authors upon request.

Table 9.2 The determinants of DCSs of the U.S. IG corporate bond index

	3cuscorporate		
	(1)	(2)	(3)
*IRPR*03	60.29	89.99	
(t)	(2.74)	(4.17)	
*d*Y5*y*	−.13	−.14	−.15
(t)	(−3.61)	(−3.60)	(−4.01)
Beta	38.34		
(t)	(2.21)		
dCCCB	−.48	−.46	−.48
(t)	(−1.80)	(−1.97)	(−2.06)
dVolCap	.05	.05	.05
(t)	(6.28)	(6.86)	(6.29)
*Russ*2	−.40	−.32	−.35
(t)	(2.06)	(−1.70)	(−1.99)
*dSwsp*5	.57	.58	.40
(t)	(3.71)	(4.02)	(3.92)
const	−.29	−9.12	45.8
(t)	(−.17)	(−2.60)	(−2.85)
R^2	.47	.50	.52
N of Obs	83	83	83
F	26.94	26.06	24.77
Prob > F	0.0000	0.0000	0.0000

estimate explains a relevant part of the variability of DCSs. Signs of regressor coefficients are those expected and are consistent with previous literature findings (Huang and Kong 2003), Becchetti *et al.* (2004), etc.). More specifically, interest rate levels (*Y*5*y*), stock returns (*Russell*2), and the macro-economic indicator (*CCCB*) are negatively correlated with DCS; while swap options (*Swsp*5) and volumes of HY trades from institutional investors (*vol*) are positively correlated with it. The negative relationship between the 5-year yield and DCSs may be interpreted by considering, as pointed out by Longstaff and Schwartz (1995), that the static effect of a higher spot rate increases the risk-neutral drift of the firm value process. A higher drift reduces the incidence of default, and, in turn, DCSs. Our findings show that when the 5-year government bond return rises by 10 bps, DCSs decrease by 13 bps. The impact of stock market performance on DCSs relies directly on structural models where positive stock returns imply a rise in firm asset value and reduce default risk, by increasing the so-called

Table 9.3 Effects of the implied equity risk premium on DCSs of IG bond indexes by ratings and maturities

	AAA	AA	A	BBB
1–3	21.99	16.85	67.75	182.19
(t)	(1.61)	(1.22)	(1.63)	(2.19)
(R^2)	(.09)	(.21)	(.27)	(.25)
3–5	19.11	33.79	80.90	182.19
(t)	(1.05)	(2.14)	(4.08)	(3.24)
(R^2)	(.24)	(.41)	(.41)	(.46)
5–7	51.01	47.36	61.31	136.98
(t)	(3.45)	(2.61)	(3.43)	(3.57)
(R^2)	(.29)	(.33)	(.44)	(.54)
7–10	28.35	45.54	70.84	119.9
(t)	(1.72)	(2.10)	(3.12)	(3.67)
(R^2)	(.29)	(.44)	(.37)	(.58)
10–15	66.71	63.87	70.84	121.39
(t)	(1.49)	(1.88)	(2.21)	(3.69)
(R^2)	(.05)	(.20)	(.37)	(.43)
15+	34.16	63.13	59.24	110.54
(t)	(1.28)	(2.35)	(2.56)	(3.45)
(R^2)	(.40)	(.34)	(.47)	(.54)

Note: The table reports coefficients and T-stats of the coefficient of the implied equity risk premium variable from estimates of Equation (4) for different bond indexes. The combination of rows and columns headings indicate the corporate bond index selected as dependent variable.

"*distance-to-loss*," which is inversely related to credit spreads.[10] This explains why a 10 bps rise in stock returns reduces DCSs by 4 bps. The *CCCB* variable measures the effect of consumers' confidence and/or perception of business cycle on DCSs. A 1 point increase of this index generates a reduction of 0.5 bps in DCSs, confirming the strong relationship between boom or recessions and DCSs. The *vol* variable representing institutional investors trades of HY U.S. bonds has a statistically significant impact on DCSs, which is not too relevant in magnitude. A 1% increase of purchases generates an increase in the spread

[10] The inverse relationship between credit spreads and *distance to loss* is clearly expressed by the equation: $cs(T) = \frac{-1}{T} ln \left[N(d_2) + \frac{V_t}{Fe^{-rt}} N(d_1) \right]$

where $\frac{V_t}{Fe^{-rt}}$ is the *distance to loss*. When V_t, the asset value directly correlated to this distance, increases, the probability of default decreases, given Fe^{-rt}, which represents the present value of the debt. Indeed, the distance to loss is the inverse of financial leverage.

Table 9.4 Effects of the implied equity risk premium on DCSs of U.S. IG industry indexes

	3cuscorporate		
	Coeff	t	R^2
FinancialCorp	90.20	4.14	.48
Brokerage	46.75	2.42	.39
Banking	50.56	2.97	.47
Finance and Investment	136.45	3.51	.40
Insurance	86.32	2.46	.75
Industrial	101.14	3.50	.47
Auto	166.84	2.83	.40
Basic Industry	30.22	1.22	.38
Consumer Cyclical	96.61	2.07	.35
Capital Goods	49.07	1.92	.17
Consumer non Cyclical	64.22	2.35	.44
Energy	84.14	2.56	.51
Media	91.77	1.46	.29
Real Estate	40.95	2.03	.31
Service Cyclical	69.70	1.11	.28
Service non Cyclical	193.73	2.70	.19
Telecommunications	213.13	1.79	.21
Technology and Electronics	13.45	.17	.26
Utilities	53.42	.90	.20

Note: The table reports coefficients and T-stats of the coefficient of the implied equity risk premium variable from estimates of Equation (4) for different bond indexes. Rows headings indicate the corporate bond index selected as dependent variable.

of .05 bps. Also the swap option variable is positively correlated with DCSs (and the magnitude of the effect is such that a 10pbs change in the regressor generates a 6 bps change in the dependent variable). This is because the variable may be considered as a liquidity indicator of both swap and corporate bond market (Collin-Dufresne *et al.*, 2001) and as a proxy of the return differential required by investors for switching from a government to a generic AA corporate bond. All the above-mentioned covariates have co efficients of similar magnitude (and statistical significance) when beta and the implied equity risk premium are added to the baseline model in regressions 2 and 3. With this inclusion, R^2 rises up to 50%, and to 52% if we also add the associated beta.[11] This

[11] This confirms the best performance of the model when we add the variable beta and support our preference for the Equation 3.

suggests that our proxies for risk factors of stock market returns are relevant in explaining the variability of DCSs not captured by traditional control factors. The positive sign of the implied equity risk premium shows that the risk of holding a stock is positively correlated with the risk of holding a corporate bond. More specifically, a 1% variation in the implied equity risk premium (whose value range in the sample time interval is from 1 to 28%) generates an increase of about .9 bps in DCS. A plausible rationale is that a common risk component affects the compensation required by investors for the marginal risk run when holding a corporate bond (instead of a government bond) and the risk run when holding a common stock (instead of a fixed income asset). Based on both our descriptive and econometric findings, we argue that this component must capture the impact on DCSs of the confidence in the future of non-institutional financial investors, given that the behavior of institutional investors should be captured by our purchases/sales variable. Consider, though, that since we are measuring this variable on a specific geographical (U.S.) market, we should be cautious in drawing general conclusions. What we observe is just risk tolerance on U.S. equity and bonds. This variable may be driven by substitution effects from other markets. To compare the magnitude of the impact of the IERP on the U.S. stocks with that of stock market returns on the same market consider that, by imputing sample average parameters to our DD model in Equation (2), a 2% reduction in the risk premium generates a 10% positive stock return. Hence, the 3.5 bps impact of a 10% stock return in the augmented specification in Table 9.2 must be compared with the 1.8 bps impact of the IERP (generated by a 2% EIRP change equivalent to a 10% stock return). In the first case, the effect may be given by a combination of different factors (changes in earnings forecasts, in expected nominal perpetual rate of growth, etc.). In the second case, investors risk tolerance should be the only driver. In Table 9.2 we run the same regression by ratings and maturities for corporate bonds, finding an increase of around .9 bps for an increase of 1 point percent in the IERP. Table 9.2 shows that DCS sensitivity to variations in implied risk premium rises as ratings decreases. The relationship with maturity for a given rating class is increasing for high rating (AAA and AA) and decreasing for lower rating (A and BBB) bonds. In particular, the implied equity risk premium has its stronger influence on the BBB 1–3 year bond index, where an EIRP variation of 1% generates a change of almost 2 bps. Regressions of industry indexes show that financial, banking, industrial, consumer cyclical and service non-cyclical, are those where DCSs are most sensitive to the implied equity risk premium.

6. CONCLUSIONS

After the optimism of the post Berlin-Wall era, peaking in the excess of confidence in financial markets which led to the burst of the bubble in March 2000, a new era of reduced optimism induced by financial scandals and by the war on terrorism war has begun. In this chapter we show that we can track these waves of pessimism and optimism, by extracting the implied equity risk premium from standard DD asset pricing formulas. In this way we may follow the evolution of financial investors' attitude to bear the risk of holding assets whose value is determined by expected future cash flow realizations. This is because the IERP captures investors overall confidence in the future coupled, of course, with their specific confidence in the future value of their financial asset. Such confidence is affected by components such as corporate governance rules and capacity of enforcing them. Our empirical findings clearly show that major economic and political shocks affect the implied equity risk premium and that the latter variable has significant impact on DCSs. We therefore conclude that a common risk tolerance component affects changes in investors perceived risk of holding stock (with respect to corporate bonds) and corporate (with respect to government) bonds.

BIBLIOGRAPHY

Adriani, F., Bagella, M. and Becchetti, L. (2004) "Observed and fundamental price earning ratios: A comparative analysis of high-tech stock evaluation in the U.S. and in Europe." *Journal of International Money and Finance*.

Bagella, M., Becchetti, L. and Ciciretti, R. (2004) *Market vs analysts' reaction: the effect of aggregate and firm specific news.* Working Paper, 2004.

Becchetti, L and Carpentieri, A. (2004) *The determinants of option adjusted delta credit spreads. a comparative analysis on U.S., U.K. and Eurozone*. Working Paper, 2004.

Claus, J. and Thomas, J. (2001) "Equity premia as low as three percent? Evidence from analysts earning forecast for domestic and international stock markets." *Journal of Finance* **56**: 1629–66.

Collin-Dufresne, P., Goldstein, R.S. and Spencer Martin, J. (2001) "The determinant of credit spread changes." *The Journal of Finance* **56**: 2177–203.

Cornell, B. and Green, K. (1991) "The investment performance of low-grade bond funds." *Journal of Finance* **46**: 29–48.

Duffie, D. and Singleton, K. (1999) "Modeling term structure of faultable bonds." *Review of Financial Studies* **12**: 687–720.

Duffie, D. and Lando, D. (2001) "Term structure of credit spreads with incomplete accounting information." *Econometrica* **69**: 633–64.

Duffie, D. and Singleton, K.J. (2003) *Credit Risk:pricing measurement and management*. Princeton: Princeton University Press.

Elton, E.J., Gruber, M.J., Agrawal, D. and Mann, C. (2001) "Explaining the rate spread on corporate bonds." *Journal of Finance* **56**: 247–77.

Fama, E. and French, K. (1989) "Business conditions and expected returns on stocks and bonds." *Journal of Financial Economics* **25**: 23–49.

Fama, E. and French, K. (1993) "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* **33**: 3–56.

Fama, E. and French, K. (1995) "Size and book-to-market factors in earnings and returns." *Journal of Finance* **50**: 1–55.

Huang, J. and Kong, W. (2003) "Explaining credit spread changes: Some new evidence from option-adjusted spreads of bond indexes." *Working Paper Series, Pennsylvania State University*, 2003.

Jagannathan, R., McGrattan, E.R. and Scherbina, A. (2001) "The declining U.S. equity premium." *Quarterly Review Federal Reserve Bank of Minneapolis*.

Jarrow, R. and Turnbull, S. (1995) "Pricing options on financial securities subject to default risk." *Journal of Finance* **50**: 53–85.

Jarrow, R., Lando, D. and Turnbull, S.A. (1997) "Markov model for the term structure of credit spreads." *Review of Financial Studies* **10**: 481–523.

Kwan, S.H. (1996). "Firm-specific information and the correlation between individual stocks and bonds." *Journal of Financial Economics* **40**: 63–80.

Leland, H. and Toft, H.R. (1996) "Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads." *Journal of Finance* **51**: 987–1019.

Longstaff, F. and Schwartz, E. (1995) "A simple approach to valuting risky fixed and floating rate debt." *Journal of Finance* **50**: 789–819.

Merton, R.C. (1974). "On the pricing of corporate debt: the risk structure of interest rates." *Journal of Finance* **29**: 449–69.

Pedrosa, M. and Roll, R. (1998). "Systematic risk in corporate bond credit spreads." *Journal of Fixed Income*.

Sarig, O. and Warga, A. (1989) "Some empirical estimates of the risk structure of interest rates. *Journal of Finance* 1351–60.

Van Horne, J. (2001) *Financial Market Rates and Flows*. 6th edn. Englewood Cliffs, NJ: Prentice-Hall.

Zhou, C. (1997) "A jump-diffusion approach modeling credit risk and valuing defaultable securities." *Finance and Economics Discussion Series Board of Governors Federal Reserve System*, **15**.

This Page is Intentionally Left Blank

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 10

ASSET AND DEFAULT CORRELATIONS BETWEEN NON-FINANCIAL CORPORATIONS: EVIDENCE FROM THE ITALIAN STOCK MARKET

Cristiano Zazzara* and Zeno Rotondi

Abstract

This chapter deals with the problem of empirically calibrating asset and default correlations for a portfolio of loans to non-financial Italian public and private firms.

Following the approach proposed by Credimetrics™ on a sample of 130 Italian listed firms, first we estimate the "systematic" portion of return variation of each company. This "systematic" component is expressed by the R^2 coefficient derived from regressing individual stock returns on the returns of a proper industry index. Second, we estimate for the same sample the relationship between the estimated R^2 coefficients and two proxies of company size, the book asset value and the turnover. Our results show that an accounting rule based on turnover may be used to derive correlation estimates for non-listed companies (and usually smaller firms), which constitute the bulk of banks' loan portfolios. In sum, our analysis suggests that stock market data could be used to calibrate internal credit risk models even for Italian non-listed firms, yielding a model that is both reliable and dynamic.

Keywords: Asset and Default Correlations, Concentration Risk, Asset Value Model, Basel 2.

JEL classification No.: G11, G21, G28

*Corresponding author

1. INTRODUCTION

The origins of quantitative credit risk modeling at the portfolio level can be traced back to 1997, when J.P. Morgan (in conjunction with several co-sponsors) launched CreditMetrics™, and Credit Suisse First Boston (1997) released its CreditRisk+™[1]. Together with KMV's Portfolio Manager™ and McKinsey's CreditPortfolioView™, these portfolio models still form the cornerstone of current industry practice[2] and their underlying concepts are at the center of ongoing theoretical debate. CreditMetrics™ and the KMV model are both asset value approaches, sometimes referred to as "structural" – in contrast to "reduced form" – models. Their common notion that default is triggered, if the market value of the firm's total assets falls below some critical threshold level, derives from the seminal work of Merton (1974).

Among the major challenges of these credit risk models is the issue of modeling the joint default behavior and the correlation structure in a portfolio of fixed-income securities, for example, corporate bonds or loans. Whereas a probability of default (PD) is comparatively easy to estimate for a single obligor firm, it is almost impossible to directly estimate probabilities for the joint default events in a loan portfolio comprising several hundreds or thousands of companies. In fact, given the relative infrequency of default events, even for small real-world portfolios there is no way to estimating joint default probabilities and correlations from loss experience data.

This chapter deals with the problem of empirically calibrating the asset and default correlation coefficients for a portfolio of loans to non-financial Italian public and private firms. To determine the correlation structure under the asset value framework, we need to estimate the portion of asset return volatility that is firm-specific (idiosyncratic). From a practitioner perspective, estimating this weight is important, probably being the most sensitive calibration problem in credit portfolio modeling. Setting the percentage portion of idiosyncratic risk has a tremendous impact on the resulting loss distribution, especially on its lower tail. If the asset value approach to estimate correlations is feasible for public firms, then difficulties arise in estimating the idiosyncratic risk for private firms, so need to be modeled differently. Even though this issue is of utmost importance for banks, since most of them are heavily involved in granting loans to private (non public-traded) firms, there is an almost total lack of academic literature on this latter issue. To our knowledge, there are only two empirical studies on this

[1] See Gupton *et al*. (1997) and Credit Suisse First Boston (1997).
[2] See Basel Committee on Banking Supervision (1999) and also Saunders and Allen (2002).

subject, one by Xiao (2002) on the U.S. market, and the other conducted by Hahnenstein (2004) on the German market.

This chapter is organized as follows: In Section 2, we briefly review the basics of the asset value model approach, describing the correlation methodology in a default mode setting similarly to that proposed by Gordy (2000). In Section 3, we concentrate on the asset return correlations. First, we describe a simplified version of the Index Model to estimate correlations for listed companies. Second, turning to the case of non-listed firms, we briefly review the CreditMetrics™ (Xiao, 2002) approach to the estimation of a non-listed company's systematic risk, which consists of a "general rule" that relates the weight of the idiosyncratic component in the index model to company size as measured by the firm's asset value. This latter case is our main concern, since in Italy (and in Europe as well) the banks' loan portfolios are mostly composed of private medium-sized enterprises. In Section 4, we provide an empirical analysis for the estimations of asset and default correlation coefficients, based on a two-step regression methodology. The data in our sample (190 weekly stock returns for 130 listed Italian companies) and our results are presented in sub-sections 4.1, 4.2, and 4.3. We use both the book asset value and the turnover as proxies for company size and analyze under what conditions there is an empirically valid relationship between systematic risk and these two variables. To conclude the empirical analysis, in sub-section 4.4 we propose a goodness test of the estimated asset return correlation coefficients, comparing these latter with the correlation coefficients derived directly from equity prices. In Section 5, we summarize our findings and consider issues for future research. Finally, in the Appendix, we show how default correlation coefficients may be used to build a simple index of credit risk concentration under a mean-standard deviation approach to credit portfolio risk.

2. THE ASSET VALUE METHODOLOGY FOR PORTFOLIO MODELING

Consider a bank's loan portfolio with n different corporate obligors. Each obligor firm i is characterized by its probability of default PD_i, which can be regarded as inferred from the bank's rating systems. We further assume that each company's asset return r_i follows a standard normal distribution. The asset return, which is in fact a latent variable, can be regarded as describing the annual percentage change in the market value of the firm's total assets. It is a one-period measure of the overall corporate business performance. The standard normal distribution is characterized by its probability density function ϕ or its cumulative probability density function Φ. Company i defaults, if and only if its realized asset return r_i falls below the critical level z_i, the so-called default threshold:

$$PD_i = \int_{-\infty}^{z_i} \phi(r_i)dr_i = \Phi(z_i) \tag{1}$$

where $z_i = \Phi^{-1}(PD_i)$.

Obviously, the "cut-off" return z_i is a function of the company's PD, which may, in turn, be derived from the firm's rating class.

For ease of exposition, we now turn to the case of a simple two-obligor portfolio. The random asset returns r_i and r_j are assumed to be drawn from a bivariate standard normal distribution with a joint density function and a known correlation coefficient $\rho_{i,j}$. This latter is equal to

$$\rho_{i,j} = \frac{Cov(r_i, r_j)}{\sqrt{Var(r_i)}\sqrt{Var(r_j)}} \tag{2}$$

and represents the *asset return correlation* between companies *i* and *j*.

Therefore, the probability that both obligors *i* and *j* default jointly, denoted by $PD_{i,j}$, is calculated using the default thresholds z_i and z_j, which result from PD_i and PD_j via Equation [1], together with the bivariate asset return density function[3]:

$$PD_{i,j} = \int_{-\infty}^{\Phi^{-1}(PD_i)} \int_{-\infty}^{\Phi^{-1}(PD_j)} \frac{1}{2\pi\sqrt{1-\rho_{i,j}^2}} \times \exp\left[-\frac{1}{2(1-\rho_{i,j}^2)}\left(r_i^2 + r_j^2 - 2\rho_{i,j}r_ir_j\right)\right] dr_i dr_j \tag{3}$$

According to Li (2000), the above bivariate standard normal distribution function $PD_{i,j}$ can be interpreted as the CreditMetrics copula function. Then, the default correlation coefficient $\tilde{\rho}_{i,j}$ can be expressed as

$$\tilde{\rho}_{i,j} = \frac{PD_{i,j} - PD_iPD_j}{\sqrt{PD_i(1-PD_i)}\sqrt{PD_j(1-PD_j)}} \tag{4}$$

Hence, the probability distribution of the potential losses in the loan portfolio is characterized completely by the individual PDs of the obligors together with their

[3] See Gupton *et al*. (1997) for details.

asset return correlation. This concept, which can be generalized from our simple two-obligor illustration to the case of an *n* obligor loan portfolio, forms the core of the CreditMetrics™ and KMV's Portfolio Manager™ asset correlation approach to the modeling of joint default events.

3. THE ESTIMATION OF ASSET RETURN CORRELATIONS

Focusing on the CreditMetrics™ model, we now turn to the description of the methodology to estimate the asset return correlations – as per the Equation [2] above – for public and private companies respectively.

3.1. The Index Model approach to estimate Asset Correlations for Public Companies

Determining the loss distribution for a portfolio of *n* obligors, as described above, requires empirical estimates of the $n \times (n-1)/2$ pairwise asset correlations. In order to reduce data requirements and simplify the parameter estimation, the CreditMetrics™ methodology deduces estimates of the obligors' individual asset correlations from stock indices by means of a factor model.[4] In the general CreditMetrics™ approach, both country and industry weights are assigned to each obligor according to its participation. For our purposes, we make the following two simplifying assumptions. First, we ignore potential calibration problems arising from cross-country diversification. Since our focus is on a portfolio of Italian corporate obligors, all country weights can simply be set to 100% for Italy, so that all other countries are ignored. Hence, the degree of concentration in such a purely national loan portfolio is driven mainly by the companies' industry composition. Therefore, mapping each firm $i = 1 \ldots n$ to its affiliated industry $k(i) = 1 \ldots n$ (where $m < n$), the Index Model can be expressed in its simplified version as

$$r_i = \sqrt{w_i \times R_{k(i)}} + \sqrt{1 - w_i \times \varepsilon_i} \tag{5}$$

[4] See Gupton *et al*. (1997). The problem of using *equity* correlations as a proxy for *asset* correlations, which has already been recognized as a potential drawback by the model's inventors themselves and which has recently been attacked on theoretical and empirical grounds by Zeng and Zhang (2002) of KMV, does not form the focus of our work.

where:

- r_i is the asset return of firm *i*;
- $R_{k(i)}$ is the return of the industry index k to which company i belongs;
- w_ie $(1 - w_i)$ represent the weights assigned to the industry and to the firm-specific influence on asset returns respectively. The greater the w_i, the closer the firm tracks its industry performance and the less it moves independently of its industry associates.
- ε_i is the company-specific noise term representing the idiosyncratic movements in asset returns.

In Equation [5] it is assumed that r_i, $R_{k(i)}$, ε_i are standard normal distributed as N(0,1), and that $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ as well as $\mathrm{Cov}(\varepsilon_i, R_{k(j)}) = 0$. These latter assumptions indicate that each obligor's noise term is uncorrelated with the noise terms of all other firms and it is also uncorrelated with the movements that affect the industry as a whole (therefore fully reflected in the respective index return).

The Index Model represented in Equation [5] enables a straightforward calculation of pairwise asset correlations. For two obligors *i* and *j* belonging not necessarily to different industries k(*i*), k(*j*), using Equation [2] we get

$$\begin{aligned}\rho_{i,j} &= \mathrm{Cov}(r_i, r_j) = E(r_i \times r_j) - E(r_i) \times E(r_j) = \sqrt{w_i \times w_j} \times E(R_{k(i)} \times R_{k(j)}) \\ &= \sqrt{w_i \times w_j} \times \rho_{k(i),k(j)} \end{aligned} \tag{6}$$

where
$\rho_{k(i),k(j)}$ = linear correlations between the k (from 1 to m) industry indices.

Whenever all the bank's obligors in the credit portfolio under consideration are companies listed on a stock exchange, individual estimates of the weights w_i can be derived from the coefficient of determination (R-squared) of the time-series regression model of Equation [5][5]:

$$w_i = R_i^2 \tag{7}$$

In the case of non-publicly traded obligor firms, typical for many medium-sized enterprises in Italy, the regression model described above cannot be fitted, because of a lack of stock price data. In the next section, we analyze if and how an alternative approach can offer a reasonable solution to this case.

---

[5] For a derivation of this result see Hahnenstein (2004).

3.2. The Accounting Rule for Private Companies

The notion that company size is an important driver of systematic risk lies at the heart of the approach already implemented in the CreditMetrics™ model. In general, systematic risk can be considered to be a function of company size. Larger companies have relatively large systematic risk because their behaviour tends to be similar to that of the overall market (often they are components of market benchmarks). Smaller companies can have smaller systematic risk, since they are more likely to behave independently of broad market trends and are less likely to be index components.

As a potential solution to this problem for private firms, the CreditMetrics™ model (Xiao, 2002) proposes an empirical relationship between R-squared and company size, based on the following logistic form[6]:

$$w_i = R^2 = \frac{1}{1 + S^{-\gamma} \times e^{-\lambda}} \tag{8}$$

where S is the size variable of a company (which may assume, for example, the form of Total Assets, Sales or Turnover). Equation [8] can be transformed into

$$\mathrm{Log}\left(\frac{1}{R^2} - 1\right) = -\gamma \times \mathrm{Log}(S) - \lambda \tag{9}$$

where the estimation of γ and λ can be obtained by means of a standard linear regression of $\mathrm{Log}\left(\frac{1}{R^2} - 1\right)$ over Log(S). This relationship can be therefore considered a "general rule" that gives an estimate of the overall weight of the systematic component in asset returns for each obligor depending on its size. The rule is simple since the only input data required to calculate this weight is the size variable of the company. According to Xiao (2002), using daily data of 741 NYSE-traded stocks from January 1, 1999 to August 2002, the above parameters of the Accounting Rule using Total Assets were found to be $\gamma = 0.5105$ and $\lambda = -12.5832$. This is the current calibration adopted in the CreditMetrics™ model for the US market.

In the next section we will empirically validate this rule for the Italian market, evaluating two firm size variables, such as the book value of Total Assets and the Turnover.

---

[6] The logistic regression is a common nonlinear functional form used when the dependent variable varies between 0 and 1.

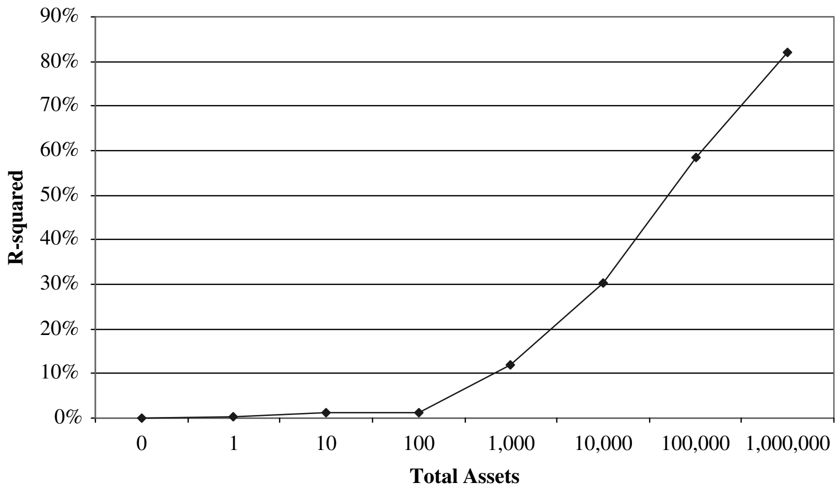


Fig. 10.1. Relationship between the book value of the firm's total assets and the portion of its systematic risk (R-squared)

Source: Our elaborations on estimates by Xiao (2002)

4. THE EMPIRICAL ANALYSIS: ASSET RETURN CORRELATIONS AND DEFAULT CORRELATIONS FOR ITALIAN NON-FINANCIAL COMPANIES

In this section we derive asset return and default correlation matrices between Italian non-financial companies, under both the approaches based on the Index Model and the Accounting Rule. To accomplish this goal, we estimate the R-squared for each company in our sample with these two approaches. Particularly, we first calibrate the weights of the systematic component (the R-squared) using the Index model reported in Equation [5], through a set of univariate OLS time-series regressions. Second, we run cross-sectional regressions between the estimated R-squared and the two size variables, Total Assets and Turnover. This two-step OLS regression approach was introduced into the field of capital market research by Fama and MacBeth (1973) and, starting with Chen *et al.* (1986), was used in a number of empirical studies dealing with the derivation of risk premia from beta coefficients in the context of the Arbitrage Pricing Theory (APT).

4.1. The Sample Data

Our sample consists of 130 shares that were listed on the Milan Stock Exchange on September 17, 2004. These data were provided by Bloomberg™ and include the stocks' ISIN codes, company names, market capitalizations, industry affiliations and weights in the respective 14 industry indices. For these 130 ISIN codes, we downloaded weekly Friday fixing prices for the period from February 2, 2001 to September 17, 2004 (190 weeks of data) and the book value of total assets and the turnover as of December 31, 2003. Further, in order to estimate the coefficient of determination (R-squared) for each of our companies according to the Index Model (Equation [5]), we also downloaded from Bloomberg™ the time series of the 14 indices for the 190 weeks selected.

In the first step of our data selection, we removed all stocks belonging to the three industry groups of banking, insurance and financial services from the sample, because our aim was to obtain a calibration for a portfolio of corporates, not for financial intermediaries with rather atypical balance sheet characteristics. Second, we eliminated all stocks with missing price entries (e.g., because of a delisting) and with a short time-series of information (recent listing).

Summary data for all the firms included in our sample are given in Table 10.1.

The market capitalization in our sample varies between 5 (Filatura di Pollone) and 73,641 mln. (ENI), with an average of 2,304 mln per company. The book value of total assets ranges from 17 (Schiapparelli 1824) to 80,497 mln. (Telecom Italia), with an average of 3,585 mln. per company. Finally, the turnover varies between 6 (Bonifica Ferraresi and Imprese Agricole) and 51,487 (ENI) mln., with an average of 2,044 per company.

Our sample covers all targeted 14 industry indices of the Milan Stock Exchange and represents almost the total market capitalization of all listed companies in these industries.

4.2. The Estimation of the R^2 and its relation with the size variables

We apply the Index Model outlined in Equation [5] to estimate the coefficient of determination (the R^2) for each company included in our sample, which is just the goodness of fit between the firm return and the index return. From the time-series regressions we get coefficients of R^2 ranging from a minimum of 0.03% (Acquedotto Nicolay) to a maximum of 91.99% (ENI), with an average of 24.01% (see Table 10.2).

The most cyclical sectors appear to be those of Publishing and Printing (49.3%), Mining, Steel, Metals (41.3%), Automobiles (36.5%), and Transportation & Tourism (30.7%), while among the least correlated with the business

Table 10.1 Sample composition per industry sector: number of firms, market capitalizations, book values of total assets and turnover (data in millions of euro)

Industry Sector	No. of firms	Market Cap.	Tot. Assets	Turnover
Automobile	6	6,616	65,313	51,349
Food	6	1,654	2,844	3,254
Building	10	6,716	16,976	9,042
Paper	1	191	706	542
Chemicals	14	2,997	5,444	3,977
Electrical	21	22,629	48,364	21,784
Plant & Machinery	7	5,309	11,495	7,465
Miscellanous Industrials	8	8,968	6,155	4,382
Mining, Steel, Metals	3	74,765	70,476	56,626
Textile	16	6,101	8,664	7,263
Retailers	2	769	1,846	1,582
Publishing and Printing	8	16,767	10,176	7,764
Public Utility	16	128,632	189,196	78,747
Transportation & Tourism	12	17,417	28,339	11,934
	130	299,532	465,995	265,709
	Min	5	17	6
	Average	2,304	3,585	2,044
	Max	73,641	80,497	51,487

Source: Elaborations on data from Bloomberg™

Table 10.2 Average R-squared per industry sector and average weight in the industry index composition

Industry Sector	Average R^2%	Average Weight in the Index %
Automobile	36.5	16.7
Food	8.3	16.7
Building	25.7	10.0
Paper	91.9	100.0
Chemicals	13.6	7.1
Electrical	22.4	4.8
Plant & Machinery	21.1	14.3
Miscellanous Industrials	14.2	12.5
Mining. Steel. Metals	41.3	33.3
Textile	17.9	6.3
Retailers	33.1	50.0
Publishing and Printing	49.3	10.9
Public Utility	21.4	6.0
Transportation & Tourism	30.7	8.3
Min	0.03	0.0
Average	24.0	10.6
Max	91.99	100.0

Source: Elaborations on data from Bloomberg™ using Equation [5] cycle emerge those of Food (8.3%) and Chemicals (13.6%). Regardless of the high value of the R-squared, we do not consider the Paper (91.9%) and Retailers (33.1%) as very cyclical sectors, since the average weights of the firms' capitalization in the industry sector index composition (the ratio between each firm's market capitalization and the total industry sector capitalization) are 100% and 50% respectively. Therefore, the higher the value of the average weight in the index, the more "suspect" is the estimation of the R-squared and the resulting correlation coefficient.[7]

We now estimate the relation between the above R-squared and two size variables (Total Asset Value and Turnover), according to the model proposed in Equations [8] and [9]. Regressing all the 130 R-squared values on our two measures of company size leads to the following estimates of parameter values of the logistic function.

As can be seen from Table 10.3, both size variables undoubtedly explain a significant portion of the cross-sectional variation in the "true" R-squared values. In the two linear cross-sectional regressions, all coefficient signs are highly significant at above the 99% level, and our estimates for the parameters y and λ are similar to those obtained in the above-mentioned Xiao (2002) study,[8] despite the many differences in the data used.

As is observed, the difference caused by the change of proxy (independent variable) in the estimation seems relevant. The turnover, in fact, shows a superior statistical significance compared to the total assets variable in explaining the firm's systematic risk. This result is also in line with the new Basel 2 regulatory framework, which considers the turnover as a criterion to calibrate the asset

Table 10.3 OLS Cross-sectional regression summary statistics for the complete sample of 130 firms (dependent variable R^2 and independent variables Total Assets and Turnover respectively)

Statistical Relation	Y			λ			R-squared of the regression
	Coeff.	t-statistics	Signific.	Coeff.	t-statistics	Signific.	
R^2 vs. Assets	0.57	8.06	0.00%	−12.99	−9.18	0.00%	33.69%
R^2 vs. Turnover	0.61	8.95	0.00%	−13.51	−10.12	0.00%	38.48%

Source: Estimations using Equation [9].

[7] In the following analysis, we will eliminate the Paper sector for this reason. The asset return correlation will, in fact, result in a coefficient of 100% only because of the presence of 1 firm in this industry sector.

[8] Recall this paper lies at the heart of the current calibration of the CreditMetrics model.

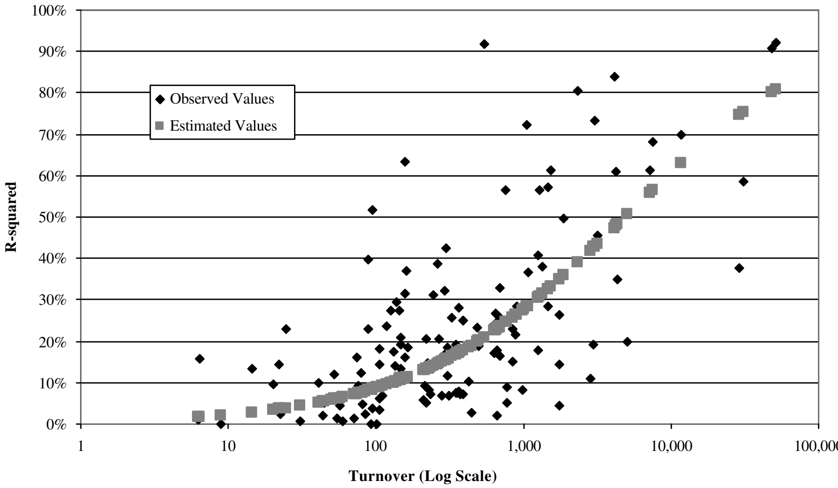


Fig. 10.2. Scatter plot of the coefficients of determination (R-squared) against turnovers, and of the estimated logistic function

Source: Our elaborations on Equation [5]

return correlation coefficients under the new capital adequacy rules (for more details, see Basel Committee on Banking Supervision, 2004). Before proceeding with further analysis, we summarize our results by plotting the 130 calculated R-squares against the Turnover, and also by drawing the logistic function based on the just estimated parameters of y and λ for the Italian market (see Figure 10.2).

The Figure 10.2 clearly reveals that the relation between the R-squared and the firm's turnover may be approximated by a logistic function of the type given in Equation [5].

We could be tempted to say that, based on these results, the calibration obtained should work well with Italian non-financial companies.[9] However, in order to use this relationship with statistical confidence for the universe of medium-small companies – which form the bulk of banks' loan portfolios – we should test the robustness of this estimation at least in one direction: the size of the turnover.

[9] We also investigated the stability of the parameters of the accounting rule in order to evaluate whether this rule should be updated periodically. Dividing the sample period in two equal parts and running a Chow test, we found evidence to reject the null hypothesis of equality of regression parameters in the two sub-samples. Therefore, we conclude that a periodic update of the rule is advisable (for the sake of brevity, these results have been omitted from the text but are available upon request from the authors).

Therefore, in the next section we test the hypothesis that this Accounting Rule, based on the logistic function of Equation [5], fits well the data even for a sub-sample of smaller companies.

4.2.1. A Robustness Test: the empirical validity of the Accounting Rule for medium-small companies

In order to further investigate the empirical validity of the above estimated Accounting Rule, we split our dataset in two sub-samples, each composed of 65 companies according to decreasing market capitalization, and compare the results. Detailed information about the two sub-samples of large- (the highest 65 market capitalizations) and medium-small companies (the lowest 65 market capitalizations) is reported in Table 10.4.

As can be seen from this table, the average index weight in the second sub-sample is only 3%, with the maximum weight of a single company in its industry index being only 6%. Therefore, we can reasonably assume that the effect that leads to a higher R-squared estimate for a stock, simply because it has a higher index weight, is successfully filtered out in the second sub-sample.

Moreover, the average turnover of about 182 mln. for the companies in the second sub-sample (in contrast to 3,906 mln. in the first one) indicates that this sub-sample is far better suited to infer calibration results with respect to medium-small sized companies.

Running the regressions on the two sub-samples, according to Equation [9], yields the result as in Table 10.5.

As shown in Table 10.5, in both sub-samples the turnover variable performs better than total assets and the estimated coefficients have the expected signs

Table 10.4 Summary descriptive statistics for the two sub-samples of large and medium-small companies (Market Cap. and Turnover values in € mln.)

Sample / Variable	Sub-Sample 1 (Large companies)			Sub-Sample 2 (Medium-Small companies)		
	Min	Average	Max	Min	Average	Max
Market Cap.	223	4,526	73,641	5	82	216
Turnover	126	3,906	51,487	6	182	844
Index Weight	0%	19%	98%	0%	3%	6%
R^2	2%	33%	92%	0%	15%	52%

Source: Elaborations on data from Bloomberg™

Table 10.5 OLS Cross-sectional regression summary statistics for the two sub-sample of 65 firms (dependent variable R^2 and independent variables Total Assets and Turnover respectively)

Large Companies							
Statistical Relation	Y			λ			R-squared of the regression
	Coeff.	t-statistics	Signific.	Coeff.	t-statistics	Signific.	
R^2 vs. Assets	0.50	5.14	0.00%	−11.64	−5.57	0.00%	29.53%
R^2 vs.Turnover	0.54	5.55	0.00%	−12.18	−6.00	0.00%	32.87%
Medium-Small Companies							
Statistical Relation	Y			λ			R-squared of the regression
	Coeff.	t-statistics	Signific.	Coeff.	t-statistics	Signific.	
R^2 vs. Assets	0.67	3.21	0.21%	−14.89	−3.79	0.03%	14.05%
R^2 vs. Turnover	0.70	4.15	0.01%	−15.21	−4.88	0.00%	21.49%

Source: Estimations using Equation [9]

and are highly significant, indicating that R-squared decreases with a decreasing value of the size variables. However, while the results for the first sub-sample that contains the large companies are in line with our results for the complete sample (see Table 10.3), the results for the second sub-sample are different. For the latter, the exclusion of the firms with high index weights removes part of the explanatory power from the cross-sectional regression, but even for this sample of smaller firms the turnover variable performs well (it explains 21.49% of the cross-sectional variation in the R-squared values). Hence, the robustness of these calibration results confirms the positive relationship between a company's turnover and its systematic risk. Further, our results suggest that this empirical relationship may be applied to non-listed medium-small firms with a certain degree of statistical confidence.

4.3. The derivation of the asset and default correlation coefficients

Our final step involves the estimation of the asset return correlation and default correlation matrices between Italian non-financial companies, according to the model presented in Sections 2 and 3. First, we use Equation [6] to derive the asset return correlation coefficients under both the Index Model (IM) and the Accounting Rule (AR) based on the turnover. Next, by aggregating companies per industry sector, we obtain the correlation matrices reported in Tables 10.6 and 10.7.

Table 10.6 Asset Return Correlation Matrix through the Index Model (CreditMetrics™ Methodology for public firms)

	Food	Automobile	Building	Chemicals	Retailers	Electrical	Plant & Machinery	Miscellaneous Industrials	Publishing and Printing	Mining, Steel, Metals	Public Utility	Textile	Transportation & Tourism
Food	7.1%												
Automobile	4.3%	32.9%											
Building	5.2%	16.2%	19.7%										
Chemicals	2.4%	11.4%	9.6%	12.4%									
Retailers	5.6%	15.9%	14.5%	10.1%	32.3%								
Electrical	2.7%	14.9%	11.9%	11.9%	12.5%	19.5%							
Plant & Machinery	3.5%	13.2%	12.4%	6.7%	8.7%	8.9%	15.2%						
Miscellaneous Industrials	0.6%	5.6%	4.9%	3.9%	5.2%	5.0%	3.5%	9.3%					
Publishing and Printing	4.2%	25.0%	19.8%	14.3%	17.5%	21.6%	14.1%	9.2%	45.2%				
Mining, Steel, Metals	3.3%	16.3%	12.9%	8.2%	10.3%	9.6%	16.6%	4.6%	15.8%	305%			
Public Utility	2.6%	12.9%	9.3%	9.1%	10.7%	14.1%	7.7%	4.4%	19.2%	11.7%	15.2%		
Textile	3.3%	12.8%	11.9%	7.7%	10.0%	10.4%	8.3%	5.0%	17.7%	11.0%	8.5%	14.6%	
Transportation & Tourism	5.0%	18.3%	16.3%	11.1%	16.6%	14.8%	13.8%	7.3%	23.2%	16.7%	13.1%	10.8%	27.7%

Source: Our elaborations using equations [6]

Table 10.7 Asset Return Correlation Matrix through the Accounting Rule (CreditMetrics™ Methodology for private firms)

	Food	Automobile	Building	Chemicals	Retailers	Electrical	Plant & Machinery	Miscellaneous Industrials	Publishing and Printing	Mining, Steel, Metals	Public Utility	Textile	Transportation & Tourism
Food	14.1%												
Automobile	5.6%	28.1%											
Building	7.0%	14.1%	17.8%										
Chemicals	3.3%	10.5%	8.9%	12.3%									
Retailers	6.6%	12.3%	11.5%	8.4%	22.3%								
Electrical	3.4%	12.0%	9.8%	10.3%	9.1%	14.7%							
Plant & Machinery	5.9%	14.3%	13.7%	7.8%	8.6%	9.0%	21.8%						
Miscellaneous Industrials	1.0%	6.4%	5.6%	4.8%	5.3%	5.3%	5.1%	14.7%					
Publishing and Printing	4.1%	16.0%	12.9%	9.8%	10.1%	13.0%	11.4%	7.7%	21.6%				
Mining, Steel, Metals	5.0%	15.9%	12.9%	8.7%	9.1%	8.8%	20.6%	6.0%	11.6%	34.5%			
Public Utility	4.4%	14.3%	10.5%	10.9%	10.7%	14.8%	10.9%	6.4%	15.9%	14.9%	22.1%		
Textile	4.5%	11.6%	11.0%	7.5%	8.2%	8.8%	9.5%	5.9%	11.9%	11.4%	10.0%	14.0%	
Transportation & Tourism	5.7%	13.7%	12.4%	8.9%	11.1%	10.3%	13.0%	7.2%	12.9%	14.2%	12.7%	8.5%	17.9%

Source: Our elaborations using equations [6] and [8]

Results show that, on average, asset correlation coefficients are around 12% and 10% for the Index Model and the Accounting Rule respectively. In terms of asset return correlation, the most cyclical sectors are Automobiles (IM: 32.9%, AR: 28.1%), Retailers (IM: 32.3%, AR: 22.3%), and Mining, Steel, Metals (IM: 30.5%, AR: 34.5%), while the least dependent on the economic cycle are Food (IM: 7.1%, AR: 14.1%) and Chemicals (IM: 12.4%, AR: 12.3%).

Second, in order to derive the default correlation matrices for the same companies, we use Equations [3] and [4] with an estimated probability of default for each company based on the KMV's CreditMonitor™ Model[10] (see Table 10.8).

Also we aggregate companies per industry sector obtaining the correlation matrices reported in Tables 10.9 and 10.10.

On average, the default correlation coefficients lie between 0.81% and 1.03% for the Accounting Rule and the Index Model respectively, yielding a ratio of asset return to default correlations of approximately 12 to 1. This result is

Table 10.8 Average probabilities of default per industry sector for the overall sample of 130 listed non-financial companies (data as of September 17, 2004)

Industry Sector	Average Probability of Default
Automobile	0.59%
Food	0.80%
Building	0.48%
Paper	0.72%
Chemicals	2.44%
Electrical	0.82%
Plant & Machinery	1.49%
Miscellanous Industrials	3.57%
Mining, Steel, Metals	0.03%
Textile	3.02%
Retailers	2.25%
Publishing and Printing	1.35%
Public Utility	0.55%
Transportation & Tourism	1.03%

Source: Elaborations on Moody's KMV CreditMonitor™ Model

[10] For methodological details, see Crosbie (1999).

Table 10.9 Default Correlation Matrix derived from the Index Model

	Food	Automobile	Building	Chemicals	Retailers	Electrical	Plant & Machinery	Miscellaneous Industrials	Publishing and Printing	Mining, Steel, Metals	Public Utility	Textile	Transportation & Tourism
Food	0.4%												
Automobile	0.2%	4.5%											
Building	0.2%	1.0%	1.3%										
Chemicals	0.1%	0.9%	0.5%	1.0%									
Retailers	0.5%	1.9%	1.2%	1.1%	7.3%								
Electrical	0.1%	1.3%	0.6%	0.9%	1.4%	2.0%							
Plant & Machinery	0.2%	1.2%	0.8%	0.5%	0.9%	0.7%	1.5%						
Miscellaneous Industrials	0.0%	0.3%	0.1%	0.2%	0.4%	0.2%	0.2%	0.4%					
Publishing and Printing	0.2%	3.0%	1.5%	1.3%	2.4%	2.4%	1.3%	0.5%	9.4%				
Mining, Steel, Metals	0.0%	0.5%	0.3%	0.2%	0.3%	0.2%	0.8%	0.1%	0.5%	1.0%			
Public Utility	0.1%	0.9%	0.4%	0.6%	0.9%	1.1%	0.5%	0.2%	1.8%	0.3%	1.2%		
Textile	0.2%	1.2%	0.8%	0.7%	1.3%	0.9%	0.7%	0.3%	2.1%	0.3%	0.6%	1.8%	
Transportation & Tourism	0.2%	1.3%	0.8%	0.6%	1.5%	1.0%	1.0%	0.3%	2.0%	0.5%	0.7%	0.7%	2.2%

Source: Our elaborations using Equations [3] and [4]

Table 10.10 Default Correlation Matrix derived from the Accounting Rule

	Food	Automobile	Building	Chemicals	Retailers	Electrical	Plant & Machinery	Miscellaneous Industrials	Publishing and Printing	Mining, Steel, Metals	Public Utility	Textile	Transportation & Tourism
Food	1.0%												
Automobile	0.3%	3.2%											
Building	0.3%	0.9%	1.0%										
Chemicals	0.2%	0.8%	0.5%	1.1%									
Retailers	0.6%	1.3%	0.9%	1.0%	4.2%								
Electrical	0.2%	0.9%	0.5%	0.8%	0.9%	1.2%							
Plant & Machinery	0.4%	1.2%	0.9%	0.6%	0.9%	0.6%	2.4%						
Miscellaneous Industrials	0.0%	0.3%	0.2%	0.3%	0.4%	0.2%	0.3%	0.6%					
Publishing and Printing	0.2%	1.4%	0.8%	0.8%	1.1%	1.0%	1.0%	0.4%	2.3%				
Mining, Steel, Metals	0.1%	0.5%	0.3%	0.2%	0.3%	0.2%	0.9%	0.1%	0.3%	2.9%			
Public Utility	0.2%	0.9%	0.4%	0.6%	0.8%	0.9%	0.6%	0.2%	1.1%	0.4%	1.7%		
Textile	0.3%	1.0%	0.7%	0.7%	1.0%	0.7%	0.9%	0.4%	1.2%	0.3%	0.6%	1.6%	
Transportation & Tourism	0.3%	0.9%	0.6%	0.6%	1.0%	0.6%	0.9%	0.3%	0.9%	0.4%	0.6%	0.6%	2.6%

Source: Our elaborations using equations [3] and [4]

also in line with the industry benchmark, as reported for example by Crouhy *et al*. (2000).[11]

Looking at the default correlation tables, it is worth noting that the Automobile (IM: 4.5%; AR: 3.2%) and Retailers (IM: 7.3%; AR: 4.2%) sectors report the highest coefficients of default correlations in both the Index Model and the Accounting Rule, confirming the asset return correlation evidence. The only exception concerns the Publishing and Printing sector that reports a very high average default correlation coefficient under the Index Model (9.4%), due to the combination of both a high level of asset return correlation (45.2%) and probability of default (1.35%). Among the least correlated sectors in terms of defaults there are instead the Food (IM: 0.4%; AR: 1%) and the Miscellanous Industrials (IM: 0.4%; AR: 0.6%) segments.

In the Appendix, we show how default correlation coefficients may be used to build a simple index of credit risk concentration under a mean-standard deviation approach to credit portfolio risk.

4.4. An assessment of the goodness of the asset return correlation coefficients

Finally, we compare the above estimated asset return correlation coefficients to the coefficients we obtain directly from equity prices. This latter exercise may be considered as a sort of in-sample back-test of our empirical models. In the following plot we compare the asset return correlations obtained from equity prices to the ones estimated with the Index Model.

The result is shown in Figure 10.3.

The three lines in the plot correspond to identical correlation and +/− 20% from the identical correlation. Therefore, points falling between the +/− 20% lines indicate that the rule-based correlation is within 20% of the actual correlation.

We find that almost all of the Index Model-based correlations are within 20% of the actual correlations, while approximately 80% of the Accounting Rule-based correlations are within the same boundaries.[12]

---

[11] The authors affirm that "the ratio of asset return correlations to default correlations is approximately 10–1 for asset correlations in the range of 20–60%".

[12] This graph has not been reported but is available upon request from the authors.

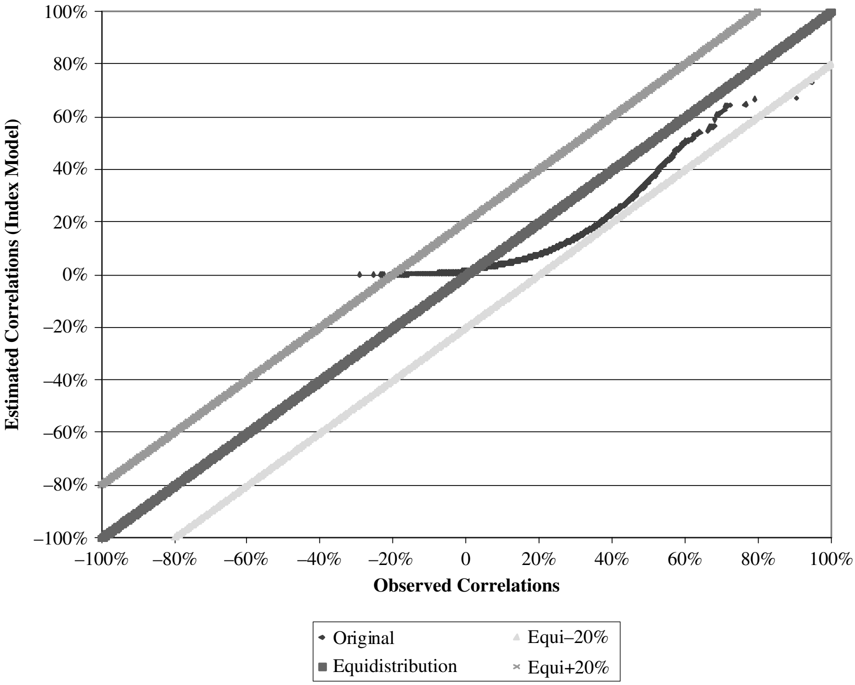


Fig. 10.3. Comparison of asset return correlations obtained from equity prices and those estimated with the Index Model
Source: Estimations using Equations [5] and [6]

5. CONCLUSIONS AND FUTURE RESEARCH

Calculating reliable credit value at risk figures requires not only valid probability of default estimates for each obligor, but also valid estimates of pairwise asset and default correlations. Therefore, financial institutions dealing with their own internal credit risk models will have to verify their correlation estimates on a regular basis.

This chapter presents the first empirical evidence on the problem of calibrating the asset and default correlation coefficients for Italian non-financial companies[13]

[13] Alternative estimates of asset and default correlation coefficients for Italian companies, based on accounting default rates, were firstly proposed by Zazzara (2002) and Sironi and Zazzara (2003).

with stock market data. Our analysis may be considered as an improvement to the standard solution for determining correlation coefficients currently offered in the credit risk management field for non-listed firms.

Our findings on the relationship between a firm's size and its systematic risk, defined as the percentage movement of a stock explained by its corresponding Milan Stock Exchange industry index (R-squared), are as follows. With respect to the complete sample of 130 companies, our estimates for the parameters y and λ, which describe the relationship between a firm's systematic risk and its total asset size variable, are similar to those obtained by Xiao (2002) for the US market, despite the many differences in the underlying data. However, the use of the turnover variable as an alternative proxy for company size is generally superior to the total assets variable in terms of its explanatory power in the cross-sectional regressions. Further, the robustness tests of our calibration results, carried out by splitting up our sample in two sub-samples of equal size according to the companies′ market capitalizations, confirm the positive relationship between firm's size and its systematic risk both for large- and medium-small companies. Our results suggest that this empirical relationship may be applied to non-listed medium-small firms with a certain degree of statistical confidence.

We then estimate the entire asset return correlation distributions under the Index Model and the Accounting Rule, and compare them to observed correlations estimated directly from equity prices.

According to our empirical results, an average asset return correlation coefficient between 10% and 12% and a corresponding average default correlation coefficient of around 1% appear to be reasonable estimates for current credit value at risk calculations for a portfolio of Italian non-financial firms.

Finally, our analysis suggests that stock market data could be used to calibrate internal credit risk models, even for Italian non-listed firms, yielding a model that is both reliable and dynamic.

ACKNOWLEDGEMENT

We thank Emiliano Santi of Bloomberg™ for providing us the data used in this study and conference participants at the XIII International Tor Vergata Conference on Banking and Finance (December 3, 2004). We also thank Prof. Giuseppe Lusignani and Lorenzo Bocchi for comments and useful discussions during the presentation of this work at Prometeia Research Institute (June 9, 2005). Our opinions do not reflect the Institutions we are affiliated with.

BIBLIOGRAPHY

Basel Committee on Banking Supervision (1999) *Credit Risk Modelling: Current Practices and Applications*. April, Basel.

Basel Committee on Banking Supervision (2003) *The New Basel Capital Accord*. Consultative Document, April, Basel.

Basel Committee on Banking Supervision (2004) *The New Basel Capital Accord*. Final Document, June, Basel.

Chen N., Roll R. and Ross S. (1986) "Economic Forces and the Stock Market." *Journal of Business* 59: 383–403.

Credit Suisse First Boston (1997) CreditRisk+. *A Credit Risk Management Framework*. Technical Document, available from Credit Suisse First Boston.

Crouhy M., Galai D. and Mark R. (2000) "A comparative analysis of current credit risk models." *Journal of Banking & Finance* 24: 59–117.

Dietsch M. and Petey J. (2002a) "The credit risk in SME loans portfolios: Modeling issues, Pricing, and capital requirements." *Journal of Banking & Finance* 26: 303–22.

Dietsch M. and Petey J. (2002b) "Are SMEs sensitive to systematic risk? A study of probabilities of default and assets correlations in French and German SMEs." Working paper, LARGE University Robert Schuman of Strasbourg.

Fama E.F. and MacBeth J.D. (1973) "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy* 81: 607–36.

Ford J.K. (1998) "Measuring Portfolio Diversification." *The Journal of Lending & Credit Risk Management* February: 50–53.

Gordy M.B. (2000) "A comparative anatomy of credit risk models." *Journal of Banking & Finance* 24: 119–49.

Gupton G.M., Finger C.C. and Bhatia M. (1997) *CreditMetrics™ – Technical Document*, J.P. Morgan & Co. Incorporated, New York.

Hahnenstein L. (2004) "Calibrating the CreditMetrics Correlation Concept – Empirical Evidence from Germany." *Financial Markets and Portfolio Management* 18(4): 358–2.

Hamerle A., Liebig T. and Rösch D. (2003) "Benchmarking asset correlations." *Risk*_November: 77–81.

Li D.X. (2000) "On Default Correlation: A Copula Function Approach." *Journal of Fixed Income* 9: 43–54.

Lopez J.A. (2002) "The empirical relationship between average asset correlation, firm probability of default and asset size." Working Papers in Applied Economic Theory 2002–05, Federal Reserve Bank of San Francisco.

Merton R.C. (1974) "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29: 449–70.

Saunders A. and Allen L. (2002) *Credit Risk Measurement – New Approaches to Value at Risk and Other Paradigms*. 2nd Edn, New York.

Sironi A. and Zazzara C. (2003) "The Basel Committee proposals for a new capital accord: implications for Italian banks." *Review of Financial Economic* 12(1): 99–126.

Xiao J.Y. (2002) *Obligor* R^2 *in Creditmetrics*, Research Technical Note, available from RiskMetrics™ Group on request.

Zazzara C. (2002) “Credit Risk in the Traditional Banking Book: A VaR Approach under Correlated Defaults.” *Research in Banking and Finance* 2: 355–84.

Zeng B. and Zhang J. (2002) “Measuring Credit Correlations: Equity Correlations Are Not Enough!” Working Paper, January, KMV, San Francisco.

Appendix

An Index of Credit Risk Concentration under Correlated Defaults

Under a Default Mode philosophy (i.e. only credit defaults are modeled, variations in the value of assets due to changes in credit quality short of default are assumed unimportant), we derive the first two moments of a credit portfolio loss distribution and build an Index of Credit Risk Concentration under two cases: I) absence of correlation between default events, and II) correlated defaults. We show the importance of default correlation coefficients for managing the risk of a credit portfolio.

I. No correlation between default events

We consider a portfolio with N exposures. The default rate for exposure i is p_i; the amount of the exposure is e_i, and the recovery rate at default is r_i $(0 \leq r_i \leq 1)$ (all values are fixed).[14] The portfolio loss L can be expressed using a random variable with either 1 or 0 as its value[15]:

$$D_i = \begin{cases} 1 & \text{(Probability } p_i) \\ 0 & \text{(Probability } 1 - p_i) \end{cases}$$

Therefore,

$$L = \sum_{i=1}^{N} D_i \times e_i \times (1 - r_i) \tag{A.1}$$

In Equation (A.1), the loss is a discrete value, but when N is sufficiently large and the interval between values is sufficiently small, it can be treated as continuously distributed. Assuming the recovery rate r_i is equal to zero and

[14] It is common to set up models so that these parameters are deterministic values, but ordinarily they will have some degree of uncertainty.
[15] This is called a Bernoulli random variable.

ignoring correlation between default events, the expected value and the variance for L can be expressed respectively as

$$E[L] = \sum_{i=1}^{N} p_i \times e_i \tag{A.2}$$

$$Var[L] = \sum_{i=1}^{N} p_i \times (1 - p_i) \times e_i \tag{A.3}$$

Therefore, the standard deviation for L is equal to

$$\begin{aligned} St.dev[L] &= \sqrt{\sum_{i=1}^{N} p_i \times (1 - p_i) \times {e_i}^2} \qquad \text{(A.4)} \\ &= \sqrt{\frac{\sum_{i=1}^{N} p_i(1 - p_i)}{N}} \times \sqrt{\sum_{i=1}^{N} e_i^2} \end{aligned}$$

From the above formula we can derive the following equivalent expression:

$$St.dev[L] = \sqrt{\frac{\sum_{i=1}^{N} p_i(1 - p_i)}{N}} \times \sum_{i=1}^{N} e_i \times \frac{\sqrt{\sum_{i=1}^{N} e_i^2}}{\sum_{i=1}^{N} e_i} \tag{A.4.1}$$

The circled ratio above can be considered as an Index of Concentration (or Diversification) of the portfolio (IC), in the absence of correlation between default events.[16] In this case, the IC ranges from 0 to 1 and depends on the amount and number of the exposures only. Therefore, this index will be lowest (diversified portfolio) when the exposures in the portfolio are of the same amount and their number is large. On the other hand, it will be highest (concentrated portfolio) when the exposure amounts are all different and there are few exposures in the portfolio. In the limit, when the portfolio is composed of only one exposure, the IC is equal to 1.

Since this index does not include any level of default correlation, the information it gives may be misleading. Thus, we can have a portfolio with a low IC,

[16] The IC is similar to the Herfindhal Index – widely used to calculate concentration of market shares – which is equal to the sum of the squared weights of the portfolio. Ford (1998) proposes a similar concentration ratio for managing the risk of a loan portfolio.

but which is not diversified. We clarify this point with the following example: Suppose we have a portfolio of 10,000 loans of the same amount, and all these loans have been granted to borrowers belonging to the same industry-group (for example, the Automobile industry sector).[17] In this case, the IC will be low even though, in practice, the loan portfolio is totally concentrated in only one industry sector. In this latter case, loans in the portfolio may be influenced by common background factors. Therefore, the IC seems to be of dubious meaning in terms of concentration risk.

Hence, let us see what happens when including some level of default correlation between exposures.

II. The case of correlated defaults

The extension to the correlated case is straightforward. In fact, the standard deviation of the loan loss random variable will be equal to

$$\mathrm{St.dev}[L] = \sqrt{\sum_{i} p_i(1-p_i)\times {e_i}^2 + 2\times \sum_{\substack{i,j=1\\ i\neq j}}^{N} \rho_{ij}\times \mathrm{St.dev}[L_i]\times \mathrm{St.dev}[L_j]} \quad \text{(A.5)}$$

where $\mathrm{St.dev}[L_i] = e_i \times \sqrt{p_i \times (1-p_i)}$

$\mathrm{St.dev}[L_j] = e_j \times \sqrt{p_j \times (1-p_j)}$

ρ_{ij} = default correlation coefficient

that, after a few steps, becomes

$$\mathrm{St.dev}[L] = \sqrt{\frac{\sum_{i=1}^{N} p_i(1-p_i)}{N}} \times \sum_{i=1}^{N} e_i \times \frac{\sqrt{\sum_{i=1}^{N} e_i^2}}{\sum_{i=1}^{N} e_i} \times \sqrt{1+\frac{2\times \sum_{\substack{i,j=1\\ i\neq j}}^{N} \rho_{ij}\times e_i \times e_j}{\sum_{i=1}^{N} e_i^2}} \quad \text{(A.5.1)}$$

The circled ratio in formula (A.5.1) can be considered as an Index of Concentration in case of Correlated default (ICC). In fact, it is expressed as the product

[17] In addition to industry data, this case can be easily extended to consider geographic-specific information.

of the previous IC and a correlation factor (CF). Obviously, the CF collapses to 1 when ρ_{ij} is equal to zero; in this latter case the IC and ICC will coincide. This extension is extremely important since the higher the default correlation between loans, the higher the degree of concentration in the portfolio, despite the amount and number of loans. In fact, when ρ_{ij} is large, even though the number of loans in the portfolio is large, the ICC is high (in the limit, when ρ_{ij} is equal to 1 also the ICC is equal to 1). Thus, the lower the default correlation between exposures, the more prevalent is the effect of diversification by number (and amount) of loans, as measured by the IC. Conversely, the greater the default correlation, the less prevalent is the effect of diversification by number (and amount) of loans.

Therefore, the Index of Concentration under Correlated Defaults (ICC) – that just depends on the amount and number of the exposures and on default correlation coefficients – may be easily applied to assess the degree of credit risk concentration at the portfolio level.

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 11

MARGIN REQUIREMENTS WITH INTRADAY DYNAMICS

John Cotter and François Longin

Abstract

Both in practice and in the academic literature, models for setting margin requirements in futures markets use daily closing price changes. However, financial markets have recently shown high intraday volatility, which could bring more risk than expected. Such a phenomenon is well documented in the literature on high-frequency data and has prompted some exchanges to set intraday margin requirements and ask intraday margin calls. This chapter proposes to set margin requirements by taking into account the intraday dynamics of market prices. Daily margin levels are obtained in two ways: first, by using daily price changes defined with different time-intervals (say from 3 pm to 3 pm on the following trading day instead of traditional closing times); second, by using 5-minute and 1-hour price changes and scaling the results to one day. An application to the FTSE 100 futures contract traded on LIFFE demonstrates the usefulness of this new approach.

Keywords: ARCH process, clearinghouse, exchange, extreme value theory, futures markets, high-frequency data, intraday dynamics, margin requirements, model risk, risk management, stress testing, value at risk.

JEL Classification No.: G15

1. INTRODUCTION

The existence of margin requirements decreases the likelihood of customers' default, brokers' bankruptcy and systemic instability of futures markets. Margin requirements act as collateral that investors are required to pay to reduce default risk.[1] However, margin committees face a dilemma in determining the magnitude of the margin requirement imposed on futures traders. On the one hand, setting a high margin level reduces default risk. On the other hand, if the margin level is set too high, then the futures contracts will be less attractive for investors due to higher costs and decreased liquidity, and finally less profitable for the exchange itself. This quandary has forced margin committees to impose investor deposits that represent a practical compromise between meeting the objectives of adequate prudence and liquidity of the futures contracts.

For products traded on the London International Financial Futures and Options Exchange (LIFFE), margin requirements are set by the London Clearing House (LCH)[2] using the London Systematic Portfolio Analysis of Risk (SPAN) system, a specifically developed variation of the SPAN system originally introduced by the Chicago Mercantile Exchange (CME). The London SPAN system is a non-parametric risk-based model that provides output of margin requirements that are sufficient to cover potential default losses in all but the most extreme circumstances. The inputs to the system are estimated margin requirements relying on price movements that are not expected to be exceeded over a day or couple of days. These estimated values are based on diverse criteria incorporating a focus on a contract's price history, its close-to-close and intraday price movements, its liquidity, its seasonality and forthcoming price sensitive events. Market volatility is especially a key factor to set margin levels. Most important, however, is the extent of the contract's price movements with a policy for a minimum margin requirement that covers three standard deviations of historic price volatility based on the higher of one-day or two-day price movements over the previous 60-day trading period. This is akin to using the Gaussian distribution, where multiples of standard deviation cover certain price movements at various probability levels.[3]

The academic literature has applied a number of alternative statistical approaches in order to compute the margin requirement that adequately protects against default at various probability levels and/or determine the probabilities associated with different margin requirements. Figlewski (1984) and Gay *et al*. (1986) classically assume that futures price movements follow a Gaussian distribution. One well-documented problem with using a particular distribution

[1] Futures exchanges also use capital requirements and price limits to protect against investor default.
[2] The LCH risk committee made up of qualified risk management members is responsible for all decisions relating to margin requirements for LIFFE contracts. Margin committees generally involve experienced market participants who have widespread knowledge in dealing with margin setting and implementation, through their exposure to various market conditions and their ability to respond to changing environments (Brenner (1981)). The LCH risk committee is independent from the commercial function of the clearinghouse.
[3] For instance, under the hypothesis of normality for price movements, two standard deviations would cover 97.72% of price movements, and three standard deviations 99.87%.

such as the Gaussian distribution is model risk. In particular, it is well known in that most cases the Gaussian distribution underestimates the weight of the tails of the distribution. Longin (1996) uses extreme value theory to quantify this statement and shows that the empirical distribution of financial asset price changes is fat-tailed while the Gaussian distribution is thin-tailed. Edwards and Neftci (1988) and Warshawsky (1989) use the historical distribution of past price changes, which overcomes the underestimation issue of assuming normality. However, the historical distribution is unable to deal with low probability levels, due to the lack of sufficient price changes available for analysis.

A distinct approach focuses on an economic model for broker cost minimization in which the margin is endogenously determined (Brennan (1986)). Another approach developed by Craine (1992) and Day and Lewis (1999) is based on the fact that the distributions of the payoffs to futures traders and the potential losses to the futures clearinghouse can be described in terms of the payoffs to barrier options. Initial margins requirements can then be related to the present value of such options.

Kofman (1993), Longin (1995 and 1999), Booth *et al.* (1997) and Cotter (2001) apply extreme value theory, a statistical theory that specifically models the tails of the distribution of futures price changes. This latter framework focuses on the main measurement issue relating to margin setting, namely trying to adequately model quantiles and probabilities of the distribution tails for future price changes. As the problem of setting margin requirements is related to the tails of the distribution of futures price changes (the left tail for a long position and the right tail for a short position), it is beneficial to examine specifically lower and upper tail percentiles. Extreme value theory does exactly this by focusing only on tail values, thereby minimizing model risk that is associated with procedures that model the full distribution of futures price changes. Extreme value theory removes the need for making assumptions of the exact distributional form of the random process under analysis, as the limiting distribution of extreme price changes is the same for many classes of distributions and processes used to describe futures price changes (Longin and Solnik (2001)). Another advantage of the extreme value approach is the parametric form that allows us to extrapolate to out-of-sample time frames, unlike the use of the historical distribution of price changes that is constrained to in-sample predictions. By having an objective likelihood function, we avoid the problem of subjectively defined stress tests that try to examine the impact of financial crises. Further, extreme value theory requires tail estimates that are time invariant due to their fractal nature. This allows for precise tail measurement incorporating a simple and efficient scaling law for different frequency intervals, for example from intraday to daily estimates.

One question that we may ask about the nature of risk management is whether the clearinghouse should care more about ordinary market conditions or more about extraordinary market conditions. In other financial institutions, such as banks, two distinct approaches are used: value at risk models for ordinary market conditions and stress testing for extraordinary market conditions (Longin 2000). The clearinghouse must also address both sets of market conditions in margin setting, so as to minimize the likelihood of investor default by examining a range of probabilities of price movements associated with common and uncommon events. The first approach is conditional reflecting the changing of market conditions over time while the second approach is unconditional trying to incorporate extreme events that occurred over a long period of time. All above studies are based on unconditional distributions and cannot reflect current market conditions. Cotter (2001) considers a conditional process by applying a GARCH specification to address issues relating to the dynamic features of futures contracts volatility.

Previous studies based on statistical models used closing prices to estimate daily margin requirements mainly due to data unavailability. However, trading on futures markets takes place on an intraday level and a complete understanding of their operations requires analysis of high-frequency intraday features (Cotter 2004). Margin setting using intraday dynamics incorporates the full information set regarding price movements over the trading day. In contrast, margin setting using closing prices only uses trading information around close of day. Intraday dynamics are important. For instance, it is well documented that daily volatility varies over time with particular characteristics (Bollerslev *et al.* (1992)). However, more recently, intraday volatility has also been examined and distinct patterns are also documented. For example, macroeconomic announcements impact volatility sharply but their impacts have a life span of less than two hours, and thereafter have a negligible influence on price movements (Bollerslev *et al.* (2000)). Thus, an analysis of daily prices alone would not take account of these intraday activities.

Intraday price movements supply the margin setter with a mechanism to adequately describe and predict the impact of futures price volatility within the appropriate timeframe. In terms of statistical modeling, the impact of futures volatility on margin requirement setting require a certain minimum number of observations for first accurately identifying the empirical feature, next developing a model that adequately describes the feature and finally testing the model to predict market occurrence. Notwithstanding, the clearinghouse must ensure that they are modeling the same economic event in their analysis of financial data. For instance, futures price changes may exhibit a structural change over time from say the 1980s to the 1990s. Thus, given the average lifespan of many futures contracts is one year, margin setting is based on analysis of price movements for this interval size in this chapter. However, in model development, this interval size may sometimes provide insufficient observations at daily frequency using various statistical techniques. Using higher frequency intraday price changes and scaling to relatively low frequency daily estimates overcomes this modeling difficulty.

In practice, clearinghouses are beginning to recognize the importance of intraday dynamics. For example, in 2002, the LCH introduced an additional intraday margin requirement that is initiated if price movements on a contract challenge the prevailing margin requirement. Specifically, an intraday margin requirement is initiated if a contract price changes by 65% of the margin requirement originally set for that contract. In this case, the clearinghouse requires an additional margin payment for falling prices on a long position or for rising prices on a short position. The possible impact of intraday price movements is now clearly, and rightly so, of concern to risk management overseers for LIFFE contracts.

The main contribution of this chapter is to take into account the intraday dynamics of futures market prices by computing margin requirements. All previous academic studies considered daily closing prices only, thus missing important information. Closing prices alone lose information regarding price movements and their associated transaction activity within the trading day. The clearinghouses modeling margin requirements should incorporate the intraday price movements in margin setting. Daily margin levels are obtained in two ways: first, by using daily price changes defined with different time-intervals (say from 3 pm to 3 pm on the following trading instead of traditional closing times); second, by using 5-minute and 1-hour price changes and scaling the results to one day following Dacarogna *et al*. (1995). As shown by Merton (1980) for risk measures (as opposed to performance measures), it is beneficial to use data with the highest frequency in order to get more precise estimates of the tail parameter. In this chapter, different statistical distributions are also used to model futures price changes: the Gaussian distribution, the extreme value distribution and the historical distribution. An ARCH-type process is also used to take into account the time-varying property of financial data. An application is given for the FTSE 100 futures contract traded on LIFFE.

The remainder of the chapter is organized as follows. The statistical models used for the distribution of futures contract price changes and the scaling methods are presented in the next section. Section 3 provides a description of the FTSE 100 futures contract data used in the application and a detailed statistical analysis of the intraday dynamics of the market prices. Section 4 presents empirical results for margins by taking into account the intraday dynamics. Finally, a summary of the paper and some conclusions are given in Section 5.

2. STATISTICAL MODELS AND SCALING METHOD

This section presents the different statistical models used to compute the margin level for a given probability. It also presents the scaling method to obtain daily margin levels from intraday price changes.

2.1. The extreme value distribution

The theoretical framework applied in this study relies on the findings of extreme value theory. According to this statistical theory, three types of asymptotic distribution can be obtained: Gumbel, Weibul and the one of concern to this study, the Fréchet distribution, which is obtained for fat-tailed distributions (Gnedenko (1943)). Weak convergence is assumed to occur for the Fréchet distribution underpinned by the maximum domain of attraction (MDA). This allows for approximation to the characteristics of the Fréchet distribution, giving rise to a semi-parametric estimation procedure. This theoretical framework offers a number of advantages to margin setting. First, the main prudence issue in determining margin requirements is to protect against default that results from extreme price movements. These price changes are extreme values and as such should be modeled with procedures specifically focused on capturing these quantile and probability estimates, and this is exactly what extreme value does. Second, modeling only the tail of the distribution as opposed to the center of the distribution, which is irrelevant for margin setting, minimizes bias in the estimation procedure. Third, tail behavior of the fat-tailed Fréchet distribution exhibits a self-similarity property that allows for an easy extension for multi-period margin estimation using a simple scaling rule.

We begin by examining the framework and by assuming that a margin requirement can be measured as futures price change, represented by a random variable, R, and that exceeding this level is estimated at various probabilities. Further, we assume that the random variable is independent and identically distributed (iid) and belongs to the true unknown cumulative probability density function F_R.[4] We are interested in the probability that the maximum of the first n random variables exceeds a certain price change, r,[5]

$$P\{M_n > r\} = 1 - F^n(r) \tag{1}$$

for n random variables, $M_n = \max\{R_1, R_2, \ldots, R_n\}$.

[4] The successful modelling of financial returns using GARCH specifications clearly invalidates the iid assumption. De Haan *et al.* (1989) examine less restrictive processes more akin with futures price changes only requiring the assumption of stationarity and this is followed in this paper.

[5] Extreme value theory is usually detailed for upper order statistics focusing on upper tail values and the remainder of the paper will follow this convention. This study also examines empirically the lower order statistics focused on lower tail values.

The probability estimator could also be expressed as a quantile where we are investigating what margin requirement is sufficient to exceed futures price changes at various probability levels.[6]

Whilst the exact distribution is unknown, assuming the distribution exhibits the regular variation at infinity property, then asymptotically it behaves like a fat-tailed distribution.

$$1 - F^n(r) \approx ar^{-\alpha} \tag{2}$$

where a represents the scaling parameter and α the shape parameter.[7]

This expression is for any given frequency and it is easy to extend the framework to lower frequencies as these extremes have an identical tail shape. For instance, taking the single period price changes, R, and extending these to a multi-period setting, kR, using the additive property of a fat-tailed distribution from Feller's theorem (Feller 1971):

$$1 - F^n(kr) \approx kar^{-\alpha} \tag{3}$$

Importantly the shape parameter, α, remains invariant to the aggregation process and also has implications for empirical benefits in its actual estimation.[8] Dacarogna *et al.* (1995) have shown that high-frequency tail estimation has efficiency benefits due to their fractal behavior. In contrast, low frequency estimation suffers from negative sample size effects. Further, for ease of computation, the scaling procedure does not require further estimation, but only involves parameters from the high-frequency analysis, shown to provide the most detailed information on futures price movements.

The regular variation at infinity property represents the necessary and sufficient condition for convergence to the fat-tailed extreme value distribution. Thus it unifies fat-tailed distributions and allows for unbounded moments:

$$. \lim_{t \to +\infty} \frac{1 - F_R(t \cdot r)}{1 - F_R(t)} = r^{-\alpha} \tag{4}$$

---

[6] For the issue at hand the probability of exceeding a predetermined margin level on a short position for n price changes is: $P_{short} = P\{M_n > r_{short}\} = \delta$, where r_{short} represents the margin level on a short position and δ is the unknown exceedance probability given by $1 - F^n(r)$.

[7] The shape parameter α is related to the tail index τ often used in the EVT literature by the relation: $\alpha = 1/\tau$.

[8] The α-root scaling law for the extreme value estimates is similar in application to the $\surd$ scaling procedure of a normal distribution.

By l'Hopital's rule it can be shown that the Student-t, and symmetric non-normal sum-stable distributions, and certain ARCH processes with an unconditional stationary distribution and even assuming conditionally normal innovations, all exhibit this condition as their tails decline by a power function. Subsequently all these distributions exhibit identical behavior far out in the tails. In contrast, other distributions such as the normal distribution, and the finite mixtures of Gaussian distributions have a tail that declines exponentially, which declines faster than a power decline and thus are relatively thin-tailed. The shape parameter, α, measures the degree of tail thickness and the number of bounded moments (see Appendix for details of the semi-parametric estimation procedure). A shape parameter greater than 2 implies that the first two moments, the mean and variance, exist whereas financial studies have cited values between 2 and 4, suggesting that not all moments of the price changes are finite (Longin 1996). In contrast, support for the Gaussian distribution would require a shape parameter equal to infinity, as all moments exist. Thus the estimate of the shape parameter distinguishes between different distributions and, for instance, α represents the degrees of freedom of the Student-*t* distribution and equals the characteristic exponent of the sum-stable distribution for $\alpha < 2$.

Given the asymptotic relationship of the random variable to the fat-tailed distribution, non-parametric tail estimation takes place, giving two related mechanisms for describing the margin estimates. The first focuses on the margin requirement and determines the probability of various price movements, r_p:

$$r_p = r_t(m/np)^{1/\alpha} \tag{5}$$

By using this estimate, we can examine different margin requirements that would not be violated at various probability levels and implicitly determine if the trade-off between optimizing liquidity and prudence is being met. Rearranging gives the probability, p, of exceeding any preset margin requirement:

$$p = (r_t/r_p)^{\alpha} m/n \tag{6}$$

Again these probabilities are used to determine if the prudence and commercial concerns of the futures exchange is reached.

2.2. *The APARCH process*

To model the time-varying behavior of price changes suggested by the previous analysis, we use the Asymmetric Power ARCH (APARCH) developed by Ding *et al.* (1993). This model nests many extensions of the GARCH process. As well as encompassing three ARCH specifications (ARCH, Non-linear ARCH and Log-ARCH), two specifications of the GARCH model (using standard deviation and variance of returns), it also details two asymmetric models (both ARCH and GARCH versions). It is given by

$$\sigma_t^d = \alpha_o + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| + \gamma_i \varepsilon_{t-i})^d + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^d \tag{7}$$

for $\alpha_0, \alpha_i, \beta_j \geq 0, \quad \alpha_i + \beta_j \leq 1, \quad -1 \leq \gamma_i \leq 1.$

The APARCH incorporates volatility persistence, β, asymmetries, γ, and flexibility of power transformations, d, in the estimation of volatility. Detailing the model, the process presents the volatility measure in the form of a Box-Cox transformation, whose flexibility allows for different specifications of the residuals process. This transformation provides a linear representation of non-linear processes. As well as describing the traditional time dependent volatility feature, the model specifically incorporates the leverage effects, γ, by letting the autoregressive term of the conditional volatility process be represented as asymmetric absolute residuals. A general class of volatility models incorporating the non-linear versions are defined by the power coefficient, d.

The APARCH (1, 1) was applied to the price series at the end of the sample during December 2000. A number of variations of the model are applied and Akaike's (AIC) and Schwarz's (BIC) selection criteria are used to determine the best fitted process. Fat-tails are accounted for by assuming that the conditional distribution is a Student-*t* distribution.

3. DATA ANALYSIS

3.1. Data

The empirical analysis is based on transaction prices for the FTSE 100 futures contract trading on the LIFFE exchange (data are obtained from *Liffedata*). This exchange has made a clear distinction between contracts that are either linked to an underlying asset or developed formally on the basis of links to the recently developed European currency, the euro, and those that remain linked to factors outside the currency area. The FTSE 100 represents the most actively traded example of the latter asset type.

Data are available on the stock index contract for four specific delivery months per year, March, June, September and December. Prices are chosen from those contracts with delivery months on the basis of being the most actively traded using a volume crossover procedure. The empirical analysis is completed for sampling frequencies of 5 minutes, 1 hour and 1 day. The first interval is chosen so as to meet the objective of analyzing the highest frequency possible and capturing the most accurate risk estimates but also avoids microstructure effects such as bid ask effects. For the daily frequency, the price changes are computed by taking different starting (and ending) times to define the day: the beginning of the "day" can start from 9 am (the opening of the trading day) to 5 pm (the closing of the trading day). Nine different time-series of daily price changes are then obtained. Log prices (or log prices to the nearest trade available) for each interval are first differenced to obtain each period's price change. The period of analysis is for the year 2000, involving 247 full trading days corresponding to an average life span of an exchange traded futures contract. The FTSE 100 futures daily interval encompasses 113 5-minute intervals and 9 hourly intervals. A number of issues arise in the data capture process. First, all holidays are removed. This entails New Year's (2 days), Easter (2 days), May Day (1 day), spring holiday (1 day), summer holiday (1 day), and Christmas (2 days). In addition, trading took place over a half day during the days prior to the New Year and Christmas holidays and these full day periods are removed from the analysis.

3.2. Basic statistics

Basic statistics are reported in Table 11.1 for price changes (Panel A) and for squared price changes (Panel B). Concentrating on the first four moments of the distribution, we study their behavior according to frequency of measurement. Most predominately, the kurtosis increases as the frequency increases. For price changes, the (excess) kurtosis is equal to 0.26 for a 1-day frequency, 1.54 for a 1-hour frequency and 254.50 for a 5-minute frequency. The high kurtosis (higher than the value equal to 0 implied by normality) gives rise to the fat-tailed property of futures price changes. It is also illustrated by the probability density function and QQ plots of the shapes of price changes for different frequencies given in Figure 11.1. The extent of fat-tails is strongest for 5-minute realizations supporting the summary statistics. Also, the magnitude of values for these realizations can be large, as indicated by the scale of the density plots. These features generally result in the formal rejection of a Gaussian distribution using the Kolmogorov-Smirnov test.[9] Deviations from normality are strongest at the highest frequency. The other moments emphasize the magnitude and scale

[9] Whilst a formal rejection of normality for the full distribution of daily price is not recorded at common significance levels the tail behaviour in Figure 11.1 clearly indicates a fat-tailed property.

Table 11.1 Basic statistics for the FTSE 100 contract price changes defined for different frequencies

Panel A. Price changes

	Frequency of price changes		
	5-minutes	1-hour	1-day
Mean	0.00	−0.02	−0.03
Standard deviation	0.11	0.30	1.30
Skewness	−1.44	−0.28	−0.15
Kurtosis	254.5	1.54	0.26
Kolmogorov-Smirnov	0.08	0.05	0.04
test of normality	(0.00)	(0.00)	(0.31)
Minimum	−5.17	−1.57	−4.38
1st quartile	−0.05	−0.18	−0.77
2nd quartile	0.00	−0.00	−0.03
3rd quartile	0.05	0.16	0.76
Maximum	4.34	1.29	3.20

Panel B: Squared price changes

	Frequency of price changes		
	5-minutes	1-hour	1-day
Mean	0.01	0.09	1.70
Standard deviation	0.21	0.17	2.55
Skewness	107.99	5.24	2.69
Kurtosis	12 815.78	46.5	10.38
Kolmogorov-Smirnov	0.47	0.29	0.25
test of normality	(0.00)	(0.00)	(0.00)
Minimum	0.00	0.00	0.00
1st quartile	0.00	0.01	0.14
2nd quartile	0.00	0.03	0.65
3rd quartile	0.01	0.09	2.21
Maximum	0.01	0.10	19.17

Note: This table gives the basic statistics and empirical quantiles for the price changes (Panel A) and the squared price changes (Panel B). It also presents the results of the Kolmogorov-Smirnov test for normality with the p-value below in parentheses. Three different frequencies are used to compute the price changes: 5 minutes, 1 hour and 1 day. Data are price changes of the FTSE 100 future contract over the year 2000.

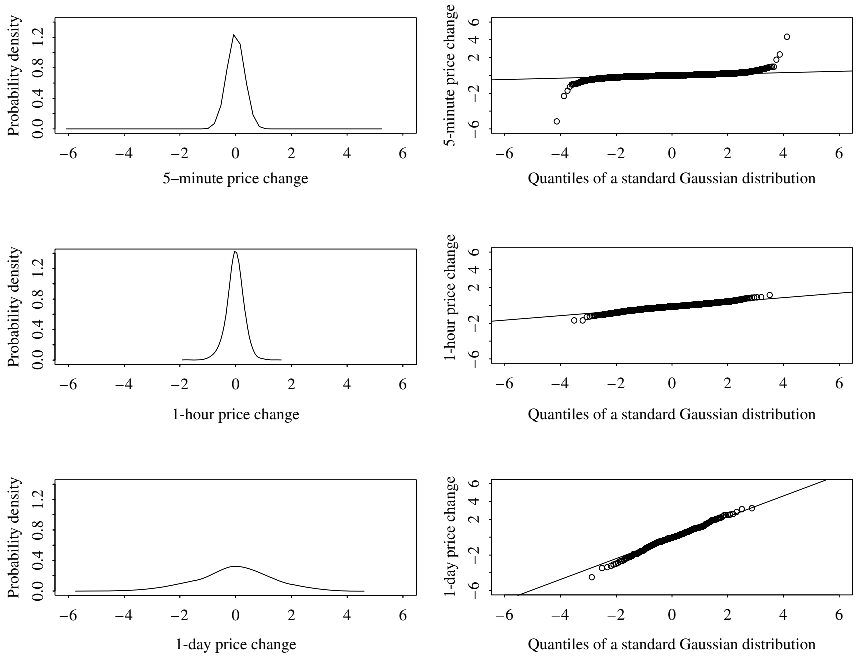


Fig. 11.1. Probability density function and QQ plot for price changes of the FTSE 100 contract

Note: These figures represent the probability density function and the QQ plots for price changes in the FTSE 100 future contract for the year 2000. Three different frequencies are used to compute the price changes: 5 minutes, 1 hour and 1 day.

of the realizations sampled at different frequencies. On average, price changes were negative during the year 2000 and unconditional volatility increases for interval size. Selected quantiles reinforce divergences in magnitude at different frequencies. Similar conclusions can be made for the proxy of volatility, the squared price changes, although the skewness and kurtosis are more pronounced.

Notwithstanding the divergence in moments for different frequencies, it is interesting to examine daily price changes and volatility as it is these estimates that are used in the statistical analysis resulting in daily margin requirements. In addition to examining daily price changes using closing prices that are the norm in margin setting through the marking to market system, daily price changes are also defined with different time-intervals. Basic statistics are reported in Table 11.2 and a time-series plot for two of these time-intervals, using opening prices and closing prices are presented in Figure 11.2. Whilst the mean price changes remain reasonably constant, other moments are more diverging. For

Table 11.2 Basic statistics for the FTSE 100 contract price changes defined with different time-intervals

Panel A. Price changes

	Open	10 am	11 am	12 pm	1 pm	2 pm	3 pm	4 pm	Close
Mean	−0.04	−0.04	−0.03	−0.03	−0.03	−0.03	−0.04	−0.03	−0.03
Standard deviation	1.32	1.23	1.20	1.23	1.18	1.29	1.22	1.16	1.30
Skewness	−0.13	−0.10	−0.30	−0.47	−0.32	−0.13	−0.14	−0.09	−0.15
Kurtosis	1.52	1.13	0.88	1.39	0.16	0.14	−0.05	−0.32	0.26
Kolmogorov-Smirnov	0.05	0.04	0.05	0.06	0.04	0.03	0.04	0.03	0.04
test of normality	(0.10)	(0.48)	(0.11)	(0.11)	(0.46)	(0.62)	(0.57)	(0.71)	(0.31)
Minimum	−5.84	−4.92	−4.74	−5.73	−4.48	−4.54	−3.60	−3.13	−4.38
1st quartile	−0.79	−0.86	−0.78	−0.76	−0.80	−0.79	−0.79	−0.80	−0.77
2nd quartile	−0.04	−0.01	0.02	−0.01	0.03	−0.02	−0.04	0.02	0.00
3rd quartile	0.78	0.74	0.73	0.81	0.80	0.86	0.78	0.76	0.76
Maximum	4.26	4.06	3.59	3.09	2.59	3.20	3.02	2.48	3.20

Panel B: Squared price changes

	Open	10 am	11 am	12 pm	1 pm	2 pm	3 pm	4 pm	Close
Mean	1.73	1.51	1.44	1.51	1.40	1.65	1.48	1.35	1.70
Standard deviation	3.24	2.66	2.44	2.79	2.06	2.42	2.06	1.74	2.55
Skewness	5.38	4.49	4.27	6.58	4.03	3.15	2.25	1.75	2.69
Kurtosis	43.77	27.90	26.24	65.77	27.84	16.08	5.94	2.90	10.38
Kolmogorov-Smirnov	0.30	0.29	0.28	0.29	0.25	0.25	0.24	0.22	0.25
test of normality	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1st quartile	0.09	0.13	0.13	0.14	0.13	0.18	0.16	0.12	0.14
2nd quartile	0.63	0.62	0.58	0.60	0.64	0.72	0.63	0.60	0.58
3rd quartile	2.05	1.70	1.74	1.94	1.85	1.86	1.95	1.76	2.21
Maximum	34.13	24.24	22.46	32.87	20.06	20.60	12.93	9.79	19.17

Note: This table gives the basic statistics and empirical quantiles for the price changes (Panel A) and the squared price changes (Panel B) over different time-intervals. It also presents the results of the Kolmogorov-Smirnov test for normality with the *p*-value below in parentheses. To define the price change, the starting time, which is equal to the ending time on the following day, varies from 9 am (opening of the market) to 5 pm (closing of the market). Data are price changes of the FTSE 100 future contract over the year 2000.

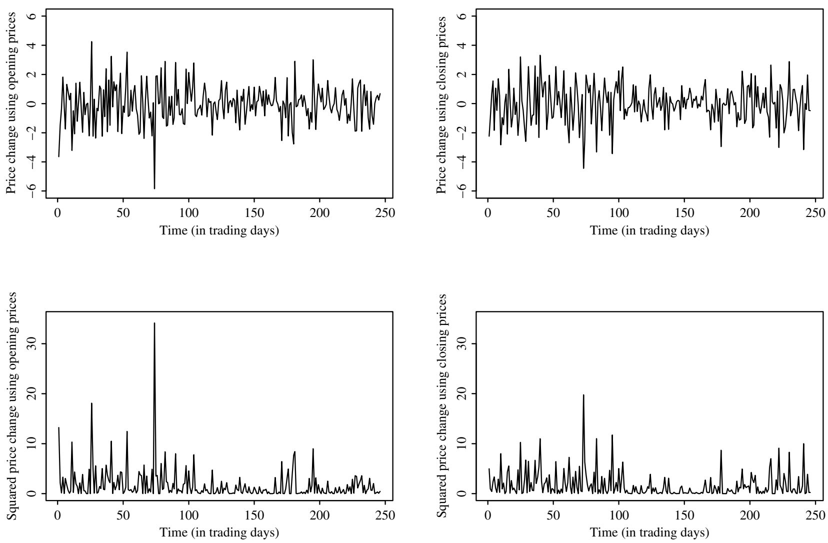


Fig. 11.2. Daily price changes and daily squared price changes of the FTSE 100 contract

Note: These figures represent the history of the price change and squared price change of the FTSE 100 future contract for the year 2000. Daily price changes are computed in two ways: from 9 am to 9 am on the following day (opening prices) and from 5 pm to 5 pm (closing prices).

instance, skewness goes from −0.09 to −0.47 and the kurtosis statistic goes from being platykurtic (−0.32) to leptokurtic (1.52). Also the dispersion of various quantiles is considerable. Again inferences for the squared price changes are similar, although greater in magnitude. However, it can be observed that both time-series have similar time-varying features evidencing volatility clustering with periods of high and low volatility, but the diverging features are clearly demonstrated as suggested by the magnitude of realizations.

Given the divergence indicated by the intraday analysis, it is interesting to incorporate these features in the margin setting process.

3.3. Extreme value analysis

Shape parameter estimates using different time-intervals to compute daily price changes are presented in Table 11.3 for the left tail (Panel A) and the right tail (Panel B). The point estimates are calculated using the weighted least squares

Table 11.3 Shape parameter estimates and test of the existence of moments

Panel A. Left tail

	Open	10 am	11 am	12 pm	1 pm	2 pm	3 pm	4 pm	Close
Shape parameter α	3.06 (0.65)	3.25 (0.69)	2.68 (0.57)	3.30 (0.70)	3.62 (0.77)	3.51 (0.75)	6.34 (1.35)	3.03 (0.65)	3.11 (0.66)
H0: $\alpha > 2$	1.63 (0.45)	1.81 (0.46)	1.18 (0.38)	1.85 (0.47)	2.10 (0.48)	2.02 (0.48)	3.21 (0.50)	1.60 (0.45)	1.68 (0.45)
H0: $\alpha > 4$	−1.43 (0.00)	−1.08 (0.00)	−2.32 (0.00)	−0.99 (0.00)	−0.49 (0.00)	−0.65 (0.00)	1.73 (0.46)	−1.50 (0.00)	−1.33 (0.00)

Panel B. Right tail

	Open	10 am	11 am	12 pm	1 pm	2 pm	3 pm	4 pm	Close
Shape parameter α	2.58 (0.55)	3.63 (0.77)	4.34 (0.93)	3.77 (0.80)	4.20 (0.90)	3.48 (0.74)	4.96 (1.06)	4.08 (0.87)	3.64 (0.78)
H0: $\alpha > 2$	1.05 (0.35)	2.11 (0.48)	2.53 (0.49)	2.20 (0.49)	2.46 (0.49)	2.00 (0.48)	2.80 (0.50)	2.39 (0.49)	2.11 (0.49)
H0: $\alpha > 4$	−2.59 (0.00)	−0.48 (0.00)	0.37 (0.14)	−0.29 (0.00)	0.22 (0.09)	−0.70 (0.00)	0.91 (0.32)	0.09 (0.04)	−0.47 (0.00)

Note: This table gives the shape parameter estimates for the left tail (Panel A) and the right tail (Panel B) of the distribution of daily price changes and a test of the existence of the moments of the distribution. The first line of the table gives the shape parameter estimate obtained with the method developed by Huisman *et al.* (2001) with the standard error below in parentheses. The second and third lines give the results of a test of the existence of the second moment (the variance) and the fourth moment (the kurtosis) with the p-value below in parentheses. As the shape parameter corresponds to the highest moment defined for the distribution, the null hypotheses are defined as follows: $H_0: \ \alpha > 2$ and $H_0: \ \alpha > 4$. To define the price change, the starting time (which is equal to the ending time on the following day) varies from 9 am (opening of the market) to 5 pm (closing of the market). Data are price changes of the FTSE 100 future contract over the year 2000.

technique that minimizes the small sample bias following Huisman *et al.* (2001). The point estimates range from 2.57 to 6.34 and the values are generally in line with previous findings (Cotter 2001). As the shape parameter is positive, the extreme value distribution is a Fréchet distribution that is obtained for a fat-tailed distribution of price changes.

We also use the shape parameter estimates to test if the second and the fourth moment of the distribution are well defined. For classical confidence level (say 5%), we are unable to reject the hypothesis that the variance is infinite in any scenario, whereas we are able to reject the hypothesis that the kurtosis is infinite in many scenarios. Advantageously the extreme value scaling law is applicable as it only requires the existence of a finite variance.

3.4. Conditional estimation

Time-varying behavior is described from fitting the APARCH model to daily price changes from different time-intervals at the end December 2000. The fat-tailed property is accounted for by assuming the error innovations belong to a Student-t distribution. The APARCH estimates consistently indicate that the conditional distributions exhibit persistence, with for example, past volatility impacting on current volatility as is typical of GARCH modeling at daily intervals.[10] Further, the conditional distributions vary according to the time intervals analyzed that will give rise to different margin requirements.

4. MODEL-BASED MARGIN REQUIREMENTS

This section presents empirical results for margin requirements obtained with daily price changes (4.1) and 5-minute and 1-hour price changes scaled to one day (4.2).

4.1. Margin requirement based on daily price changes

Table 11.4 presents margin requirements obtained with daily price changes for a long position (Panel A) and for a short position (Panel B). Margin requirements are computed for a given probability. Four different values are considered: 95%, 99%, 99.6% and 99.8% corresponding to average waiting periods of 20, 100, 250 and 500 trading days. Thinking of risk management for financial institutions, probabilities of 95% and 99% would be associated with ordinary adverse market events modeled by value at risk models, and probabilities of 99.6% and 99.8% with extraordinary adverse market events considered in stress testing programs. In the margin setting context, the probability reflects the degree of prudence of the exchange: the higher the probability, the higher the margin level, the

---

[10] For instance the parameter estimates based on closing prices are: $\alpha_0 = 0.014$, $\alpha_1 = 0.011$, $\beta_1 = 0.962$, $\gamma_1 = -0.999$ and $d = 1.855$. Further details and coefficient estimates are available on request.

less risky the futures contract for market participants, but the less attractive the contract for investors. Margin requirements are also computed with various statistical models: three unconditional distributions (Gaussian, extreme value and historical) and a conditional process (the Asymmetric Power ARCH process).

For the presentation of the results, the extreme value distribution will be the reference model as it presents many advantages (parametric distribution, limited model risk, limited event risk) and as the problem of margin setting is mainly concerned with extreme price changes. Beginning with the analysis of extreme value estimates, we first note that variation occurs in the estimates based on the different time-intervals to define daily price changes. For example, for a long position and a probability level of 95%, the estimated margin level ranges from 1.83% to 2.05% of the nominal position. For the most conservative level of 99.8%, it ranges from 2.77% to 5.32%, almost double. Also there does not seem to be a systematic pattern to these deviations. For instance, for a probability of 95%, the minimum is obtained with 2 pm prices and the maximum for closing

Table 11.4 Margin levels for given probabilities based on daily price changes

Panel A. Long position

Probability (waiting period)	Model	Open	10 am	11 am	12 pm	1 pm	2 pm	3 pm	4 pm	Close
	Gaussian	2.21	2.06	2.00	2.05	1.97	2.15	2.05	1.94	2.17
95%	Extreme value	1.85	1.95	1.89	1.84	2.04	1.83	1.85	1.95	2.05
(20 days)	Historical	1.90	1.87	2.23	2.08	2.34	2.14	2.04	2.28	2.28
	APARCH	2.05	2.22	2.63	2.55	3.19	2.94	2.65	2.90	2.91
	Gaussian	3.11	2.90	2.82	2.89	2.78	3.03	2.88	2.73	3.05
99%	Extreme value	2.94	3.22	3.12	2.70	2.78	2.42	2.26	2.74	2.93
(100 days)	Historical	2.98	3.23	3.06	2.76	2.90	2.89	2.51	3.19	3.25
	APARCH	3.62	3.62	3.62	3.12	3.90	3.85	3.29	4.38	4.39
	Gaussian	3.54	3.30	3.21	3.29	3.16	3.45	3.28	3.11	3.48
99.60%	Extreme value	3.83	4.29	4.15	3.35	3.32	2.84	2.54	3.32	3.59
(250 days)	Historical	3.59	3.39	3.41	3.01	3.01	3.10	2.71	3.31	3.45
	APARCH	4.13	3.92	3.85	3.73	4.88	4.02	3.55	4.77	4.67
	Gaussian	3.84	3.58	3.48	3.57	3.43	3.74	3.55	3.37	3.77
99.80%	Extreme value	4.67	5.32	5.15	3.95	3.79	3.20	2.77	3.84	4.18
(500 days)	Historical	na	na	na	na	na	na	na	na	na
	APARCH	4.88	4.61	6.51	4.99	6.51	4.91	3.63	5.51	5.43

(*Continued*)

Table 11.4 (*Continued*)

Panel B. Short position

Probability (waiting period)	Model	Open	10 am	11 am	12 pm	1 pm	2 pm	3 pm	4 pm	Close
	Gaussian	2.13	1.98	1.94	1.99	1.91	2.09	1.97	1.88	2.11
95%	Extreme value	1.70	1.76	1.80	1.65	1.96	1.74	1.77	2.06	1.94
(20 days)	Historical	1.85	1.72	1.75	1.73	2.06	2.03	1.92	2.19	2.10
	APARCH	2.28	2.14	2.33	2.16	3.12	2.86	2.66	3.33	3.24
	Gaussian	3.03	2.82	2.76	2.83	2.72	2.97	2.80	2.67	2.99
99%	Extreme value	2.69	2.91	2.98	2.41	2.67	2.31	2.16	2.89	2.77
(100 days)	Historical	2.76	2.82	2.67	2.47	2.82	2.50	2.37	2.78	2.77
	APARCH	3.51	3.38	3.68	3.13	5.22	3.97	3.33	4.51	4.51
	Gaussian	3.42	3.46	3.22	3.15	3.23	3.10	3.39	3.20	3.05
99.60%	Extreme value	3.87	3.87	3.97	2.99	3.18	2.71	2.42	3.51	3.40
(250 days)	Historical	3.70	3.01	2.90	2.58	2.97	2.70	2.48	2.96	3.20
	APARCH	4.50	4.45	3.83	3.22	5.56	4.55	3.47	4.93	5.40
	Gaussian	3.76	3.50	3.42	3.51	3.37	3.68	3.47	3.31	3.71
99.80%	Extreme value	4.80	4.80	4.93	3.53	3.63	3.05	2.63	4.06	3.96
(500 days)	Historical	na	na	na	na	na	na	na	na	na
	APARCH	4.94	4.55	3.96	3.25	5.87	4.60	3.54	5.14	5.76

Note: This table gives the margin level for a long position (Panel A) and a short position (Panel B) for different probability levels ranging from 95% to 99.8% or equivalently different waiting periods ranging from 20 trading days (1 month) to 500 trading days (2 years). Different statistical models are used: three unconditional distributions (the Gaussian distribution, the extreme value distribution and the historical distribution) and a conditional process (the Asymmetric Power ARCH or APARCH). The historical estimates are not available (na) for out of sample inferences due to data unavailability. To define the price change, the starting time (which is equal to the ending time on the following day) varies from 9 am (opening of the market) to 5 pm (closing of the market). Data are price changes of the FTSE 100 future contract over the year 2000.

prices, and for a probability of 99.8%, the minimum is obtained with 3 pm prices and the maximum for 10 am prices. The same remarks apply to a short position. These findings suggest that the daily price change distributions vary to some extent, based on different time-intervals sampled suggesting separate tail behavior for each price series.

Turning to the estimates obtained under normality, some key insights are obtained. First, the measures are almost identical for long and short positions due to the assumption of a symmetric distribution of futures price changes and an average price change close to zero over the period considered. In contrast, the extreme value distribution and the historical distribution take account of the possibility of non-symmetric features in line with the oft-cited stylized facts of financial time series, and verified for the FTSE 100 futures contract of diverging upper and lower distribution shapes. However, in line with all the estimates, diverging margin estimates occur according to the time-intervals used to define price changes. For example, for a long position and a probability of 95%, the estimated margin varies from 1.83% using 3 pm prices to 2.05% using closing prices. Traditional comparisons of extreme value and normal risk estimates suggest the latter underestimates tail behavior due to its exponential tail decline that results in relatively thin-tailed features. These findings hold for the FTSE 100 contract for high probability levels of 99.6% and 99.8%. In contrast, for the relatively low probability level of 95%, this conclusion cannot be sustained and is due to this confidence level representing a common rather than extreme threshold. For instance, the probability of this event occurring using daily data is once every 20 trading days, representing a typical event rather than an extreme one, although it is the latter events that need to be guarded against to avoid investor default.

Then turning to the historical estimates, diverging margin requirements again occur according to the time-interval chosen with the largest (smallest) estimate on a long position at the 95% level happening at 1 pm (10 am). These estimates are based on using the historical price series gathered for the year 2000. The historical estimates are confined to in-sample inferences, due to the limited number of price observations. This implies that margin setting, using the historical distribution that tries to avoid investor default, may not be able to model the events that actually cause the default, whereas in contrast, extreme value theory specifically models these tail values.

The margin requirements based on the unconditional distributions may be compared to the other estimates, such as the conditional estimates using the APARCH process. Again it is clear that estimation at different time-intervals necessitates diverging margins. For instance, the out-of-sample estimates measured at 11 am and 1 pm (3 pm) represent the largest (smallest) possible margin requirements for a long position. Comparing the extreme value and APARCH estimates provides information on the distinction between unconditional and conditional environments facing margin setters. Distinct patterns occur based on the volatility estimation for the last trading day of the sample (December 29, 2000).

An alternative way to present the results is to compute the probability for a given margin level. Results for a large and a very large futures price change, ±5% and ±10%, are given in Table 11.5. These results can be thought of as margin requirements that would be violated at certain probabilities. The results

Table 11.5 Extreme value probabilities for given margin levels

Panel A. Long position

Margin level	Open	10 am	11 am	12 pm	1 pm	2 pm	3 pm	4 pm	Close
−5%	0.39 (2.57)	0.60 (1.66)	0.54 (1.84)	0.18 (5.51)	0.12 (8.60)	0.04 (26.29)	0.00 (237.08)	0.14 (7.03)	0.22 (4.53)
−10%	0.03 (28.73)	0.06 (15.39)	0.06 (16.90)	0.01 (102.89)	0.00 (318.81)	0.00 (1418.71)	0.00 (62485.51)	0.01 (187.70)	0.01 (103.83)

Panel B. Short position

Margin level	Open	10 am	11 am	12 pm	1 pm	2 pm	3 pm	4 pm	Close
−5%	0.29 (3.49)	0.43 (2.30)	0.47 (2.11)	0.11 (8.84)	0.09 (10.68)	0.03 (34.60)	0.00 (349.89)	0.19 (5.40)	0.17 (5.81)
−10%	0.03 (38.96)	0.05 (21.34)	0.05 (19.44)	0.01 (165.15)	0.00 (395.93)	0.00 (1867.17)	0.00 (92216.90)	0.01 (144.21)	0.01 (133.13)

Note: This table gives the extreme value distribution probability levels and the corresponding waiting periods below in parentheses for given margin levels for a long position (Panel A) and a short position (Panel B). Two margin levels are considered: ±5% and ±10%. To define the price change, the starting time (which is equal to the ending time on the following day) varies from 9 am (opening of the market) to 5 pm (closing of the market). Data are price changes of the FTSE 100 future contract over the year 2000.

indicate a number of characteristics about the inherent risk in futures contracts. For instance, if a very large margin level of 10% is imposed, the probability of it being violated on any individual day is very low. For example, the probability of exceeding a price change of 10% for a long position using 10 am prices is 0.06 in contrast to 0.01 using closing prices. In terms of average waiting time-period, these extreme price movements based on 10 am prices would occur approximately once every 15 years whereas in contrast, the occurrence for close of day prices is much less likely estimated at about every 103 years. Obviously the probability of exceeding a price movement increases as the price changes decrease, so the likelihood of occurrence increases for 5% price moves. These results again imply that the starting point for the time interval used is an important factor in the setting of sufficient margin requirements as, regardless of trading position, there is a general finding that estimates taken using close of day prices are dominated by greater price movements at other intervals. In fact, there is substantial variation in the excess probability estimates for different daily intervals.

4.2. Daily margin requirement based on high-frequency price changes

Table 11.6 presents daily margin requirements obtained with 5-minute and 1-hour price changes for a long position (Panel A) and for a short position (Panel B). Margin levels are scaled to one day (see Section 2 for the presentation of the scaling method) and compared to the ones obtained directly from daily price changes. The general lack of divergence of tail estimates for different frequencies supports the invariant with respect to aggregation property. Margin estimates are presented using the extreme value scaling procedure coupled with the average estimates based on daily estimates measured at different hourly intervals. Concentrating on the more extreme 99% level, the events that occur once every 100 trading days, the scaling procedure provides robust estimates in line with the average daily values.

5. SUMMARY AND CONCLUSIONS

This chapter proposes a method to incorporate the intraday dynamics of futures prices changes in daily margin setting, thereby including lost information that is unavailable with the traditional approach of using closing prices in a marking to market system. The intraday futures price movements are relied on in two ways. First, daily prices movements and associated margins are measured using different time-intervals to define price changes, and second high-frequency 5-minute

Table 11.6 Daily margin levels obtained with the extreme value distribution based on 5-minute, 1-hour and 1-day price changes

Panel A. Long position

Probability (waiting period)	Frequency of price changes		
	5 minutes	1 hour	1 day
95% (20 days)	1.87	1.92	1.92
99% (100 days)	3.09	2.91	2.79
99.60% (250 days)	3.34	3.68	3.47
99.8% (500 days)	4.05	4.39	4.10

Panel B. Short position

Probability (waiting period)	Frequency of price changes		
	5 minutes	1 hour	1 day
95% (20 days)	1.81	1.54	1.82
99% (100 days)	3.03	2.41	2.64
99.60% (250 days)	3.12	2.99	3.32
99.8% (500 days)	3.78	3.53	3.93

Note: This table gives the daily margin levels obtained with the extreme value distribution for a long position (Panel A) and a short position (Panel B) for different probability levels ranging from 95% to 99.8% or equivalently different waiting periods ranging from 20 trading days (1 month) to 500 trading days (2 years). Three different frequencies are used to compute the price changes: 5 minutes, 1 hour and 1 day. Margin levels obtained with 5-minute price changes and 1-hour price changes are scaled to obtained daily margin levels. Margin levels obtained from daily price changes correspond to the average over the margin levels obtained with different time-intervals. Data are price changes of the FTSE 100 future contract over the year 2000.

and 1-hour price changes are used to compute margins that are then scaled to give daily estimates.

Margin requirements by definition are collateral to avoid investor default, but must also be set by the clearinghouse at a level that ensures the competitiveness of an exchange. This chapter examines margin setting in the context of investor default through statistical analysis of extreme price movements. In practice, margin setting for the FTSE 100 contract uses a customized version of the SPAN system developed by the CME. In particular, the minimum margin requirement incorporates implicitly the assumption of a Gaussian distribution for a contract's price movements, as they must be able to match three standard deviations of price changes over the previous 60-day trading period.

Alternative statistical approaches are available for margin setting with varying degrees of attractiveness, including assuming a Gaussian distribution, estimation based on the historical distribution of past price changes, conditional modeling with a GARCH process and unconditional estimation with extreme value theory. The key feature in separating out the approaches is to examine their ability in dealing with the fat-tailed characteristic of futures price movements. Model risk arises with any approach that assumes a particular distribution for price changes. For instance, conditional estimation that incorporates the time-varying properties characteristic of financial price changes still requires assumptions for the conditional price generating process. Further, the supposition of normality incorporates a relatively thin-tailed distribution and leads to an underestimation of margin levels. The historical distribution of past price changes is incapable of dealing with the extreme price movements that result in investor default focusing only on in-sample probability levels. Finally, the approach advocated here, using extreme value theory, minimizes these problems by focusing exclusively on tail price movements thereby avoiding making inappropriate assumptions on a futures contract's price generating process, and also allowing for out-of-sample extrapolation. Advantageously, this chapter merges the theoretical benefits of extreme value theory to the empirical benefits of analyzing intraday dynamics that include scaling from high to low frequency margin levels.

After identifying the fat-tailed property of the futures price changes that becomes more pronounced for relatively high-frequency realizations, this chapter identifies a number of key factors in margin setting. First and most important is the finding that intraday dynamics should be a key component in margin estimation. Daily price movements measured at different intervals can have a very tenuous relationship, suggesting that the common procedure of using only close-of-day prices neglects the dynamics that investors face in trading futures. In addition, using high-frequency intraday realizations negates this problem, even if estimating at a daily frequency through a simple scaling law of extreme value theory. Second, we illustrate the relative dominance of extreme value theory over alternative statistical methods in margin setting. The weaknesses of the other approaches, including the underestimation of Gaussian estimates in extreme price movement modeling, the inability to deal with relatively low probability levels using the historical distribution and the over reliance on a particular period of time associated with conditional estimation, are all documented.

ACKNOWLEDGEMENT

The authors would like to thank Lianne Arnold, Leonardo Becchetti John Knight, Andrew Lamb, Alan Marcus and Hassan Tehranian, participants at the Tor Vergata Annual Conference 2004 and a centre for Financial Markets research seminar for their helpful comments on this chapter. The first author acknowledges financial support from a Smurfit School of Business research grant.

BIBLIOGRAPHY

Bollerslev T., R.Y. Chou and K.F. Kroner (1992). "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence." *Journal of Econometrics* **52**: 5–59.

Bollerslev T., J. Cai and F.M. Song (2000) "Intraday Periodicity, Long Memory Volatility, and Macroeconomic Announcement Effects in the US Treasury Bond Market." *Journal of Empirical Finance* **7**: 37–55.

Booth G.G., J.-P. Broussard, T. Martikainen and V. Puttonen (1997) "Prudent Margin Levels in the Finnish Stock Index Market." *Management Science* **43**: 1177–88.

Brennan M.J. (1986) "A Theory of Price Limits in Futures Markets," *Journal of Financial Economics* **16**: 213–33.

Brenner T.W. (1981) "Margin Authority: No Reason for a Change." *Journal of Futures Markets* **1**: 487–90.

Cotter J. (2001) "Margin Exceedances for European Stock Index Futures using Extreme Value Theory." *Journal of Banking and Finance* **25**: 1475–502.

Cotter J. (2004) "Minimum Capital Requirement Calculations for UK Futures." *Journal of Futures Markets* **24**: 193–220.

Craine R. (1992) "Are Futures Margins Adequate?" Working Paper, University of California – Berkley.

Dacarogna M.M., O.V. Pictet, U.A. Muller and C.G. de Vries (1995) "Extremal Forex Returns in Extremely Large Data Sets." Mimeo, Tinbergen Institute.

Danielson J., L. de Haan, L. Peng and C.G. de Vries (2001) "Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation," *Journal of Multivariate Analysis* **76**: 226–48.

Day T.E. and C.M. Lewis (1999) “Margin Adequacy and Standards: An Analysis of the Crude Oil Futures Markets.” Working Paper, Owen Graduate School of Management, Vanderbilt University.

Ding Z.C., C.W.J. Granger and R.F. Engle (1993) “A Long Memory Property of Stock Returns.” *Journal of Empirical Finance* **1**: 83–106.

Drost F.C. and T.E. Nijman (1993) “Temporal Aggregation of GARCH Processes.” *Econometrica* **61**: 909–27.

Edwards F.R. and S.N. Neftci (1988) “Extreme Price Movements and Margin Levels in Futures Markets.” *Journal of Futures Markets* **8**: 639–55.

Feller W. (1971) *An Introduction to Probability Theory and its Applications*. New York: John Wiley.

Figlewski S. (1984) “Margins and Market Integrity: Margin Setting for Stock Index Futures and Options.” *Journal of Futures Markets* **4**: 385–416.

Gay G.D., W.C. Hunter and R.W. Kolb (1986) “A Comparative Analysis of Futures Contract Margins.” *Journal of Futures Markets* **6**: 307–24.

De Haan L.S., I. Resnick, H. Rootzen and C.G. de Vries (1989) “Extremal Behavior of Solutions to a Stochastic Difference Equation with Applications to ARCH processes.” *Stochastic Process and their Applications* **32**: 213–24.

De Haan L.S., D.W. Jansen, K. Koedijk and C.G. de Vries (1994) “Safety First Portfolio Selection, Extreme Value Theory and Long Run Asset Risks,” in Galambos Ed., *Proceedings from a Conference on Extreme Value Theory and Applications.* Dordrecht: Kluwer Academic Publishing. pp. 471–87.

Gnedenko B.V. (1943) “Sur la Distribution Limite du Terme Maximum d’une Série Aléatoire.” *Annals of Mathematics* **44**: 423–53.

Hall P. and A. Welsh (1984) “Best Attainable Rates of Convergence for Estimates of Parameters of Regular Variation.” *Annals of Statistics* **12**: 1072–84.

Huisman R., K. Koedijk, C.J.M. Kool and F. Palm (2001) “Tail Index Estimates in Small Samples.” *Journal of Business and Economic Statistics* **19**: 208–15.

Kearns P. and A. Pagan (1997) “Estimating the Density Tail Index for Financial Time Series.” *Review of Economics and Statistics* **79**: 171–5.

Kofman P. (1993) “Optimizing Futures Margins with Distribution Tails.” *Advances in Futures and Options Research* **6**: 263–78.

Longin F.M. (1995) “Optimal Margins in Futures Markets: A Method Based on Extreme Price Movements.” *Proceedings of the CBOT Conference* “*Futures and Options.*” Bonn, Germany.

Longin F.M. (1996) “The Asymptotic Distribution of Extreme Stock Market Returns,” *Journal of Business* **63**: 383–408.

Longin F.M. (1999) “Optimal Margin Levels in Futures Markets: Extreme Price Movements.” *Journal of Futures Markets* **19**: 127–52.

Longin F.M. (2000) “From Value at Risk to Stress Testing: The Extreme Value Approach.” *Journal of Banking and Finance* **24**: 1097–130.

Longin F.M. and B. Solnik (2001) “Extreme Correlation of International Equity Markets.” *Journal of Finance* **56**: 651–78.

Merton R.C. (1980) "On Estimating the Expected Return on the Market." *Journal of Financial Economics* **8**: 323–61.

Warshawsky M.J. (1989) "The Adequacy and Consistency of Margin Requirements: The Cash, Futures and Options Segments of the Equity Markets." *Review of Futures Markets* **8**: 420–37.

Appendix

Estimation of the shape parameter

This Appendix describes the semi-parametric estimation procedure for the shape parameter of the extreme value distribution.

The widely used Hill (1975) moment estimator is used to determine tail quantiles and probabilities. The Hill estimator represents a maximum likelihood estimator of the tail index, the inverse of the shape parameter:

$$\gamma = 1/\alpha = (1/\mathrm{m})\Sigma[\log \mathrm{r}_{(\mathrm{n}+1-\mathrm{i})} - \log \mathrm{r}_{(\mathrm{n}-\mathrm{m})}] \quad \text{for } \mathrm{i} = 1. \ldots .\mathrm{m} \tag{A1}$$

focusing on the maximum upper order statistics. This tail estimator is asymptotically normal (de Haan *et al.* 1994):

$$(\mathrm{m})^{1/2}/(\mathrm{r}_{\mathrm{m}+1} \log(\mathrm{m}/\mathrm{np}))(\mathrm{r}_\mathrm{p} - \mathrm{E}\{\mathrm{r}_\mathrm{p}\}) \approx \mathrm{N}(0, \gamma^2) \tag{A2}$$

An estimation issue is determining the optimal number of tail values, m (Danielson *et al.* 2001, for a discussion). The dilemma faced is that there is a trade-off between the bias and variance of the estimator, with the bias decreasing and variance increasing with the number of values used. The approach introduced by Huisman *et al.* (2001) is applied here that performs well under simulation. The use of the Hill estimator in the literature is due to a number of factors. The estimator is the most widely used with the most desirable time series properties (Hall and Welsh 1984), with specific support for its application to financial time-series from simulation studies of it versus other estimators based on order statistics (Kearns and Pagan (1997)). Also, the Hill estimator does not require the existence of a fourth moment, a characteristic that is strongly debated for financial data. Most importantly, the Hill estimator is the intrinsic part of a larger procedure used in this study that examines tail behavior. In fact, Dacarogna *et al.* (1995) show, that by applying the highest frequency data, possibly ensures that the shape parameter provides the most efficient estimator of tail behavior exploiting the fractal nature of extremes. Intuitively a large (high) frequency data set has more observable extremes that a small (low) frequency one over the same time interval, thereby allowing for stronger inferences of these rare events. Thus, estimation of relatively low frequency margins is best achieved by estimating shape parameter values at high frequencies and using a simple scaling law to extend for these aggregated price changes. A simple scaling factor, similar to the $\sqrt{n}$ used for normal distribution, is applicable. The high-frequency margin estimates are adjusted by an α-root scaling law scaling ($k^{1/\alpha}$), with no additional estimation of extra parameters required.

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 12

NONPARAMETRIC TECHNIQUES TO VALIDATE CREDIT CLASSIFICATION MODELS: AN EMPIRICAL ANALYSIS

S. Borra, S. Caiazza and S. Di Giacomo

Keywords: bankruptcy risk, validation, cross-validation, holdout, classification tree, bagging, boosting, logit

1. INTRODUCTION[1]

There is a rich literature on empirical classification models within a credit risk framework, which predict the default event of firms or banks. Each author applies one or more models in an attempt to fit the non-random component of default.

Many authors have proposed a wide variety of classification methods to forecast the default event. In the first contribution proposed by Altman (1968), default is modeled using linear discriminant analysis. Altman *et al.* (1972) use quadratic discriminant analysis. Moody's (2001), Krainer and Lopez (2001) and Demirgüc-Kunt and Detragiache (1999) estimate the logit model, while Eichengreen and Rose (1998) and Eichengreen and Arteta (2000) use the probit model. Tam and Kiang (1993) and Galindo and Tamayo (2000) apply classification trees (CARTs) and Gonzales and Hermosillo (1999) use survival analysis. Borra and Caiazza (2002) aggregate predictors, while Alfò, *et al.* (2005) use the mixture

[1] We are indebted to Michele Bagella for guidance and encouragement. We also thank Iftekar Hasan and participants to the XII International Tor Vergata Conference on Banking and Finance: "Transparency, Governance and Markets", held in Rome 1–3 December 2004, for helpful comments and suggestions. The usual disclaimer applies.

logit model. Each of these contributions tries to predict default using qualitative and quantitative input variables. However, the empirical literature rarely tackles the problem on how to validate these models.

The importance of the validation process is related to the high probability of nonparametric models in finding a perfect fit to data without any real generalization capability. This is a well-known overfitting problem. Moreover, validation of parametric models plays an important role, since the prediction capability could depend on the specific sample drawn. In this case, the attempt to generalize results to new observations that could lead to poor results. Economic analyses based on the sign and the magnitude of the biased model parameters could lead to wrong policy choices and, in the framework of credit risk analysis, wrong lending decisions.

The relevance of validation process, for internal rating purposes, has been stressed by the Basel Committee. Their document issued in 2003 says that "institutions employing internal models for regulatory capital purposes are expected to have in place a robust system to validate the accuracy and consistency of the model and its inputs."[2]

Sobehart *et al.* (2001) emphasizes the importance of the experiment design to validate classification models. The prediction capability of models should be verified on sample(s) with different statistical units with respect to the estimation sample. Soberhart *et al.* propose to validate the model using out-of-sample, out-of-universe and out-of-time testing approaches on panel or cross-sectional datasets.

In the *out-of-sample approach*, the strategy is to split data into two mutually exclusive sub-samples, using usually 605 to 75% to estimate the model (the *in sample set*) and the remainder (the *out of sample set*) to validate it. Observations of both samples, belonging to the same population, are detected at the same time.

In the *out-of-universe approach* the *in* and *out of sample* sets belong to different populations over the same period of time. In the *out-of-time approach*, *in* and *out of sample* sets, belong to different period of time.

The main shortcoming of the latter approach is related to the underlying and never tested assumption of stationarity. When lacking this, estimates are biased with poor *out of sample* forecasts. This approach has been applied in an Italian economic context by Fabi *et al.* (2002), to test the effects of default on manufacturing firms in severe macroeconomic conditions. They estimate a logit model, from data collected in 1999–2000 replacing the validation analysis of input variables (balance-sheet ratios) with the same ratios observed in 1993–94,

[2] Basel Committee (2003) "*The New Basel Capital Accord*", April 2003, p. 97.

the years of a severe slow down in Italy. In this case there is a high probability of reaching an overestimation of error rate for defaulted firms.

The out-of-universe approach could lead to explicability and predictability problems. A good *in sample* estimation on observations drawn from a population of manufacturing firms and a poor *out of sample* prediction on observations drawn from a population of tourist firms, does imply a bad performance of the model or a different dynamic of the default event between the two populations? Are input variables and estimated parameters able to explain and predict default of two mutually exclusive populations?

In this study we focus on the prediction capability of the classification model. The aim of this analysis is to compare several validation techniques applied to classification models, both parametric and nonparametric, analyzing the influence of sampling variability on misclassification errors. In Section 2, we describe how it is possible to evaluate the prediction capability of a classification model. In Section 3, we describe the dataset used in the analysis and the design of the experiment chosen for different validation strategies. In Section 4, we show results. The last section is devoted to conclusions.

2. HOW EVALUATE PREDICTION CAPABILITY

To compare the prediction capability of several parametric and nonparametric classification models, we use a dataset with a nominal variable Y, for example a dummy variable $Y = \{1, 0\}$, indicating default (unsound firm) and not default (sound firm), and a vector of K explanatory variables $\mathbf{X} = (X_1, \ldots X_K)$.

It is possible to obtain an estimate of the prediction capability considering two different samples drawn from the same observed dataset: the *in sample* set, *in sample* $= \{(y_i, x_{i1}, x_{i2}, \ldots, x_{iK})\}$ $i = 1, 2, \ldots, n$, and the *out of sample* set, *out of sample* $= \{(y_h^*, x_{h1}^*, x_{h2}^*, \ldots, x_{hK}^*)\}$ $h = 1, 2, \ldots, m$. Then, given a classification model C, the predicted class for the i-*th* unit is $\hat{y}_i = C(\mathbf{x}_i)$.

A measure of the prediction capability is given by the estimated prediction error, obtained as proportion of *out of sample* cases misclassified by the model C:

$$P_e(C) = \frac{1}{m} \left|\{k : C(\mathbf{x}_k^*) \neq y_k^*\}\right|.$$

Whenever an extremely large dataset is available (related to the number of explanatory variables and complexity of the problem), it is possible to randomly split the original set into two large datasets, *in* and *out of sample* sets. The estimated classification model is stable (in the sense that it does not vary, so changing few cases) and $P_e(C)$ is a reliable estimation of the true prediction error.

In the presence of small to medium datasets concerning complex classification problems, the previous estimator is not the best choice. In this case, it is possible to use the same dataset for model building and assessment. There are a wide variety of empirical methods proposed in the literature to estimate the prediction error. The simplest method is the *apparent error*, corresponding to the proportion of misclassified cases by the model, which is estimated on all cases of the dataset. However, the estimated error is too optimistic, the *k-fold cross-validation* producing a more realistic estimation of error rate. When adopting this approach, cases are randomly divided into k $(1 \leq k \leq n)$ mutually exclusive validation partitions of approximately equal size. Cases not included in each partition are independently used for estimating the model. The average error rate over all k partitions is the k-fold cross-validation error rate. Friedman *et al.* (2001) show that variance of the k-fold cross-validation estimator decreases rising the parameter k, but bias becomes larger. However Kohavi (1995) remarks that, in small sample size, bias and variability increase, so decreasing the value of k. In the *stratified k-fold cross-validation*, the same distribution of the classification variable Y observed in the dataset is maintained in each of the k partitions.

Different estimators of prediction error are based on bootstrap techniques, by resampling with the replacement of several samples from the dataset. The holdout technique randomly splits the dataset in two mutually exclusive subsets (e.g., *in sample* set containing 67% of the cases and *out of sample* set with 33% of cases). The *in sample* set is used to estimate the classification model and the *out of sample* set to evaluate the prediction capability. In the *holdout random resampling*, holdout is repeated many times, randomly splitting the dataset and averaging the estimates. Comparing different classification models requires the evaluation of bias and variability of the prediction error estimator. Davison and Hall (1992) and Kohavi (1995) show, in simulation analyses, that k-fold cross validation estimator has higher variance but lower bias than bootstrap. Stratified holdout random resampling is biased but variance is lower than k-fold cross validation for large numbers of replications (Borra and Di Ciaccio, 2004). In this analysis, we consider two estimators, stratified holdout random resampling and repeated stratified k-fold cross-validation. The last estimator is obtained repeating T times the k-fold cross-validation and averaging over the T estimates.

3. EMPIRICAL DATASET AND DESIGN OF THE EXPERIMENT

In this section, we describe the dataset used in our experiment, classification models and validation techniques applied.

3.1. Data

The data considered in our experiment are collected by Bureau van Dijk S.A. on 230,000 Italian firms of varying size. Selecting for juridical status, default information and activity sector, we derived a subsample of 39,237 manufacturing firms for the years 1999–2001.

The event of default (unsound firms) is recorded when firms are bankrupted or are going to be liquidated. The dummy variable *status* assumes value 0 for all sound firms and value 1 for the unsound ones. In Table 12.1 we report the distribution of the default event through years 1999–2001.

We select 22 balance-sheet indicators (liquidity, profitability and financial ratio) reported in Appendix 2, and three qualitative variables; *Area* (North-East, North-West, Center, South and Isles), *Size* (small, medium and large firms, classified using the number of employees), *Districts* (based on the classification produced by Istat – National Institute of Statistics – of the Italian areas that can be classified as industrial districts).

Tables 12.2 and 12.3 report on firms included in the analyzed sample, by locality and size.

Table 12.1

Area	Freq.	Percent	Cum.
NorthEast	5776	14.72	14.72
NorthWest	17,810	45.39	60.11
Centre	8389	21.38	84.49
South	7262	18.51	100
Total	39,237	100	

Table 12.2

Size	Freq.	Percent	Cum.
$[0, 15)$	10,072	25.67	25.67
$[15, 50)$	19,124	48.74	74.41
$[50, 250)$	7180	18.30	92.71
$[250, +)$	2860	7.29	100
Total	39,237	100	

Table 12.3

	1999	2000	2001
Sound	39,237	38,788	38,626
Unsound		449	162

3.2. Design of the experiment

We compare the prediction performance of four different classification models, logit, CART and two aggregate classifiers.[3] The use of different models avoids the dependency of validation results to the particular choice of model. Logit is a common parametric approach widely described in literature for classification purposes, as reported in Barniv and McDonald (1999) – 178 articles in accounting and finance journals between 1989 and 1996 used this model. For prediction purposes, nonparametric models present excellent results due to their robustness in picking up nonlinear relationships in the data. We than estimate a parametric model, the logit, and nonparametric models as CARTs and two aggregate classifiers, obtained with bagging (Breiman 1996) and boosting (Freund and Schapire 1997) techniques (described in Appendix 1) combined to CART. The last two classifiers take better account the overfitting problem, improving the prediction capability as shown, for example, in Borra and Caiazza (2002).

We focus on the prediction capability of the model, i.e. its capacity to correctly predict future status of debtors, rather than the explanatory capability. In a credit classification framework, both in monitoring and screening procedures, it is crucial to correctly evaluate the reliability of the model because of the different costs of misclassification between sound and unsound firms.

The goodness of these models is evaluated considering the type I error (classify a firm as sound while it is unsound), type II error (classify a firm as unsound while it is sound) and total error rate in the *out of sampl*e set (the rate of misclassified observations in the *out of sample*), using the repeated stratified k-fold cross validation, the stratified hold-out random resampling and the out-of-time sample.

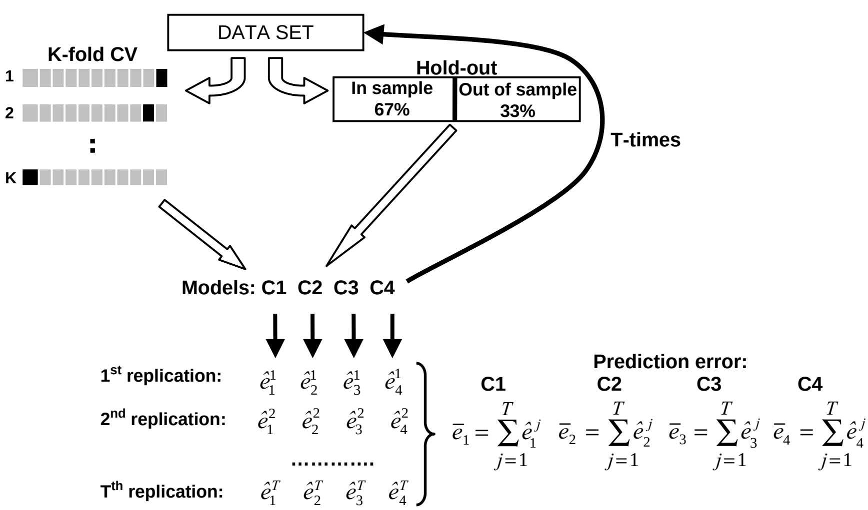
We apply these techniques since, from a theoretical point of view, the credit classification problem is a complex challenge due to lack of an underlying microeconomic theory that explains the default event. From an empirical point of view, credit classification analyses are based on datasets that are not clean and

[3] We used Weka package which is open source software issued under the GNU General Public License.

completely reliable, due to measuring errors in the data and to window-dressing policies on balance sheets.

In the *stratified holdout random resampling* (SHRR) approach we randomly split the dataset into two mutually exclusive subsets (*in sample* and *out of sample*), holding the same proportion between the two classes in both subsets (stratification).[4] We use the *in sample* set to estimate the classifiers, evaluating the prediction capability on the *out of sample* set. We repeat the holdout T times (T = 1, 10, 20, 30), randomly splitting the dataset. The holdout random resampling estimate is obtained by averaging the T estimates.[5]

In the *repeated stratified k-fold cross validation* (SKCV) approach, we randomly split the dataset into k mutually exclusive subsets holding the proportion between classes in all the k subsets and then using $k-1$ subsets to estimate the classifiers, evaluating the prediction error(s) on the k^{th} subset. We estimate the prediction capability on each subset to get an average error over k ($k=3$, 10, 20, 30). We repeat the procedure T times (T = 1, 10, 20, 30), randomly splitting the dataset. The repeated stratified k-fold cross validation is obtained averaging the T estimates. In Figure 12.1 we show both procedures, where C1–C4 indicate



$$\hat{e}_1^1 \quad \hat{e}_2^1 \quad \hat{e}_3^1 \quad \hat{e}_4^1$$
$$\hat{e}_1^2 \quad \hat{e}_2^2 \quad \hat{e}_3^2 \quad \hat{e}_4^2$$
$$\hat{e}_1^T \quad \hat{e}_2^T \quad \hat{e}_3^T \quad \hat{e}_4^T$$
$$\bar{e}_1 = \sum_{j=1}^{T} \hat{e}_1^j \quad \bar{e}_2 = \sum_{j=1}^{T} \hat{e}_2^j \quad \bar{e}_3 = \sum_{j=1}^{T} \hat{e}_3^j \quad \bar{e}_4 = \sum_{j=1}^{T} \hat{e}_4^j$$

Fig. 12.1.

[4] The use of stratification usually reduces the estimator's variability (Borra and Di Ciaccio, 2004).
[5] When T = 1, we are simply splitting data into the *in sample* and *out of sample* sets and, of course, it is not possible to evaluate the estimator variability.

the different classification models, $\hat{e}_j^t$ the misclassification error estimated at the t^{th} replication for the j^{th} model, and $\bar{e}_j$ the estimated prediction error for the j^{th} model.

We apply the SHRR and the SKCV to a sample of 22,385 sound and 449 unsound firms during years 1999 and 2000. The unsound event is observed in 2000, while covariates are observed in 1999 to avoid a possible endogeneity problem.

Both methods, based on resampling, allow us to estimate sampling variability of the estimators. This aspect points out how the sole use of a single *out of sample* set can lead to misleading validation results in term of prediction capability.

Supposing the observed series are generated by a stationary process, we can use the *out-of-time* validation methodology as a further procedure to validate models. In this approach we use a period of time to estimate models and a different (following) period of time to validate them. We apply the out-of-time methodology estimating the models in 1999–2000 (covariates in 1999, default event in 2000; 22,385 sound firms and 449 unsound firms) and predicting results in years 2000–01 (2000 covariates, default event in 2001, 10,240 sound and 162 unsound firms).

The *in sample* set is made up by all firms with unsoundness in 2000, while the *out of sample* set is made up by all firms who failed in 2001. To guarantee the independence of sound firms, we randomly draw firms for the *in sample* and *out of sample* sets, in order to leave unchanged the percentage of sound firms across the two sets.[6]

4. RESULTS

Evaluating SHRR and SKCV results show it to be necessary to consider the values of fold (k) and the number of replications (T).

Graphs in Appendices 3–7 show estimated error rates for values of k considered in the cross validation (3, 10, 20, 30) and for the number T of replications (1, 10, 20, 30). So, for example, the error rate associated to T = 10 and $k = 3$ represents the total average error rate on 10 replications of a 3-fold cross validation. For a given value of k, and the higher the value of T, the higher is the

[6] Our approach is slightly different from Sobehart *et al*. (2001) and Fabi *et al*. (2002) out of time analyses. In the *out of sample* set they use the same cases of the *in sample* set observed in a different period of time. We use, in the *out of sample*, different cases in respect to *in sample set* observed one year ahead, randomly drawn from the same population, to enforce the independency between the two samples.

number of estimates produced, while increasing the value of k decreases the size of each k^{th} in the *out of sample* set.

The relevant parameter of holdout procedure is the value of T and the percentage of splitting between *in* and *out of sample*. Following the literature, we set the latter parameter equal to 67% of observations in the *in sample* set and the remaining in the *out of sample* set. However, we tried with different percentages of splitting (90% and 95% in the *in sample set* without showing any relevant difference of results).

SKCV results, reported in Appendix 3, show the effect of the number T of replications on the estimated error rate. With no replication (T = 1, just one *out of sample*) we get a predicted error rate highly different with respect to predicted error rates evaluated at higher values of T. In this case, the number k of folds is relevant for the logit model (the error rate decreases at higher values of k), while nonparametric models are less sensitive. At a higher number of replications, estimated error rates are more stable. The three nonparametric models do not show different behavioral patterns in the error rates. This result seems to indicate that, in this particular analysis, the problem is not complex and there is not an effective improvement when using aggregate predictors. Logit model is more sensitive to the number of folds and the total error rate is always higher than nonparametric models.

For type I error rate (Appendix 4), the number of folds becomes more important. Again, without any replication, the type I error rate strongly depends on the value of k, from the classification model. It is higher than the total prediction error rate, with an impact of k particularly relevant for logit model. The difference between the worst (3-fold) and the better (10-fold) result is almost twofold. There is no evidence that a higher number of replications reduces type I error rate for a given fold value.

SKCV allows the estimation of the standard deviation of the k-fold cross validation estimator. In the first graph in Appendix 5, for $k = 20$, we compare the estimated standard deviation of k-fold cross validation for each model and each value of T. The smaller standard deviation is reached by CART while the highest by logit. In the second graph, by fixing the number of replication to 30, it is possible to evaluate the effect of folds on standard deviation. It is stable for a value of k greater than 10 independently by classifiers.

SHRR shows a deeper effect of number of replications T on error rates for all models (Appendix 6): total prediction error rate increases with the number of replications. In other words, by validating the model by just using one *out of sample* leads to underestimate the error rate. The rate of variation of error decreases with the rising the number T and, from T = 3, becomes stable. As shown with SKCV, nonparametric models achieve lower error rates than logit model.

The behavior of standard deviation is similar for all models (Appendix 7) and is stable when increasing the number of replication.

The comparison between SHRR and SKCV does not suggest a relevant difference in the predicted error rates. The choice of parameter k in SKCV is important for a low value of T, while increasing the number of replications estimates become more stable.

Both validation methodologies are time-consuming, which depends on the value of k, T, the total number of observations and the number of variables. Further comparison between SHRR and SKCV can be obtained by working out the machine-time needed for estimations. However, it is evident that the two methodologies have to be compared on the same number of equal size subsamples. For example, SHRR threefold 10 replications can be compared to SKCV 67/33% 30 replications and SHRR threefold 1 replication with SKCV 67/33% 3 replications.

In Table 12.4 we report the machine-time in seconds for computations on P4 processor CPU 3.20 GHZ, 1 GB Ram.

Estimating a larger number of samples is more time-consuming. The logit model requires more time than the nonparametric models. It is five times greater than CART for SKCV $k = 3$ and $T = 10$.

There is no particular difference when using the two validation methodologies for nonparametric models, while SHRR is faster for logit model.

Results of out of time validation are reported in Table 12.5. In Table 12.6 are shown results of SKCV $k = 30$ and $T = 30$.[7]

The error rates under the out-of-time validation procedure are similar for the three nonparametric models and much larger for logit.

Table 12.4

	Logit	**CART**	**Bagging-Cart**	**Boosting-Cart**
SKCV $k = 3$, $T = 10$	906	177	526	498
SHRR 67/33 $T = 30$	769	169	410	448
SKCV $k = 3$ $T = 1$	134	19	51	61
SHRR 67/33 $T = 3$	83	18	55	61

[7] We consider the SKCV $k = 30$ $T = 30$ the best result in term of estimation reliability.

Table 12.5

Out of Time	Total Prediction Error (%)	Type I Prediction Error (%)
Logit	0.141	9.845
CART	0.070	2.070
Bagging-Cart	0.070	2.070
Boosting-Cart	0.060	2.070

Table 12.6

SKCV $k = 30$ T $= 30$	Total Prediction Error (%)	Type I Prediction Error (%)
Logit	0.100	2.810
CART	0.040	1.670
Bagging-Cart	0.040	1.540
Boosting-Cart	0.050	1.540

More interesting, results are not very dissimilar from SKCV, with the relevant exception of type I predicted error rate in logit model, which in the out of time (1 sample) is four times greater than under SKCV.

5. CONCLUSIONS

In this study, in a credit risk framework, we compared the performance of three different validation techniques, stratified k-fold cross validation (SKCV), stratified random hold-out (SHRR), out-of-time evaluated using four models, one parametric, the logit model, and three nonparametric, CART, bagging-cart and boosting-cart. The choice of nonparametric models is due to the relevance, in our study, of the prediction capability of the model rather than the explanatory aspect.

Two out of three approaches, SKCV and SHRR, are obtained by averaging several estimates of error rate. These two validation techniques show that the accuracy and reliability of the predicted error rate is higher that one single replication estimates. This result is independent of the classifier chosen. Another important result, independent by the validation techniques, is the poor performance of logit, since error rates are always greater than error rates of nonparametric models. Results of the latter models are similar and aggregated predictors do not improve the prediction capability. With SHRR, logit shows higher sensitivity to replication T, since total predicted error rates grow while being stable for $T = 3$ onwards in nonparametric models.

In SKCV, the effect of replications and folds is more relevant for logit than nonparametric models. The standard deviation of the predicted error rates is stable with respect to the number of the folds and replications. The choice between these two validation techniques has to consider several aspects. SHRR implies the choice of the numerousness of the *in* and *out of sample* sets (67/33 in this study) and the number of replications T. The higher the latter parameter, the more reliable are the estimated error rates, but computational cost becomes relevant. Predicted error rates are the error rates of each single replication averaged over the total number of replications. SKCV implies the choice of folds, i.e. the number k of samples to estimate in a k-fold cross validation, and the number of replications T. Larger value of k implies a higher number of smaller size *out of samples*. SKCV predicted error rate is an average of T k-fold cross validations. Given an equal number of estimations, computational cost does not vary between SKCV and SHRR, with the exception of logit model (SKCV requires more machine-time than SHRR), which is always more time-consuming than the other classifiers.

The out-of-time validation technique, which is the predicted error rates on a single *out of sample* formed with new observations detected at a future period of time (one year ahead in this study), always shows higher error rates with respect to the "best" SKCV. In this case, it is important to take into account the problem of the stationarity of the series, which should be explicitly tested.

The main result of this study is the great importance of the validation strategy based on replicated estimations to achieve a more reliable estimated prediction error rates. In our opinion the use of these techniques in a credit risk framework is crucial because of availability and reliability problems of banks' datasets. Frequently reliable data are simply small sets of the whole dataset, while it is relevant to reach a reliable estimation of error rates for the whole dataset. Then validation methodologies based on replicated estimations can fit better for this requirement.

BIBLIOGRAPHY

Alfò, M., Caiazza, S. and Trovato, G. (2005) "Enhancing Logistic Approach for Credit Risk Modelling: the Mixed Effect Approach." *Journal of Financial Services Research*, forthcoming.

Altman, E.I. (1968) "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." *The Journal of Finance* **23**: 589–609.

Altman, E., Haldeman, R. and Narayanan, P. (1972) "Zeta analysis. A new model to identify bankruptcy risk of corporations." *Journal of Banking and Finance* **1**(1): 29–54.

Barniv, R. and McDonald, J.B. (1999) "Review of Categorical Models for Classification issues in Accounting and Finance." *Review of Quantitative Finance and Accounting* **13**: 39–62.

Basel Committee on Banking Supervision (2003) *The New Basel Capital Accord*. Consultative Document, April, Basel.

Bauer, E. and Kohavi, R. (1999) "An empirical comparison of voting classification algorithms: bagging, boosting, and variants." *Machine Learning* **36**: 105–13.

Borra, S. and Caiazza, S. (2002) "Comparative performance of credit scoring models using aggregated predictors," *Data Mining III*, WIT Press, pp. 747–56.

Borra, S. and Di Ciaccio, A. (2001) "Reduction of prediction error by bagging projection pursuit regression," in: Borra S. *et al*. (eds), *Advances in Classification and Data Analysis*. Berlin: Springer-Verlag.

Borra, S. and Di Ciaccio, A. (2004) "Methods to compare nonparametric classifiers and to select the predictors," in M. Vichi *et al*. (eds) *New Developments in Classification and Data Analysis*, Studies in Classification, Data Analysis and Knowledge Organization. Berlin: Springer-Verlag. pp. 11–20.

Breiman, L. (1996) "Bagging Predictors." *Machine Learning* **26**: pp. 123–40.

Breiman, L. (1998) "Half and half bagging and hard boundary points." *Technical Report* No. 534, Department of Statistics, University of California.

Breiman, L. (1999) "Using adaptive bagging to debias regressions." *Technical Report* No. 547, Department of Statistics, Stanford University, February.

Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984) *Classification and Regression Trees*. Wadsworth.

Burman, P. (1989) "A comparative study of ordinary cross-validation, hold cross-validation and repeated learning testing methods." *Biometrika* **76**: 503–14.

Caiazza, S. (2004) "The comparative performance of credit scoring models: an empirical approach," in *Monetary Integration, Markets and Regulation, Research in Banking and Finance*, **4**, Elsevier Science Ltd. pp. 17–65.

Davison, A. and Hall, P. (1992) "On the bias and variability of bootstrap and cross-validation estimates of error rate in discriminant problems." *Biometrika* **79(2)**: 279–84

Demirgüc-Kunt, A. and Detragiache, E. (1999) "Monitoring Banking Sector Fragility: A Multivariate Logit Approach." IMF Working Paper, 47, Washington: International Monetary Fund.

Dietterich, T.G. (2000) "An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization." *Machine Learning* **40**: 139–58.

Eichengreen, B. and Arteta, C. (2000) "Banking Crises in Emerging Markets: Presumptions and Evidence." *Haas School of Business, University of California Berkeley, Center for International and Development Economic Research*, Working Papers, 115.

Eichengreen, B. and Rose, A.K. (1998) "Staying Afloat When the Wind Shifts: External Factors and Emerging-Market Banking Crises." *NBER Working Paper*, 6370.

Freund, Y. and Schapire, R. (1997) "A decision-theoretic generalization of on-line learning and an application to boosting." *Journal of Computor System Science* **55(1)**: 119–39.

Friedman, J. (2002) "Stochastic Gradient Boosting." *Computational Statistics & Data Analysis* **38(4)**: 367–78.

Galindo, J. and Tamayo, P. (2000) "Credit Risk Assessment using Statistical and Machine Learning: Basic Methodology and Risk Modelling Applications." *Computational Economics* **15**: 107–43.

Gonzales, B. and Hermosillo, B. (1999) "Determinants of Ex-Ante Banking System Distress: A Macro-Micro Empirical Exploration of Some Recent Episodes." IMF Working Papers, 33.

Kohavi, R. (1995) "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 1137–43. San Mateo, CA: Morgan Kaufmann.

Krainer, J. and Lopez, J.A. (2001) "Incorporating Equity Market Information into Supervisory Monitoring Models." *Federal Reserve Bank of San Francisco Working Paper*, 14.

Moody's, (2001) RiskCalcTM Public – Europe, www.moodys.com

Quinlan, J.R. (1996) "Bagging, boosting, and C4.5." In: *Proceedings of the 13th National Conference on Artificial Intelligence*, pp. 725–30.

Sobehart, J., Keenan, S. and Stein, R. (2001) "Benchmarking Quantitative Default Risk Models: a Validation Methodology." *Algo Research Quarterly* **4**: 55–69.

Tam, K. and Kiang, M. (1993) "Managerial applications of neural networks: the case of bank failure prediction," in *Neural Networks in Finance and Investing*, Probuis Publishing Company.

Appendix 1

Bagging and Boosting

The *bagging* technique was proposed by Breiman (1996). Given a in sample dataset with N cases $\mathbf{t} = \{y, \mathbf{x}\}_1^{\mathrm{N}}$ and Y, a J-class response variable, we indicate with $h(\mathbf{x}|\mathbf{t})$ a classification function estimated on the in sample dataset $\mathbf{t}$. Bagging combines, with same weights, K classification functions $h_k(\mathbf{x})$ calculated on a set of K bootstrapping samples of in sample dataset. The bagging aggregated estimator $h(\mathbf{x})_{bagg}$ is a J-vector $(p_1, \ldots, p_J)$, with p_i equal to the proportion of classification functions predicting class i at $\mathbf{x}$. The predicted class is the one corresponding to maximum p_i, hence it uses the majority vote criterion.

Many simulation studies (Quinlan, 1996; Breiman, 1998, 1999; Bauer and Kohavi, 1999; Dietterich, 2000; Borra and Di Ciaccio, 2001) showed the efficiency of bagging in the reduction of the prediction error. One possible interpretation is related to the bias/variance trade-off, since it reduces the variance of the aggregated predictor, maintaining almost constant bias.

The second technique, proposed by Freund and Schapire (1997), is "AdaBoost.M1" and belongs to the class of *boosting* methods. Let us consider a two-class problem, with classes coded as $\{-1, +1\}$. This method proposes to apply the classification function to a sequence of bootstrapped samples obtaining K classification functions $h_k(\mathbf{x})$. At each step, a bootstrapped sample is obtained assigning different probabilities of inclusion of the observations. It gives a higher probability of inclusion to badly predicted observations at previous steps. The boosting predictor combines the predictions from all of them through a weighted majority vote:

$$h(\mathbf{x})_{boos} = sign\left(\sum_{k=1}^{K} \alpha_k h_k(\mathbf{x})\right)$$

The weight α_k is higher when corresponding to the more accurate classification capability of $h_k(\mathbf{x})$. Many simulation studies proved that using bagging and boosting conjointly to nonparametric models (for instance CARTs) it is possible to reach a lower misclassification rate, and generally boosting works better than bagging (Breiman, 1999). Recently, an explanation has been described for the boosting performance in terms of a gradient descent algorithm with respect to a loss function (Friedman, 2002).

For this analysis we used an unpruned CART (J48 algorithm), boosting with four aggregations combined with unpruned CART (AdaboostJ48 algorithm), and bagging with four aggregations combined with unpruned CART (BaggJ48 algorithm).

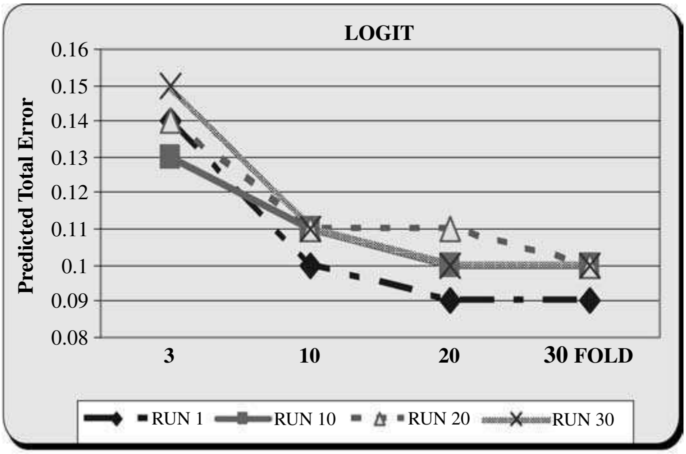
Appendix 2

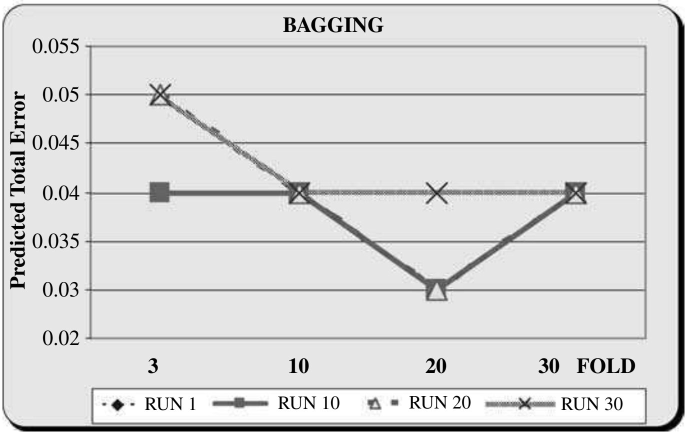
Balance Sheet Ratios

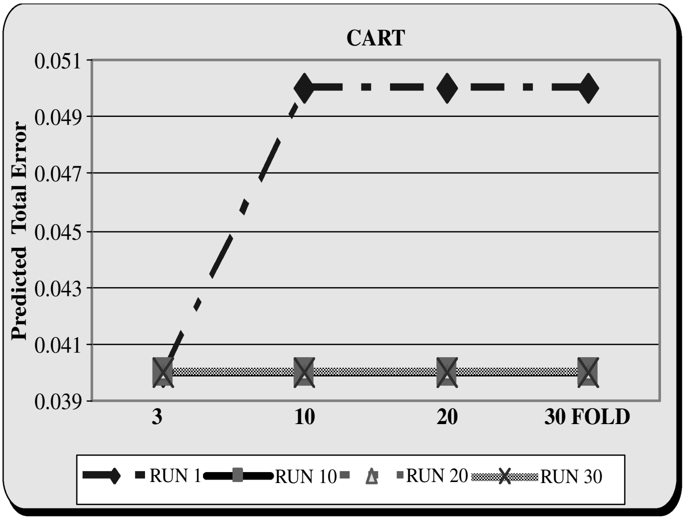
Ros
Roi
Roe
Return on capital employed
Leverage
Short-term debts index
Long-term debts index
Short-term banks debt/Total assets
Long-term banks debt/Total assets
Debt v/banks/Sales
Cost of borrowing
Average cost of borrowing/Total Debts
Financial charges/Sales
Financial charges/Debts
Liquidity Index
Liquid assets/Total Assets
Added value/Total assets
Added value/Sales
Labour cost pro capita
Total production costs/Sales
Total production costs/Profits
Cover margin/Total assets

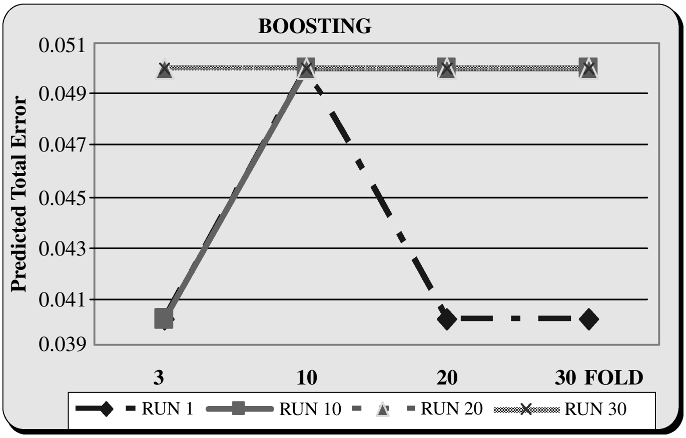
Appendix 3

SKCV Total Error Rate by *k* and T



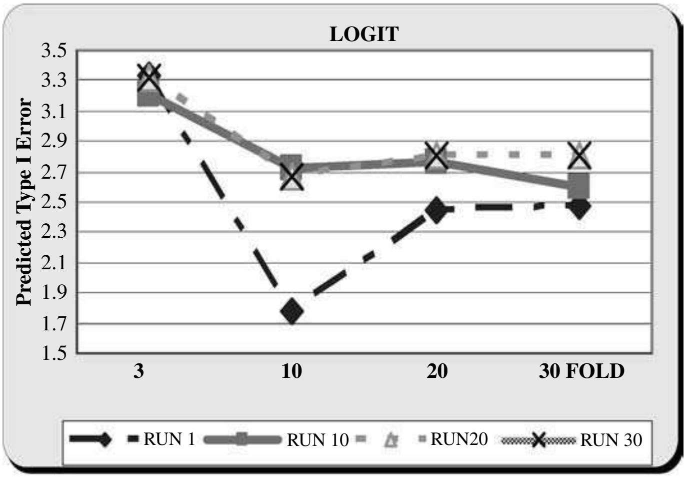


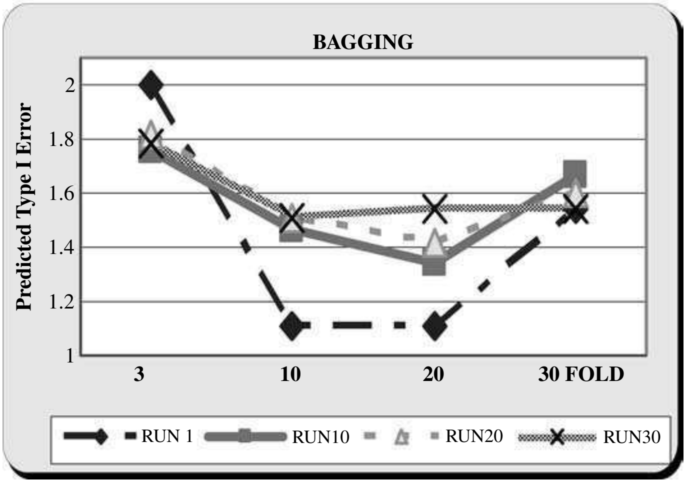

CART
Predicted Total Error
0.051
0.049
0.047
0.045
0.043
0.041
0.039
3
10
20
30 FOLD
RUN 1
RUN 10
RUN 20
RUN 30

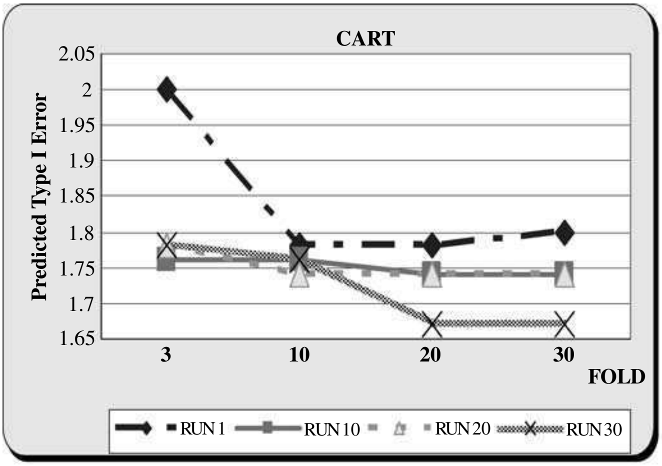

BOOSTING
Predicted Total Error
0.051
0.049
0.047
0.045
0.043
0.041
0.039
3
10
20
30 FOLD
RUN 1
RUN 10
RUN 20
RUN 30

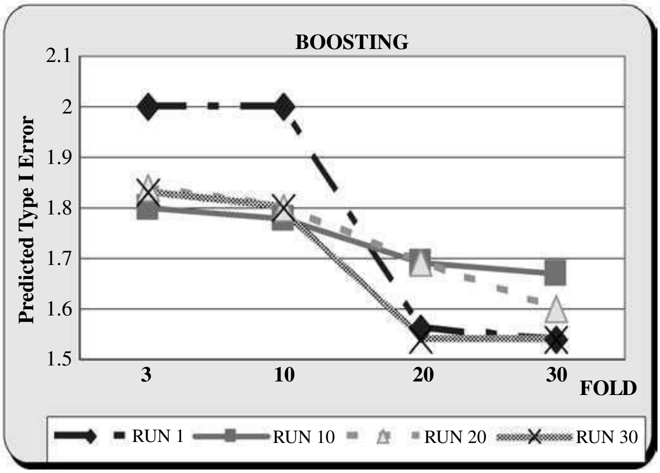
Appendix 4

SKCV Type I Error Rate by *k* and T



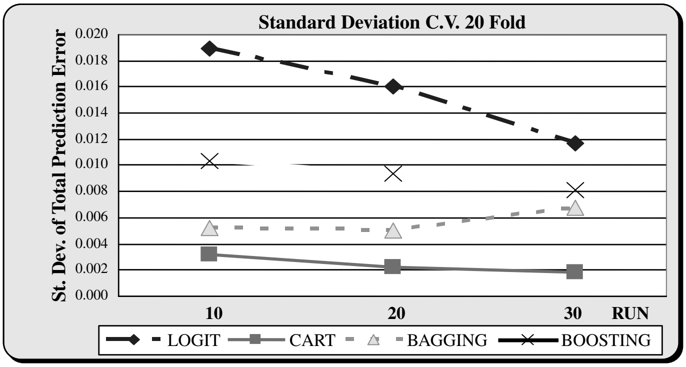


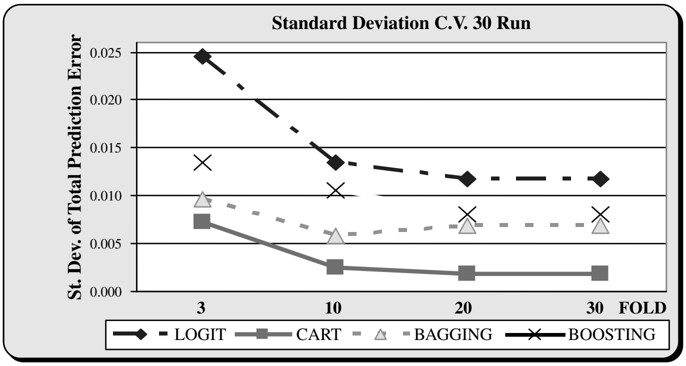

CART
2.05
2
1.95
1.9
1.85
1.8
1.75
1.7
1.65
Predicted Type I Error
3
10
20
30
FOLD
RUN 1
RUN 10
RUN 20
RUN 30


BOOSTING
2.1
2
1.9
1.8
1.7
1.6
1.5
Predicted Type I Error
3
10
20
30
FOLD
RUN 1
RUN 10
RUN 20
RUN 30

Appendix 5

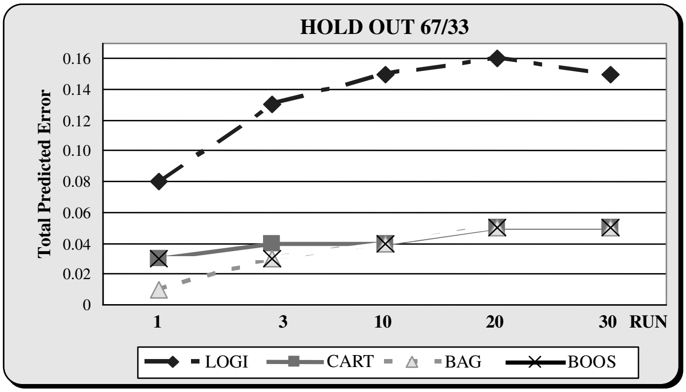
SKCV Standard Deviation of Total Prediction Error Rate by *k* and T

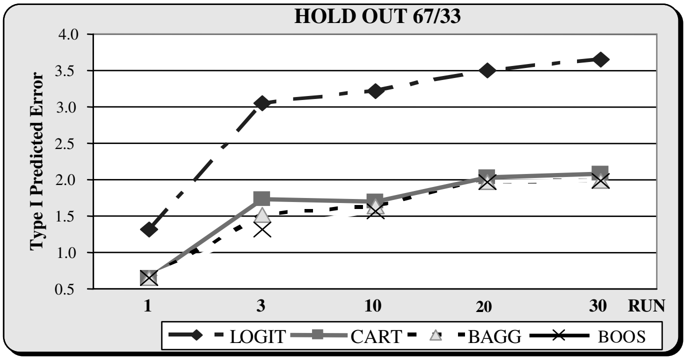




Appendix 6

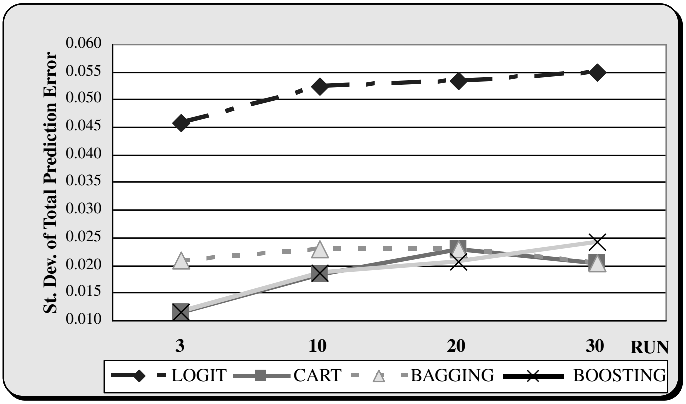
SHRR of Total Prediction Error Rate by T





Appendix 7

SHRR Standard Deviation of Total Prediction Error Rate by T



Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 13

TRANSPARENCY, INSTITUTIONAL FRAMEWORK AND CAPITAL STRUCTURE: INTERNATIONAL EVIDENCE FROM INDUSTRY DATA

Natalia Utrero-González

Abstract

This chapter investigates the degree to which transparency and institutional environment affects corporate finance choices. La Porta *et al.* (1997, 1998) have shown the effects of investor protection on financing decisions. We extend these measures of investor protection and develop a new measure for transparency. We find that transparency has a negative effect on leverage volume. We also find a significant affect of investor protection on leverage decisions. Further, we find that creditor protection makes it easier for firms not investing in fixed assets to access credit markets and that when more transparency is present, high growing sectors present lower debt levels.

Keywords: transparency, investor protection, capital structure

JEL Classification: G32, G38, G18.

1. INTRODUCTION

The empirical literature on corporate finance has shown that financial decisions depend on firm characteristics (Titman and Wessels 1988, Barclay and Smith 1995, Cleary 1999 and Wald 1999). More recently, legal rules and the degree of investor protection have proved to be important determinants of capital structure (La Porta 1997 and 1998, Demirguc-Kunt and Maksimovic 1998 and 1999, and Lombardo and Pagano 1999).[1] However, despite the growing interest of this area of research, thus far there is room to deepen the analysis. First, research has not paid much empirical attention to other institutional arrangements that may also affect financing decisions, namely transparency in the economy. This is surprising since the theoretical literature suggests that asymmetric information conditions credit relationships. Second, conclusions rely on institutional measures that are incomplete. Both questions are tackled in this chapter.

The objective of the chapter is therefore twofold. On one hand, we empirically explore the role of disclosure requirements in influencing capital structure decisions. On the other hand, we complete existing measures of investor protection.

In addition to this, previous work has mostly relied on aggregated data or large firms, which have easier access to international capital markets and therefore are less subject to the institutional constraints imposed by domestic markets. Instead, we use industry-level panel data drawn from a cross-section of nine European countries, United States and Japan, therefore our data do not have the aforementioned problems. Industry-level data have important advantages over both firm-level and country-level data, for the purpose of this research. Because agency problems vary systematically across industries, the institutional developments on which we focus may affect leverage levels differently within each industry. In this case, country-level data may mask the effects of interest because of aggregation. On merely statistical grounds, industry-level data are preferable to firm-level data, mainly because of the "survivorship bias". This bias arises when firms with a long history are included in the sample (Lombardo, 2000). Since the sudden disappearance of an industry is a rare event, this problem is drastically alleviated by the use of industry indices.

We find that investor protection influences leverage decisions, consistent with existing results. Disclosure requirements affect indebtedness of economic sectors in a significant way. In countries with better protection of creditors, problems associated with the lack of collateral assets are lessened, consistent with the hypothesis that agency problems are reduced in such scenarios. Further, disclosing regulation has a negative impact on leverage levels. There are two possible reasons. First, costs associated with disclosure overcome the benefits of more transparency and so firms are more conservatively financed. Second, more transparency enhances new equity issuances and consequently debt levels

[1] See in particular Giannetti (2003), who explores this relationship at firm level in eight European countries and also distinguishes between company size.

are reduced. Beyond their academic interest, these conclusions can be of interest to policymakers engaged in institutional design.

This chapter is organized as follows. Section 2 reviews the related literature and the relevant theoretical relations found between firm leverage and institutional variables as well as the correlations with industry attributes. Section 3 describes the construction of the legal variables and the dataset. The statistical model and the results are presented in Section 4. Section 5 concludes.

2. RELATED LITERATURE AND HYPOTHESIS

A main branch of theoretical finance literature analyses the effect that asymmetric information has on financial decisions, and the difficulties caused by incomplete contracts.[2] In empirical studies, firm attributes have been traditionally used as proxies of asymmetric information in credit relations. Table 13.1, panel A summarizes the standard results and collects the usual proxy. The institutional framework claim, however, that institutional arrangements affect the agency problems associated to credit relationships (La Porta *et al.* 1997). Creditor rights play a role in determining how much the presence of collateral can favor the choice of debt over equity. Given a certain level of asymmetric information, lender ability to recover their loans (collateral) is important. If creditors do not have the right to require collateral or cannot effectively repossess collateralized assets, interest rates will raise and if that is not enough to cover the opportunity cost of lending, the credit market may even collapse. Strict protection of creditor rights leads to cheaper credit. Consequently, many valuable investment projects, which would not be funded because of moral hazard problems, may be financed (Padilla and Requejo 1998). Fabbri (2001), using a general equilibrium model, shows that firms located in countries that provide stronger legal protection to creditor rights, have access to a larger external financing, therefore affecting positively debt maturity.

Further, strict protection of creditors also raises efficiency *ex post*. This effect is connected to the bankruptcy procedure. Strict protection of creditor rights to be repaid with absolute priority eliminates the possibility that liquidation may not take place even when it is efficient (Gertner and Scharfstein 1991). Henceforth, strict protection has positive effects on leverage levels. However, collateral rights might also lead to underinvestment in project evaluation by

[2] See Harris and Raviv (1991) for a survey.

Table 13.1 Theoretical background. Impact on leverage volume

Panel A

VARIABLE	LEVERAGE VOLUME
Size (proxied by logarithm of total assets)	↑ Rajan and Zingales (1995) Myers (1984)
Internal funds (proxied by profitability ratio)	↓ Myers and Majluf (1984) ↑ Jensen (1986)
Collateral (proxied by tangible assets)	↑ Assymetric information theory
Growth opportunities (proxied by growth rate of value added)	↓ Myers (1977) ↑ Giannetti (2003)

Panel B Institutional factors

VARIABLE	LEVERAGE VOLUME
Creditor Rights	↑ Ortodox theories (Gertner and Scharfstein, 1991, La *Porta et al.* 1998, Fabbri 2001) ↓ Critical theories (Bebchuk and Fried 1996, Manove *et al.* 1998)
Shareholder Rights	↓ Jensen and Meckling (1976), Grossman & Hart (1988), La Porta *et al.* (1999)
Non-debt tax shields (proxied by depretiation on total assets)	↓ Titman and Wessels (1988)
Disclosure Requirements	↑ Stock market development theory ↓ Weil (2002) ↓ Subrahmanyan and Titman (1999) ↓ Almazan *et al.* (2004)

banks (Manove *et al.* 2001) and collateral repossession may reduce project control by lenders (Bebchuk and Fried 1996). Consequently, many unworthy projects may be funded and bankruptcy cases will increase. In case of default, if creditor rights are strictly enforced, they will have no incentives to allow their debtors to restructure financially, although it may be efficient to keep assets in debtor hands. The empirical evidence seems to agree with the former view (Giannetti 2003). Therefore, the effect of creditor protection on leverage is not obvious.

The theoretical literature has also analyzed the effects of shareholder rights. Shareholder rights are related to capital market development. Investor protection is crucial because expropriation of minority shareholders by the controlling shareholders is generally present in modern corporations. Expropriation is connected to agency problems, as well as the legal system (Jensen and Meckling 1976). Grossman and Hart (1988) point out the relevance of residual power of investors (guaranteed by legal framework) in order to partly mitigate distortion practices from insiders. Legal protection of shareholders makes expropriation practices less efficient. When investors are protected from expropriation, they pay more for securities, making this form of external finance more attractive for entrepreneurs, who will issue more securities. Shareholder protection also promotes ownership diversification and risk reduction. Better protection of shareholders implies lower firm costs in participating to equity markets (Pagano 1989). Therefore, higher shareholder protection will encourage firms to float capital and reduce debt levels. In this context, resource allocation in fixed assets reduces its relevance to obtain external finance. The other institutional factor that has been considered to affect leverage levels is tax distortion. Higher corporate tax levels tend to favor the use of debt, while non-debt tax shields such as depreciation deductions can be used as substitutes for debt tax advantage and so reduce the leverage level of firms.

Apart from these variables, we examine the effects of legal disclosure requirements. Legal information disclosure is also related to the functioning of financial markets. Requiring information disclosure as a regulatory tool implies the existence of an underlying information asymmetry. In particular, frictions associated with incomplete information in financial markets may generate instability problems that can be reduced with the adequate regulation (Bhattacharya and Thakor 1993). Therefore, an effective regulatory system compels the disclosing party to provide information to end-users that redress this asymmetry. Further, disclosure policies may arise alongside other regulatory interventions, namely creditor and shareholder protection, and seek to complement them.

At the core of financial disclosure, firms provide information to investors regarding key financial flows and balance sheet results. This information allows investors to better assess the return and risks of firms and make informed capital allocation decisions. Hence, the disclosure of the corporation's contractual structures may reduce uncertainties for investors and help lower capital costs by decreasing related risk premia. The provision of information shifts investment patterns toward those more accurately reflecting the true value of firms, which in turn allocates investment resources more efficiently. In addition, since more transparent firms are likely to be more efficiently priced, they are also likely to make better investment choices, which make them more valuable on average.

When more information is public, its cost decreases, investment profitability increases, and consequently investors are encouraged to invest (Subrahmanyan and Titman 1999). Further, agency problems between managers and shareholders reduce, as managers' incentives to extract rents decrease (Pagano and Röell 1998). Therefore, firms should have easier access to external financing in countries with stricter information requirements.

However, when the information about the firm affects the terms under which the firms transacts with its stakeholders, transparency can have an offsetting negative effect. Accordingly, the potential costs of disclosure, *ex ante*, are high. Risky firms that might have been able to mask their true financial status are forced after disclosing either to modify firm strategies to attract more external capital at the previous rates or provide investors with higher returns to compensate for higher level of risks. Therefore, there exist strong incentives to keep required disclosure to a minimum (Weil 2002). Almazan *et al.* (2004) claim that good news improves terms of trade less than bad news worsens them, implying that increased transparency can lower firm value. In particular, transparency might be especially costly when firms have specific investment opportunities that require external financing. In this case, to avoid the cost of information revelation, firms may choose to pass up positive net present value investments. Therefore, under these circumstances, transparency is costly and firms will tend to choose lower leverage ratios (Almazan *et al.* 2004). Table 13.1, Panel B collects these theoretical results.

3. DATA AND VARIABLES INCLUDED

To capture the empirical relationship between the institutional environment and leverage decisions, we use data from different sources.

3.1. Institutional Variables

La Porta *et al.* (1998) develop an index for creditor protection and another for investor protection. These indexes (LLSV) are derived from bankruptcy and company regulations. We incorporate merger rules to the standard LLSV indexes. Further, we study merger law and auditing norms to measure legal disclosure requirements respectively.

The method of construction is described in the works by La Porta *et al.* (1998). Using accounting and legal literature,[3] we define different categories which summarize public intervention in banking and the quantity of information that has to be public. For every country and each category we sum one unit when the feature is present and zero otherwise. Afterwards, we sum for each index and obtain the result for every country. We have taken into account legal innovations, in order to obtain time series differences and not only cross-country variations.

Henceforth, shareholder protection index, *share*, has nine categories, which include:

the LLSV ones;
1. mail voting allowed for general meeting;
2. no need to deposit the shares before voting;
3. cumulative voting allowed;
4. protection to small shareholders;
5. pre-emptive right to buy new issues of stock;
6. the required percentage to attend a shareholder meeting is inferior to 10%;
   those associated to merger law:
7. equal treatment of shareholders;
8. control of directors' activities
9. forbidden manager protection practices[4] (such as poison pills when they impose restrictions to shareholder rights[5]).

The new creditor index, *cred* ranges between 0 and 5, being the categories:

1. "stay on assets" procedure allowed;
2. no priority to other stakeholders (employees, government or public entities);
3. managers are not allowed to begin the reorganisation process without the consent of creditors;

---

[3] On accounting: Alexander and Archer (1992) and Blake and Amat (1993).
On banking and financial markets: Interbank Research Organization (1978), Moreiro (1992), Parejo *et al.* (1993), Campbell and Moore (1993), Forestieri and Mottura (1998) and Katayama and Makov (1998). On merger and corporate law: Hawkins and Morton (1990) and Raybould and Firth Ed. (1991).

[4] In this case we add one unit when the protective tactics are not permitted, in order to maintain the internal coherence of the index: higher value indicates higher protection.

[5] The case of United States is complex because poison pills are allowed, although its use is not absolute: "managers can use defensive tactics only to the extent that is reasonable in relation to the threat posed and always protecting shareholder rights (Unocal Test, 1985). The Delaware courts also indicated that they protect against managerial moves to impede voting by shareholders to remove them. On the contrary, some European countries allow shareholder voting rights restrictions (France) or the "creative use of share capital" (Netherlands). Therefore, we decide to sum one unit in the case of United States in order to capture these differences.

4. creditors have the right to impose an external administrator; and
5. explicit protection in merger procedures.

For both indexes, higher values are associated to better protection. The United States has the maximum shareholder protection, scoring 8. Belgium presents the minimum value of 1. In relation to creditor index, Denmark presents a score of 4 being the maximum, and France with a 0 value is the minimum. Panel A of Table 13.2 presents the construction of these variables.

Disclos captures the information firms are required to disclose. A disclosure system normally affects three aspects of information:

1. the degree of information made public (the use of information);
2. the accuracy of the information; and
3. the scope of the information (Weil 2002).

We analyse some aspects related to the information scope and information accuracy. In particular, we evaluate disclosure rules related to the going public decision and mergers and acquisition of companies on the one hand and balance sheet and auditing norms, on the other. Generally speaking, merger regulation requires a public communication both to shareholders and national authorities. However, in some European countries this information is first confidentially provided to national authorities (Hawkins and Morton 1990). This fact can increase asymmetric information. We will sum one point to the variable only when the information is public at the same time that it is communicated to the pertinent authorities. We also analyze the information requirements to issue equity. We add one point when firms have to provide detailed information about financial results. In this aspect, there is no difference in the countries included in the sample.

Related to annual accounts there are national accounting norms that allow not to include all available information in firm balances, but only in the memory. This fact may be difficult to understand and reduces transparency of public information. Blake and Amat (1993) and Alexander and Archer (1992) classify annual accounts according to their formality.[6] We sum one unit to *disclos* when national rules are said to be strict, such that represents higher transparency. In connection to auditing rules, Blake and Amat (1993) and Alexander and Archer (1992) account for differences in auditing reporting. For instance, in Italy, effective auditing is done only in larger firms. Nevertheless, it is also admitted that larger firms that float internationally, have their accounts audited, although home country regulation is not strict in this sense. Following Blake and Amat (1993) and Alexander and Archer (1992), the index will be increased, in one unit if norms

[6] Both taxonomies are coincident. Japan and United States present strict rules on public information, meanwhile the Netherlands and Denmark present the most flexible norms.

Table 13.2 Institutional variable

Panel A Shareholder and Creditor Protection

We extend LLSV indexes to include merger regulation. First, we sum one point to LLSV shareholder index, if control of boards of Directors is present (Dir. Cont). Second, when equal treatment to all shareholders is guaranteed, we also sum one point. Third, we increase the index by one unit more if defensive tactics are forbidden.

Regulation	**Shareholder protection**					**Creditor protection**		
	LLSV	**Dir cont**	**Eq. treat**	**def tacts**	**total**	**LLSV**	**merger**	**Total**
Austria	2	0	1	1	**4**	3	0	**3**
Belgium	0	0	1	0	**1**	2	1	**3**
Denmark	2	0	1	1	**4**	3	1	**4**
France	3	1	1	0	**5**	0	0	**0**
Germany	1	0	1	1	**3**	3	0	**3**
Italy	1	0	0	0	**1**	2	0	**2**
Japan	4	0	0	0	**4**	2	0	**2**
Netherlands	2	0	0	0	**2**	2	0	**2**
Portugal	3	1	1	0	**5**	1	0	**1**
Spain	4	0	1	0	**5**	2	0	**2**
United States	5	1	1	1	**8**	1	0	**1**

Panel B Information requirements

Disclos captures the information available to investors. Features analyzed: strict patterns of annual accounts, if merger information has to be made public at the same time it is communicated to authorities, information requirements to participate in stock markets and compulsory auditing for all large firms (not only for floating firms).

	disclosure requirements (disclos)				
	Merger	**stock mkt**	**Strict**	**auditing**	**total**
Austria	1	1	1	1	**4**
Belgium	1	1	1	1	**4**
Denmark	1	1	0	0(1)a	**2(3)a**
France	1	1	1	1	**4**
Germany	1	1	1	1	**4**
Italy	1	1	1	0	**3**
Japan	1	1	1	1	**4**
Netherlands	0(1)a	1	0	1	**2(3)a**
Portugal	1	1	1	1	**4**
Spain	1	1	1	1	**4**
United States	1	1	1	1	**4**

Note: *Lowest limitation to non-financial firm participation, therefore very weak control (almost non-existence). a: the value in brackets is valid from 1993 onwards. b: the value in brackets is valid from 1997 onwards.

and their enforcement are rigorous. *Disclos* ranges from 0 to 4. Accordingly, higher values of *disclos* imply more information available to markets and investors. When comparing *disclos* values, a great uniformity is observed among European countries.[7] Panel B of Table 13.2 presents the score and time changes.

La Porta *et al*. (1998) use an index to account for the level of public information (acc_{LLSV}). This index is constructed through the analysis of 1990 annual accounts of a sample of companies for each country.[8] This index reflects firm revelation of financial information, "the use of information" rather than the scope or accuracy of information. Therefore, we consider this index complementary to the one we have developed. The main difference is that ours include legal innovations passed during the period and the one by La Porta *et al*. (1998) is constant for the whole period. We also include this index in the analysis to obtain a more complete picture of legal information disclosure effects.

These three institutional factors (altogether) have not been empirically investigated yet. Further, with the inclusion of these variables, we have come closer to the real institutional environment faced by firms throughout the 1990–99 period.[9] The analysis of the legal innovations in the 1990–99 period is especially relevant for disclosure requirements. It can be observed that they have become stricter for Denmark and The Netherlands since 1993.[10] The remaining norms have remained stable since the 1980s, when major changes took place.

3.2. Leverage and sector attributes

Our sample includes 11 developed countries.[11] Data originate from the BACH[12] database, created within the European Committee of Central Banks. The main advantage of this data is its comparability, so that the robustness of the results

---

[7] The constitution treaty of the European Community (1952) established the mutual recognition of national firms. Harmonisation, however, has not finished yet, even though there are some fields where it has evolved more quickly. Information requirement is one of the most homogeneous fields. We have also controlled for the information asked to floating firms. Since differences in this point disappeared in the countries anlysed we decided not to include it.

[8] For more details about the computation of the index see La Porta *et al*. (1998).

[9] We have not included any variable for the origin of the law, although present in previous articles. As Rajan and Zingales (2001) point out, legal origin does not have a constant explanatory power over time, due to legal innovations and reforms. Therefore, we are not sure that the legal origin still has an explanatory power, when legal systems continuously change to adapt to new scenarios.

[10] European Market Review, 1993.

[11] Germany, France, Denmark, Italy, Spain, Portugal, Austria, Belgium, Netherlands, United States, and Japan.

[12] Bank for the Accounts of Companies Harmonized. It contains sector data since 1985.

is assured. All data comes from book information, hence it is not possible to evaluate the market values of debt ratios. However, it is generally admitted that the book value of leverage is the result of the management's financial decisions. Moreover, previous empirical papers (Rajan and Zingales 1995, Boot *et al*. 2000) do not find significant differences in factors correlated with debt to book and market capital. The period considered is 1990–99.

The ratios included are:

- *Leverage* is calculated as total debt on assets. Total debt includes banking credit, trade creditors and debenture loans.
- *Maturity* is calculated as the ratio of long-term debt on total debt.
- *Profitability* is the standard return on assets ratio: Profit before tax plus interest and depreciation over total assets.
- *Collateral* captures tangible assets to total assets.
- *Non-debt tax shields*: Depreciation and provision[13] level over total assets.
- *Growth opportunities*: The growth rate of value added at time t, between t and t – 1.

Summary statistics presented in Table 13.3 confirm the cross-country differences accounted in previous papers (Wald 1999 and Giannetti 2003). The United States presents a leverage ratio of 50%, lower than the European average of 59%. However, the European average of collateralized assets in European countries is smaller than in the United States. In particular, Italy, Germany and Belgium ratios lie in the lowest quartile and present higher protection to creditors than the United States. On the other hand, Spain, Portugal and The Netherlands have a higher level of collateral than the United States but their degree of creditor protection is below the average (according to Table 13.2.A). This preliminary evidence suggests that collateral assets may be less important when creditor rights are better protected.

4. ESTIMATION METHOD AND RESULTS

The two previous sections suggest that financial decisions are affected by the institutional settings where sectors are operating in. In particular, we want to test whether agency problems are mitigated in protected and more transparent

[13] Depreciation and provisions of non financial fixed assets.

Table 13.3 Summary statistics

Lever is leverage ratio. Mat is the maturity ratio. Roa is return on assets ratio. Collat is the percentage of fixed assets on total assets. Grop accounts for growth opportunities.

Panel A Summary statistics for Countries and all economic sectors

		AUS	**BEL**	**DNK**	**FRA**	**GER**	**ITA**	**JPN**	**NTLS**	**POR**	**SPA**	**US**	**avg**	**stdev**
	Mean	0.63	0.59	0.60	0.62	0.60	0.66	0.69	0.54	0.54	0.56	0.50	**0.59**	**0.06**
LEVE	Std	0.15	0.11	0.10	0.08	0.16	0.08	0.11	0.10	0.11	0.10	0.05	**0.10**	**0.03**
	min	0.27	0.13	0.33	0.33	0.22	0.36	0.41	0.26	0.14	0.23	0.33	**0.27**	**0.09**
	max	1.14	0.85	0.93	0.81	0.86	1.53	0.96	0.91	0.82	0.92	0.74	**0.95**	**0.22**
	Mean	0.34	0.25	0.28	0.30	0.23	0.26	0.34	0.34	0.39	0.32	0.31	**0.31**	**0.05**
COLLLAT	Std	0.14	0.12	0.16	0.11	0.08	0.12	0.11	0.14	0.15	0.14	0.07	**0.12**	**0.03**
	min	0.07	0.04	0.00	0.02	0.07	0.05	0.05	0.08	0.10	0.06	0.17	**0.06**	**0.04**
	max	0.88	0.67	0.63	0.78	0.40	0.76	0.78	0.85	0.95	0.82	0.57	**0.74**	**0.16**
	Mean	0.00	−0.06	0.01	0.01	−0.01	0.00	−0.01	0.00	0.00	0.01	0.00	**0.00**	**0.02**
GROP	Std	−0.17	0.53	−0.09	−0.16	0.06	−0.13	−0.07	−0.09	−0.42	−0.13	−0.07	**−0.07**	**0.23**
	min	0.77	−7.75	0.28	0.53	−0.24	0.53	0.28	0.42	5.46	0.58	0.36	**0.11**	**3.03**
	max	1.37	2.56	0.54	2.57	0.26	1.19	0.47	0.42	3.41	1.58	0.30	**1.33**	**1.09**
	Mean	0.11	0.14	0.16	0.14	0.15	0.10	0.11	0.14	0.14	0.11	0.15	**0.13**	**0.02**
ROA	Std	−0.04	0.80	−0.05	−0.03	0.03	−0.03	0.03	0.04	−0.24	−0.05	0.04	**0.05**	**0.26**
	min	0.03	−3.49	0.11	0.01	0.02	0.05	0.01	−0.03	1.26	0.36	0.03	**−0.15**	**1.17**
	max	0.37	20.29	0.47	0.25	0.25	0.19	0.25	0.28	4.77	0.28	0.26	**2.51**	**6.05**

Table 13.3 (*Continued*)

Panel B Summary statistics for countries and manufacturing sectors

		AUS	BEL	DNK	FRA	GER	ITA	JPN	NTL	POR	SPA	US	AVG	STD
	Mean	0.59	0.56	0.60	0.59	0.56	0.64	0.66	0.53	0.52	0.55	0.50	**0.57**	**0.05**
LEV	sdev	0.13	0.09	0.10	0.06	0.16	0.07	0.11	0.10	0.10	0.09	0.05	**0.10**	**0.03**
	min	0.35	0.19	0.33	0.33	0.22	0.43	0.41	0.26	0.14	0.25	0.33	**0.29**	**0.09**
	max	0.94	0.85	0.93	0.73	0.83	1.53	0.84	0.91	0.82	0.88	0.74	**0.91**	**0.22**
	Mean	0.35	0.25	0.28	0.30	0.25	0.26	0.35	0.33	0.40	0.32	0.31	**0.31**	**0.05**
COLLAT	Sdev	0.11	0.08	0.16	0.07	0.06	0.09	0.09	0.12	0.12	0.10	0.07	**0.10**	**0.03**
	Min	0.14	0.09	0.00	0.02	0.11	0.11	0.18	0.08	0.16	0.15	0.17	**0.11**	**0.06**
	Max	0.88	0.64	0.63	0.61	0.40	0.70	0.78	0.85	0.95	0.72	0.57	**0.70**	**0.16**
	Mean	0.00	−0.02	0.01	0.02	−0.01	0.00	−0.01	−0.01	0.02	0.01	0.00	**0.00**	**0.01**
GROP	Sdev	0.13	0.14	0.09	0.19	0.06	0.11	0.08	0.09	0.20	0.14	0.07	**0.12**	**0.05**
	Min	−0.53	−0.55	−0.28	−0.53	−0.24	−0.53	−0.28	−0.42	−0.75	−0.58	−0.36	**−0.46**	**0.16**
	Max	0.94	0.92	0.54	2.57	0.26	1.07	0.47	0.24	1.76	1.58	0.30	**0.97**	**0.74**
	Mean	0.12	0.13	0.16	0.14	0.15	0.11	0.12	0.14	0.14	0.12	0.15	**0.13**	**0.02**
ROA	Sdev	0.03	0.03	0.04	0.02	0.03	0.03	0.02	0.04	0.08	0.05	0.04	**0.04**	**0.02**
	Min	−0.03	−0.11	−0.11	−0.01	0.04	−0.05	0.05	−0.03	−0.57	−0.36	0.03	**−0.10**	**0.19**
	Max	0.23	0.25	0.27	0.20	0.25	0.19	0.20	0.28	0.36	0.28	0.26	**0.25**	**0.05**

Table 13.3 (*Continued*)

Panel C Summary statistics for economic sectors

Lever is leverage ratio, calculated as total debt (banking credit, trade creditors and debenture loans) on total assets. Profit is return on assets ratio (profit before tax plus interest and depreciation over total assets). Collat is the percentage of tangible assets on total assets. Grop accounts for growth opportunities (growth rate of value added between t and t + 1). Averg presents the arithmetic average across country and size for the period 1990–99. Std. dev. is the standard deviation for the same period and median the median across sector and size for 1990–99. Book data from BACH database.

	LEV				**ROA**				**COLLAT**				**GROP**			
Sect	**Mn**	**Sd**	**Min**	**Max**	**Mn**	**Sd**	**Min**	**Max**	**Mn**	**Sd**	**Min**	**Max**	**Mn**	**Sd**	**Min**	**Max**
100	**0.53**	0.17	0.19	0.93	**0.11**	0.09	0.57	0.21	**0.53**	0.17	0.11	0.95	**0.02**	0.29	−0.75	2.57
200	**0.58**	0.09	0.33	0.82	**0.13**	0.02	0.05	0.20	**0.28**	0.08	0.00	0.45	**0.00**	0.05	−0.17	0.21
210	**0.53**	0.10	0.25	0.80	**0.14**	0.04	0.03	0.29	**0.32**	0.09	0.00	0.51	**0.01**	0.07	−0.20	0.37
211	**0.59**	0.14	0.31	1.53	**0.11**	0.07	−0.36	0.30	**0.34**	0.13	0.00	0.72	**0.02**	0.25	−0.58	1.95
212	**0.52**	0.11	0.14	0.83	**0.16**	0.05	0.06	0.36	**0.34**	0.09	0.00	0.55	**−0.01**	0.11	−0.31	0.66
213	**0.52**	0.11	0.22	0.79	**0.15**	0.03	−0.03	0.24	**0.29**	0.08	0.00	0.44	**−0.01**	0.09	−0.35	0.34
220	**0.59**	0.09	0.35	0.82	**0.13**	0.03	0.02	0.23	**0.24**	0.07	0.00	0.40	**0.00**	0.06	−0.18	0.27
221	**0.60**	0.09	0.24	0.82	**0.13**	0.03	−0.02	0.27	**0.23**	0.08	0.00	0.41	**0.00**	0.07	−0.25	0.29
222	**0.58**	0.10	0.32	0.86	**0.13**	0.04	−0.03	0.27	**0.21**	0.07	0.00	0.39	**0.00**	0.13	−0.53	1.11

Table 13.3 (*Continued*)

	LEV				ROA				COLLAT				GROP			
Sect	**Mn**	**Sd**	**Min**	**Max**	**Mn**	**Sd**	**Min**	**Max**	**Mn**	**Sd**	**Min**	**Max**	**Mn**	**Sd**	**Min**	**Max**
223	**0.62**	0.11	0.27	0.89	**0.11**	0.04	−0.11	0.24	**0.27**	0.09	0.00	0.58	**0.02**	0.21	−0.53	1.76
230	**0.59**	0.09	0.35	0.85	**0.14**	0.02	0.08	0.20	**0.30**	0.08	0.00	0.50	**−0.01**	0.04	−0.22	0.17
231	**0.59**	0.10	0.35	0.86	**0.14**	0.03	0.07	0.24	**0.32**	0.09	0.00	0.56	**−0.01**	0.07	−0.26	0.22
232	**0.59**	0.09	0.33	0.81	**0.13**	0.03	0.03	0.21	**0.25**	0.09	0.00	0.45	**0.00**	0.07	−0.20	0.27
233	**0.59**	0.10	0.26	0.86	**0.14**	0.03	0.05	0.22	**0.33**	0.10	0.00	0.66	**−0.01**	0.07	−0.28	0.56
234	**0.60**	0.09	0.38	0.87	**0.15**	0.04	−0.10	0.28	**0.30**	0.08	0.00	0.48	**0.00**	0.08	−0.32	0.61
300	**0.72**	0.10	0.40	0.91	**0.09**	0.03	0.01	0.27	**0.16**	0.08	0.03	0.53	**0.00**	0.10	−0.52	0.36
400	**0.69**	0.09	0.43	0.92	**0.10**	0.04	−0.13	0.18	**0.20**	0.08	0.08	0.43	**−0.05**	0.55	−7.75	0.30
410	**0.69**	0.09	0.38	0.86	**0.10**	0.04	−0.13	0.18	**0.15**	0.05	0.05	0.27	**−0.04**	0.32	−3.60	0.35
420	**0.72**	0.09	0.42	0.98	**0.10**	0.04	−0.04	0.23	**0.18**	0.08	0.05	0.40	**0.01**	0.14	−0.40	0.71
430	**0.68**	0.11	0.40	0.96	**0.12**	0.03	0.03	0.24	**0.28**	0.09	0.13	0.57	**−0.01**	0.09	−0.39	0.37
440	**0.63**	0.17	0.13	1.14	**0.13**	0.05	0.05	0.47	**0.53**	0.18	0.04	0.86	**0.01**	0.16	−0.33	1.09
500	**0.59**	0.12	0.23	0.97	**0.13**	0.32	−1.26	4.77	**0.54**	0.14	0.21	0.82	**0.00**	0.55	−5.46	3.41
600	**0.62**	0.12	0.33	0.96	**0.18**	1.45	−3.49	2.03	**0.27**	0.13	0.05	0.74	**−0.05**	0.69	−7.65	2.56

environments and therefore incentives to invest in collateralized assets decrease. For the empirical test, we estimate the following equation:

$$Leverage_{i,j,t} = \alpha + \beta \times industry\ attributes_{i,j,t} + \gamma \times INST_{i,t} + \eta_{ij} + \psi_t + \varepsilon_{i,j,t} \quad (1)$$

where $i = 1, \ldots, n$ refers to countries, $j = 1, \ldots, m$ refers to economic sectors and $t = 1, \ldots, T$ to time periods. The error term, $\varepsilon_{i,j,t}$ is identically distributed and uncorrelated across observations and with exogenous variables, but cov $(\varepsilon_{i,j,t}, \varepsilon_{i,j,s})$ may be different from zero if $t = s$.

In some specifications we also include some interaction terms in the spirit of Rajan and Zingales (1998) and Claessens and Laeven (2001). The interaction terms are formed by one institutional factor and one industry characteristic. In order to facilitate the interpretation of the results, we use dummy variables associated with the institutional variables, which group countries above and below the average.

From an econometric point of view, the estimation of the coefficients α, β and γ must take into account the structure of the error terms ε_{ijt}. When there are individual effects, the ordinary least squared estimation (OLS) of the panel data model may produce errors and biases in the coefficient values and can induce a mistake in the serial correlation degree. The specific effects can be treated as fixed or random. The problem is not if effects are fixed or random, but whether the effects are correlated to the observable variables. When effects can be considered random, the OLS estimator is consistent but inefficient. Provided the effects are fixed and correlated with the explanatory variables, the OLS estimator is not consistent. Therefore, to get consistent estimations, it is necessary to have an estimator that makes these individual effects disappear. A consistent estimator is the within-group estimator. The fixed effect estimator provides unbiased estimates by taking all the variables in deviation from the individual mean and exploiting only the time-series variability.[14] We perform the Hausman test, to test whether individual fixed effects are correlated with the explicative variables. When correlation is present, conditional inference must be done (fixed effect estimation) (Arellano and Bover 1990), otherwise random effect estimation is applied.

Moreover, the model to be estimated may have problems of endogeneity in the regressors. On the one hand, some industry variables can be determined simultaneously with the debt ratio. On the other hand, legal arrangements can

[14] In some of the specifications, however, we only include industry effects in order to be able to estimate γ, the parameter of interest. In these cases, we include country dummies to account for potential unobservable effects.

also be endogenous. This potential endogeneity may seriously affect the results. The within estimator controls for unobservable heterogeneity, but it is derived from the strict exogeneity of the dependent variables. Therefore, the estimations would be biased if there were endogenous variables in the model. In order to tackle this problem, we apply the Sargan test that compares the coefficients estimated through OLS and instrumental variables. As instruments, we used the industry variables lagged from t – 1 to t – 2 and as legal variables do no change much over time, we use the origin of the legal system, the rule of law, gdp and population as instruments (Rajan and Zingales 1998, Claessens and Laeven 2004). La Porta *et al.* (1997) argue that legal systems have a long history and have shaped the development of accompanying institutions. Legal origin and rule of law can therefore be treated as exogenous variables in analyzing modern economic regulation. In the presence of economies of scale in financial institutions and systems, the size of the country (population) and the economic level (gdp) will affect financial structure and financial regulation. The test is reported for each specification. If the p-value is below 10%, then IV estimates are reported. Otherwise, within estimates are reported. In most of the regressions, the test rejects the null hypothesis and so IVs are reported.

In Table 13.4, we present results for Equation 1 without interaction effects. We report for brevity only the coefficients of interest, namely the institutional variables and the industry attributes said to influence leverage levels. Standard errors and t-statistics are corrected for heteroskedasticity. We introduce one institutional feature at a time and in the last columns we introduce them all.

Mostly the data supports the traditional theories of corporate finance and industry characteristic affect significantly leverage levels. Profitability, measured by the return on assets, presents a positive and significant coefficient in all runs. This is coherent with Jensen's theory of free cash flow (1986). No support for the pecking order theory is found in any of the specifications. Collateral assets seem to ease access to external credit. Therefore, under markets with asymmetric information, offering a guarantee to the lender is important to attract external financing and therefore may affect investment decisions and resource allocation. Future growth opportunities can be considered as intangible assets that requires external financing. Previous studies show a negative relationship between growth opportunities and leverage, meaning that future growth opportunities may be financed through stock markets. However, this result may be driven by the excessive weight of listed firms in the sample used. Instead, we find a positive and significant coefficient, meaning that industries become more leveraged as their growth opportunities improve. This effect may be stronger in countries with a less developed stock market, where firms with growth opportunities are not able to issue new capital to fund growth or in illiquid capital markets with

Table 13.4 Determinants of leverage

The dependent variable, lever, is the leverage ratio on industry k and country i at time t. Industry attributes are profitability (roa), size (logsales), tangible assets (collat) and non-debt tax shields (ndts). Legal variables included are banking regulation (*bank*), creditor protection (*cred*), shareholder protection (*share*), and disclosure requirements (*disclos*). Standard errors (in parenthesis) corrected for heteroskedasticity. All specifications include time effects.

	(1)	(2)	(3)	(4)	(5)	(6)
Roa	.1285***	.1286***	.0207***	.1286***	.1222***	.1287***
	(.0102)	(.0102)	(.0072)	(.0102)	(.0071)	(.0102)
collat	.2895***	.2894***	.0047	.2893***	.2524***	.2894***
	(.0217)	(.0217)	(.0123)	(.0216)	(.0166)	(.0217)
Size	.1427***	.1426***	.0057	.1426***	.1041***	.1427***
	(.0153)	(.0153)	(.0056)	(.0153)	(.0063)	(.0153)
Ndts	−.7565***	−.7566***	−.1351***	−.7565***	−.7934***	−.7566***
	(.0380)	(.0380)	(.0360)	(.0380)	(.0289)	(.0380)
groppor	.4878***	.4878***	.0113***	.4875***	.0936***	.4876***
	(.1368)	(.1368)	(.0046)	(.1368)	(.0451)	(.1368)
share	−.0358***				−.0198***	−.0089***
	(.0021)				(.0043)	(.0021)
Cred		.0196***			.1302***	.0126***
		(.0032)			(.0092)	(.0039)
disclos			−.0136***		−.0219***	
			(.0038)		(.0123)	
Acc_{LLSV}				−.1790***		−.0117*
				(.0106)		(.0075)
Constant	.7275***	.5612***	0.5948***	.6201***	.2910***	.5323***
	(.0186)	(.0118)	(0.0061)	(.0148)	(.0146)	(.0194)
Obs	4100	4100	5499	4100	4373	4100
Hausman test	708.17a	253.39a	103.13a	1261.18a	52.01a	104.1a
Sargan test	194.11a	71.61a	5.08	97.12a	171.21a	293.48a
Sector effects	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	No	No	Yes	No	No	No
Country dummies	Yes	Yes	No	Yes	Yes	Yes

*significant at 10%; ** significant at 5%; *** significant at 1%.
Hausman test and Sargan test: a significant at 1%, b significant at 5% and c significant at 10%.

high degree of ownership concentration where controlling shareholders do not want to loose control financing new projects through equity. Size, proxied by the logarithm of total assets, presents a positive and significant coefficient, coherent with the results obtained by Rajan and Zingales (1995). Finally, the coefficient of non-debt tax shield is negative and significant as expected. Hence, when the tax advantage of debt reduces, debt is less attractive for entrepreneurs who may look for alternative financing. This evidence is consistent with previous findings.

With respect to the legal variables, all coefficients are significant, both one at a time and when they are introduced altogether. Shareholder rights affect negatively leverage. As we have previously claimed, when external investors are protected from expropriation, they are willing to invest and securities become more attractive for entrepreneurs as costs of floating decreases and the discipline of debt can be avoided (Pagano 1989). Creditor protection has a positive effect on debt levels, as expected by the main stream of financial literature (Padilla and Requejo 1998), that is, better protection provides cheaper credit that seems to drive to higher debt levels (column 2). The quantity of public information, *disclos* presents a negative significant coefficient, meaning that the quantity of information affects leverage negatively, consistent with the results by Almazan *et al*. (2004), who claim that transparency can reduce the incentives of firms to undertake specific investments, i.e., firms can pass up positive net present value investment that require external financing and therefore choose more conservative capital structures than they would otherwise choose. Column 4 presents the results for the acc_{LLSV} index of transparency. The coefficient is negative and significant. The depressing effect of transparency in leverage is confirmed. Hence, this effect is robust to the measure used, information accuracy and scope or quantity of information. An alternative interpretation of these results is that more transparency implies lesser costs to float in the stock market and being debt and equity, somehow substitutes more transparency incentives equity issuances in detriment of indebtedness. With the present data, we cannot discern which hypothesis is more appropriate.

Columns 5 and 6 collect together the results for the estimation with all institutional variables. All coefficients maintain sign and significance. Therefore, creditor protection has a promoting effect on leverage ratio, whereas the factors more related to the development of stock market, shareholder protection and disclosure requirements seem to discourage debt decisions. These results are more interesting, since it is generally accepted that legal disclosure requirements complement other regulatory interventions that are passed to reduce agency problems associated with asymmetric information.

Interaction terms allow us to deepen the analysis of legal arrangements. We first analyze whether the necessity of collateral assets is affected by investor protection or information requirements. In particular, we want to see if collateral needs are less demanding in countries where creditors and shareholders are protected and investors enjoy a higher degree of transparency. We then interact growth opportunities with the legal variables. First, the features associated to stock market development (shareholder protection and disclosure requirements) and second, protection to creditors. In this case, we control if high growth sectors present lower indebtedness levels in countries with more transparency and better protection to shareholders and creditors, that is, if they are more capable of issuing new capital to finance growth and the benefits of disclosing information offset the costs. Results are collected in Table 13.5.

The interaction terms reported are calculated with dummy variables associated to the legal factors.[15] Results show that the legal environment affects the extent of agency problems. The results, when we introduce the interaction terms with the value of the legal variables (not reported), remain unchanged. Creditor protection (columns 1 and 2) confirms the positive significant effect on leverage previously obtained. Further, the collateral interaction term (*cred2*) is negative and significant, that is, good creditor protection makes collateral assets less relevant. Effective creditor protection eases access to credit to those sectors investing in intangible assets, i.e., industries operating in countries with good creditor protection need less collateral to access external financing. The growth opportunities interaction effect is positive but not statistically significant. Therefore, in countries with good protection to creditors, growth opportunities do not have a different effect on leverage decisions.

Shareholder protection presents a negative and significant coefficient. Better shareholder protection reduces agency problems between insiders and outsiders and helps to mitigate distortion practices from insiders. It promotes incentives to float in the stock market. Therefore, better protection to external shareholder makes it more attractive to resort the stock market to raise capital and therefore has a depressing effect on debt levels. The collateral interaction term is negative, meaning that it is less important to invest in fixed assets when firms operate in countries with high protection to shareholders. The growth opportunities interaction term is also negative, therefore when shareholders are protected and stock markets are more liquid and developed, more fast growing firms will finance through stock markets. However, none of these coefficients are statistically significant. Hence, firm investment policy is not significantly different to leverage decisions in those countries with effective shareholder protection.

[15] We construct legal dummy variables in order to classify countries above or below the average value of each institutional feature.

Table 13.5 Determinants of leverage. Interaction effects

The dependent variable, lever, is the leverage ratio on industry k and country i at time t. Industry attributes are profitability (roa), size (logsales), tangible assets (collat), growth opportunities (groppor) and non-debt tax shields (amo). Legal variables included are banking regulation (*bank*), creditor protection (*cred*), shareholder protection (*share*), and disclosure requirements (*disclos*). Standard errors (in parenthesis) corrected for heteroskedasticity. All specifications include time effects.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	lever	Lever		lever	lever	lever		
Roa	.1327***	.1358***	.1276***	.2315	.1282***	.0773***	.1192***	.1147***
	(.0104)	(.0138)	(.0100)	(.1876)	(.0104)	(.0158)	(.0082)	(.0087)
Collat	.4445***	.2945	.2884***	.3154***	.2802***	.2265***	.2111***	.2699***
	(.0652)	(.0238)	(.0300)	(.0837)	(.0696)	(.0343)	(.0190)	(.0192)
Size	.1464***	.1472***	.1397***	.2071	.1464***	.1925***	.1176***	.1255***
	(.0156)	(.0178)	(.0157)	(.1660)	(.0199)	(.0266)	(.0129)	(.0171)
Groppor	.4899**	.4339***	.4599***	2.5615	.5023***	.7665	.3064**	.4615**
	(.1370)	(.1449)	(.1393)	(3.933)	(.1680)	(.5043)	(.1317)	(.1949)
Ndts	−.7819***	−.7460***	−.7563***	−.7817***	−.7547***	−.4328***	−.7416***	−.6849***
	(.0393)	(.0416)	(.0372)	(.1003)	(.0394)	(.0410)	(.0333)	(.0420)
Cred	.0354***	.0197***						
	(.0068)	(.0031)						
Share			−.0357***	−.0386***				
			(.0027)	(.0071)				
Disclos					−.0604	−.0602**		
					(.1132)	(.0267)		
Acc_{LLSV}							−.1892***	−.1110***
							(.0114)	(.0123)
Cred2	−.0632***							
(cred*collat)	(.0246)							
Cred3		.2558						
(cred*grop)		(.2766)						
Shar2			−.0004					
(share*collat)			(.0350)					

Table 13.5 (*Continued*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Shar3				−2.222				
(share*grop)				(3.599)				
Dis2					.0495			
(disclo*collat)					(.0497)			
Dis3						.1794		
(disclo*gropp)						(.5192)		
Acc2							.0573**	
(Acc*collat)							(.0232)	
Acc3								−.3588*
(Acc*groppor)								(.2153)
Constant	.5308***	.5572***	.7286***	.6892***	.6060***	.7386***	.000	.6130***
	(.0175)	(.0136)	(.0200)	(.0961)	(.04238)	(.0973)	(.000)	(.0166)
Obs	4100	4100	4100	4100	4100	4100	4813	4100
Hausman test	133.53a	66.53a	138.04a	341.86a	102.67a	46.50a	16.77	68.69[a]
Sargan test	63.12a	66.34a	177.31[a]	51.04a	45.64a	32.95[a]	576.76a	50.99a
Sector effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	No	No	No	No	Yes	Yes	No	No
Country dummies	Yes	Yes	Yes	Yes	No	No	Yes	Yes

*significant at 10%; **significant at 5%; *** significant at 1%.
Hausman test and Sargan test: a significant at 1%, b significant at 5% and c significant at 10%.

The scope and accuracy of information (*disclos*) present a negative and significant coefficient (column 6). Again, there are two alternative explanations. On the one hand, this negative sign implies that the costs associated with unfavorable information exceed the benefits associated with favorable information, thus transparency is costly and firms choose lower leverage ratios. On the other hand, more transparency helps stock market development, equity issuances increase and consequently leverage ratios reduce. The collateral and growth opportunities interaction terms present a positive but insignificant coefficient (columns 5 and 6). Therefore, neither sectors with more collateral assets nor high growing sectors (with more growth opportunities) operating in countries with ampler information scope and accuracy in the economy present distinctive debt ratios.

The coefficient for "the use of information" (acc_{LLSV}) is also negative and significant (columns 7 and 8). Therefore it confirms the results obtained for scope and accuracy of information. The interaction term of collateral assets is positive and significant. Hence, when firms operate in more transparent environments and the costs of disclosing are high, the presence of fixed assets mitigates the negative effect of disclosing on debt ratios. That is, assets that can be collateral reduce the adverse effects on transaction terms (ex. interest rates) when firms disclose unfavorable news. In the case of the growth opportunities interaction term, the coefficient is negative and significant. Therefore, the negative effects of disclosing may become more detrimental for leverage. The negative sign of the interaction term can be interpreted as a lesser effect of bad news disclosure. If future prospects are positive, the costs of disclosing do not offset the benefits, as bad news is less probable. Hence, firms are willing to finance new projects and as equity issuances are less costly, choose to issue equity (we observe less debt). The second interpretation is more straightforward. If more transparency implies more developed stock markets, more firms with future growth opportunities will be willing to make new equity issuances and therefore, we observe a reduction in debt levels. So the general message is that in countries with less asymmetric information, fast growing firms will present lower debt levels.

5. CONCLUDING REMARKS

Recent contributions claimed with reasonable confidence that institutional environment matters for financial decisions. Beyond this general characterization of institutions, there is a complex structure of rules and economic norms, which shape capital structure. The extent of disclosure regulation is, in our opinion, part of the most important determinants of such relationship. Further, economic rules evolve and innovations should also be taken into account. We expand previous studies to test these effects. First, we use international standardized data for a wide sample, and we test the main capital structure results jointly with the institutional effect that incorporates disclosure requirements. Second, we improve the construction of the institutional variables by taking into account legal innovations when available.

Our findings suggest a significant impact of disclosure requirements. In particular, legal disclosure requirements have a negative effect on leverage. On the one hand, this result can be derived from the costs associated to disclosing, especially when disclosing implies revealing bad news. On the other, transparency helps to develop stock markets, therefore more firms issue new equity and leverage levels reduce. This is the first study that analyzes the effect of legal disclosure requirements on capital structure. However, more research is needed in order to disentangle the viability of these competing theoretic explanations. Further, if firms have good future expectations and transparency increases, they present lower levels of debt. Transparency eases access to external finance if firms do not experience downturns. Provided firms are doing poorly (under these circumstances, equity issuance is not feasible), disclosure requirements can be too costly and they may decide to pass up good investment opportunities.

Besides, we find evidence that creditor protection has a positive effect on firm indebtedness. This finding is consistent with the theoretical predictions and previous empirical work. Further, we find that in countries with higher protection to creditors, agency problems are alleviated, since industries' need of collateral to secure credits is lower. Shareholder protection has a negative effect of leverage. Hence, when shareholders are effectively protected, stock market is more attractive for investors and firms can find external investors more easily. However, firm investment policy does not alter this result. The results just commented upon provide additional support for the institutional effect as a complementary explanation for finance decisions and confirms the importance of designing an adequate institutional environment to alleviate financial constraints.

ACKNOWLEDGEMENT

I am specially grateful to my thesis advisor Professor Sandro Brusco for guidance and encouragement. I also thank Julio Pindado for helpful comments on previous versions. The author acknowledges the financial support provided through the European Community's Human Potential Programme under contract HPRN-CT-2000-00064 [UFA].

BIBLIOGRAPHY

Alexander, D. and Archer, S. (1992). *The European accounting guide*. London: Academic Press London.

Allen, F. and Gale, D. (1995) "A welfare comparison of intermediaries and financial markets in Germany and the US." *European Economic Review* **39**: 179–209.

Almazan, A., Suarez, J. and Titmam, S. (2004) *Stakeholder, transparency and capital structure*. CEPR, Discussion Paper 4181.

Arellano, M. and Bover, O. (1990) "La econometría de datos de panel." *Investigaciones Económicas (Segunda Época)* **14(1)**: 3–45.

Bhattacharya, S. and Thakor, A.V. (1993) "Contemporary banking theory." *Journal of Financial Intermediation* **3**: 2–50.

Blake, J. and Amat, O. (1993) *European accounting*. London: Pitman.

Campbell, T. (1979) "Optimal investment financing decisions and the value of confidentiality." *The Journal of Financial and Quantitative Analysis* **14**: 913–24.

Campbell, D. and Moore, M. (1993) *Financial services in the new Europe. The comparative law yearbook of international business*. Special issue, 1992. London: Graham and Trotman.

Claessens, S. and Laeven, L. (2001) "Financial Development, Property Rights, and Growth." *The Journal of Finance* **58(6)**:

Cleary, S. (1999) "The relationship between firm investment and financial status." *The Journal of Finance* **54(2)**: 673–92.

Comisión Europea (1989) *Segunda Directiva* 89/646/CEE: Acceso a la actividad de las entidades de crédito y a su ejercicio.

Comisión Europea (1993) *El mercado interior de la comunidad*.

Delbreil, M., Cano, J.R., Friderichs, B.G., Paranque, B., Parsch, F. and Varetto, F. (1998) *Net Equity and Corporate Finance in Europe*. European Committee of Central Balance Sheets.

Demirguc-Kunt, A. and Maksimovic, V. (1998) "Law, finance and firm growth." *The Journal of Finance* **53(6)**: 2107–37.

——— (1999) "Institutions, financial markets, and firm debt maturity." *The Journal of Financial Economics* **54(3)**: 295–336.

Dewatripont, M. and Tirole, J. (1994) "The prudential regulation of banks." Cambridge, MA: The MIT Press.

Fabbri, D. (2001) *Legal institutions, corporate governance and aggregate activity: theory and evidence*, manuscript.

Frankel, A.B. and Montgomery, J.D. (1991) "Financial structure: an international perspective." *Brookings Papers on Economic Activity* **1**: 257–97.

Gertner, R. and Scharfstein, D. (1991) "A theory of workouts and the effects of reorganization law." *The Journal of Finance* **46**: 1189–222.

Giannetti, M. (2003) "Do better institutions mitigate agency problems? Evidence from corporate finance choices." *Journal of Financial and Quantitative Analysis* **38(1)**: 185–212.

Grossman, S. and Hart, O. (1988) "One share-one vote and the market for corporate control." *Journal of Financial Economics* **20**: 175–202.

Harris, M. and Raviv, A. (1990) "Capital structure and the informational role of debt." *The Journal of Finance* **45**: 321–49.

Hawkins, D. and Morton, C. (1990) *European corporate financial law: a guide to M&A and corporate restructuring legislation*. London: Euromoney publications.

Jensen, M.C. (1986) "Agency costs of free cash flow, corporate finance and takeovers." *American Economic Review* **76**: 323–39.

Jensen, M.C. and Meckling, W.H. (1976) "Theory of the firm: managerial behaviour, agency costs, and ownership structure." *Journal of Financial Economics* **3(4)**: 305–60.

Katayama, T. and Makov, R. (1998) "Deregulation of financial markets in Japan." *Journal of International Business Law* **4**: 128–33.

La Porta, R., López de Silanes, F., Shleifer, A. and Vishny, R. (1997) "Legal determinants on external finance." *The Journal of Finance* **52**: 1131–50.

——— (1998) "Law and finance." *Journal of Political Economy* **106(6)**: 1113–55.

——— (1999) *Investor protection and corporate valuation*. NBER, Working Paper 7403.

Lombardo, D. (2000) *Is there a cost to poor institutions?* Job Market Paper, October, Stanford University.

Lombardo, D. and Pagano, M. (1999) *Legal determinants of the return on equity*. CSEF, Working Paper 24.

Mayer, C. (1988) "New issues in corporate finance." *European Economic Review* **32**: 1167–89.

Myers, S. (1977). "The determinants of corporate borrowing." *Journal of Financial Economics* **5**: 147–75.

——— (1984) "The capital structure puzzle." *Journal of Finance* **39**: 575–92.

Myers, S. and Majluf, M. (1984) "Corporate financing and investment decisions when firms have information that investors do not have." *Journal of Financial Economics* **5**: 147–75.

Padilla, A.J. and Requejo, A. (1998) "The cost and benefits of the strict protection of creditor rights: theory and evidence." manuscript.

Pagano, M. (1989) "Endogenous market thinness and stock price volatility." *Review of Economic Studies* **56**: 269–88.

Pagano, M. and Röell, A. (1998) "The choice of stock ownership structure: agency costs, monitoring, and the decision to go public." *The Quarterly Journal of Economics* **113**: 187–225.

Rajan, R.G. and Zingales, L. (1995) "What do we know about capital structure? some evidence from international data." *The Journal of Finance* **50**: 1421–60.

——— (1998) "Financial dependence and growth." *American Economic Review* **88**: 559–86.

Raybould, D.M. and Firth, A. (1991) *Law of monopolies, competition law and practice in the USA, EEC, Germany and the UK*. London: Graham & Trotman.

Subrahmanyam, A. and Titman, S. (1999) "The going-public decision and the development of financial markets." *The Journal of Finance* **54(3)**: 1045–82.

Titman, S. and Wessels, R. (1988) "The determinants of capital structure choice." *Journal of Finance* **63(1)**: 1–19.

Wald, J.K. (1999) "How firm characteristics affect capital structure: an international comparison." *The Journal of Financial Research* **22(2)**: 161–87.

This Page is Intentionally Left Blank

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 14

MONEY LAUNDERING AND FINANCIAL OFFSHORE CENTRES: A POLITICAL ECONOMY APPROACH

Donato Masciandaro

1. INTRODUCTION

After the dramatic attack on the United States on the September 11, 2001, growing attention has been paid to the role of Non-Cooperative Countries and Territories (NCCT) in money laundering and terrorist financing.[1] Policymakers concentrate their attention on the possibility that NCCT jurisdictions might facilitate the task of terrorists as well as criminal organizations (*black money*).

Two interacting principles commonly feature in the debate on the relationship between money laundering and NCCTs: money laundering is facilitated by lax financial regulation;[2] countries adopting lax financial regulation do not co-operate in the international effort aimed at combating money laundering.[3] These two principles characterized the mandate of the Financial Action Task

[1] As Norgren (2004) noted, money laundering is defined as the processing of criminal proceeds to disguise their illegal origin in order to legitimize the gains of crime, while terrorist finance can be characterized as the direct or indirect provision of funds – illegal or legal – with the intention that they should used in terrorist acts. But the techniques are similar, or at least overlapping. On similarities and differences between money laundering and terrorism finance (or money dirtying) see the following section and von Furstenberg 2004; see also Rider (2003). On the key role of the U.S. legislation in promoting the international financial war against terrorism see Wasserman (2002), Banoun *et al*. (2002), Preston (2003), Van Cleef (2003); see also Davis (2003).

[2] On the relationship between money laundering and lax financial regulation see Section 3.

[3] See International Monetary Fund (1998), Holder (2003).

Force (FATF)[4] for the prevention of money laundering. To address the problems associated with money laundering risks, it is fundamental to develop legal standards for rules and regulations. The FAFT standards (Recommendations) became the benchmark for measuring the degree of laxity of financial regulation in every country setting. On the other hand, to monitor the compliance of countries with international standards, the FAFT used a list of specific criteria – consistent with the standards – to determine the NCCT jurisdictions.[5]

The FAFT produces periodic reports on the NCCTs, commonly described as blacklists. From June 2000 to February 2004, nine NCCT lists have been published and the FATF has monitored a total of 45 countries, selected for their potential regulation weakness. Using a worldwide data set on the main 130 countries, we can highlight that these 45 countries represent 8% of total GDP, 15% of total population, and 25% of foreign bank deposits worldwide. Obviously these figures understate the overall relevance of the problem, given the relationships between the non-co-operative attitude, on the one hand, and the global economic and social costs due to the growth of the money laundering risks, on the other.[6]

Therefore the blacklist instrument represents the cornerstone of the international effort to reduce the risks that single countries or territories became havens for money laundering activities. But is this institutional device effective?

It has been argued[7] that the overall result of the blacklisting mechanism is positive, since transparency regarding which countries do not comply has important

---

[4] The Financial Action Task Force on Money Laundering (FATF) is an inter-governmental organization that seeks to develop and promote policies at both national and international levels to combat money laundering. The FATF was established following the G7 Summit held in Paris in 1989. G7 members are: Canada, France, Germany, Italy, Japan, the United Kingdom (UK) and the United States (US). Initially, the FATF was convened from the G7 member States, The European Commission (EC) and 8 other countries, but it now has a membership of 29 jurisdictions, with the EC and the Gulf Cooperation Council as international member organizations. The 29 member jurisdictions are: Argentina, Australia, Austria, Belgium, Brazil, Canada, Denmark, Finland, France, Germany, Greece, Hong Kong, Iceland, Ireland, Italy, Japan, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, Turkey, the United Kingdom (UK) and the United States (US). The FATF has a small Secretariat that is housed in the headquarters of the OECD in Paris, but the FATF is a separate international body and not part of the OECD. See also Alexander (2001).

[5] On differences and similarities between NCCT jurisdictions and offshore centres see Mitchell (2003), Alworth and Masciandaro (2004); on the offshore centres issues see also Errico and Musalem (1999), Hampton and Christensen (2002), Masciandaro (2004a).

[6] On the qualitative and quantitative aspects of money laundering see Tanzi (2000).

[7] Norgren (2004). An economic analysis on the FAFT effects is performed by Johnson and Lim (2002). On the first different country reactions to the blacklisting process see Johnson (2001a) and (2001b).

effects in the financial markets, increasing the market pressures on the NCCT countries. Then why is it that various jurisdictions, notwithstanding the blacklist threat, delay or fail to change their rules, confirming their non-co-operative attitude (*reluctant friend effect*)? Further, it is true that most jurisdictions placed on the blacklist have enacted regulatory measures in an effort to be removed from it. But is regulatory reform sufficient to prove that a country has really changed its non-co-operative attitude (*false friend effect*)?

Perhaps the key problem is that discussions on these often take as a *given*, that some countries offer financial services to terrorism and organized crime by adopting lax financial regulations. In other words, lax financial regulation is treated as an independent variable. Therefore, any regulatory reform consistent with the international standards is sufficient to prove that the country is attempting to become a co-operative jurisdiction, while it fails to explain, for example, why specific countries continue in their non-co-operative attitude, notwithstanding the blacklist stigma.

This chapter takes a different perspective. We develop the assumption that lax financial regulation may be a strategic dependent variable for national lawmakers seeking to maximize the net benefits produced by any public policy choice. Therefore, given the structural features and endowments of their own countries, lawmakers may find it profitable to adopt financial regulations that attract capital of illicit origin (money laundering services) or destination (terrorism finance services), therefore choosing to be a NCCT jurisdiction.

From a methodological point of view, we develop the classic intuitions a' la Becker, using the new political economy approach, basing our work on three hypotheses:

1) the definition of regulatory policy is not independent, as in conventional economics, but endogenous;
2) policy is not determined by maximizing a social welfare function but by taking into account the political cost-benefit payoff;[8] and
3) lawmaker maximization is constrained and influenced by the structural framework, economic as well as institutional.

The chapter proceeds as follows. The second section proposes a simple model to describe, through the lawmaker payoff maximization, the relationships between specific country features and endowments, on the one hand, and lax financial regulations, on the other. Given that in the real world relatively lax regulation means a non-co-operative attitude in the international fight against

[8] For the new political economy see Drazen (2000) and Persson and Tabellini (2000).

black money, in the third section we empirically test the above theoretical relationship in the case of NCCT jurisdictions. The policy consequences on the pros and cons of international blacklisting procedures are discussed in the conclusive section.

2. LAX FINANCIAL REGULATION: KEY CONCEPTS

Therefore, we can identify four different categories of actors potentially interested in regulation:

1) the lawmakers;
2) terrorist and criminal organizations, deriving utility from the possibility of black money;
3) those who bear the costs of black money; and
4) the financial community and, in general, the citizens that receive benefits from the inflow of foreign black and grey capital.

Starting with this last category, it seems difficult to predict which side the financial community will take. In general, we tend to think that the utility function of financial intermediaries does not appear to be affected by whether profits stem from legal or illegal financial activities (*pecunia non olet*). We think that they simply maximize the expected revenues and that, given the asymmetric information issues, they are not able to clearly distinguish the customers' nature, legal or illegal.

The interests of 2) and 3) are obviously incompatible, as the gains of the former depend on the losses of the latter. 1) appears to be caught in the middle, having to decide which demand schedule to follow.

Note that we are not assuming that 2) and 3) are necessarily based *outside* the country where the lawmaker we are concerned with is based. This is not an assumption, but rather the consequence of our line of argument.

We have thus limited our attention to lawmakers that are based in countries different from those where the other actors, who are potentially interested in the regulations, are based in. From this starting point, the confrontation between those who benefit from black money and those who suffer from it, is almost a "win-win" game for criminal and terrorist organizations.

Organized crime and terrorism enjoy huge asymmetrical organizational advantages over those who bear the costs of money laundering. A small, powerful group opposes a large, dispersed group, thus making the outcome predictable.

It is certain that black money regulation could be opposed, and indeed is opposed, by political authorities that represent the public interest. However, the dispersion of the costs makes black money a low salience issue for the public, and consequently low on the political agenda. The man on the street simply does not feel the bite of black money, and political actors will act as a consequence.

3. COUNTRY ENDOWMENTS, LAWMAKERS AND LAX FINANCIAL REGULATION: A PRIMER MODEL

The relationship between black money and national financial regulation is a key issue in the international debate. To discuss this issue from an economic point of view it can be useful to treat the regulation against money laundering and terrorism finance as a product, with a demand and supply schedule. But whose demand schedule is driving the system?

We assume that the lawmaker in a given country has not yet decided the direction that it will impose on its financial regulation, with specific regard to black money. The lawmaker may decide to implement regulations that create serious obstacles to black money, and thus to terrorism and organized crime. It may decide – at the other extreme – to make the opposite choice, devising lax regulations that facilitate black money.

Black money generates costs as well as benefits for the parties involved. The costs for society depend on the fact that more predicate offences will be committed by terrorist or criminal organizations if black money is possible, and on the possible negative impact that black money will have on the economic system.

The benefits of black money accrue, first of all to terrorist and criminal organizations. Such organizations can employ the proceeds of crime and avoid the threat of prosecution for predicate offences (money laundering in the strict sense), or can use legal capital to finance illegal activities (money dirtying).

On the other side of the transaction, black money offers the host country the possibility to earn a "commission" in exchange for its services – what we can call *the expected national benefits* due to lax financial regulation.

But, as with all policy issues, as long as the costs and benefits of a decision fall within the boundaries of the area of influence of the lawmaker, we expect to have an efficient decision. Lawmakers in countries where crime or terrorism is pervasive will tend to bear at least some of the costs associated with a decision to favor black money.

Countries where organized crime or terrorism is pervasive might appear to play a minor role in the offer of black or grey financial services at the international level, because they are sensitive to terrorism- and crime-related national costs. This might be so because the widespread presence of organized crime or terrorism in the country increases, for the lawmaker, the costs of regulations that favor black money.

The public will bear the costs of the decision and will hold the lawmaker responsible. Entering the international market for black money services has a greater potential for countries that are immune from terrorist or criminal activities. By definition, such countries will almost be able to externalize the costs associated with the increase of predicate offences. A negative correlation between crime rate or terrorist episodes in the country and the role played in the offering of black money services appears likely.

As a result of this process, some countries that do not bear the costs associated with black money become predisposed to adopting lax regulations. The other side of the coin is that both criminal and terrorist organizations, and those who bear the costs stemming from black money, will "naturally" tend to be situated in countries other than the one where the regulations are adopted.

To design the key elements of our approach using the above economic concepts, we shall use a simple model, in order to present the economic intuitions in a compact and consistent framework. [9] Our goal is to discuss the possible relationships between specific country features, lawmaker payoff and lax financial regulation against black money.

Let us assume that a lawmaker is aware that a potential demand for black money exists on the part of one or more criminal or terrorist organizations,[10] for a total amount equal to W. We analyze a situation in which the international market for black money is demand-driven, as it is likely to be in the real world. Therefore every potential lax regulation jurisdiction is a relatively "small country".

The lawmaker can decide to launder an amount of money Y, $0 < Y < W$. For the sake of simplicity in our model, the decision on the optimal level of black money services is equivalent to the choice of the optimal degree of laxity in financial regulation. Calling U the payoff function of the lawmaker, it is obvious that the expected payoff from unlaundered liquidity is zero, whatever the amount:

$$U(W - Y) = 0 \tag{1}$$

Black money has a positive expected value for the lawmaker, if his country derives benefits from offering financial services. In particular, the lower the national income and the higher the proportion of that income that depends on

[9] For an in-depth analysis of the model see Masciandaro (2005a) and (2005b).

[10] For a general microeconomic analysis of the money laundering demand see Masciandaro (1996, 1998). For the peculiar relationship between money laundering demand and tax evasion see Yaniv (1994, 1999); see also Alldridge (2001).

the financial industry, the greater will be the propensity to offer black money services, all other things being equal. In general, we define those expected benefits as laxity national benefits.

To be more precise, black money provides B expected profit to the lawmaker:

$$B = mY \tag{2}$$

where $m > 0$ is the expected net rate of return on black money services. The inflow of black and grey foreign capital produces national revenues, increasing the activity of the financial industry and then throughout the traditional macroeconomic multiplier effects.[11] On the contrary, the implementation of a severe regulation against black money generates high compliance costs.[12]

If the decision to launder were cost free, $Y = W$. But other elements intervene.

First, lawmakers may face international reputation costs. To be more attractive to criminal or terrorist organizations, a country must make legislative and regulatory choices that increase its credibility as a lax financial regulation (LFR) jurisdiction.[13] Second, the activity of black money implies the strengthening of organized crime and terrorism.

Within our framework, we do not separate expected crime costs from expected terrorism costs. From the theoretical standpoint, we prefer to stress the different sensitivity of the lawmaker to expected international costs and expected national costs, based on a clearly different political cost-benefits analysis. Further, for each country, it should not be difficult to introduce in Equation (3) a specific parameter for each expected national cost factor. The chosen cost specification, C, consists of two parts. The first is reputational cost, captured by parameter $c > 0$. The second is the cost against crime or terrorism that rises as the amount of black money increases, captured by γ^2. Let us assume that for political-electoral reasons the lawmaker, all other things being equal, is more sensitive to the crime and/or terrorism costs, which can weigh directly on the country's citizens, than to the international reputation costs, whose effect on the citizens-voters is probably less perceptible and direct. We have:

$$C = cY + \gamma^2 Y \tag{3}$$

Finally, we must consider that a lax financial regulation jurisdiction is a source of economic, political and social risk for the international community

[11] For a macroeconomic analysis of the interrelationships between money laundering, banking industry, legal and illegal economic sectors see Masciandaro (2000). For the peculiar vulnerability of securities markets see Jayasuriya (2003).

[12] Masciandaro (1999).

[13] Masciandaro and Portolano (2003).

and prompts possible sanctions and punitive countermeasures. Let S denote the monetary value of sanctions and p the associated probability of discovering black money.[14] S must at least equal Y. In reality, the damage from a sanction is a multiple, because of the value of collateral damages related to the sanction:

$$S = tY^2 \tag{4}$$

where t denotes the degree of international political enforcement.

The lawmaker, modeled as a risk-neutral agent, is thus faced with the problem of deciding on the optimal level of laxity. The lawmaker's expected payoff E can now be better specified as:

$$E(U) = [(1-p)(B-C) - p(C+S)] \tag{5}$$

But since we have defines $B = mY$ and $C = cY + \gamma^2 Y$, then (5) becomes:

$$E(U) = (1-p)\left\{mY - cY - \gamma^2 Y\right\} - p\left(cY + \gamma^2 Y + tY^2\right) \tag{6}$$

and the optimal level of laxity is:

$$Y^* = \frac{m(1-p) - c - \gamma^2}{2pt}$$

For $Y^* > 0$, $m(1-p) - c - \gamma^2 > 0$ or the expected benefit from black money must exceed the cost of loss of reputation and the cost of fighting crime and terrorism. It is easy to check that Y^* rises as m rises and falls as c, γ, and p rise.

4. AN EMPIRICAL INVESTIGATION OF LAX FINANCIAL REGULATION AND NON-CO-OPERATIVE COUNTRIES

In this section we will test the implication of the simple model developed in the previous section. In the real world, the international community considers LFR countries as potential non-co-operative jurisdictions (NCCTs) in the fight against black money laundering. We assume that NCCTs share common structural features and can test for this using econometric techniques. In fact, financial

[14] For sanctions and enforcements, see the classic Becker (1968).

regulatory regimes can be viewed as resulting from a continuous, unobserved variable – the optimal degree of financial laxity, consistent with the lawmaker payoff. Each regime corresponds to a specific range of the optimal financial laxity, with higher discrete index values corresponding to a higher range of financial laxity. Since we will use a qualitative ordinal variable as laxity indicator, the estimation of a model for such a dependent variable necessitates the use of a specific technique.

In particular, given a constant international environment, we assume that an NCCT jurisdiction has scant physical resources to spend in international trade, which gives an incentive for lax financial regulation. The potential for developing financial services and can gain from lax financial regulation, social characteristics that shield it to some extent from the risks of terrorism and/or of organized crime, and thus reduce the expected cost of lax financial regulation.

Since June 22, 2000, the FATF has been publishing a periodic report on the NCCT jurisdictions – the blacklist. The report lays down 25 criteria, plus 8 recent special recommendations on terrorist financing that, if violated, identify the national rules that in each country are detrimental to international co-operation in the fight against black money. From June 2000 to February 2004, 45 countries have been monitored, and 9 blacklists have been published, indicating the jurisdictions that fail to conform to the criteria.

Using a worldwide data set on the main 130 countries,[15] we do a probit analysis. The dependent variable is a Binary Probit variable equal to 1 for the 45 potential NCCTs and 0 otherwise.

The estimated equation[16] is as follows:

$$(BinaryLI)_t = \beta_1 + \beta_2(A1)_t + \beta_2(C1) + \beta_4(E1) + \varepsilon_t t = 1 \ldots N \qquad (7)$$

where:

$A1$ = Land Use;[17]

$B1$ = GDP per capita;[18]

---

[15] Given the 267 world countries (UN members = 180), our 130 countries (BRI sample) represent the 98% of the world GDP and the 90% of the world population.

[16] Masciandaro (2005a) and (2005b).

[17] Landuse: This entry contains the percentage shares of total land area for five different types of land use: *arable land* – land cultivated for crops that are replanted after each harvest like wheat, maize, and rice; *permanent crops* – land cultivated for crops that are not replanted after each harvest like citrus, coffee, and rubber; *permanent pastures* – land permanently used for herbaceous forage crops; *forests and woodland* - land under dense or open stands of trees; *other* – any land type not specifically mentioned above, such as urban areas. Source: Central Intelligence Agency.

[18] Gdp-capita: This entry shows GDP on a purchasing power parity basis divided by population (year 2001). Source: Central Intelligence Agency.

$C1$ = Foreign deposits per capita;[19]
$E1$ = Terrorism and organized crime[20] Index.[21]

The results of Table 14.1 confirm that the probability of being an NCCT jurisdiction depends on specific country endowments. The probability that a country is a NCCT jurisdiction tends to be higher, the lower the level of economic development – measured by per-capita GDP and degree of land exploitation. The higher the flow of foreign deposits, the lower the extent of terrorism and organized crime. Given data limitation, we could not test for the role of international reputation sensitivity.

We can go a step further step if we hypothesize three different levels of non-co-operation:

1) level one non-co-operation for countries recently monitored by FAFT;
2) level two non-co-operation for countries with at least one presence in the blacklist; and
3) level three non-co-operation for countries that permanently stay in the blacklist.[22]

---

[19] Fordepositscapita: The data on foreign deposits are derived from reporting as such or calculated by subtracting separately reported data on positions other than deposits from total external assets and liabilities. The only exception is the Netherlands Antilles, which does not provide this information separately (year 2001). Source: BRI. The deposit data are then divided by the popolation (year 2001).

[20] Regarding the Organized Crime Dummy, the size of the drug market dimension is evidently an indirect and imperfect indicator of the organized crime problem. At the same time, the drug market has given organized crime its massive resources. It has been correctly noted that during the 1970s the drug trade became far too profitable and easy for even traditional and "conservative" organized crime organisations to ignore (see Rider (2002), p. 17), Further, it is also noted there that even terrorist groups entered the market and by so doing became virtually indistinguishable from "ordinary" organized crime.

[21] Terrorism and Organized Crime Index: we built this variable by summing two separate variables for each country: Organized Crime Dummy = 1 if there is drug production and/or drug markets in the country, 0 otherwise (Source: CIA); Normalized Terrorism Indicator = average number of terrorist episodes in the country (years 1968–91)/max average number of terrorist episodes in a country (1968–91); the Terrorism indicator therefore ranges from 0 to 1 (Source: Blomberg). Consequently, our Index ranges from 0 to 2.

Data Sources; Central Intelligence Agency – www.cia.gov/cia/publications/factbook; Democracy Index – www.geocities.com/CapitolHill/Lobby/3535/country/list-di.htm; Foreign Bank Deposits: Bank for International Settlements – www.bri.org/publ/qtrpdf/r_qa0206.pdf#page=44; Terrorism Indicators, see Blomberg *et al*. (2002) and ITERATE Data Set.

[22] The following list of NCCTs is current and was last changed in February 2004: Cook Islands, Guatemala, Indonesia, Myanmar, Nauru, Nigeria, Philippines.

Table 14.1 Binary Laxity Index determinants (130 countries and territories)

Dependent variable	Binary Laxity Index
Land Use	0.0079108****
	(0.003060)
Gdp capita	−0.0000723****
	(0.0000190)
Foreign deposit capita	3.18E-06****
	(1.36E-06)
Terrorism and org. crime	−0.5737521****
	(0.2436112)

Standard Errors in parentheses. Superscript asterisks indicate statistical significance at 0.01 (****), 0.02 (***), 0.05 (**), 0.10 (*).

These rankings, shown in Table 14.2, can be used as an ordered probit variable (complying countries are set equal to zero). The estimates of the ordered probit are shown in Table 14.3.

The regressions confirm the robustness of the two channels of national laxity benefits, while the proxy of the terrorism and organized crime risks has the right sign but is not statistically significant. If we split the organized crime dummy from the terrorism dummy, the former is statistically significant and the latter is not.

Further, it should be noted that non-co-operation is not associated with tax competition. While there is a theoretical presumption that international tax evasion and black money through *offshore centres* should overlap,[23] this is not necessarily the case.

We also explored the possibility that offshore financial centres are more prone to regulatory laxity than non-offshore centres (see Table 14.4). The dependent variable acquires a value of unity when a country is listed as an offshore centre by the OECD, otherwise it is zero.[24] With the exception of the crime and terrorism index, none of variables have any explanatory power. This seems to suggest that the underlying economic characteristics of offshore centres and our NCCTs tend to differ. In general, we can reject the hypothesis that the causes of lax financial regulation decisions and of offshore activities are exactly the same.

In conclusion, non-co-operation seems to be dependent on the key structural features of the country. Now what are the consequences of our analysis on the debate concerning the effectiveness of blacklisting procedures?

---

[23] Yaniv (1994) and (1999), Alworth and Masciandaro (2004).

[24] Alworth and Masciandaro (2004), Masciandaro (2005b).

Table 14.2 Ordered Laxity Index (OLI)

	Countries	**OLI**
1	Antigua	1
2	Bahamas	2
3	Barbuda	1
4	Belize	1
5	Bermuda	1
6	British Virgin I.	1
7	Cayman I.	2
8	Cook I.	3
9	Cyprus	1
10	Czech Republic	1
11	Egypt	2
12	Dominica	1
13	Gilbratar	1
14	Grenada	2
15	Guatemala	3
16	Guernsey	1
17	Hungary	2
18	Indonesia	3
19	Isle of Man	1
20	Israel	2
21	Jersey	1
22	Lebanon	2
23	Liechtenstein	2
24	Malta	1
25	Marshall I.	2
26	Mauritius	1
27	Monaco	1
28	Myanamar	3
29	Nauru	3
30	Nigeria	3
31	Niue	2
32	Panama	2
33	Philippines	3
34	Poland	1
35	Russia	2
36	Samoa	1
37	Seychelles	1
38	Slovak Rep.	1
39	St Kitts Nevis	2
40	St Lucia	1
41	St. Vincent	2
42	Turk Caicos	1
43	Ukraine	2
44	Uruguay	1
45	Vanuatu	1

Table 14.3 Ordered Laxity Index determinants (130 countries and territories)

Dependent variable	Ordered Laxity Index	
Land Use	0.0135717****	0.0144398****
	(0.0049385)	(0.0049597)
Gdp capita	−0.0000523****	−0.0000527****
	(0.0000155)	(0.0000161)
Foreign deposit capita	8.86E-08***	9.04E-08***
	(3.98E-08)	(4.05E-08)
Terrorism and org. crime	−0.3313072	
	(0.2245221)	
Organized crime		−0.4018445*
		(0.2414516)
Terrorism		(0.0099674)
		(0.0293882)

Standard Errors in parentheses. Superscript asterisks indicate statistical significance at 0.01 (****), 0.02 (***), 0.05 (**), 0.10 (*).

Table 14.4 Comparing Binary Offshore Index and Binary Laxity determinants (130 countries and territories)

Dependent Variable	Binary Laxity Index	Binary Offshore Index
Land Use	0.007***	−0.002
	(0.003)	(0.005)
Gdp capita	−7.07E-05****	−2.04E-07
	(1.92E-05)	(2.60E-07)
Foreign deposit capita	3.18E-06****	1.71E-06
	(1.36E-06)	(1.33E-08)
Terrorism and org. crime	−0.508***	−1.888****
	(0.224)	(0.448)

Standard Errors in parenthesis. Superscript asterisks indicate statistical significance at 0.01 (****), 0.02 (***), 0.05 (**), 0.10 (*).

5. CONCLUSIONS ON WHETHER BLACKLISTING IS AN EFFECTIVE DEVICE?

In this chapter we theoretically discuss and empirically test the relationships between specific country features, lawmaker choices toward lax financial regulation, and national non-co-operative attitudes with respect to the international effort to combat black money phenomena. Our results suggest two main prescriptions for designing international policies aimed at reducing the global risks of terrorism and organized crime.[25]

First, a pure and just formal "name and shame" approach may even prove counterproductive. Assuming that the international community is capable of effectively singling out NCCT jurisdictions that are indeed involved in black money schemes, a cautious approach is still deemed necessary. When the international community points the finger at a given country as a leading supplier of black money financial services, it may also be certifying, to the benefit of the country itself, that the country is indeed specialized in that business. The signaling effect embedded in the "name and shame approach" should not be underestimated. The main difficulty for a genuine LFR country is credibly solving the commitment problem. Then, what is a better choice for an LFR country than having the international community – not exactly its closest friends – solving that problem through a public statement certifying a non-co-operative attitude (reluctant friend effect)?

It is a matter of fact that the blacklisting procedures do not seem to have an influence on the flows of foreign financial assets toward the NCCT jurisdictions. Tables 14.5 and 14.6 show that financial figures, on Indonesia, Myanmar, Nigeria and Philippines (data on Cook Islands, Guatemala and Nauru are not available) to be in a blacklist do no cause any visible effect.

In other words, listing should also be regarded as a sort of third-party bonding, which is likely to generate two interacting effects. First, it is capable of cementing the commitment by the LFR country. Second, listing increases the transaction-specific nature of investments in reputation. Inclusion in a blacklist could increases the value of the sunk investment in reputation. In terms of our analysis, blacklisting could raise the expected benefits rather than improve international political enforcement.

Further, a blacklisted country will find it even more difficult to switch course and decide to exit the market, thus being encouraged to compete more aggressively in that market.

The second conclusion that can be reached, based on the empirical evidence we have examined, is that we must not exclude the possibility that there are LFR countries not presently included in the FATF monitoring action. This is true, perhaps, because they are highly effective in bringing their formal rules in line with international precepts, while in fact they remain lax. Similarly, by modifying the formal rules, NCCT countries may not shake off their acquired reputation for laxity (false friend effect). The "name and shame" approach, separated from

[25] On the possible specific role for the G8 countries in combating black money see Masciandaro (2005c).

Table 14.5 IMF Monetary Survey: Deposit Money Banks: Foreign Asset billions assets LCU

	1994	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
Indonesia	26351	33737	51633	60331	111844	116141	177941	229202	233736	254321
Myanmar	−11.124	−10.143	−11.332	−10.96	−8.74	−9.856	−10.195	−10.384	−8.974	−9.135
Nigeria	55.869	107.768	237.359	228.494	234.954	662.507	1275.072	1433.027	1387.199	1475.694
Philippines	125.86	117.92	70.23	−48.93	92.51	278.29	302.37	329.81	480.66	573.46

Table 14.6 IMF Monetary Survey: Foreign Asset (Net) billions assets LCU

	1994	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
Indonesia	5.851818	7.407279	8.736886	10.06667	14.41209	16.96669	10.64919	10.55519	10.08356	9.136444
Myanmar	0.000172	0.000524	0.001352	0.003172	0.000165	0.059524	0.000459	0.00192	0.001438	na
Nigeria	1.15693	3.4902	2.898657	3.179658	4.394773	1.651394	2.035491	2.700567	3.150396	3.206293
Philippines	6.035711	6.402304	8.184723	8.878049	9.153332	10.11411	8.180927	7.46829	8.074431	8.238586
	1994	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
Myanmar	0.000172	0.000524	0.001352	0.003172	0.000165	0.059524	0.000459	0.00192	0.001438	na

other initiatives, may not be effective. Countries can be blacklisted, but only if blacklisting goes hand-in-hand with other measures.

Appropriate countermeasures that increase the actual level of international political enforcement and/or the level of international reputation costs should be grounded on the premise that in a global world even the most efficient LFR country will still need to be integrated into the world financial markets. This implies that no matter how many layers of transactions cover the targeted offence, terrorism or criminal organizations will still need to place that money within the lawful financial sector. This step is necessary, at a minimum, to exploit the capital in lawful uses, once it has been laundered. Black money is by definition instrumental to a later use.

In this regard, there is one fundamental feature of the initiative taken by the FATF that appears to be pivotal for its success. The FATF has not limited its initiative to a mere recognition of "non-co-operative countries and territories." FATF member states have also applied "Recommendation 21"[26] to the countries included in the list. Recommendation 21 requires a higher scrutiny by financial intermediaries in evaluating the suspect nature of transactions with counter parties, including legal persons, based in a country listed as non-co-operative. As a result of the FATF initiative, many countries included in the list have already taken initiatives aimed at overcoming the serious deficiencies observed by the FATF.[27]

These initiatives need to be evaluated over the long term because some of the enacted laws, for example, will require the issue of secondary regulations to become effective or, more generally, the initiatives taken at the legislative level will need to be followed by concrete actions. It can be argued, however, that the threat of being crowded out by the international community has played a key role in spurring the adoption of the above-mentioned initiatives. However, it may be the case to go beyond that. The international community could consider the possibility to introduce effective punitive measures as a *financial quarantine* for every country that did not adhere to the international standards.[28] Finally, the above conclusions imply a constant effort on the part of international organizations, particularly the FATF, to update the criteria and monitor the countries.

[26] See FAFT (1990, 2000). In addition, on June 2001, FAFT agreed to a process of stricter countermeasures for reluctant NCCTs; see Norgren (2004).
[27] See FAFT press communiqué of October 5, 2000.
[28] On the possible features of a financial quarantine see Tanzi (2000).

ACKNOWLEDGEMENT

The author wishes to thanks Michele Bagella, Leonardo Becchetti, Gerald Dwyer, Bob Eisenbeis, Giuseppe Maresca for helpful discussions and useful suggestions. The usual disclaimer is applied.

BIBLIOGRAPHY

Alexander, K. (2001) "The International Anti-Money Laundering Regime: The Role of the Financial Action Task Force." *Journal of Money Laundering Control* **4(3)**: 231–48.

Alldridge, P. (2001) "Are Tax Evasion Offences Predicate Offences for Money Laundering Offenses." *Journal of Money Laundering Control* **4(4)**: 350–9.

Alworth, J. and Masciandaro, D. (2004) "Public Policy: Offshore Centre and Tax Competition: The Harmful Problem," in D. Masciandaro (ed.), *Global Financial Crime. Terrorism, Money Laundering and Offshore Centres*. Aldershot: Ashgate (forthcoming).

Banoun, R., Cephas, D. and Fruchtman, L.D. (2002) "US Patriot Act and Other Recent Money Laundering Developments Have Broad Impact on Financial Institutions." *Journal of Taxation of Financial Institutions* **15(4)**.

Basel Committee on Banking Supervision (1988) *Prevention of Criminal Use of the Banking System for the Purpose of Money-Laundering*, Basel.

Becker, G. (1968) "Crime and Punishment: an Economic Approach, *Journal of Political Economy* **2**: 169–217.

Blomberg, B.S., Hess, D.G. and Weerapana, A. (2002) *Terrorism From Within: An Economic Model of Terrorism.*

Boeschoten, W. and Fase, M. (1992) "The Demand for Large Bank Notes." *Journal of Money, Credit and Banking* **2(3)**.

Davis, K. (2003) "Legislating against the Financing of Terrorism: Pitfalls and Prospects., *Journal of Financial Crime* **10(3)**: 269–74.

Drazen, A. (2000) *Political Economy in Macroeconomics*, Princeton: Princeton University.

Errico, L. and Musalem, A. (1999) *Offshore Banking: An Analysis of Micro- and Macro-Prudential Issues, Working Paper of the International Monetary Fund*, no. 5.

Financial Action Task Force (1990) *The Forty Recommendations*.

Financial Action Task Force (2000) *Review of Non-Co-operative Countries or Territories: Increasing the Worldwide Effectiveness of Anti-Money Laundering Measures*.

Hampton, M.P. and Christensen, J. (2002) "Offshore Pariahs? Small Island Economies, Tax Havens, and the Re-configuration of Global Finance." *World Development* **30(9)**: 1657–73.

Holder, W.E. (2003) "The International Monetary Fund's Involvement in Combating Money Laundering and the Financing of Terrorism." *Journal of Money Laundering Control* **6(4)**: 383–7.

International Monetary Fund (1998) *Money Laundering. The Importance of International Countermeasures*, address by Michel Camdessus, Plenary Meeting of the FATF, pp. 1–4, mimeo.

Jayasuriya, D. (2003) "Money Laundering and Terrorism Financing: The Role of Capital Market Regulators." *Journal of Financial Crime* **10(1)**: 30–6.

Johnson, J. (2001a) "In Pursuit of Dirty Money: Identifying Weaknesses in the Global Financial System." *Journal of Money Laundering Control* **6(1)**: 122–32.

Johnson, J. (2001b) "Blacklisting: Initial Reactions, Responses and Repercussions." *Journal of Money Laundering Control* **6(3)**: 211–25.

Johnson, J. and Lim, Y.C.D. (2002) "Money Laundering: Has the Financial Action Task Force Made a Difference?" *Journal of Financial Crime* **10(1)**: 7–22.

Masciandaro, D. (1996) "Pecunia Olet? Microeconomics of Banking and Financial Laundering." *International Review of Economics and Business*, **October**: 817–44.

Masciandaro, D. (1998) "Money Laundering Regulation: the Micro Economics." *Journal of Money Laundering Control* **2(1)**: 49–58.

Masciandaro, D. (1999) "Money Laundering: the Economics of Regulation." *European Journal of Law and Economics* **3(May)**: 245–240.

Masciandaro, D. (2000) "The Illegal Sector, Money Laundering and Legal Economy: A Macroeconomic Analysis." *Journal of Financial Crime* **2**: 103–12.

Masciandaro, D. (2004a) (ed.) *Global Financial Crime. Terrorism, Money Laundering and Offshore Centres*, Aldershot: Ashgate.

Masciandaro, D. (2004b) "Migration and Illegal Finance." *Journal of Money Laundering Control* **7(3)**: 264–71

Masciandaro, D. (2005a) "False and Reluctant Friends? National Money Laundering Regulation, International Compliance and Non-Cooperative Countries." *European Journal of Law and Economic* (forthcoming).

Masciandaro, D. (2005b) "Could Sticks Become Carrots? Money Laundering, Black Lists and Offshore Centres." *Finance India* (forthcoming).

Masciandaro, D. (2005c) "Combating Black Money: International Cooperation and the G8's Role," in M. Fratianni, J. Kirton and A. Rugman, *New Perspectives on the G8*, Aldershot: Ashgate (forthcoming).

Masciandaro, D. and Portolano, A. (2003) "It Takes Two to Tango: International Financial Regulation and Off-shore Centres." *Journal of Money Laundering Control* **6(4)**: 311–31.

Masciandaro, D. and Portolano, A. (2004) "Financial Policies: Offshore Centres and Competition in Regulation: The Laxity Problem," in D. Masciandaro (ed.), *Global Financial Crime. Terrorism, Money Laundering and Offshore Centres*. Aldershot: Ashgate (forthcoming).

Mitchell, D.J. (2003) "US Government Agencies confirm that Low Tax Jurisdictions are not Money Laundering Havens." *Journal of Financial Crime* **11(2)**: 127–33.

Norgren, C. (2004) "The Control of Risk Associated with Crime, Terror and Subversion." *Journal of Money Laundering Control* **7(3)**: 201–6.

Persson, T. and Tabellini, G. (2000) *Political Economics: Explaining Economic Policy*, Cambridge MA: MIT University Press.

Preston, E. (2003) "The US Patriot Act: New Adventures in American Extraterritoriality." *Journal of Financial Crime* **10(1)**: 104–16.

Rider, B.A.K. (2002) "Weapons of War: the Use of Anti – Money Laundering Laws against Terrorist and Criminal Enterprises." *International Journal of Banking Regulation* **4(1)**: 13–31.

Rider, B.A.K. (2003) "Financial Regulation and Supervision after 11th September, 2001." *Journal of Financial Crime* **10(4)**: 336–58.

Rogoff, K. (1997) *Foreign and Underground Demand for Euro Notes: Blessing or a Curse?*, Working Paper, 1997.

Romano, R. (1985), "Law as a Product: Some Pieces of the Incorporation Puzzle," *Journal of Law, Economic, and Organisation* **1**: 225.

Romano, R. (1993) *The Genius of American Corporate Law*, Washington, DC.

Romano, R. (1999) "Corporate Law and Corporate Governance, 365," in G.R. Carroll and D.J. Teece (eds), *Firms, Markets, and Hierarchies*, New York.

Sinn, H.W. and Westermann, F. (2001) *Why has the Euro been falling?*, CESIfo Working Paper No. 493, May.

Tanzi, V. (2000), *Money Laundering and the International Financial System, in Policies, Institutions and the Dark Side of Economics*. Cheltenham: Edward Elgar. pp. 186–200.

Van Cleef, C. (2003) "US Patriot Act: Statutory Analysis and Regulatory Implementation." *Journal of Financial Crime* **11(1)**: 73–101.

Von Furstenberg (2004) "Terrorist Finance: Within the Grip of the G8," *this volume*.

Wasserman, M. (2002) "Dirty money." *Regional Review*, Federal Reserve Bank of Boston **12(1)**: 14–21.

Yaniv, G. (1994), "Taxation and Dirty money laundering." *Public Finances/Public Finance* **49**(Supplement): 40–51.

Yaniv, G. (1999) "Tax Evasion, Risky Laundering, and Optimal Deterrence Policy." *International Tax and Public Finance* **6**: 27–38.

This Page is Intentionally Left Blank

Transparency, Governance and Markets,
Edited by M. Bagella, L. Becchetti and I. Hasan.

Chapter 15

NETWORKS AND STOCK MARKET INTEGRATION: EMPIRICAL EVIDENCE

Iftekhar Hasan* and Heiko Schmiedel

Abstract

Increased competition, regulatory initiatives, globalization, and technological developments have altered the business strategies of stock exchanges around the world. This chapter investigates whether the adoption of network strategies by stock exchanges creates additional value in the provision of trading services. Using unbalanced panel data from all major European exchanges over the period 1996–2000, we examine the consequences of network co-operation on a number of stock market performance measures. The evidence reveals that adopting a network strategy is significantly associated with higher market capitalization, growth and lower transaction costs among sample markets. Network initiatives are found to foster European stock market integration.

Keywords: stock exchanges, network externalities, remote access, Europe

JEL classification No.: F36, G15, O52

1. INTRODUCTION

In recent years, stock exchanges have been experiencing a challenging and unprecedented environment. Globalization and integration of all types of financial markets, the continuous emergence of innovative technology, new regulatory initiatives, and the adoption of alternative corporate governance systems are

*Corresponding author

among some of the key issues faced by exchanges around the world. The integration phenomenon increased the popularity of mergers, especially implicit mergers or network deals among exchanges. As companies seek to broaden their shareholder base and raise capital beyond local markets (Domowitz *et al.* 1998, Pagano *et al.* 2001, and Steil 2001), such implicit mergers[1] are preferred by investors as an alternative to multiple listings across markets. Exchanges prefer this type of deal, which allows them to avoid direct competition from stronger markets and the fragmentation of liquidity. This type of arrangement is likely to develop a competitive environment, where the most efficient exchanges will eventually win the confidence of investors, traders and companies (Cybo-Ottone *et al.* 2000).

The emergence of these types of consolidation provides a common trading platform among exchanges who are willing to open up to each others' markets for cross listing and trading purposes, with ample freedom for brokers and traders to operate across markets. Network arrangements will help in gaining new demands for exchange products and are also likely to bring efficiency gains through economies of scale (Economides 1993 and 1995; Hasan and Malkamäki 2001). Hagel III-Armstrong (1997) and Saloner and Shepard (1995) emphasize the role of critical mass and time dimensions in evaluating the true impact of network scope.

Shapiro and Varian (1999) point out that computer technology, i.e. networks, will dominate the trading business. Networks will provide investors with options to choose from alternative preferences. The recent success of EUREX is a good example of how networks can replace a trading floor in another country.[2] European exchanges, historically local monopolies, are the most active players in adopting such a network or common trading platform. Taking their cue from NASDAQ's proposed and partially implemented global plan to list and trade across markets, the European exchanges have taken the lead in forming and joining in active network co-operation among European markets. In fact, the majority of the 100 executed or potential merger-related deals in the world are in Europe (Cybo-Ottone *et al.* 2000). Today, there are four inter-exchange co-operation models that link security markets within and outside European boundaries (Figure 15.1).

---

[1] A definition also used by Di Noia (2001) and Domowitz (1995) for equity and derivative markets respectively.

[2] An additional example is the emergence of network externalities especially in the United States, where there has been a huge invasion of new equity routing/matching/trading systems, e.g., Instinet, POSIT, AZ, and Attain etc. These systems have gained increasing volume, especially in stocks listed on NASDAQ as well as many NYSE-listed stocks. This situation has opened increased pressure and possibilities for exchanges to cooperate and compete for market share.

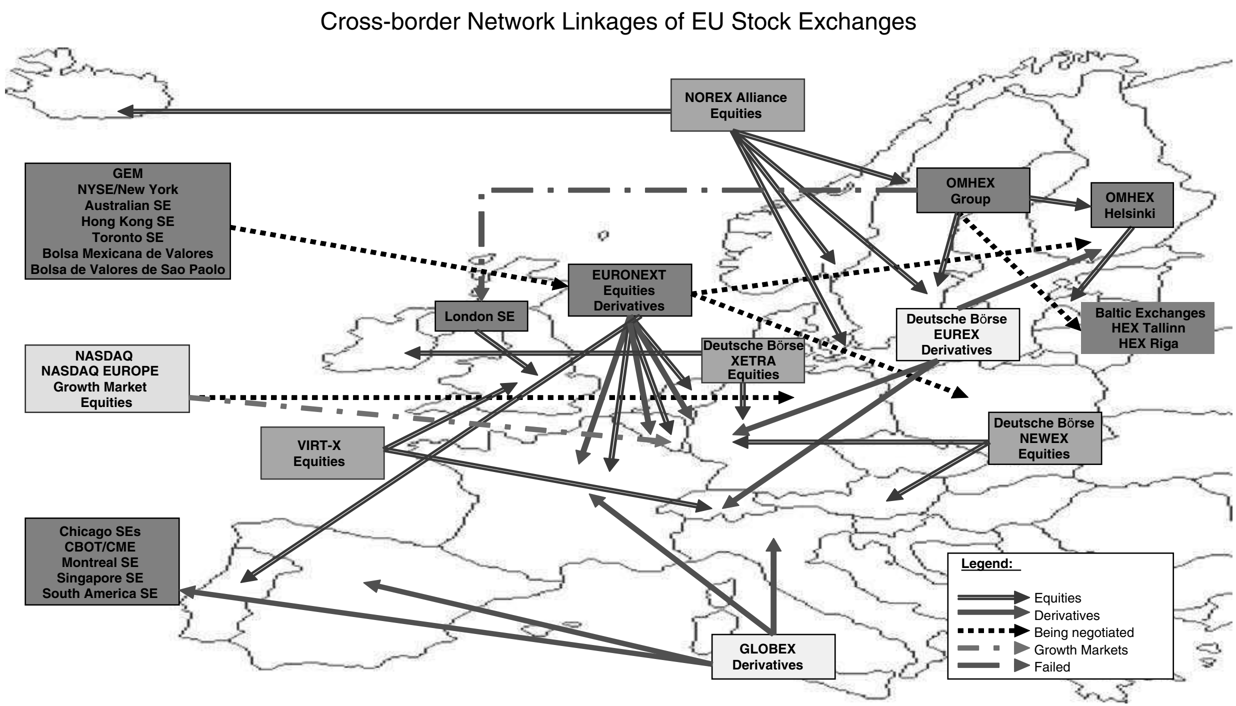


Fig. 15.1. Network of European stock and derivative exchanges

While the finance literature is abundant in introducing and describing the potential benefits of network arrangements in terms of increased participation, liquidity, efficiency and transaction costs, no article discusses the potential consequences or impact of adopting such network co-operation. Cybo-Ottone *et al.* (2000) provide the first descriptive approach to understanding mergers and cooperation across exchanges. However, their study was focused primarily on the factors associated with consolidation efforts. A separate volume of papers focused on the motives as well as on the consequences of cross-border listings and cross-listed stocks (Blass and Yafeh 2001; Chaplinsky and Ramchand 2000; Foerster and Karolyi 1998; Karolyi 1998; and Pagano *et al.* 2003). These papers, however, are more focused on the motivations and consequences among the companies rather than on the impact of cross listings on markets. Importantly, for our purpose, none of the papers deals with issues associated with networks or implicit mergers.

In this chapter, we attempt to fill this gap in the literature, not only by introducing details on the landscape of network co-operation among exchanges in Europe, but also by showing the potential impact of such inter-exchange co-operative initiatives on the performance, growth and transaction cost of the sample exchanges.[3] Our evidence shows that even after controlling for pertinent variables, the network co-operation decision, represented by several alternative network proxy variables, is significantly associated with stock exchange market capitalization, its growth, and to a lesser extent with lower transaction cost.

The chapter is organized as follows: Section 2 introduces networks, alliances and co-operation among European stock exchanges, enriched with a brief literature review. Section 3 introduces the data and descriptive statistics. Section 4 reports the results, and the conclusions are presented in Section 5.

2. NETWORKS, ALLIANCES, AND CO-OPERATION AMONG EUROPEAN STOCK EXCHANGES

Evidence for the presence of network externalities is starting to develop in various ways and can be seen in types of international alliances and co-operative arrangements between exchanges. The overall goal is to provide investors with

---

[3] Arnold *et al.* (1999), Domowitz and Steil (1999), and Pirrong (1999) stress the importance of assuming that exchanges are actually operative firms and argue that the industrial structure of market places cannot be explained by focusing on the demand side alone, as in financial market microstructure studies that concentrate on the characteristics of trading systems and the demand side of trading services, i.e., the traders. It is equally important to know more about the provision of alternative technologies for trading services.

the opportunity to trade shares of globally listed firms on a continuous 24/7 basis at the lowest possible cost of trading. In this scenario, the implications of electronic trading play a pivotal role and are far-reaching for the entire securities industry. However, in financial exchange markets, the innovation and implementation of new electronic trading technologies varies considerably by geography, culture and the organizational structure of the exchanges, which have been undergoing enormous transitions in recent years (Hasan *et al*. 2003). For example, there is evidence that North American stock exchanges operate most efficiently in order to serve the best interests of the marketplace, and in particular those of investors. However, Europe has been quicker and ambitious to respond to the rise of electronic trading by adopting it and creating several co-operative market linkages between stock and derivative exchanges (Schmiedel 2001).

Amongst the anticipated benefits of co-operative projects and strategic alliances among exchanges, were that they would give exchanges the opportunity to gain advantage over their competitors, mostly by extending trading hours, allowing for remote membership, modifying prices, and thereby lowering costs. It is crucial for the success of networked electronic trading platforms that increasing efficiency, transparency, faster executions and lowering costs can attract a critical mass of order-flows and generate additional liquidity in the market. The liquidity effect, in turn, is determined by the scope and size of the network, requiring compatible trading technologies.

A range of the most recent market linkages and co-operative initiatives proposed and undertaken by various stock exchanges are represented in Figure 15.1. It illustrates the architecture of market linkages and co-operation proposed and undertaken by various stock and derivative exchanges, forming complex and networked European securities trading landscape.[4] Tracing back the development of these linkages, shows that a large number of deals among exchanges were a recent phenomena, which have been mostly negotiated between 1997–2002. It seems evident that financial exchanges use different means of coping with investor demands for lower trading costs, improved liquidity and immediate access to international trading. However, some structural patterns can be derived as to how European stock exchanges create inter-connections between co-operating exchanges, as they were displayed in Figure 15.1. Although some of the deals among stock exchanges have failed or were abandoned, it seems apparent that Europe is increasingly a favorable environment in which stock

[4] A good survey of historical deals among stock exchanges illustrating various aspects of cooperation is presented in a number of studies, including the work of Cybo-Ottone *et al*. (2000), Domowitz (1995), Domowitz and Steil (1999), Lee (1998), and Licht (1998).

exchanges pursue co-operative strategies in order to build up networked markets and create additional value in the provision of their trading services.

The economic theory of network externalities provides a rationale for such co-operation. In the literature, network externalities are defined as a production or consumption positive size externality.[5] Formally, networks consist of links that connect nodes. In a typical network, the addition of a new consumer (or network node) increases the willingness to pay for network services among all participants. This effect is called network effects or network externalities. Several authors apply the concept of networks to financial intermediation and securities markets. Regarding a financial exchange network, Domowitz (1995) and Domowitz and Steil (1999) state that an exchange or a trading system is analogous to a communication network, as the benefit to one trader transacting in a given trading system increases when another trader chooses to transact there as well.

Economides (1996) points out that there are two ways in which financial exchange networks exhibit network externalities. First, the act of matching buys and sells for goods or assets generate a composite good, namely the exchange transaction. It is important that a critical mass of counter offers is available. In financial terms, minimal liquidity is required for the transaction to succeed. Second, network effects may also stem from different vertically related services necessary for a financial transaction, i.e. the matching services of brokers. However, the first type of externality seems to be more pronounced in financial markets.

Strong network externalities force exchange markets to create formal or informal linkages. The exact design of such inter-connections is less important. They are likely to occur in the form of implicit and explicit acquisitions and mergers, strategic alliances, simply pooling order-flows, or even information sharing agreements as discussed in Domowitz and Steil (1999). Financial exchanges that are less active in forming alliances, or linkages are likely to lose competitive ground vis-à-vis their counterparts engaging in network strategies.

The existing literature on networks, that relates to stock exchanges or to financial intermediaries, is theoretical or descriptive in nature. We are not aware of any empirical literature particularly dealing with network economics among the exchanges. A number of articles – as mentioned earlier – focused on the impact

---

[5] The concept of network externalities is developed in the New Theory of Industrial Organization and represents an important field in economics, as it applies to a variety of industries, such as telecommunications, airlines, railroads etc. (Shy 2001). An interactive bibliography on the network-externalities literature and related issues applied to finance can be found online at http://www.stern.nyu.edu/networks/biblio.html.

of cross-listing across exchanges and evaluated its impact on stock prices.[6] Additionally, Cybo-Ottone *et al*. (2000) outlined the merger of exchanges during the 1990s. However, they did not investigate any likely association between networks or implicit mergers with different elements of exchange-specific firm performance, volatility and efficiency. Thus, there is an obvious need for empirical research in this area. This study analyzes empirically the implications of network externalities for liquidity, trading costs, and growth in securities markets in Europe.

3. DATA AND METHODOLOGY

The empirical approach in this chapter is to trace the potential relationship between network variable(s) and several measures of exchange performance. These performance measures include market capitalization, the growth of market capitalisation, and the transaction costs of trading of the respective exchanges. The estimations control for other pertinent variables that are likely to affect stock exchange performance, such as the local economic environment, the relative importance of the private sector, accounting or disclosure standards, market monopoly by the largest firms, the costs of trading, market competition and size.

The data used in this study come from a variety of sources, including annual reports of stock exchanges, various issues of the International Federation of Stock Exchanges (IFBV), IMF International Financial Statistics (IFS), Elkins/McSherry, and information from exchange Internet sites. Most of the data were collected from annual balance sheets, income statement reports, and the Internet pages of all major operating stock and derivative exchanges covering a five-year time period (Annual Reports 1996–2000). In some cases, additional information was obtained from the exchanges through correspondence. Also various issues of the MSCI Handbook served as an important source of information on exchange-specific characteristics, such as the concentration of market share of the top three companies in each market (a proxy for market monopoly by largest firms) as well as the number of additional exchanges in the country (market competition) where the sample exchange is located.

A consistent data set has been constructed including all necessary information on 24 individual exchanges' key balance sheet and income statement items, of which 120 observations over the period 1996–2000 finally entered into the

[6] See Blass and Yafeh (2001), Chaplinsky and Ramchand (2000), Foerster and Karolyi (1993), Karolyi (1998) and Pagano *et al*. (2001).

estimations. All national currencies are converted into USD and are inflation-adjusted using data from IFS. All variables other than qualitative proxies are expressed in natural logarithms.[7,8] The accounting or disclosure standard is constructed by using the CIFAR index to measure the quality of accounting disclosure, a method used previously by researchers. The CIFAR index used in the existing literature represents the average number of 90 specific items disclosed in the annual reports of at least 3 companies per country, including items from the company's income statement, balance sheet, statement of cash flows and notes to the financial statements. The maximum score a country can obtain is 90.[9]

In order to examine network effects among stock exchanges, a dataset has been compiled, including all major inter-market connections along different types of exchange markets in the European Union (EU). Since network among exchanges is more frequent and plays an important role in European markets, we focus in this study on EU linkages. Accordingly, the network linkages in our data set include two or more entities where at least one entity is a European exchange. Figure 15.1 portrays all strategic cooperation, network experiences, and announcements among European stock and derivative exchanges by the year 2002. Building on this diagram, we traced back the development of each network to its year of implementation and establishment. The experiences of European exchanges from the mid-1990s to 2002 shows that network strategies are only recent phenomena. The total number of such linkages considerably increased after 1997–98.

A classification of network linkages has been made according to different market categories, in order to control for compatibility among different types of networks. This is particularly important, since stock exchanges are engaged in multiple transaction and trading services in various stock and derivative markets. As already mentioned in Cybo-Ottone *et al.* (2000), the classification of networks is not a straightforward exercise, given only limited access to information and details in respect of announcements, implementation status and network members. Against this background, the underlying categorization in this chapter may, however, slightly differ from schemes employed in related studies or official views stated by the exchanges themselves.

Different NETWORK variables were constructed in order to examine network externalities in financial exchange markets. The first variable included in this study controls whether an exchange generally pursues any kind of network strategy. If an exchange is engaged in networks and maintains/offers network access the variable ACCESS takes a value of one, otherwise zero. Secondly and

---

7 In constructing the growth variable, we have also used 1995 data.
8 See Schmiedel (2001) for more details on the European sample exchanges.
9 La Porta *et al.* (1997, 1998) have used this source to identify the accounting standard.

more specifically, the total number of different types of networks, NDN, captures the fact that exchanges build up various connections with varying network partners. Therefore, the variable NDN proxies the overall network activity of such exchanges that have successfully established different and not necessarily fully compatible network connections with other participating exchanges. Based on the theoretical considerations proposed in Section 2, however, the value of a network increases exponentially with each new participant that enters the network. Accordingly, the third variable, NNM, accounts for all members that are connected via each market's network.

Further, a key factor for analyzing these networks is to distinguish them alongside different types of securities segments. In respect to the total number of stock exchanges linked through networks, these market interconnections were classified using three criteria: blue chip equity markets, derivative markets, and new markets for innovative and mostly high-tech oriented companies. Equity markets account for inter-linkages and co-operation among exchanges that were established primarily for trading in all major blue chips. Derivative markets capture networked trading platforms for options and financial futures, while new growth and tech-oriented markets comprise interconnections of markets with newly listed high-growth and innovative-oriented firms. Figure 15.1 plots all major established network connections of European exchanges by 2002, classified according to the criteria discussed above.

Transaction costs data for each European exchange market come from Elkins/McSherry (E/M) Universe. This is a rolling four quarter compilation of data comprising current and historical information on 700 global managers and 800 global brokers, containing average commissions, fees, market impact and stock price information from 208 exchanges in 42 countries. Although an assessment of the quality of trading is beyond the scope of this trade execution data, it does enable a comparison of commissions, fees, and market impact to a universe of costs in different countries.

The E/M system calculates the cost of trade execution on the basis of the volume weighted average price and the spreads of the stocks.[10] The E/M data contains all items of each trade including the high, low, open and close, volume traded, volume weighted average price and average spread. The market impact, being considered as a major cost component of the transaction cost, is calculated by E/M as the difference between the trade execution price and the average price (high, low, open, and close) for every stock in 42 countries daily. Commissions, fees and market impact costs are compared to the average institutional costs

[10] Consult http://www.elkins-mcsherry.com/edata.html for an example of volume weighted average price and spread calculations.

in each country and then broken down by portfolio manager, account, client and broker. Finally, the summary costs for each institution enter into the E/M Universe of average costs. The total trading cost is measured in basis points representing the average sum of commission, fees and market impact based on trade data on all global trades executed by large institutional investors in a given market. Finally, macroeconomic information such as GDP per capita, and concentration of private sectors, is taken from the IFS data bank.

As mentioned earlier, the estimation model in this chapter investigates the potential relationship between the NETWORK variable(s) and exchange PERFORMANCE measures as portrayed by Equations 1 and 2. As evident, we employ a series of ordinary least squared regressions to capture these potential relations. First, we investigate the relationship with a number of simple single variable regressions (1), later followed by multivariate estimations (2) incorporating other control variables that are pertinent to the exchange performance measures. Market capitalisation (MKTCAP), the growth of market capitalisation (GMKTCAP), and the transaction costs of trading (TCOSTR) are used as proxies for the dependent variables. Hasan and Schmiedel (2004) tested the performance of stock exchanges using a different set of indicators.

$$PERFORMANCE_{it} = \alpha_0 + NETWORK_{it} + \varepsilon_{it} \tag{1}$$

$$PERFORMANCE_{it} = \alpha_0 + NETWORK_{it} + \sum CONTROL\ VARIABLES_{it} + \varepsilon_{it} \tag{2}$$

The NETWORK variable is represented by alternative variables. The first three estimates are based on the variables that trace (1) ACCESS; (2) NDN; and (3) NNM respectively. The next three regressions follow the definition of NETWORK portrayed in (3), i.e., NNM, except in each case, it considers the total number of other exchanges linked with an individual exchange via an (4) Equity or Blue Chip Network (ENNM); or (5) Tech. or Growth Network (TNNM); or (6) Derivative Network links (DNNM). Although our interest is primarily focused on the first four estimates, it is interesting and informative, when detailed information is available, to investigate the relative importance of specific types of network or impact in connecting with other exchanges.

Control variables considered are:

- GDP per capita in the country where the exchange is located (local economic environment);
- total Private Sector Accumulation to GDP ratio (relative importance of private sector);

- Disclosure Index (accounting or disclosure standards in CIFAR);
- Concentration of Ownership by the top three firms in the Exchange (extent of influence of larger firms in the exchange);
- Transaction Cost (cost of trading);
- Number of Exchanges within the Domestic Borders (competition in the exchange business); and
- Market Capitalisation (market size).

These control variables are selectively added to each regression, given what is considered as exchange performance (dependent variable) in a particular estimation, and are consistent with the relevant literature.[11]

4. EMPIRICAL EVIDENCE

Table 15.1 provides the names and a number of key statistics for each of the sample exchanges. These statistics include average market capitalization, growth of market, the transaction cost of trading and the extent of their involvement in exchange networks during the sample years. It reveals that the exchanges are of different sizes of market capitalization. An interesting observation in this respect is that the transaction costs and network involvement are not necessarily always proportional to market size. Many smaller exchanges report lower transaction costs and higher involvement in network co-operation.

The Riga exchange of Latvia has the smallest market with a market capitalisation of USD 289 million, while the London stock exchange represents the largest market of USD 2,474,579 million in a given sample year respectively. On average, the markets are growing at a rate of almost 29%. The transaction costs range from as low as 23.80 (Paris exchange) to as high as 161.01 (Czech Republic) in a given sample year. The maximum number of network links available to exchanges in Europe is 4 and the total number of stock exchanges linked

[11] GDP Per Capita and Private Sector Accumulation to GDP ratio are taken from International Financial Statistics and are adjusted for inflation and converted into US dollars. Concentration of Ownership, and Number of Exchanges are taken from the MSCI Handbook; Disclosure Index has been taken from La Porta *et al.* (1997) and cross-checked with the CIFAR Index; Transaction Cost, which is used as a dependent variable in some estimates and as an independent variable in others, is from Elkins-McSherry. As mentioned in the text, the Network variable is constructed by tracing the developments of stock exchanges over the sample period from different public information sources and on some occasions by writing to the exchanges directly. Additionally, the dependent variables, Market Capitalization and Growth of Market Capitalization are taken from the FIBV.

Table 15.1 Correlation Coefficient Matrix

Variables/Ratios	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Market Capitalisation	1	0.40**	−0.25	0.47**	0.63**	0.53**	0.30*	0.84**	0.40**	0.36**	0.59**
2. Market Capitalisation Growth		1	−0.43**	0.46**	0.37*	0.30*	0.04	0.34*	−0.02	−0.21	0.33*
3. Transaction Cost of Trading			1	−0.57**	−0.40**	−0.43**	−0.31	−0.05	−0.21	0.15	−0.25*
4. Access to Network (Yes = 1 No = 0)				1	0.71**	0.63**	0.50**	0.38**	0.13	−0.11	0.41**
5. Extent of Network Involvement					1	0.92**	0.45**	0.51**	0.09	−0.18	0.44**
6. Total Number of Stock Exchanges Linked through Network						1	0.38**	0.24	−0.05	−0.09	0.42**
7. GDP Per Capita (000)							1	0.21	0.50**	0.12	0.09
8. Concentration of Private Sector to GDP								1	0.23	−0.37*	0.51**
9. Accounting Standard									1	0.09	0.05
10. 3-Firm Concentration on the Exchange										1	−0.30*
11. Number of Exchanges in the Country											1

Note: ** and * portray significance at the 1% and 5% levels respectively.

through networks as high as 19 exchanges. These sample exchanges are from countries with a wide range of GDP per capita, private sector involvements and accounting standards.

These initial sets of single equation estimates are reported in Tables 15.2a to Table 15.2c. In each table, we provide results of the possible impact of all alternative NETWORK variables (or components of it) on one of the exchange PERFORMANCE measures. To illustrate, Table 15.2a reports the potential relationship between the logarithms of market capitalization with five different independent variables in five separate estimates. The evidence portrayed here reveals overwhelmingly a positive and significant association between NETWORK variable(s) and market capitalization. Interestingly, we observe that in each of the reported regressions, the model statistics, i.e. adjusted R-squared and F-Statistics, are high and significant. For example, the first regression of Table 15.2a shows that over 35% of the market capitalization variability of the sample is captured by a simple bivariate independent variable.

In Table 15.2b market growth is considered as the dependent variable, calculated by taking the annual growth of market capitalisation of the respective exchanges.[12] The evidence shows a strong association between NETWORK variable(s) and market growth.

We then focus our attention on the possible relationship between the NETWORK variable and the TCOSTR (transaction cost of trade) in respective markets. Exchanges with higher network linkages are expected to be associated with lower trading costs. Evidence in Table 15.2c reports high model statistics, and importantly for our purposes, most NETWORK variables are found to be negatively and significantly associated with TCOSTR. In line with the results of the regression results portrayed in the previous tables, the model statistics of these regressions in Table 15.2c are highly significant.

We follow up estimations in Table 15.2a–c with another set of estimations, as portrayed in Table 15.3a and 15.3b, with the exception that we proceed with reporting only the first 4 estimates (rather than the 7 represented in 4s) as these represent the most general proxy for NETWORK variables. In these regressions, we also control for additional variables that may be pertinent in explaining all the dependent variables used in our regressions. These variables were selected based on similar use of these variables in the literature in different research contexts. Most of the independent variables used in Table 15.3a and 15.3b are similar across regressions, except for an additional size variable (market capitalisation)

---

[12] We include 1995 market capitalization data for the sample exchanges in order to calculate the growth variable.

Table 15.2a Ordinary Least Square Estimates – Network on Market Performance

Independent Variable/Ratios Model	Dependent Variable Market Capitalisation					
	1 Parameter (t-statistics)	2 Parameter (t-statistics)	3 Parameter (t-statistics)	4 Parameter (t-statistics)	5 Parameter (t-statistics)	6 Parameter (t-statistics)
Intercept	6.784 (30.16)**	7.314 (44.03)**	11.43 (65.01)**	17.965 (70.52)**	17.819 (78.53)**	17.550 (76.93)**
Access to Network (Yes = 1 – No = 0)	2.772 (7.89)**	–	–	–	–	–
Extent of Network	–	1.618	–	–	–	–
Involvement	–	(6.84)**				
Total Number of Exchanges	–	–	0.304	–	–	–
Linked through Network	–	–	(5.98)**	–	–	–
Total Number of Stock Exchanges Linked through the Equity Network	–	–	–	0.413 (2.62)**	–	–
Total Number of Stock Exchanges Linked through a networked market for Growth or Tech-oriented Companies	–	–	–	–	0.566 (3.49)**	–
Total Number of Stock Exchanges Linked through Derivative Network	–	–	–	–	–	0.401 (4.81)**
Adjusted R-Squared	0.352	0.302	0.239	0.105	0.118	0.144
F-Statistics	52.60**	39.03**	28.68**	9.80**	12.19**	23.17**
Number of Observations	120	120	120	120	120	120

Note: ** and * portray significance at the 1% and 5% levels respectively.

Table 15.2b Ordinary Least Square Estimate – Network on Market Growth

Independent Variable/Ratios Model	Dependent Variable Market Capitalization Growth					
	1 Parameter (t-statistics)	2 Parameter (t-statistics)	3 Parameter (t-statistics)	4 Parameter (t-statistics)	5 Parameter (t-statistics)	6 Parameter (t-statistics)
Intercept	13.600	16.480	9.892	11.450	8.265	9.704
	(5.65)**	(4.98)**	(4.59)**	(3.91)**	(3.53)**	(4.46)**
Access to Network	1.650	–	–	–	–	–
(Yes = 1 – No = 0)	(3.59)**	–	–	–	–	–
Extent of Network	–	3.582	–	–	–	–
Involvement	–	(2.83)*	–	–	–	–
Total Number of Exchanges	–	–	1.364	–	–	–
Linked through Network	–	–	(2.17)*	–	–	–
Total Number of Stock	–	–	–	0.127	–	–
Exchanges Linked through the Equity Network	–	–	–	(1.68)	–	–
Total Number of Stock	–	–	–	–	0.309	–
Exchanges Linked through a networked market for Growth or Tech-oriented Companies	–	–	–	–	(1.75)	–
Total Number of Stock	–	–	–	–	–	1.082
Exchanges Linked through Derivative Network	–	–	–	–	–	(2.19)*
Adjusted R-Squared	0.202	0.122	0.072	0.054	0.048	0.031
F-Statistics	12.92**	7.54**	4.70*	3.65	3.83	3.79
Number of Obs.	120	120	120	120	120	120

Note: ** and * portray significance at the 1% and 5% levels respectively.

Table 15.2c Ordinary Least Square Estimates – Network on Transaction Cost of Trading

Dependent Variable Independent Variable/Ratios Model	Total Transaction of Trading					
	1 Parameter (t-statistics)	2 Parameter (t-statistics)	3 Parameter (t-statistics)	4 Parameter (t-statistics)	5 Parameter (t-statistics)	6 Parameter (t-statistics)
Intercept	4.628 (9.04)**	3.463 (10.18)**	4.514 (11.31)**	6.245 (8.16)**	5.104 (5.65)**	3.102 (4.43)**
Access to Network (Yes = 1– No = 0)	−0.063 (1.98)*	– –	– –	– –	– –	– –
Extent of Network Involvement	– –	−0.026 (1.95)	– –	– –	– –	– –
Total Number of Exchanges Linked through Network	– –	– –	−0.038 (1.93)	– –	– –	– –
Total Number of Stock Exchanges Linked through the Equity Network	– –	– –	– –	−0.091 (2.00)*	– –	– –
Total Number of Stock Exchanges Linked through a networked market for Growth or Tech-oriented Companies	– –	– –	– –	– –	−0.164 (1.80)	– –
Total Number of Stock Exchanges Linked through Derivative Network	– –	– –	– –	– –	– –	−0.346 (2.02)*
Adjusted R-Squared	0.2643	0.2518	0.2561	0.1813	0.1455	0.1539
F-Statistics	4.08**	3.93**	4.11**	3.85**	4.16**	4.29**
Number of Obs.	109	109	109	109	109	109

Note: ** and * portray significance at the 1% and 5% levels respectively.

Table 15.3a Impact of Network Access on Market Capitalisation and Market Growth

Dependent Variables	Market Capitalization			Market Capitalization Growth		
Independent Variables/Ratios Model	Parameter (t-statistics) 1	Parameter (t-statistics) 2	Parameter (t-statistics) 3	Parameter (t-statistics) 1	Parameter (t-statistics) 2	Parameter (t-statistics) 3
Intercepts	14.637	15.500	14.810	3.085	4.094	3.478
	(7.08)**	(8.15)**	(7.76)**	(2.75)*	(3.09)**	(2.94)**
Access to Network	0.227	–	–	0.206	–	–
(Yes = 1 – No = 0)	(1.98)*	–	–	(2.94)**	–	–
Extent of Network	–	0.631	–	–	1.450	–
Involvement	–	(2.76)*	–	–	(1.90)	–
Total Number of Exchanges	–	–	0.181	–	–	0.905
Linked through Network	–	–	(1.96)*	–	–	(1.77)
GDP Per Capita	0.023	0.033	0.031	0.045	0.032	1.908
(thousands of USD)	(1.52)	(1.94)	(1.79)	(1.89)	(1.44)	(1.07)
Concentration of Private	1.603	1.508	1.499	0.894	0.832	1.685
Sector to GDP	(4.87)**	(4.90)**	(4.77)**	(5.09)**	(1.98)*	(2.32)*
Accounting Standard	0.065	0.060	0.063	0.058	1.47	0.953
	(5.05)**	(4.71)**	(5.00)**	(4.29)**	(3.03)**	(1.76)
3-Firm Concentration on	–0.026	–0.014	–0.016	–0.029	–0.054	–0.095
the Exchange	(2.51)*	(1.84)*	(2.03)*	(2.44)*	(2.94)**	(2.46)*
Transaction Cost	–0.030	–0.124	–0.002	–0.027	–0.075	–0.745
of Trading	(1.47)	(1.30)	(0.94)	(1.56)	(1.33)	(1.21)
Number of Exchanges	–0.700	–0.061	–0.084	–0.574	–0.316	–0.286
in the Country	(1.44)	(1.19)	(1.49)	(1.50)	–(0.73)	(1.31)
Adjusted R-Squared	0.739	0.715	0.723	0.242	0.208	0.217
F-Statistics	14.06**	16.88**	16.33**	5.98*	4.47*	3.86*
Number of Observations	106	106	106	106	106	106

Note: ** and * portray significance at the 1% and 5% levels respectively.

Table 15.3b Impact of Network Access on Network on Market Efficiency and Transaction Cost

Dependent Variable Independent Variables/Ratios Model	Transaction Cost of Trading		
	Parameters (t-statistics) 1	Parameters (t-statistics) 2	Parameters (t-statistics) 3
Intercepts	4.296	4.284	4.232
	(14.54)**	(13.84)**	(14.03)**
Access to Network	−0.126	–	–
(Yes = 1 – No = 0)	(1.88)	–	–
Extent of Network	–	−0.032	–
Involvement	–	(1.80)	–
Total Number of Exchanges	–	–	−0.042
Linked through Network	–	–	(1.50)
GDP Per Capita	−0.012	−0.023	−0.019
(thousands of USD)	(2.40)*	(3.09)**	(2.82)*
Concentration of	−0.252	−0.236	−0.242
Sector to GDP	(1.93)	(1.90)	(1.90)
Accounting	0.001	0.001	0.002
Standard	(0.18)	(0.34)	(0.47)
3-Firm Concentration	0.005	0.004	0.006
on the Exchange	(1.54)	(1.46)	(2.05)*
Transaction Cost of Trading	–	–	–
Market Capitalisation	−1.065	−1.120	−0.901
	(2.02)*	(2.23)*	(1.96)*
Number of Exchanges	−0.042	−0.051	−0.035
in the Country	(1.80)	(2.09)*	(1.87)
Adjusted R-Squared	0.331	0.332	0.321
F-Statistics	3.79**	4.05**	4.24**
Number of Observations	106	106	106

Note: **and *portray significance at the 1% and 5% levels respectively.

used in the cost regressions in Table 15.3b. Once again, these independent variables were controlling for the macroeconomic environment, incorporating:

- GDP per capita;
- the relative importance of the private sector in the economy, considering the total private sector accumulation to GDP ratio;
- accounting or disclosure standards of the economy where an exchange is located;
- relative concentration of the top three firms in the exchange;
- the cost of trading (as relevant for specific dependent variables); and

- the number of exchanges within the domestic borders, a proxy for market competition.

In summary, even after adding all other independent variables in our estimations, we find our key focus variables represented by NETWORK (ACCESS, NDN, and NNM) are still significantly associated with dependent variables in most estimations. In some multivariate regression models, the network variables show a least significance at the 10% level. Indeed, their relative significance – or t-statistics – were not as strong as the ones reported in Table 15.2a–c, where no control variables were added to the NETWORK variables. Nonetheless, they are relevant and significant in explaining the variability of dependent variables. Moreover, the marginal increase in model statistics due to the addition of several new independent variables reveals that the R-squared represented by NETWORK variable(s) takes the lead in explaining the variability of exchange performance.

The empirical results of this study bear important implications for the future development and integration of European equity markets. European stock and derivative exchanges reorganize their business and their operations by forming alliances, takeovers, or other forms of co-operation in order to maintain market shares and leverage themselves into a better position vis-à-vis their competitors. In this light, such co-operations among European securities markets are mainly motivated to become more efficient and productive if trading were centralized, not necessarily on a few or eventually on only one physical base. It may be simply a technological agreement between exchanges to use standardized technologies ensuring high compatibility in different or even one centralized trading system, so as to maximize scale economies and network externalities.

As it can be seen from Figure 15.1, advances of technology allows exchanges companies to overcome differences in location, thereby transforming their competitive framework and business targets of their trading services from a traditional domestic into a cross-border context. As a potential market outcome, the way toward further European equity integration may take place at three parallel levels. The first level refers to horizontal integration. Here, stock exchanges create horizontal alliances in order to expand services to other products or equity markets, for example, Euronext Stock Exchange. As a result, European securities markets are likely to benefit from a greater extent of cross-border integration and co-operation between trading service providers. The second level is vertical integration. This process foresees a closer integration of the institutions in charge of securities trading, clearing, and settlement and custody services. Accordingly, vertical mergers facilitate exploiting synergies along the transaction value chain. Prominent examples for so-called silo systems are the Deutsche Boerse and Clearstream International, and Helsinki Stock Exchange and APK. Finally, the third level refers to lateral consolidation. In view of this concept, stock exchanges laterally promote their IT services by selling new trading systems and support to other trading service providers. This could enhance increased standardization, greater compatibility, and better connectivity between systems.

All these integration processes between institutions in charge of trading services and products have opened new strategic scenarios in which economies of scale and expectations of efficiency gains in the provision of trading services lead to improved integration of European equity markets.[13] The final outcome of the consolidation process shaping the European trading landscape is still to be seen. However, it is likely that interconnected clusters or groups of systems with a different degree of integration will emerge at an intermediate stage.

5. CONCLUSIONS

The topic of networks has been popular in the academic literature. Whether it is theoretical or descriptive in nature, no empirical attempt has been made to understand and investigate the actual structure of the network and its impact on market performance and market integration. The increasing involvement of stock exchanges in different trading network modules, especially in Europe, warrants further investigation as to whether the adoption of network strategies adds additional value in the provision of trading services. This paper investigates the network externalities among stock exchanges by constructing and quantifying the network strategy and the extent of networks adopted by the European stock exchanges in recent years. This is one of the very first empirical initiatives to explore whether network linkages or common trading platforms among exchanges matter in affecting individual exchange performance. Tracing the experiences of all major European exchanges over the 1996–2000 period, this chapter examines the impact of the network effect on market liquidity, growth, and transaction costs of trading.

All alternative NETWORK variables constructed reveal a strong and significant association with exchange performance. In summary, the empirical evidence clearly reveals that the adoption of a network strategy by stock exchanges is significantly associated with performance measures. As the stock exchanges around the globe are increasingly moving toward a more network-linked market set-up, stock exchanges seem to take action to enhance their overall performance. Building on the analytical framework of this study, the adoption of network strategies

[13] This view is supported by recent theoretical academic literature, e.g. Shy and Tarkka (2001), where alliances among stock exchanges are very likely to improve total welfare as well as to increase profits for stock exchanges.

pursued by stock exchanges seem to play a crucial role for future market design and integration of European equities market infrastructure. Therefore, further empirical attempts are warranted on the impact of network economics on the exchange industry and financial markets.

ACKNOWLEDGEMENT

Part of this research was carried out when Heiko Schmiedel was visiting the Bank of Finland Research Department in Helsinki. The views expressed herein do not necessarily reflect the official views of the Bank of Finland and the European Central Bank.

BIBLIOGRAPHY

Annual Reports (1996–2000) Annual Reports of 24 Stock and Derivative Exchanges.

Arnold, T., Hersch, P., Mulherin, J.H. and J. Netter (1999) "Merging Markets." *Journal of Finance* **52**: 655–81.

Blass, A. and Y. Yafeh (2001) "Vagabond Shoes Longing to Stray: Why Foreign Firms List in the United States." *Journal of Banking and Finance* **25**: 555–72.

Chaplinsky, S. and L. Ramchand (2000) "The Impact of Global Equity Offerings." *Journal of Finance* **55(6)**: 2767–89.

Center for International Financial Analysis and Research (CIFAR) (1995) *Level of Accounting Disclosure, 1995*.

Cybo-Ottone, A., Di Noia, C. and M. Murgia (2000) *Recent Development in the Structure of Securities Markets*, Brookings-Wharton Papers on Financial Services.

Di Noia, C. (2001) "The Stock-Exchange Industry: Network Effects, Implicit Mergers, and Remote Access." *European Financial Management* **7(1)**: 39–72.

Domowitz, I. (1995) "Electronic Derivatives Exchanges: Implicit Mergers, Network Externalities and Standardization." *The Quarterly Review of Economics and Finance* **35(2)**: 163–75.

Domowitz, I. and B. Steil (1999) *Automation, Trading Costs, and the Structure of the Trading Services Industry*, Brookings-Wharton Papers on Financial Services, 1–52.

Domowitz, I., Glen, J. and A. Madhavan (1998) "International Cross-Listing and Order Flow Migration: Evidence from an Emerging Market." *Journal of Finance* **53**: 2001–27.

Economides, N. (1993) "Network Economics with Application to Finance, Financial Markets." *Institutions & Instruments* **2(5)**:

Economides, N. (1995) "How to Enhance Market Liquidity," in Schwartz, R. (ed.), *Global Equity Markets*. New York: Irwin Professional.

Economides, N. (1996) "The Economics of Networks." *International Journal of Industrial Organization* **16(4)**: 675–99.

Elkins/McSherry (1995–2001) Global Trading Cost Analysis, New York.

Foerster, S. and A. Karolyi (1998) "Multimarket Trading and Liquidity: A Transaction Data Analysis of Canada-U.S. Interlistings." *Journal of International Financial Markets, Institutions and Money* **8**: 393–412.

Hagel III-Armstrong, A.G. (1997) *Net Gain*. Boston, Massachusetts: Harvard Business School Press.

Hasan, I. and M. Malkamäki (2001) "Are Expansions Cost Effective for Stock Exchanges? A Global Perspective." *Journal of Banking and Finance* **25(12)**: 2339–66.

Hasan, I., Malkamäki, M. and H. Schmiedel (2003) "Technology, Automation, and Productivity of Stock Exchanges: International Evidence." *Journal of Banking and Finance* **27**: 1743–73.

Hasan, I. and H. Schmiedel (2004) "Networks and Equity Market Integrations." *International Review of Financial Analysis* **13**: 601–19.

International Federation of Stock Exchanges (FIBV) (1996–2000) *Annual Reports 1996–2000*, Paris.

International Monetary Fund, various issues. *International Financial Statistics*. Washington.

Karolyi, A. (1998) "Why do Companies List Shares Abroad? A Survey of the Evidence and Its Managerial Implications." *Financial Markets, Institutions & Instruments* **7(1)**: 1–60.

La Porta, R., Lopez-de-Silanes, F., Shleifer A. and R.W. Vishny (1997) "Legal Determinants of External Finance." *Journal of Finance* **52**: 1131–50.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and R.W. Vishny (1998) "Law and Finance." *Journal of Political Economy* **106**: 1113–55.

Lee, R. (1998) *What is an Exchange: The Automation, Management and Regulation of Financial Markets*. Oxford University Press.

Licht, A. N. (1998) *Regional Stock Market Integration in Europe, Harvard Institute for International Development, Consulting Assistance on Economic Reform II* Discussion Papers, No. 15.

MSCI (1995–2000) *Handbook of World Stock, Derivative, and Commodity Exchanges*. UK: Mondo Visione.

Pagano, M., Randl, O., Röell, A. and J. Zechner (2001) "What makes Stock Exchanges Succeed? Evidence from Cross-Listing Decisions." *European Economic Review* **45**: 770–82.

Pagano, M., Röell, A. and J. Zechner (2003) *The Geography of Equity Listing: Why do Companies List Abroad? Centre for Studies in Economics and Finance*. Working Paper No. 28, *Journal of Finance* (forthcoming).

Pirrong, C. (1999) "The Organization of Financial Exchange Markets: Theory and Evidence." *Journal of Financial Markets* **2**: 329–57.

Saloner G. and A. – Shepard (1995) "Adoption of Technologies with Network Effects: An Empirical Examination of Adoption of Automated Teller Machines." *Rand Journal of Economics* **26**: 479–501.

Schmiedel, H. (2001) *Technological Development and Concentration of Stock Exchanges in Europe*. Bank of Finland Discussion Paper Series, 21/2001.

Shapiro, C. and H. R. Varian (1999) *Information Rules. A Strategic Guide to the Network Economy*. Boston, Massachusetts: Harvard Business School Press.

Shy, O. (2001) *The Economics of Network Industries*. Cambridge University Press.

Shy, O. and J. Tarkka (2001) *Stock Exchange Alliances, Access Fees and Competition*. Bank of Finland Discussion Paper Series, 22/2001.

Steil, B. (2001) "Creating Securities Markets in Developing Countries: A New Approach for the Age of Automated Trading" **4(2)**: 257–78.

This Page is Intentionally Left Blank