tinbergen institute

Essays on Systemic Risk

An analysis from multiple perspectives

Sander Muns

Erasmus University Rotterdam

Essays on Systemic Risk

An analysis from multiple perspectives

ISBN: 978 90 5170 767 0

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul
picture: FlippyCat (https://www.youtube.com/watch?v=7BVr6LaC_HQ&t=4m49s)

This book is no. 644 of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Essays on Systemic Risk:
An analysis from multiple perspectives

Essays over systeemrisico:
een analyse vanuit meerdere perspectieven

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board

The public defence shall be held on
28 January 2016 at 11:30

by

Sander Muns
born in Leidschendam, The Netherlands

Erasmus University Rotterdam

Doctoral Committee

Promoter:	Prof.dr.	C.G. de Vries
Co-promoter:	Dr.	C. Zhou
Other members:	Prof.dr.	P. Hartmann
	Prof.dr.ir.	G. van der Laan
	Prof.dr.	W. Wagner

Acknowledgements

Although writing a dissertation is a proof of being an independent researcher, one cannot finish a dissertation in complete isolation. I would like to take the opportunity to thank those who have helped me during my dissertation.

Firstly, I am grateful to my promoter Casper de Vries and my co-employer CPB Netherlands Bureau for Economic Policy Analysis for offering me the opportunity to work on a PhD thesis. It was a great experience to write a thesis without any binding restriction except for the broad subject systemic risk. Studying a wide range of aspects of systemic risk enabled me to exploit my interest in both economics and mathematics.

Working on this thesis has been an excellent opportunity to work with Casper de Vries and my co-promoter Chen Zhou. It was a privilege to trigger their enthusiasm, not to mention to share a common interest in the intersection of economics and mathematics. The quick thinking of Chen is unprecedented. Particularly Chapter 4 benefited from Chen's supervision. The comments and suggestions of Michiel Bijlsma (Chapter 3), Gerard van der Laan (Chapter 5) and Ludolf Meester (Chapter 6) are greatly acknowledged. As the main author of all chapters, the usual disclaimer applies that all remaining errors are mine.

My colleagues at CPB and the Erasmus School of Economics (ESE) created an excellent research environment. Working on this thesis would not be as pleasant without my colleagues Arie, Gijsbert, Jan, Jasper, Marcel, Nick, Peter, Riemer, Rob, Thomas (CPB), Jonathan, Kyle, Ona, Pengfei, Suzanne, Victor, Wei, Xudong (ESE), and many others at both institutions.

I am also indebted to the support of the administrative staff of the Department of Economics, Tinbergen Institute, and CPB. Last but not least, I would like to thank my friends and family. Their support enabled me to focus on my thesis during the busiest times.

Sander Muns
December 2015

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> One of the most feared events in banking is the cry of systemic risk. It matches the fear of a cry of "fire!" in a crowded theater or other gatherings. But unlike fire, the term systemic risk is not clearly defined. Moreover, unlike firefighters, who rarely are accused of sparking or spreading rather than extinguishing fires, bank regulators at times have been accused of contributing to, albeit unintentionally, rather than retarding systemic risk.
>
> Kaufman and Scott (2003)

Since the bankruptcy of Lehman Brothers in September 2008, policy makers and supervisors are increasingly aware of the economy-wide danger of a systemic financial crisis. This concern for systemic risk has led to several reforms in supervisory tools. A tangible result is the adoption of the Single Supervisory Mechanism in November 2014. This mechanism makes the European Central Bank the chief supervisor of banks in the eurozone, a major milestone in the creation of a banking union.

Another recent event with a prominent role for systemic risk is the third bailout package for Greece in August 2015. The reforms linked to this package are very similar to the rejected proposals in a Greek referendum held only one week before the agreement. Nonetheless, Greece's policy makers accepted the conditions attached to the rescue package, thereby neglecting the outcome of the referendum, and violating their promises at the elections less than six months before. However, a sovereign default, or even worse, Greece leaving the eurozone (a Grexit), could very well trigger a series of cascading events, particularly in the economy of Greece. More specifically, a sudden reintroduction of a national currency would be expected to bring about a new capital outflow, an uncontrolled depreciation of the new currency, bank runs, trials on contractual obligations, and other systemic misery.

Ultimately, it may even lead to a breakup of the eurozone.

Nevertheless, cross-border systemic risk related to a hypothetical bankruptcy of Greece has decreased over time, at least in the perception of the financiers in the Trojka (EC, ECB and the IMF). In late 2009, the Trojka preferred to avoid a Grexit at any cost with the systemic Lehman event fresh in mind and markets behaving very nervously. Since then foreign direct investments in Greece have plunged, markets have priced a potential breakup of the euro, and Greek banks and the Trojka are the main investors in sovereign debt of Greece. This makes the cross-border effect of a Greek default better predictable. Indeed, financial markets outside Greece reacted moderately in response to the recent uncertainty around Greece. For instance, Portuguese ten year bonds temporarily increased to 3.2%, a muted response compared to the spike at 17.3% in January 2012. Accordingly, policy institutions are confident that spillover effects are now in control such that the systemic impact of a Greek default and a Grexit would be mainly local. This explains why the Greeks found themselves with their backs against the wall during the negotiations.

Despite its importance, the concept of systemic risk lacks a uniform definition as noted in Kaufman and Scott (2003). Generally speaking, the concept of systemic risk is on the risk of the collapse of an entire system. To assess this risk, an in-depth analysis of several characteristics of the system is essential. Obviously, multiple fragile components with strong linkages correspond to a very fragile system. It is however empirically less clearcut to find out beforehand whether linkages between the components are strong *conditional* on a nontrivial shock. In addition, what is the impact on systemic risk when we add components to the system with strong linkages to the original components of the system? For instance, would a euro entry of countries as Poland or, less likely, Switzerland stabilize the euro economy as a whole? Is the mutual trade ban between Russia and lots of European countries helpful in stimulating domestic production and thereby stabilizing the economies? Addressing such issues involves an assessment of contagious effects from possibly highly complex linkages within the system (see e.g. Haldane and May (2011)). Indeed, the concept of systemic risk is not limited to the financial system. It extends to an assessment of other systemic problems including the social-economic impact of a cyber attack, a power breakdown or an outbreak of a virus disease such as Ebola.

Systemic risk differs from the related concept of *systematic* risk. In a finance context, systematic risk refers to the risk that cannot be eliminated by diversification within an equity portfolio. For instance, stocks tend to have a common exposure to natural disasters, terrorist attacks, and other catastrophes. Unless one implements a hedging strategy, investing in multiple stocks cannot completely eliminate the exposure to systematic risks. Thus, systematic risk captures the common exposure of the individual components to some external shock. As such, systematic risk can trigger a systemic event, but it is no more than one possible cause of such an event.

This thesis focuses on systemic risk in the financial industry. In this context, Taylor (2010) defines systemic risk as

> Any definition of systemic risk must be based on three considerations. The first is the risk of a large triggering event. The second is the risk of financial propagation of such an event through the financial sector by contagion or chain reaction. The third is the macroeconomic risk that the financial disruption will severely affect the whole economy.

The first two considerations are discussed in this thesis from multiple perspectives. In future research, these perspectives may contribute to an analysis of the third consideration, the link to the whole economy.

To establish systemic risk in the financial industry, we *measure* systemic risk in Europe (Chapter 2) and the U.S. (Chapter 3) with two different methods. Having measured a substantial level of systemic risk in the financial industry, we consider the *effects* of systemic risk on stock pricing in Chapter 2. We identify *determinants* of systemic risk, empirically (Chapter 3) as well as theoretically (Chapter 4). Chapter 5 considers a fair *allocation* of systemic risk to the contributing institutions. Finally, we present in Chapter 6 a more accurate method to *compute* and *simulate* systemic risk using the multivariate normal distribution.

The chapters are ordered by the link to real life economics. The first two chapters have an empirical focus, while the other three Chapters are increasingly theoretical. Though the chapters share their relation with systemic risk, each chapter can be read and understood independently. Next, we discuss the contents of each chapter in more detail.

Chapter 2 provides evidence that big financials in Europe have significantly lower risk-adjusted returns than other European financials. This pattern is absent in any other industry. We interpret this as evidence for latent government guarantees. This safety net protects big financials from tail events, and is quite strong. Even during the debt crisis, higher probabilities of distress dominated doubts on government guarantees. Another finding is that our financial risk factor improves upon the momentum factor in explaining financial stock returns, and the anomalies on big and value stocks in the European Fama and French (2012)-model.

In Chapter 3, we use tail dependence as a measure for systemic risk in financial markets at the industry level. In particular, we measure tail dependence in the 48 U.S. Fama and French (1997) industries by comparing comovements in extreme negative returns within these industries. The industries that score highest on tail dependence are banking, petroleum & natural gas, utilities, financial trading and insurance. All other industries score significantly lower. We identify a large market beta, a large market cap and a small volatility as significant determinants of tail dependence. In addition, tail dependence is a significant predictor for tail dependence in the next year.

Chapter 4 analyzes in a theoretical model how social welfare is jointly affected by bank size, banks' capital structure, and asset dependence across banks. The model suggests that banks always prefer a capital ratio below the socially optimal level, while banks prefer an extra large size when cross-sectional dependencies are typically low. More stringent capital requirements simultaneously result in larger banks under low bankruptcy costs, high financing costs and high taxes. To enhance social welfare, policies on capital are nevertheless more effective than policies on size. This is particularly true when dependencies are low, i.e., during economic booms. Our results are in support of the countercyclical capital buffers in the Basel III proposals and imply a negative association between the stringency of monetary and prudential policy.

Chapter 5 identifies a simple sufficient condition to fairly allocate systemic risk according to the Shapley value concept. Our condition implies that two allocation measures in Segoviano and Goodhart (2009) and Zhou (2010) are a Shapley value. The widely used MES in Acharya et al. (2012), and $\Delta CoVaR$ in Adrian and Brunnermeier (2011) are however not a Shapley value.

The tails of the multivariate normal distribution play a pivotal role in simulating systemic risk, either directly or indirectly. For instance, the t-distribution is an infinite mixture of normal distributions. Even for the multivariate normal distribution, existing methods lack an accurate modeling of the multivariate tails. In Chapter 6, we propose a new analytical method to bound the probability of the multivariate normal distribution. Our method combines a Taylor expansion with a change of measure. The bounds are (i) tighter than bounds of existing methods, (ii) in a substantial proportion of simulated low-dimensional cases more precise than the simulation method of Genz (1992), and (*iii*) tightest if the condition number of the covariance matrix is close to one. Since the relative error of the bounds declines in the tails, our method is particularly useful in the multivariate tail region. This is of particularly interest for systemic risk as systemic events are associated with events in this tail region.

Chapter 2

The size anomaly in European financial stock returns

2.1 Introduction

Recent research shows that tail risk is important in asset pricing (e.g., Gabaix (2012), and Wachter (2013)). The financial crisis has shown that the impact of this tail risk is potentially huge for society. The bailouts during the crisis reduced the potential tail risk of protected financials and led to a corresponding reduction in spillover effects. Such bailouts have led investors to suppose that governments bail out the systemically most important financials during periods of serious distress.

Moral hazard may induce the protected financials to opt for more risky investments. On the one hand, the standard risk-return tradeoff in the CAPM predicts a higher expected return on the assets of such financials. On the other hand, investors of the protected financials are forward-looking. They price ex-ante the higher return on assets, thereby indicating that the return on debt and equity of a protected financial is not necessarily higher. Even stronger, investors may anticipate the implicit embedded put option of the protection. Similar to an insurance premium, investors may demand a *lower* return on an investment in a protected financial. When sufficiently large, such government guarantees increase sovereign risk, which leads investors to demand a higher return on sovereign bonds. As a consequence, the lower financing cost of protected financials is an implicit government subsidy at the expense of tax payers. It is therefore important to identify ex-ante the protected financials.

Though not always explicitly formalized, the "too-big-to-fail" concern is

I thank Nico van der Sar and Casper de Vries for comments on an earlier version.

generally accepted, and suggests that size is an indicator of systemic importance (see e.g. Huang et al. (2012)), and thus a potential indicator of government protection. The size effect may act in several ways. A default of a big financial results in a default on a large amount of debt, which has a large direct impact on the financial system and a large indirect impact through spillover effects and banking panics (see Diamond and Dybvig (1983)). Relatedly, big financials are more involved in interbank activities (see Furfine (2003)). As such, the default of a big financial may induce a cascading effect of defaults in the banking system. In addition, the assets of big financials tend to be more diversified. This makes the exposure of big financials more similar (see Ibragimov et al. (2011)). As a consequence, a default of a big financial is more likely to be associated with multiple distressed financials.

In any case, regulators aim to prevent a default of a big financial by a bailout in case of distress. In contrast, small financials do not have a similar protection. The higher tail risk of small financials should therefore translate into an additional size risk premium for small stocks in the financial industry. This risk premium is on top of the size premium in equity returns that is well documented in the literature (Banz (1981), Chan and Chen (1991), Fama and French (1993), and Van Dijk (2011)).

Our approach is related to the following literature. Viale et al. (2009) consider several asset pricing models for U.S. bank stocks during the period 1985–2003. The pricing of bank stocks differs from non-financials as shocks to the yield curve have a significant effect on bank stocks. Our setup and results differ from this study. After correcting for standard risk factors, we find limited evidence for an effect of the yield curve on average stock returns in the European financial industry. Rather than finding an empirically optimal asset pricing model, we establish and explain the link between average financial stock returns and government guarantees. A similar approach for the U.S. is in Zeng et al. (2014), who create a banking factor long in big banks and short in small banks during the sample period 1980–2007. They find a significant return for the U.S. banking factor. In several asset pricing models, this factor helps in pricing stocks, including non-financial stocks.

Gandhi and Lustig (2015) adjust returns of bank stocks using the three risk factors from Fama and French (1993) and two bond factors. They construct a banking factor in an indirect way from a principal component analysis on the residuals of a standard asset pricing model for ten financial size portfolios. During the period 1970–2013, the biggest U.S. bank stocks had significantly lower risk-adjusted returns than smaller sized bank stocks. In contrast, the risk-adjusted returns of portfolios of non-financial stocks tend to increase in size. Notably, big banks had a similar volatility, and a higher leverage than smaller banks. The findings in Gandhi and Lustig (2015) are thus consistent with government guarantees that protect big U.S. banks in disaster states.

We report smaller risk-adjusted returns for big financials, which provides evidence on a safety net for big financials. We find a higher volatility of big financials which is in support of a moral hazard story. These findings motivate us to construct a European financial risk factor, SMBfin, with a methodology closely related to Zeng et al. (2014). The significant return of this factor provides European evidence for implicit government guarantees to the financial industry by the unexplained return difference in size. We extend the existing literature by showing that such a return difference is absent in any other industry. Moreover, the momentum risk factor (Carhart (1997)) is not a substitute for our financial risk factor because it cannot fully explain the anomaly.

Our sample period ends in June 2013. Thus, it contains the most recent crisis period. This period has highlighted the importance of, and time variation in, cross-country differences within Europe. Therefore, we study time variation in the factor loadings on our risk factor SMBfin and the 3FF factors of the three-factor model in Fama and French (1993). We detect a large increase in the HML loading of big financials in Southern Europe. We show that this time variation in the HML loading is mitigated if we exclude financials from the 3FF factors, and include SMBfin as a separate factor.

We test in a cross-sectional setup for the determinants of the exposure on our financial risk factor SMBfin. In particular big financials in less prosperous countries tend to benefit from the guarantees as they are more exposed to adverse shocks. This effect dominates doubts on the national safety net, even after mid 2008. Interbank lending raised the guarantee up to the start of the crisis, afterwards investments in government securities is a significant determinant for government protection.

In addition, we address the pricing of the whole universe of European stocks, including non-financials, in the framework of Fama and French (FF 2012). FF (2012) find a value pattern in the risk-adjusted returns of big stocks, and a size pattern in the risk-adjusted return of value stocks. We show that an important driver of the anomalies is the increasing weight over time of financials in the value portfolio of the HML factor. The anomalies are non-existent in an asset pricing model with SMBfin and the 3FF factors constructed without financials.

The banking landscape in Europe and the U.S. differ from each other.

More specifically, Europe abounds with financial conglomerates engaged in several business lines within the financial industry. For instance, Europe's largest banks also offer insurance products. However, a major component of systemic risk is the risk on simultaneous bank runs at multiple institutions, either physically or electronically. Compared to the insurance business, bank runs make banking operations more prone to a systemic crisis, and government guarantees. Isolating the effect of banking operations on stock returns of conglomerates is infeasible from a practical perspective. Nonetheless, one could argue that our results are a lower bound on the impact of government guarantees for institutions specialized in banking. In addition, the biggest European institutions operate on a global scale, which may indicate that foreign affiliates may not enjoy a government protection. Still, our cross-sectional results are significant with this distorting feature. In other words, the results could be more pronounced when eliminating the effect of non-banking businesses and foreign operations on the stock returns. As this exercise is infeasible from a practical perspective, we simply use stock returns that include the effect of non-banking businesses and foreign operations.

Since we construct a single factor for whole Europe, our results are conditional on an integrated stock market in Europe. Evidence for the latter is in Fratzscher (2002), Kim et al. (2005), Hardouvelis et al. (2006), and Bekaert et al. (2009). Although our sample is restricted to publicly held banks, it is likely that our results extend to privately owned banks or cooperative banks. For instance, Barry et al. (2011) find no significant differences in default risk between publicly held and privately owned banks.

The paper proceeds as follows. Section 2.2 describes the data and variables. Section 2.3 lays out the methodology. In Section 2.4 we address time variation in both factor loadings and factor compositions. Section 2.5 explains the factor loading on our risk factor SMBfin for individual stocks. We discuss the implications for the Fama and French (2012)-model in Section 2.6. Section 2.7 presents the conclusions.

2.2 Data and variables

Our sample runs from July 1990 to June 2013. The market return (RM) and the European risk factors SMB (size factor, Small Minus Big), HML (value factor, High Book-to-Market (value) Minus Low B/M (growth)), and WML (momentum factor, previous Winners Minus previous Losers) are taken from Kenneth French's website.[1] The monthly U.S. dollar denominated returns of

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Developed

the risk factors are converted to euro denominated returns (or its predecessor ecu). We use the German 1-month Treasury bill rate from Datastream for the risk-free rate. We further collect the total return index (Datastream code RI), price-to-book ratio (PTBV), market value of equity (MV) and unadjusted prices (UP) of active and dead stocks listed on the main exchange of the following 16 European countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. This is the same set of countries used in FF (2012). We drop stocks without a positive price-to-book-ratio.

Returns are in ecu before 1999, afterwards returns are converted to euros if necessary. We winsorize the monthly returns outside the 0.1% to 99.9% range.[2] Our results are robust to the following adjustments: (*i*) a minimum price of €1 at the end of the previous month to minimize potential biases arising from low-price and illiquid stocks, (*ii*) the correction for Datastream data in Ince and Porter (2006) who suggest to drop a monthly return greater than 300% if it is reversed within one month, and (*iii*) dropping secondary listings. We omit these corrections in our baseline model to keep the data selection process simple and easy to replicate.

Size is the market value of equity at the end of June of year t. We use the inverse of the price-to-book ratio for the B/M ratio. To ensure that this accounting ratio is known before the returns, we match the year-end financial statement data of year $t-1$ with monthly returns from July of year t to June of year $t+1$, which is standard practice (e.g., FF (1993, 2012), Hou et al. (2011)). Following Jegadeesh and Titman (1993) and FF (2012), momentum in month t is the cumulative raw return from month $t-12$ to month $t-2$, skipping month $t-1$. This avoids the short-term reversal effect first documented in Jegadeesh (1990), and mitigates the impact of microstructure biases such as bid-ask bounce or non-synchronous trading. This is standard in momentum tests.

For each stock, we identify the industry by the Datastream Industrial Classification Benckmark (Datastream code INDC). This classification structure is based on the Industry Classification Benchmark (ICB) jointly created by FTSE and Dow Jones, and is used in e.g. Bekaert et al. (2009), and Hou et al. (2011). Detailed descriptives are in Appendix 2.A.

[2]The quantiles are taken from the complete set of returns because there is no reason to believe that data errors are equally spread among stock markets or over time.

2.3 Methodology

We start with the standard three-factor model from Fama and French (1993) (3FF model) having the excess market return (RM – RF), Small-Minus-Big (SMB) and High-Minus-Low Book-To-Market (HML) as risk factors.[3] For each industry, we test for the size dependence in the risk-adjusted returns of five equally-weighted size portfolios.[4] Following FF (2012), the size breakpoints are at the 3rd, 7th, 13th, and 25th percentile of aggregate market capitalization. The time-average of the market caps at the breakpoints roughly correspond to the time-average of the market caps of the NYSE size quintiles used in Fama and French (1993). Table 2.1 shows that the risk-adjusted (r-a) returns are monotonically decreasing in size in the financial industry. A standard risk-return trade-off cannot explain this pattern. Namely, the volatilities of the portfolios *increase* in size. The volatilities of the excess returns are 4.0% for the portfolio of smallest financials, 4.7%, 5.0%, 5.5%, and 6.6% for the portfolio of big financials. For the risk-adjusted returns we find volatilities equal to 1.6%, 1.5%, 1.8%, 1.9%, and 2.2%, respectively. This pattern is consistent with the hypothesis that government protection induces moral hazard for the big protected financials.

The results in Table 2.1 suggest that the 3FF model, including the SMB factor, misses an important effect related to size in the financial industry. As in Gandhi and Lustig (2015), we interpret the size pattern in risk-adjusted returns as a compensation for, possibly unobserved, financial tail risk. Big financials are supposed to be too-big-to-fail, and are therefore protected from default risk by implicit or explicit government guarantees. Investors trade on this latent put option by demanding a lower return for big financials. The r-a return difference of 0.63% ($= 0.34\% + 0.29\%$) between the two most extreme size groups is significant ($t = 3.87$) in the financial industry. None of the other industries exhibits a similar size pattern. The size pattern in volatilities is also consistent with a risk protection from tail risk. Big financials anticipate the larger government guarantees by investing in more volatile assets. Thus, our results are in line with the hypothesis that the largest firms in the financial industry exhibit the largest government guarantees.

[3] We added (*i*) the two bond factors from Fama and French (1993), (*ii*) the return of a portfolio long in a long term bond index and short in a risk free bond, (*iii*) the return of a portfolio long in a corporate investment grade bond index and short in a risk free bond, (*iv*) the monthly change of the spread between the 1-month Treasury bill rate and the 10-year government bond yield. This does not change our results qualitatively. For parsimony, our baseline model does not include any of these risk factors. Moreover, general European bond index data is only available since 1993.

[4] Value-weighted portfolios give qualitatively similar results.

	r-a return by size group (3FF)					
	1	2	3	4	5	t(1–5)
Financials	0.29*	-0.09	-0.12	-0.29**	-0.34**	3.78**
Basic Materials	0.02	-0.08	0.03	-0.06	-0.01	0.12
Industrials	0.17	0.03	0.04	-0.11	-0.07	1.66
Consumer Goods	0.11	0.07	0.01	-0.02	0.29	-1.01
Health Care	0.32	0.59**	0.56*	0.45*	0.59**	-1.40
Consumer Services	0.12	-0.10	-0.00	-0.06	0.08	0.24
Telecommunic.	0.28	0.27	0.30	0.81*	0.39	-0.29
Utilities	0.41*	0.55**	0.21	0.34	0.14	1.40
Technology	0.89**	0.35	0.34	0.42	0.37	1.75
Oil & Gas	0.46	-0.12	-0.22	0.36	0.23	0.86

Table 2.1: **Size-Industry returns on 3FF factors.**
Regressions of equally weighted (size, industry) portfolios on standard 3FF factors. The size groups increase in size with breakpoints at the 3rd, 7th, 13th, and 25th percentile of aggregate market capitalization. The rightmost column contains the t-value of a portfolio long in the risk-adjusted returns of the smallest stocks and short in the largest stocks. Industry classifications are based on the Datastream Industry Classification Benchmark. *, ** = significant at 5%, 1%.

The results in Table 2.1 indicate that the standard 3FF model misprices stocks in the health care, utilities, financial and technology industry. Nonetheless, the financial industry is the only industry with a significant difference between the risk-adjusted returns of the two extreme size portfolios (rightmost column). As a consequence, a factor related to the size of a firm in the financial industry is missing. Therefore, we construct for each industry a self-financing portfolio that takes an equally-weighted long position in the bottom 10% of aggregate market cap, and an equally-weighted short position in the top 80%. The breakpoints at the 10% and 80% percentiles maximize the average monthly return of this portfolio.

Value-weighted returns produce qualitatively similar results. That being said, a value-weighted industry portfolio may highly depend on a single firm if the total market cap of that industry mainly consists of one big firm. Further, the biggest firms within the bottom 10% have the largest weight in a value-weighted portfolio of small financials, while precisely those firms may benefit more from government guarantees.

We conjecture that government guarantees hardly exist in non-financial industries, and for small financials. In contrast, guarantees have a substantial

	mean N_S	mean N_B	mean SMB return (NW t-value)	mean r-a SMB return (NW t-value)
Financials	523.2	76.3	0.18 (0.83)	0.60** (4.65)
Basic Materials	255.9	34.4	-0.13 (-0.55)	0.00 (0.01)
Industrials	937.8	173.8	0.08 (0.61)	0.21 (1.68)
Consumer Goods	600.1	49.3	-0.19 (-0.99)	-0.16 (-1.06)
Health Care	208.1	11.8	-0.14 (-0.38)	-0.18 (-0.64)
Consumer Services	535.2	80.6	0.01 (0.03)	0.02 (0.15)
Telecommunic.	41.2	8.8	-0.18 (-0.43)	-0.11 (-0.33)
Utilities	75.2	23.9	0.16 (0.83)	0.22 (1.37)
Technology	418.8	24.4	0.25 (0.95)	0.40 (1.49)
Oil & Gas	129.8	8.4	0.23 (0.55)	0.09 (0.38)

Table 2.2: **SMB returns on 3FF factors.**

Risk-adjusted (r-a) returns are adjusted for risks in the standard 3FF model. NS and NB are the number of stocks in the industry portfolios with small stocks and big stocks, respectively. The SMB is equally weighted and long in bottom 10% industry market cap and short in biggest 80%. Industry classifications are based on the Datastream Industry Classification Benchmark. *, ** = significant at 5%, 1%.

effect on the returns in the top 80% market cap of financials. The descriptive statistics of the constructed industry-specific SMB portfolios in Table 2.2 are in support of our conjecture. It follows that the self-financing portfolio of the financial industry has a monthly return of 0.18%. The 3FF model predicts a negative return of –0.42% (not reported in Table 2.2), which leaves a significant 0.60% (= 0.18% + 0.42%) unexplained. None of the other industries has a risk-adjusted return difference close to 0.60% per month. Thus, the 3FF model cannot explain the returns of the financial size portfolio. The average number of stocks in the portfolios of small and big financials are 523.2 and 76.3, respectively. This demonstrates that the size pattern is present in a large sample of stocks, and excludes outliers in the cross-section as an explanation.

During our sample period, the European banking system was mainly nationally regulated while some countries tend to have larger financials than others. To correct for this effect, we construct *for each country* a self-financing portfolio that takes an equally-weighted long position in the bottom 10% of aggregate financial market cap, and an equally-weighted short position in the top 80%. We weigh the national financial size portfolio returns by the monthly number of stocks and refer to this portfolio as SMBfin. After adjusting the weighted portfolio returns for the 3FF risks, the risk-adjusted SMBfin return equals 0.62% ($t = 4.74$).

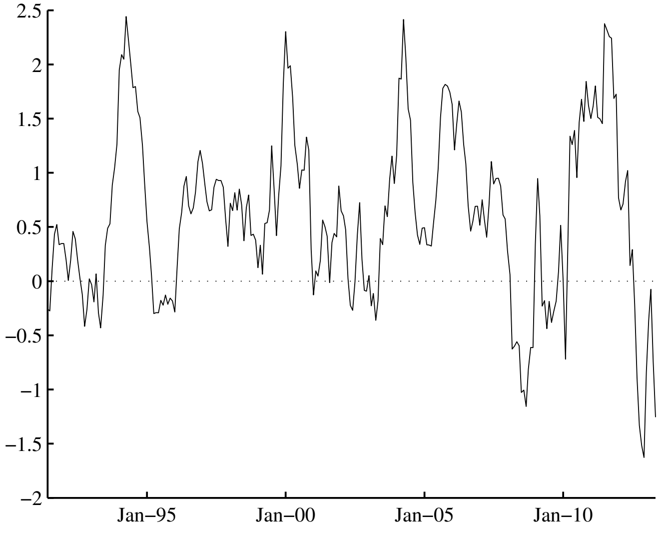
The SMBfin factor is only based on financial stocks. As such, it may have additional explanatory power on top of the standard SMB factor in the 3FF model. Accordingly, we examine the additional explanatory power of the SMBfin factor to regressions on the 3FF factors. It follows from the Frisch–Waugh–Lovell theorem that the projection of the SMBfin factor on the 3FF factors can be subtracted from the SMBfin factor without altering the regression results. In other words, the regression results are unchanged if we correct SMBfin for the risk in the 3FF factors. Therefore, we regress the returns of the SMBfin portfolio on the 3FF factors with intercept, and define the risk factor ESMBfin as the intercept plus residual. This risk factor is by construction uncorrelated with the 3FF factors. FF (1993) adopt a similar approach to measure the effect of the market return on common variation in returns left by the other factors. Similar to Gandhi and Lustig (2015), the ESMBfin factor represents returns unexplained by the 3FF factors. Figure 2.1 is the 12-month moving average of ESMBfin. It shows that during normal times, small financials have a higher risk-adjusted return than big financials to compensate for, e.g., the smaller potential benefit from government guarantees.

Considering the spike in Figure 2.1 at the start of our sample period, U.S. interest rates increased considerably during 1994. For instance, the federal funds rate doubled from 3% to 6%. The spike suggests that small European financials were initially less exposed to this increase than big European financials. A possible explanation is that the financing of Europe's biggest financials depends more on short-term U.S. debt. On the asset side, big financials have a smaller exposure to interest rate changes if they invest more in long-term assets such as mortgages. This maturity mismatch can make big financials more exposed to an unanticipated spike in the Fed rate. In addition, the more uncertain prospects of the U.S. economy are likely to have an additional upward effect on the discount rates of big European financials than on the discount rates of their smaller counterparts. This can further raise the return of the risk-adjusted ESMBfin.

The ESMBfin factor dropped sharply in 2008 and 2011 when conditions of the European financial system turned out to be worse than expected. Big financial stocks were protected by government guarantees, either implicit or explicit. This made investors less willing to hold the more risky small financial stocks. This put a downward pressure on the long small-short big

Figure 2.1: **12 month moving average ESMBfin**

12 month moving average return of monthly risk-adjusted return of SMB portfolio of financials as obtained by a regression on the 3FF factors.



portfolio.

Table 2.3 presents the residuals if we regress the industry size portfolios on the 3FF factors and the financial risk factor SMBfin (or equivalently, ESMBfin derived from the FWL theorem). The size pattern in risk-adjusted returns of the financial industry vanishes. That is, SMBfin explains the size pattern in financials. The same holds true for the spread of the two extreme size financial portfolios. This spread is now an insignificant 0.03% ($t = 0.40$). In addition, none of the financial size portfolios earns a significant risk-adjusted return.

The loadings on the SMBfin factor are in Table 2.4. As expected, the exposure to SMBfin decreases monotonically in size, starting from 0.29 ($t = 5.28$) for the smallest financials to −0.68 ($t = -13.8$) for the biggest financials.[5] The SMBfin portfolio is long in the 10% smallest national market cap such that it tends to be long in the second financial size group, the second

[5]Appendix B shows that a t-test on the risk-adjusted returns of a portfolio long in a portfolio in Table 2.3 and short in the corresponding portfolio in Table 2.1, tests on the risk-adjusted returns of SMBfin for any portfolio in the table. Thus, all portfolios would show identical test statistics. Hence, we report the t-statistics of the SMBfin coefficients rather than the t-statistics of the effect of ESMBfin on the risk-adjusted return.

	r-a return by size group (3FF+SMBfin)					
	1	2	3	4	5	t(1–5)
Financials	0.11	-0.03	0.08	-0.03	0.08	0.40
Basic Materials	-0.08	-0.03	0.07	-0.07	-0.10	0.08
Industrials	0.05	0.00	0.04	-0.09	-0.09	1.00
Consumer Goods	0.04	0.08	0.08	-0.04	0.16	-0.64
Health Care	0.25	0.69**	0.51*	0.39	0.43*	-0.68
Consumer Services	-0.01	-0.13	0.04	-0.01	0.03	-0.20
Telecommunications	0.15	0.22	0.17	0.67	0.24	-0.25
Utilities	0.36*	0.48**	0.21	0.32	0.12	1.30
Technology	0.76**	0.43	0.46	0.56*	0.44	1.08
Oil & Gas	0.31	-0.14	-0.40	0.22	-0.04	1.31

Table 2.3: **Returns on 3FF and SMBfin.**
Regressions of equally weighted size portfolios on standard 3FF factors and the SMBfin factor. The size groups increase in size with breakpoints at the 3rd, 7th, 13th, and 25th percentile of aggregate market cap. The rightmost column contains the t-value of a portfolio long in the risk-adjusted returns of the smallest stocks and short in the largest stocks. Industry classifications are based on the Datastream Industry Classification Benchmark. *, ** = significant at 5%, 1%.

size group has a negative exposure of -0.10 ($t = -2.12$) to SMBfin. This coincides with the negative coefficient of this group in Table 2.1, and indicates that government guarantees are widespread in the financial industry. None of the other industries shows a similar size pattern in risk-adjusted returns.

Our financial risk factor may shed some light on the so-called low risk anomaly in financial stocks (Baker and Wurgler (2015)). This anomaly states that banks with a smaller market beta have higher returns, even after correcting for risks in the 3FF model. Since large financials are more diversified than small financials, their market beta coefficients are indeed larger in our regressions (not reported). That is, large beta financials are large financials with less tail risk by the government guarantees. The smaller tail risk of the large beta financials is thus captured by our SMBfin factor.

The model does however still a poor job in explaining the returns of the health care, utilities and technology industry. This is not surprising since other industries are not protected to the same extent by government guarantees. Nonetheless, the standard SMB factor should already captures guarantees to big non-financial firms that are absent for small firms. For instance, the U.S. government bailed out General Motors in 2009 to avoid

	SMBfin loading by size group (3FF+SMBfin)				
	1	2	3	4	5
Financials	0.29**	-0.10*	-0.33**	-0.41**	-0.69**
Basic Materials	0.16	-0.09	-0.10	0.02	0.14
Industrials	0.18**	0.05	-0.03	-0.03	0.02
Consumer Goods	0.11*	-0.02	-0.06	0.03	0.22**
Health Care	0.11	-0.16	0.08	0.10	0.26*
Consumer Services	0.20**	0.05	-0.07	-0.09	0.08
Telecommunications	0.21	0.08	0.20	0.23	0.24*
Utilities	0.07	0.10	0.00	0.03	0.04
Technology	0.21	-0.12	-0.18*	-0.23*	-0.13
Oil & Gas	0.24	0.04	0.29*	0.23*	0.44**

Table 2.4: **SMBfin factor loading.**
Regressions of equally weighted size portfolios on standard 3FF factors and the SMBfin factor. The size groups increase in size with breakpoints at the 3rd, 7th, 13th, and 25th percentile of aggregate market capitalization. Industry classifications are based on the Datastream Industry Classification Benchmark. *, ** = significant at 5%, 1%.

a mass unemployment. The financial risk factor SMBfin captures the additional guarantees to the financial industry. Thus additional guarantees to the financial industry are rational in light of the systemic nature of this industry due to interbank relationships, interactions with the real economy, and the similar risk profiles of big financials.

2.4 Time variation and the crisis

We argued that the safety net reduces the tail risk of big financials. Market participants anticipate this lower tail risk by demanding a lower return for big financials. Our results in Table 2.1 and Table 2.2 suggest that this mechanism is not captured by the standard 3FF model. That is, the 3FF model cannot explain the lower return of big financials as their risk-adjusted return is significantly negative. This has led us to construct a financial SMB portfolio with a significant risk-adjusted return of 0.62% per month (after separation by country, see p.15).

A well-known additional risk factor to the 3FF model is the momentum risk factor which is long in previous year's winners and short in previous year's losers (Carhart (1997)). It is well documented that this factor captures some additional risk as it earns a significant average risk-adjusted return in the 3FF model. The significant risk-adjusted return of our financial SMB portfolio suggests that small financials tend to be included in the long winner portfolio, while big financials are in the short loser portfolio. Indeed, our financial risk factor is significantly correlated with the European momentum factor ($\rho = 0.45$, $t = 4.36$). This raises the question whether the momentum factor also explains the size anomaly in financial stock returns.

Table 2.5 shows that the 3FF model augmented with the momentum factor partly explains the size pattern in financials. Nevertheless, the spread in risk-adjusted returns is still a significant 0.43% (0.35% + 0.08%, $t = 2.56$). Thus, a momentum factor partly corrects for government guarantees in the financial industry, but in a more indirect way compared to the SMBfin factor. The factor loadings of the five financial size portfolios on the momentum factor decrease monotonically in size (details available upon request), and are all significantly negative ($t = -2.05$ for the smallest stocks). In other words, each financial size group is more exposed to the loser portfolio than to the winner portfolio. This gives a lower average return, and a negative correlation with the momentum portfolio. This effect is strongest for big financials since they are safer by the supposedly larger government guarantees. Similar to our SMBfin factor, the momentum factor cannot explain the returns of size portfolios in health care, utilities and technology.

Time variation in the coefficients of asset pricing models explains some of the momentum returns (see, e.g., Chordia and Shivakumar (2002)). Such time variation would of course also translate into incorrect inferences on our SMBfin factor. It is thus of interest to see if time variation can explain the return of the SMBfin factor. To study this, we considered (*i*) daily data with annual regressions starting in July, (*ii*) monthly data with 5 year rolling windows, and (*iii*) a state-space model using monthly data. All three methods give very similar results. We report the results with daily data (method (*i*)), since daily data gives the most timely estimation of changes in the coefficients. In addition, the setup does not rely on the parametric assumptions needed in state-space models.

The annual coefficients of the daily SMBfin portfolio regressed on the daily 3FF factors are in Figure 2.2. The loadings are moderately volatile with one exception, the large tilt in the HML loading in July 2007. One could argue that the significant return of our SMBfin portfolio stems from this drop. However, our results do not change qualitatively if we end the sample period at the end of June 2007, before the crisis. That is, the r-a return on the SMBfin portfolio remains significant with an average monthly return of 0.56%.

	r-a return by size group (3FF + Momentum)					
	1	2	3	4	5	t(1–5)
Financials	0.36**	0.04	-0.01	-0.11	-0.07	2.56*
Basic Materials	0.17	0.08	0.05	0.03	0.02	0.57
Industrials	0.33**	0.17*	0.17*	-0.03	0.05	1.93
Consumer Goods	0.28*	0.19	0.07	0.08	0.25	0.17
Health Care	0.56*	0.61**	0.57*	0.39	0.50*	0.21
Consumer Services	0.28*	0.03	0.16	0.06	0.12	0.90
Telecommunications	0.57	0.25	0.28	0.84	0.33	0.65
Utilities	0.42*	0.51**	0.14	0.26	0.01	2.17*
Technology	1.02**	0.46	0.43	0.61*	0.51	1.67
Oil & Gas	0.50	0.02	-0.28	0.23	0.17	1.21

Table 2.5: **Returns on 3FF and momentum.**
Regressions on standard 3FF factors and the momentum factor. The size groups increase in size with breakpoints at the 3rd, 7th, 13th, and 25th percentile of aggregate market capitalization. The rightmost column contains the t-value of a portfolio long in the risk-adjusted returns of the smallest stocks and short in the largest stocks. Industry classifications are based on the Datastream Industry Classification Benchmark.
*, ** = significant at 5%, 1%.

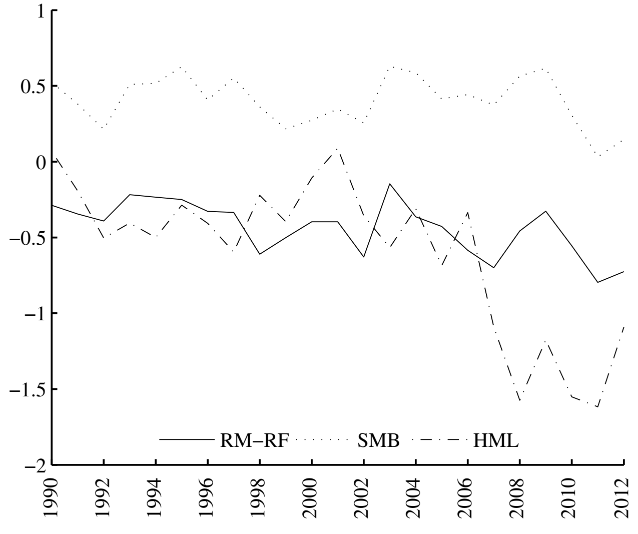
Using daily data, we also tested if the mean r-a SMBfin return is zero in each year. Even when allowing for annual changes in regression coefficients, the F-test clearly rejects zero returns ($p < 0.1\%$). Thus, time variation in the HML loading cannot explain the significant returns of our financial size portfolio. This is in line with Ang and Kristensen (2012) who find for U.S. stocks a significant variation in factor loadings, but also overwhelming evidence that a conditional version of the Fama and French (1993) model cannot account for the momentum effect.

The more negative HML loading during July 2007–June 2009 before the European debt crisis are, however, a statistical artefact due to (*i*) the large positive net weight of financials in the value-weighted HML factor (top plot Figure 2.3) as a substantial proportion of financials has become distressed, and (*ii*) the high volatility of financials during the crisis period, thereby determining the dynamics of the HML portfolio (bottom plot Figure 2.3). That is, since the crisis period the HML factor increasingly represents the dynamics of financials.

The annual tilts in the percentage of the market cap of financials are attributable to the annual rebalancing of the portfolios. To correct for this composition effect, we redo the regression of SMBfin on the 3FF factors,

Figure 2.2: $\mathrm{SMB_{fin}}$ **on 3FF**

Factor loadings of the financial risk factor SMBfin on the three Fama and French (1993) factors. Daily data from July 1990 to June 2013. Regressions run from July of year t to June of year $t+1$, year t is on the horizontal axis.



where the SMB and HML factors are now constructed without the financials. We find again a significant monthly r-a SMBfin return, now 0.54% ($t = 2.80$). Compared to the results with the original daily 3FF model in Figure 2.2, the major new finding is the postponed drop in HML loading from July 2007 to July 2009 in Figure 2.4. The composition effect thus explains a substantial part of the drop in the original HML loading of SMBfin between July 2007 and June 2009.

There is a very sharp size distinction in the drop. The HML loading of small financials is stable over time (top plot Figure 2.5). The effect of the crisis on the loadings of small financials is thus comparable to firms in other industries composing the non-financial risk factors, while the higher HML loading of big financials completely explains the drop in HML loading over the most recent years (bottom plot Figure 2.5). Big financials are thus more like distressed firms as of July 2009, even after correcting for the composition effect of the HML portfolio.

From late 2009, investors started to fear a sovereign debt crisis as a result of rising private and sovereign debt levels, mainly in Southern Eu-

Figure 2.3: **HML characteristics over time**

Top: Percentage of market cap financials in total market cap of high B/M, low B/M, and HML portfolio, bottom: 12 month moving average volatility of HML constructed without financials, and HML exclusively with financials.

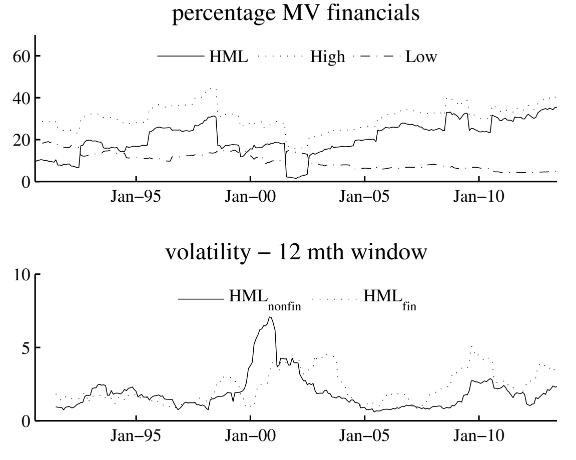


Figure 2.4: SMB_{fin} **on** $\text{3FF}_{\text{nonfin}}$

Factor loadings of the financial risk factor SMBfin on the three Fama and French (1993) factors constructed without financials. Regressions run from July of year t to June of year $t+1$, year t is on the horizontal axis.

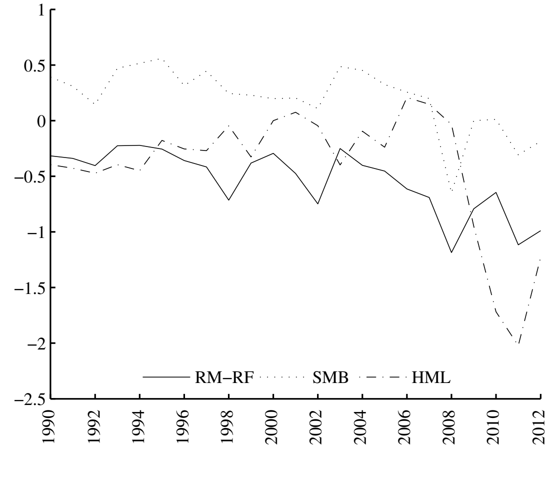


Figure 2.5: **Components** SMB_{fin} **on** $3\text{FF}_{\text{nonfin}}$

Factor loadings of the components of SMBfin on the three Fama and French (1993) factors constructed without financials. Regressions run from July of year t to June of year $t+1$, year t is on the horizontal axis.

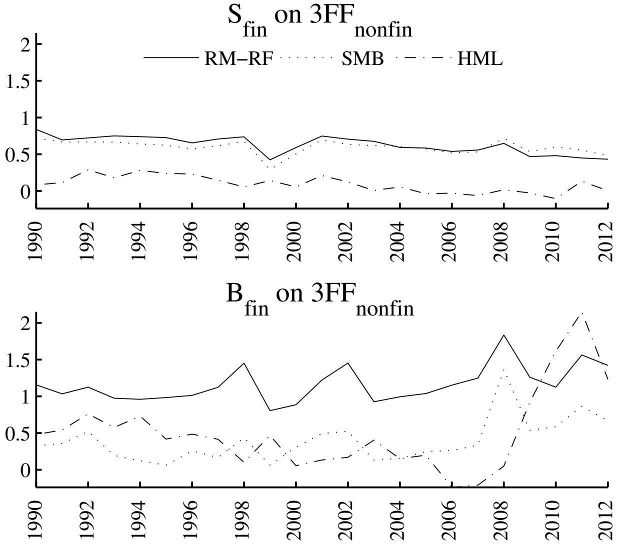
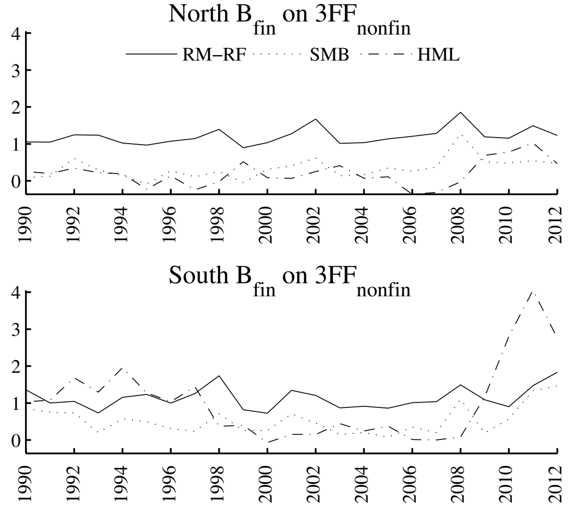


Figure 2.6: B_{fin} **by region on** $3FF_{nonfin}$

Top: Northern Europe (Austria, Denmark, Germany, Netherlands, Norway, Sweden, Switzerland, U.K.), bottom: Southern Europe (Greece, Italy, Portugal, Spain). Regressions run from July of year t to June of year $t+1$, year t is on the horizontal axis.

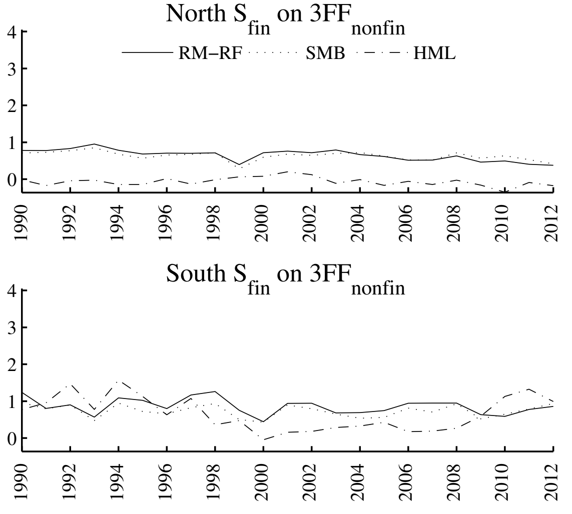


rope. The quick rise of sovereign debt was induced by several bailouts in the banking sector. European banks in turn own a significant amount of the sovereign debt, which forced bank's to depreciate on their assets. The resulting deplorable state of the banking system made a bailout more likely, thereby further depressing sovereign debt prices. This mechanism explains why solvency concerns regarding banks and sovereigns were reinforcing. The feedback loop suggests that the European debt crisis split Europe into two parts, with financials in the Southern countries much more affected. When the crisis unfolded, financials in the Northern countries were in general only moderately hit by indirect contagious effects in the financial industry. Indeed, the time variation in the European HML loading of our risk factor is mainly attributable to big financials in Southern Europe (Figure 2.6 and Figure 2.7). The HML loading of big financials in Northern Europe as well as the HML loading of small financials in Southern Europe exhibits only a moderate increase. This indicates that big financials in Southern Europe faced additional adverse shocks during the European debt crisis. Possibly, such financials had more sovereign debt of distressed countries on their balance sheets, or investors started to doubt the safety net in Southern Europe.

Nonetheless, the different shape of the HML loading of small financials

Figure 2.7: S_{fin} **by region on** $3FF_{nonfin}$

Top: Northern Europe (Austria, Denmark, Germany, Netherlands, Norway, Sweden, Switzerland, U.K.), bottom: Southern Europe (Greece, Italy, Portugal, Spain). Regressions run from July of year t to June of year $t+1$, year t is on the horizontal axis.



across Europe suggests that a safety net explanation is not the whole story. This motivates us to assess the impact of the following two mechanisms that could both explain the increased distress of Europe's biggest financials. On the one hand, we discussed the feedback loop that default risk of sovereign debt may induce investors to doubt the reliability of the safety net through depreciations on bank balance sheets. On the other hand, sovereign distress and unexpected macroeconomic shocks directly raises credit risk on the assets of financials. As a consequence, solvency risk of financials increases, which could also result in a distressed status. In the next section, we address the relative impact of both mechanisms by considering the factor loadings on our financial risk factor.

2.5 Explaining the factor loading

A more negative loading on our financial risk factor SMBfin indicates higher expected government guarantees, and thus a lower tail risk. It is of particular interest for investors and regulators to identify which financials have a substantially negative exposure to our financial risk factor. Using daily data, we regress the stock returns of each financial on SMBfin, and the

3FF factors without financials. We use annual windows to account for time variation in both factor loadings and characteristics. Each window starts in July to ensure that year-end balance sheet data is at the centre of each window. The regressions provide us panel data with the loading $b_{i,t}$ on the financial risk factor of financial stock i in annual window t. We estimate the between-effects panel regression[6]

$$\overline{b_i} = \alpha_i + \bar{\gamma}_i + \sum_j \beta_j \overline{X_{ij}}$$

where α_i is the intercept of stock i, γ_t are stock-independent time dummies for each window, β_j are regression coefficients of the explanatory variables X_{ij} of stock i, and an upper bars denotes a stock-specific time series average over the annual windows where stock i has a complete series of daily returns. The time dummies account for time-varying effects of omitted variables in the unbalanced panel.

We tested several explanatory bank-specific variables from Bankscope and macro variables from Eurostat. In our final model, we selected variables that meet the following requirements (*i*) a clear economic interpretation, (*ii*) a large number of observations, (*iii*) significant in a univariate model with time dummies included, (*iv*) significant in the final multivariate model, (*v*) a similar coefficient estimate in both models (*iii*) and (*iv*), and (*vi*) a low correlation with the other included variables in a multivariate setting. For instance, we do not use the debt-to-GDP ratio and real GDP per capita simultaneously because of the high negative correlation (–0.52). Although debt is a stock variable and GDP a flow variable, both variables are indicators for the state of the national economy. Noteworthy, the debt-to-GDP ratio is for the whole sample only available since 1995.

Panel A in Table 2.6 presents the univariate regressions. We find that common equity, total assets, the leverage ratio, the proportion of government securities, the proportion of bank loans, and the debt-to-GDP ratio have a negative effect on the financial risk factor loading $b_{i,t}$. A negative loading on SMBfin means that the r-a returns are more similar to the r-a returns of big financials. Thus, we interpret variables with negative coefficients as indicators of more government guarantees. Real GDP per capita affects the loading positively.

[6] A fixed (within) effects model differences the results of subsequent periods. It focuses on the time series dimension instead of the cross-section, our point of interest. A random effects model provides inconsistent estimates because the Hausman test is rejected. This basically means that the stock-specific error series is correlated with the explanatory variable.

	Panel A: univariate				Panel B: multivariate							
	coeff. (t-value)	N	Nfin	R^2	coeff (t-value)							
common equity	–14.0**	3,115	397	0.36	–14.4**	–14.2**	–16.5**	–16.9**	–15.2**	–15.7**	–11.2	–8.5
	(–12.7)				(–5.1)	(–12.0)	(–9.7)	(–9.9)	(–14.0)	(–14.6)	(–1.5)	(–1.1)
TA	–9.8**	3,131	398	0.31	0.36							
	(–11.2)				(0.16)							
leverage	–9.1**	3,121	398	0.11		0.95						
	(–3.7)					(0.43)						
gov. sec. /TA	–7.6**	1,185	206	0.17			–2.7	–3.3				
	(–3.0)						(–1.3)	(–1.7)				
bank loans /TA	–6.4**	2,842	353	0.09					–6.7**	–6.7**	–21.3**	–23.0**
	(–2.9)								(–3.9)	(–3.9)	(–2.7)	(–2.7)
debt /GDP	–46.0**	2,899	398	0.18			–38.8**		–37.3**		–44.8*	
	(–7.5)						(–5.1)		(–6.9)		(–2.6)	
real GDP /cap	58.2**	3,102	398	0.19				47.5**		48.8**		49.1
	(7.1)							(4.4)		(7.2)		(1.9)
N					3,115	3,114	1,077	1,153	2,601	2,801	427	443
Nfin					397	397	205	206	350	351	68	68
Size threshold					-	-	-	-	-	-	3	3
R^2					0.36	0.36	0.46	0.49	0.49	0.52	0.46	0.45

Table 2.6: **Determinants SMBfin loading.**
Between effects panel regressions of annual stock specific SMBfin factor loadings. The factor loadings are first estimated in annual regressions of daily financial stock returns on the 3FF factors without financials and SMBfin. In the second regression, the factor loadings are regressed on the logarithms of the explanatory variables. The coefficients are multiplied by 100, t-values are in brackets. N is the number of (stock, year) observations, Nfin is the number of different stocks, size threshold is threshold (bln €) on common equity, common equity and TA (total assets) are in (logs of) mln €, leverage is the ratio of total assets to common equity. Sample period is July 1990–June 2013. *, ** denotes significant at the 5% and 1% level, respectively.

When controlling for common equity, the regressions in Panel B in Table 2.6 report an insignificant coefficient of total assets. We suggest that this follows from the construction of our financial risk factor which uses equity instead of total assets as a measure for size. Similarly, the leverage ratio is insignificant if common equity is included as a regressor. It is natural that a high leverage, i.e., a large amount of debt, is only feasible for healthy financials. Firms in more moderate conditions experience harder conditions when attracting more debt. However, financials with a high leverage tend to be systemically more important because of the relatively large amount of outstanding debt. A high leverage could thus be due to better prospects (less observed guarantees), and could also indicate more systemic linkages (more observed guarantees). This makes it hard to relate leverage to our risk factor which measures implicit guarantees. We explain the significant univariate result on leverage by the positive relation between leverage and size.

The panel regression weighs each financial stock equally. This assigns a relatively large weight to the more abundant small financials, which have smaller government guarantees. Nevertheless, the two rightmost columns in Table 2.6 show that our results are robust under different size groups. A minimum of €3 bln of common equity does not change our results qualitatively. The smaller number of included financials affects the standard error of common equity, not the sign of the point estimate. Alternatively, the less negative point estimate of big financials on common equity may indicate a nonlinear effect in the expected benefit from the safety net.

What are the economic mechanisms behind the significant variables? Financials with a higher level of common equity are in general bigger and more levered and thus have more debt on their balance sheet. The corresponding larger number of linkages with other financials makes such financials systemically more important, which explains the negative factor loading of common equity to our risk factor, and thus the higher guarantees. More government securities, more bank loans, and a higher debt-to-GDP ratio increase the likelihood of being hit by a systemic crisis. This increases the measured guarantee from the safety net. A high real GDP per capita may indicate a favourable state of the national economy or a stable banking system with a well-performing supervision. Both can explain the negative relation between GDP per capita and observed implicit guarantees. Thus, we find that financials in countries with a poor state of the economy, as measured by debt-to-GDP or GDP per capita, are more likely to benefit from a safety net rather than to be penalized by a less reliable safety net.

We elaborate on the previous result in a more formal way. Let G denote the government guarantee to financial firm i. This guarantee is not necessarily explicitly formalized in supervisory regulations. If not, market participants have implicit beliefs on the size of the guarantee to firm i. This size could depend on the state of the domestic economy of this financial. Let Ω represent some indicator of this state. A larger Ω is associated with a more favourable state. This makes the probability of default decrease as Ω increases. In contrast, the expected guarantee conditional on a default increases as Ω increases, since a better state of the economy is associated with a more reliable safety net. More formally, the law of total probability states that

$$\mathbb{E}[G] = \mathbb{P}(D)\,\mathbb{E}\left[G \mid D\right] + \mathbb{P}\big(D^C\big)\,\mathbb{E}\left[G \;\middle|\; D^C\right]$$

where D^C is the complement of the default event D. By using that guarantees are only allocated in case of a default, $\mathbb{E}\left[G \;\middle|\; D^C\right] = 0$, and allowing the terms to depend on the state of the economy Ω:

$$\mathbb{E}[G\,\text{of firm}\,i\,;\,\Omega] = \mathbb{P}(\text{default firm}\,i\,;\,\overbrace{\Omega}^{-}\,) \times \mathbb{E}[G\,\text{given default firm}\,i\,;\,\overbrace{\Omega}^{+}\,] \tag{2.1}$$

We find in Panel B in Table 2.6 a significantly positive relation between the state of the economy and factor loadings. The negative of the factor loading is a measure for the unconditional size of the government guarantees $\mathbb{E}[G\,\text{of firm}\,i\,;\,\Omega]$, thus the left-hand side (LHS) in (2.1). Thus, the LHS decreases in Ω. This negative dependence on Ω implies that the right-hand side (RHS) must also decrease with the state variable Ω. That is, the state of the national economy has more impact on default probabilities ($\mathbb{P}(D)$) than on the safety net ($\mathbb{E}\left[G \mid D\right]$). The smaller effect through the safety net suggests that investors assume a latent aggregate European safety net when pricing European financials. This result holds for the complete sample (Table 2.6), before (Panel A in Table 2.7) and after the start of the financial crisis in 2008 (Panel B in Table 2.7).

Up to the crisis, bank loans had a significant impact on the factor loading, while government securities were insignificant (Panel A in Table 2.7). Interestingly, since the start of the crisis, government securities rather than bank loans are a significant determinant of the factor loading on the SMBfin factor (Panel B in Table 2.7). Although the government securities data is far from complete, the pattern is clear from the available data. Having government securities is increasingly associated with bad performance and financial distress risk. This made the holders of government securities potentially benefit from a safety net as the negative loading in Table 2.7 indicates.

	Panel A: July 1990–June 2008						Panel B: July 2008–June 2013					
	univariate				multivariate		univariate				multivariate	
	coeff. (*t*-value)	N	Nfin	R^2	coeff. (*t*-value)		coeff. (*t*-value)	N	Nfin	R^2	coeff. (*t*-value)	
common equity	–13.1** (–11.4)	2,369	363	0.34	–14.0** (–12.4)	–14.4** (–13.0)	–14.4** (–9.2)	746	210	0.31	–17.2** (–6.3)	–16.1** (–5.7)
gov.sec./TA	–4.4 (–1.7)	927	180	0.13			–16.4** (–4.2)	258	85	0.19	–7.8* (–2.4)	–9.3** (–2.9)
bank loans /TA	–7.7** (–3.5)	2,211	327	0.12	–8.3** (–4.7)	–7.6** (–4.3)	–2.3 (–0.6)	631	178	0.01		
debt/GDP	–31.8** (–4.9)	2,150	364	0.13	–27.1** (–4.8)		–56.7** (–7.2)	749	210	0.21	–49.4** (–3.9)	
real GDP /cap	44.0** (5.2)	2,353	364	0.16		39.4** (5.6)	89.5** (7.7)	749	210	0.23		60.2** (2.9)
N					1,972	2,172					256	258
Nfin					324	325					85	85
R^2					0.44	0.49					0.52	0.47

Table 2.7: **Determinants SMBfin loading by period.**

Between effects panel regressions of annual stock specific SMBfin factor loadings. The factor loadings are first estimated in annual regressions of daily financial stock returns on the 3FF factors without financials and SMBfin. In the second regression, the factor loadings are regressed on the logarithms of the explanatory variables. The coefficients are multiplied by 100, *t*-values are in brackets. N is the number of (stock, year) observations, Nfin is the number of different stocks, common equity and TA (total assets) are in (logs of) mln €, leverage is the ratio of total assets to common equity. *, ** denotes significant at the 5% and 1% level, respectively.

Another striking feature is that the state of the economy has become more important since the start of the crisis period. Namely, the coefficient estimates of real GDP per capita and the debt-to-GDP ratio are both more than 50% higher since the start of the crisis in 2008, despite the steady growth in both variables. Again, investors expect more benefits from a safety net when national economic conditions are less favorable. Thus, compared the higher default probability in less prosperous economies dominates any doubts on the reliability of the national safety net, and this dominance has become stronger since the crisis period.

We added interaction terms to each of the six multivariate regressions in Table 2.6 and Table 2.7 that have only significant coefficients. More specifically, using three different time periods the interaction term captures the interaction between standardized common equity and the standardized employed national economic state variable, either debt-to-GDP or real GDP per capita. Out of these six regressions, only the pre-crisis regression with debt-to-GDP ratio ($t = 2.66$) as state variable yields a significant interaction term. The other explanatory variables in this regression have very similar coefficients and p-values as in the original regression without an interaction term.

The significant positive coefficient of the interaction term indicates that before the crisis, a large size was associated with a less negative slope on the national debt-to-GDP ratio. That is, a large size had a mitigating impact on the positive relation between the debt ratio and expected government guarantees. Since the start of the crisis, this interrelation has disappeared. That is, debt-to-GDP has become more important for big financials.

The stronger relations in Panel B in Table 2.7 suggest an enhanced effect of debt-to-GDP and common equity on expected guarantees. The latter coincides with the results in Section 2.4 where we identified a size effect in the impact of the crisis. We argued that the crisis raised doubts on the safety net for big financials in Southern Europe, i.e., the high debt countries. Here, we find that the positive relation between debt-to-GDP and the measured government guarantees of big financials is stronger since the crisis, whilst the mitigating interacting effect through size and the domestic economy has become weaker.

2.6 Extensions to non-financials

We discuss the effect of our financial risk factor on the whole universe of European stocks, including non-financials. In doing this, we closely follow the approach in Fama and French (2012). Some average returns are poorly explained in their basic European pricing model. To illustrate the pricing ability of our factor, we perform similar regressions on the 5×5 portfolios from Kenneth French' website. The 5×5 value weighted portfolios contain equity from 16 European countries.[7] The size breakpoints are at the 3rd, 7th, 13th, and 25th quantile, while the breakpoints are at quintiles for B/M and momentum. As in FF (2012), the sample period runs in this section from November 1990 to March 2011. If necessary, we convert the returns to euro denominated returns rather than the dollar denominated returns in FF (2012), which explains some of the small differences with their results.

2.6.1 Size-B/M portfolios

Panel A in Table 2.8 shows the intercepts and Newey-West t-values for the 25 size-B/M portfolios on the three-factor model. A value pattern is present in the portfolios of big stocks. Investing in the big growth (Big, Low)-portfolio by shorting the big value (Big, High)-portfolio creates a self-financing portfolio with an expected return of 0.37% ($t = 2.81$). Similarly, a size pattern is in the portfolios of value stocks. Investing in the small value portfolio by shorting the big value portfolio generates an expected return of 0.46% ($0.21\% + 0.26\%$, $t = 3.38$).

To address time variation in the HML loading of the financials (see Figure 2.2) we exclude financials from the SMB and HML portfolio. The corresponding $3FF_{NF}$ factors improve the explanatory power of big stocks. Specifically, Panel B in Table 2.8 shows that the return of the Big HML portfolio has an expected risk-adjusted return of 0.29% ($0.21\% + 0.09\%$, $t = 1.84$). Still, the value SMB generates a significant expected return, now equal to 0.49% ($0.29\% + 0.21\%$, $t = 2.76$).

The model can neither explain the size pattern in value stocks and the value pattern in big stocks. We find in Panel C in Table 2.8 that the SMBfin factor reduces the r-a returns of these two portfolios to an insignificant 0.02% ($0.02\% - 0.00\%$, $t = 0.13$) and 0.12% ($0.14\% - 0.02\%$, $t = 0.78$). In particular the returns of big stocks are well explained. The returns are insignificant and there is no value pattern in big stocks. Though there is still a size pattern in the value stocks, none of the risk-adjusted returns is significant.

Panel D in Table 2.8 shows that the spreads of the extreme portfolios of big and value stocks are more pronounced when we use the $3FF_{NF}$ factors and the momentum factor as regressors. Thus, our SMBfin factor has better

[7] A detailed description of the data is in Appendix 2.A.

	a					*t(a)*				
	Low	2	3	4	High	Low	2	3	4	High
Panel A: 3FF										
Small	-0.30	-0.07	0.01	0.10	0.21	-2.11	-0.68	0.16	1.32	3.08
2	-0.12	-0.09	-0.04	0.07	0.08	-1.36	-1.07	-0.49	1.23	1.34
3	-0.03	-0.02	-0.04	-0.11	0.02	-0.26	-0.18	-0.39	-1.19	0.22
4	0.08	-0.01	-0.01	-0.17	-0.07	0.91	-0.08	-0.06	-1.71	-0.74
Big	0.11	0.06	0.05	0.00	-0.26	1.52	0.88	0.64	0.02	-2.22
Panel B: $3FF_{NF}$										
Small	-0.15	0.06	0.13	0.17	0.29	-0.98	0.51	1.22	1.70	3.25
2	0.01	0.00	0.05	0.15	0.16	0.13	0.00	0.67	1.96	2.23
3	0.09	0.07	0.01	-0.05	0.08	0.77	0.66	0.13	-0.54	0.84
4	0.16	0.01	0.01	-0.15	-0.02	1.52	0.10	0.13	-1.67	-0.17
Big	0.09	-0.01	0.01	0.00	-0.21	1.08	-0.14	0.10	-0.05	-1.50
Panel C: $3FF_{NF}$ + SMBfin										
Small	-0.38	-0.13	-0.03	-0.01	0.14	-3.26	-1.25	-0.35	-0.17	1.81
2	-0.14	-0.11	-0.04	0.06	0.10	-1.52	-1.19	-0.50	0.79	1.45
3	-0.04	0.01	-0.04	-0.10	0.09	-0.40	0.06	-0.45	-0.98	0.88
4	0.09	-0.04	-0.01	-0.12	0.04	0.85	-0.41	-0.09	-1.34	0.38
Big	0.00	-0.06	0.02	0.12	0.02	0.04	-0.96	0.22	1.54	0.16
Panel D: $3FF_{NF}$ + WML										
Small	-0.26	0.01	0.10	0.11	0.23	-1.74	0.05	0.98	1.04	2.40
2	-0.06	0.06	0.00	0.09	0.10	-0.62	0.60	0.05	1.13	1.25
3	0.05	0.00	-0.03	-0.08	0.09	0.42	-0.04	-0.36	-0.84	0.90
4	0.18	0.00	0.02	-0.07	0.01	1.57	0.01	0.21	-0.81	0.07
Big	0.11	-0.05	-0.02	0.03	0.00	1.46	-0.66	-0.20	0.29	0.02
Panel E: $3FF_{NF}$+ SMBfin + WML										
Small	-0.39	-0.10	0.00	0.01	0.14	-3.46	-1.08	0.02	0.06	1.80
2	-0.15	-0.03	-0.05	0.04	0.07	-1.48	-0.30	-0.60	0.53	0.96
3	-0.02	-0.03	-0.06	-0.11	0.09	-0.21	-0.31	-0.66	-1.08	0.92
4	0.13	-0.03	0.00	-0.06	0.04	1.27	-0.32	0.05	-0.72	0.36
Big	0.06	-0.08	-0.01	0.11	0.12	0.78	-1.08	-0.08	1.19	0.88

Table 2.8: **(Size, B/M)-portfolios.**

(Size, B/M)-portfolios in different pricing models, the left panel contains the intercepts, the right panel NW *t*-values.

pricing abilities for big and value portfolios compared with the momentum factor. Panel E reports regressions with the $3FF_{NF}$ factor model augmented with a momentum factor and our financial risk factor. The risk-adjusted returns are on average comparable with the models in Panel C and Panel D. Nevertheless, the size and value patterns are non-existent in this model. To summarize, our financial risk factor improves upon the momentum factor in explaining the value pattern in big stocks, and the size pattern in value stocks. Moreover, it has a clear economic interpretation. The poorly explained returns of tiny stocks could be explained by liquidity issues. Other explanations involve time variation in risk loadings, nonlinearities or other misspecifications for tiny stocks.

2.6.2 Size-momentum portfolios

Panel A in Table 2.9 shows the intercepts and Newey-West t-values of the 5×5 size-momentum portfolios on the three-factor model. The portfolio formation on momentum controls for annual changes in momentum. As it does not control for the substantial time variation in HML loadings, we observe a large magnitude of the r-a returns. Removing the financials from the SMB and HML portfolio reduces most pricing errors somewhat, though most pricing errors are still significant (details are available upon request).

In Panel B in Table 2.9 we add our SMBfin factor to the $3FF_{NF}$ factors. The regression results show that the pricing of the winner and loser portfolios improve. This suggests that our SMBfin factor explains the anomaly to some extent. In addition, none of the portfolios with big firms earns a significant risk-adjusted return and the spread in big firms decreases substantially. This confirms the result in Table 2.8 that the SMBfin factor is good at explaining returns of big stocks. Nonetheless, the bias in the pricing of the loser and winner portfolios is smaller with a momentum factor instead of the SMBfin factor (Panel B and C in Table 2.9). Remarkably, including the momentum factor reverses the momentum pattern in big stocks as shown in Panel C in Table 2.9. Referring to Panel B and Panel C, our price factor reduces the momentum pattern in big stocks and the size pattern in loser and winner stocks. That being said, the spreads are still a significant 0.58%, 0.57% and 0.94%, respectively. Furthermore, the momentum factor does a better job in terms of a smaller average pricing error.

	a					$t(a)$				
	L	2	3	4	W	L	2	3	4	W
Panel A: 3FF										
Small	-0.96	-0.20	0.04	0.59	1.39	-5.50	-1.80	0.48	5.74	5.40
2	-0.98	-0.30	0.05	0.33	1.04	-4.65	-3.14	0.69	3.58	6.10
3	-0.66	-0.31	-0.02	0.23	0.67	-3.23	-3.15	-0.20	2.24	3.90
4	-0.66	-0.20	-0.02	0.22	0.66	-2.91	-1.80	-0.18	2.02	4.29
Big	-0.63	-0.25	0.08	0.21	0.49	-3.05	-2.23	1.29	2.06	2.26
Panel B: $3FF_{NF}$ + SMBfin										
Small	-0.87	-0.23	-0.02	0.46	1.22	-4.32	-2.13	-0.24	4.71	4.58
2	-0.73	-0.25	0.01	0.24	0.94	-3.03	-2.38	0.18	2.43	4.86
3	-0.44	-0.21	0.00	0.16	0.59	-1.96	-1.85	0.03	1.66	3.06
4	-0.38	-0.11	0.02	0.17	0.50	-1.52	-0.96	0.19	1.51	2.96
Big	-0.30	-0.11	0.08	0.08	0.28	-1.27	-0.83	1.15	0.74	1.33
Panel C: $3FF_{NF}$ + WML										
Small	-0.24	0.04	0.07	0.42	0.94	-1.94	0.33	0.67	3.68	4.62
2	-0.12	0.01	0.11	0.15	0.58	-1.13	0.06	1.25	1.56	4.28
3	0.19	0.03	0.04	-0.02	0.11	1.44	0.31	0.44	-0.16	0.72
4	0.25	0.15	0.02	-0.04	0.04	2.04	1.88	0.17	-0.39	0.31
Big	0.38	0.13	0.04	-0.25	-0.33	3.09	1.06	0.44	-3.05	-2.31

Table 2.9: **(Size, Mom)-portfolios.**

(Size, Mom)-portfolios in different pricing models, the left panel contains intercepts, the right panel NW t-values. L and W refer to Loser and Winner portfolio, respectively.

2.6.3 Model evaluation

Following Fama and French (2012), Table 2.10 reports the pricing errors with the F-test of Gibbons et al. (GRS 1989) and statistics of the regression intercepts that help us interpret the GRS test. The additional statistics include the average absolute intercept $|a|$ of the 25 regressions, the average of the 25 regression adjusted R^2, the average of the standard errors $s(a)$ of the intercepts, and $SR(a)$ which is loosely speaking a Sharp ratio: $SR(a) = (a^T S^{-1} a)^{1/2}$ where a is the column vector of the 25 regression intercepts produced by a model when applied to the 25 portfolios, and S is the covariance matrix of regression residuals.

The absolute intercept $|a|$ and $s(a)$ represent the magnitude and the precision of the intercepts, respectively. The $SR(a)$ statistic aims to correct the magnitude of the intercepts by their precision. It is the core of the GRS statistic and recommended in Lewellen et al. (2010). GRS (1989) show that this statistic is the maximum Sharpe ratio for excess returns on portfolios of the LHS assets constructed to have zero slopes on the RHS returns.

The GRS test cannot reject any of the four considered (Size, B/M)-models. The model with financials in the 3FF factors and our SMBfin factor outperforms the other models in terms of the GRS statistic. Accordingly, removing the financials from the SMB and HML factors only improves on the pricing of financials (Table 2.4) and big stocks (Panel B in Table 2.8). Obviously, portfolios with big stocks contain more financials.

However, any time variation in the HML loading of the explained LHS portfolios remains unaddressed in these setups. This is problematic as we find large tilts in HML loadings of big financials. Indeed, pricing errors are large in this static model with constant factor loadings. As a consequence, the GRS test rejects all models with (Size, Mom)-portfolios. This indicates that time variation in HML loadings is a key property of stock returns because the (Size, Mom)-portfolios do not control for time variation of loadings along the HML dimension. Noteworthy, by controlling for momentum, the momentum factor has superior explanatory power compared to the SMBfin factor.

In summary, our SMBfin factor improves on the pricing of financials, and the pricing of big and value stocks of stock portfolios that include non-financials. The removal of financials from the 3FF factors tends to increase average pricing errors of other stock portfolios. The pricing of momentum portfolios is poor, which we explain by the lack of control for time variation in HML loadings.

	Standard 3FF					3FF without financials				
	GRS	$\|a\|$	adj.R^2	$s(a)$	$SR(a)$	GRS	$\|a\|$	adj.R^2	$s(a)$	$SR(a)$
5×5 (*Size*, *B/M*)										
3FF	1.19	0.09	0.93	0.09	0.38	1.19	0.09	0.91	0.09	0.38
3FF + SMBfin	0.97	0.07	0.93	0.09	0.36	1.31	0.08	0.92	0.09	0.41
3FF + Mom	1.05	0.07	0.93	0.09	0.37	1.06	0.07	0.91	0.10	0.37
3FF + Mom + SMBfin	0.91	0.06	0.93	0.09	0.35	1.22	0.08	0.92	0.09	0.40
5×5 (*Size*, *Mom*)										
3FF	4.75**	0.45	0.84	0.13	0.75	4.42**	0.44	0.83	0.14	0.73
3FF + SMBfin	3.82**	0.35	0.85	0.13	0.71	3.79**	0.34	0.84	0.13	0.69
3FF + Mom	3.48**	0.18	0.92	0.10	0.68	3.31**	0.19	0.90	0.10	0.66
3FF + Mom + SMBfin	3.14**	0.19	0.92	0.10	0.66	3.11**	0.18	0.91	0.10	0.65

Table 2.10: **Model evaluation.**

The GRS statistic tests whether all intercepts in a set of 5×5 regressions are zero, |a| is the average absolute intercept; adj. R^2 is the average adjusted R^2, $s(a)$ is the average standard error of the intercepts, and $SR(a)$ is the Sharpe ratio for the intercepts.

*, ** denotes significant at the 5% and 1% level, respectively.

2.7 Conclusion

Our study shows that big financials in Europe have significantly lower risk-adjusted returns than other European financials. This pattern is not observed in any other industry. Since government guarantees, either explicitly or implicitly, protect shareholders of big financials from tail events, we conjecture that the lower returns of big financials are similar to an insurance premium. The higher observed volatility of big financials is in line with a moral hazard explanation where the safety net induces big financials to invest in more risky assets.

Our size-dependent financial risk factor measures potential tail risk in the financial industry, and thereby the resulting protection from government guarantees. In contrast to the standard SMB risk factor in the 3FF model, our financial SMB risk factor can explain the size pattern in European financial stocks. In addition, this factor captures some of the time variation in HML loadings of financials.

In a model with our financial risk factor and the 3FF factors without financials, we find that particularly big financials in Southern Europe were much more distressed after the start of the European debt crisis in late 2009. This suggests that the crisis increased doubts on the safety net for big financials in Southern Europe. The expected guarantee, as indicated by the financial risk factor loading, increases in the size of financials as measured by common equity. Interbank lending raised the expected guarantee only before the crisis, afterwards government securities are a significant determinant. The unconditional expected guarantee decreases in the state of the national economy. This suggests that the state of the domestic economy has more impact on default probabilities than on bailout probabilities.

Removing financials from the 3FF factors explains some mispricing of portfolios with big stocks by the presence of financials. Still, we need to control for tilts in the HML loading. Our factor captures this time variation and is able to explain some of the anomalies in the European model in Fama and French (2012). Though the momentum portfolio partly captures the guarantee in an indirect way, our factor explains the anomalies more intuitively. Our factor is particularly successful in pricing financials, big stocks, and value stocks.

Appendix

2.A Descriptives

Table 2.11 reports for each country the number of stocks by industry category. The largest number of stocks are traded in the United Kingdom (UK). France (FRA), Germany (DEU) and Sweden (SWE) follow. The order for financial stocks is UK, Germany, Italy (ITA) and France.

Descriptive statistics of several risk factors are in Table 2.12. The excess market returns are positive though insignificant. Following FF (2012) and the U.S. results of FF (1993) and Zeng et al. (2014), we find no size premium. The average SMB return is also close to zero for growth (time-varying bottom 30% B/M) and value (time-varying top 30% B/M) stocks. As argued in FF (1993), a small size premium does not mean that the size risk factor is irrelevant in asset pricing tests. The substantial volatility makes this risk factor a good candidate to predict patterns in the cross-section of average returns.

The return on the HML portfolio has a monthly return of 0.38% per month ($t = 1.78$). As in FF (2012) and in the U.S. results of FF (1993), and Loughran (1997), value premiums are higher for small stocks (bottom 10% market cap). Like FF (2012) and Asness et al. (2013), we find a significant momentum return WML, in our case 0.94% ($t = 3.20$). There is also size dependence in momentum returns. Small stocks have a significant momentum return, whilst big stocks (top 90% market cap) have not. An explanation is that liquidity has a positive relation with size and a negative relation with momentum.

The correlation matrix in Table 2.13 shows that higher excess market returns are associated with a higher return on big stocks than on small stocks. A straightforward explanation is that RM is a value-weighted return. The momentum return tends to be higher when the excess market return and the value premium are lower. That is, monthly returns exhibit more autocorrelation during periods of distress.

	Basic materials	Industrials	Consumer goods	Health care	Consumer services	Telecommunications	Utilities	Financials excl.REITs	Technology	Oil & Gas	Total
AUT	16	54	32	1	14	2	3	34	6	2	164
BEL	24	53	33	18	29	2	9	50	31	1	250
CHE	23	130	60	48	56	1	21	100	29	3	471
DEN	7	104	59	25	31	3	6	78	23	3	339
DEU	81	263	162	66	131	22	36	206	213	48	1,228
ESP	26	45	30	10	33	4	18	38	5	7	216
FIN	18	66	28	11	32	6	2	23	30	2	218
FRA	61	332	245	81	240	21	25	145	226	28	1,404
GRE	42	92	119	12	62	4	6	42	41	3	423
IRL	10	18	20	5	17	3	1	13	6	6	99
ITA	35	151	109	17	63	18	31	147	32	9	612
NLD	15	85	41	9	33	7	1	25	38	9	263
NOR	24	133	44	17	34	4	4	53	54	106	473
POR	16	34	22	1	28	5	3	25	5	1	140
SWE	52	240	79	86	87	13	11	74	119	22	783
UK	302	953	383	228	802	63	69	468	391	213	3,872
Total	752	2,753	1,466	635	1,692	178	246	1,521	1,249	463	10,955

Table 2.11: **Descriptive statistics by country and by industry.**
Number of stocks for each of the 16 countries and for each of the 10 Datastream ICB industries.

	mean (%)	std (%)	N-W t-stat
RM–RF	0.38	4.52	1.12
SMB	-0.07	2.31	-0.50
SMB_G	-0.18	2.91	-1.00
SMB_V	0.09	2.71	0.60
HML	0.38	2.45	1.78
HML_S	0.52	2.73	2.03
HML_B	0.25	2.98	1.17
WML	0.94	4.20	3.20
WML_S	1.36	3.88	4.66
WML_B	0.52	4.95	1.62

Table 2.12: **Descriptive statistics of several risk factors.**
RM–RF is the market return in excess of the risk-free rate, SMB is the return of the small minus big portfolio, HML is the return of the high minus low book-to-market portfolio, and WML is the return of the winner minus loser portfolio. The subscripts G, V, S and B denote growth, value, small and big, respectively. Returns are on a monthly basis.

	correlation		
	RM–RF	SMB	HML
SMB	-0.22**		
HML	0.09	-0.08	
WML	-0.32**	0.08	-0.29**

Table 2.13: **Correlations.**
Correlation coefficients of some risk factors in Table 2.12.
*, ** = significant at 5%, 1%.

2.B Proof equivalence

We show that the t-statistic on the significance of the mean difference between the risk-adjusted returns in Table 2.1 and Table 2.3 (or Table 2.5) is identical for all return series y.

Let 1 and I denote an all-ones vector and the identity matrix, respectively, both of appropriate dimension. The matrices $X = \begin{bmatrix} 1 & X_{FF} \end{bmatrix}$, and $\tilde{X} = \begin{bmatrix} 1 & X_{FF} & v \end{bmatrix}$ represent the nonsingular regression matrices of Table 2.1 and Table 2.3, respectively. Decompose the return series y into three components:

$$y = P\tilde{P}y + (I - P)\tilde{P}y + \left(I - \tilde{P}\right)y$$

where $P = X\left(X^TX\right)^{-1}X^T$ and $\tilde{P} = \tilde{X}\left(\tilde{X}^T\tilde{X}\right)^{-1}\tilde{X}^T$ are the projection matrices on $W = \text{Span}(X)$ and $\tilde{W} = \text{Span}\left(\tilde{X}\right)$, respectively. Because $W \subseteq \tilde{W}$, we have for the first component of y,

$$y_A := P\tilde{P}y = Py \in W$$

For the second component, it follows that

$$y_B := (I - P)\tilde{P}y = (\tilde{P} - P)y \in W^\perp \cap \tilde{W}$$

Using $\dim\left(W^\perp \bigcap \tilde{W}\right) = 1$ gives $y_B = \lambda_y(\tilde{P} - P)v$ with coefficient

$$\lambda_y = \frac{y^T(\tilde{P} - P)^T(\tilde{P} - P)v}{\left|y^T(\tilde{P} - P)^T(\tilde{P} - P)v\right|}\frac{\left\|(\tilde{P} - P)y\right\|}{\left\|(\tilde{P} - P)v\right\|} = \frac{y^T(\tilde{P} - P)v}{\left|y^T(\tilde{P} - P)v\right|}\sqrt{\frac{y^T(\tilde{P} - P)y}{v^T(\tilde{P} - P)v}}.$$

The first and second term ensure that λ_y has the correct sign, and the correct length, respectively. For the third component of y,

$$y_C := (I - \tilde{P})y \in \tilde{W}^\perp \subseteq W^\perp$$

thus

$$X^Ty_C = \tilde{X}^Ty_C = 0 \qquad (2.2)$$

Define

$$X_{(0)} = \begin{bmatrix} 0 & X_{FF} \end{bmatrix} \qquad Q = X_{(0)}\left(X^TX\right)^{-1}X^T \quad R = Q - \tilde{Q}$$
$$\tilde{X}_{(0)} = \begin{bmatrix} 0 & X_{FF} & v \end{bmatrix} \quad \tilde{Q} = \tilde{X}_{(0)}\left(\tilde{X}^T\tilde{X}\right)^{-1}\tilde{X}^T$$

The risk-adjusted return is the intercept plus the unexplained return. In other words, the risk-adjusted return is the return minus the correction for risk. This correction is the projection on X and $\tilde{X}$ exclusive the projection on the intercept term. That is, $e = (I - Q)y$ for the risk-adjusted returns in Table 2.1 and $\tilde{e} = (I - \tilde{Q})y$ for the risk-adjusted returns in Table 2.3 (or Table 2.5). Consequently, the series of differences in risk-adjusted returns equals $d := \tilde{e} - e = Ry$.

Since $\tilde{X} = \begin{bmatrix} X & v \end{bmatrix}$ and $(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T \begin{bmatrix} X & v \end{bmatrix} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{X} = I$,

$$(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T X = \begin{bmatrix} I & 0 \end{bmatrix}^T \qquad X^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1} = \begin{bmatrix} I & 0 \end{bmatrix}$$

where 0 is a zero vector of appropriate dimension. This implies that the columns in X are orthogonal to the rows of R:

$$RX = (Q - \tilde{Q})X = X_{(0)} - \tilde{X}_{(0)} \left(\tilde{X}^T\tilde{X}\right)^{-1} \tilde{X}^T X = X_{(0)} - \tilde{X}_{(0)} \begin{bmatrix} I & 0 \end{bmatrix}^T = 0$$

It now follows from $y_A \in W$ that $Ry_A = 0$. By (2.2), $Qy_C = \tilde{Q}y_C = 0$ holds such that $Ry_C = (Q - \tilde{Q})y_C = 0$. Using

$$QP = X_{(0)} \left(X^T X\right)^{-1} X^T X \left(X^T X\right)^{-1} X^T = Q$$

$$Q\tilde{P} = X_{(0)} \left(X^T X\right)^{-1} X^T \tilde{X} \left(\tilde{X}^T\tilde{X}\right)^{-1} \tilde{X}^T = Q$$

$$\tilde{Q}P = \tilde{X}_{(0)} \left(\tilde{X}^T\tilde{X}\right)^{-1} \tilde{X}^T X \left(X^T X\right)^{-1} X^T = \tilde{X}_{(0)} \begin{bmatrix} I & 0 \end{bmatrix}^T \left(X^T X\right)^{-1} X^T = Q$$

$$\tilde{Q}\tilde{P} = \tilde{X}_{(0)} \left(\tilde{X}^T\tilde{X}\right)^{-1} \tilde{X}^T \tilde{X} \left(\tilde{X}^T\tilde{X}\right)^{-1} \tilde{X}^T = \tilde{Q}$$

we obtain

$$R(\tilde{P} - P) = (Q - \tilde{Q})(\tilde{P} - P) = Q - \tilde{Q} = R$$

Combining the previous results gives

$$d = Ry = Ry_B = \lambda_y R(\tilde{P} - P)v = \lambda_y Rv,$$

which means that λ_y is a linear scalar for the mean of the elements of d as well as for the standard deviation of d. Hence, the t-statistic of d does not depend on λ_y and so the return series y. Instead, it is a measure for the excess return series Rv of the additional regressor v in $\tilde{X}$.

We have proved our result when risk-adjusted returns include an intercept term, just as in our research. It is straightforward to extend this result to a more general setting by replacing the regressor 1 in X and $\tilde{X}$ by an arbitrary (possibly empty) set of linearly independent regressors.

Chapter 3

Tail dependence: A cross-industry comparison

Joint work with Michiel Bijlsma.

This Chapter is published as S.Muns and M.Bijlsma (2015), Tail Dependence: A Cross-Industry Comparison, Journal of Portfolio Management 41*, volume 3, pp.109–116,*

3.1 Introduction

Traditional risk management practices typically assume that stock returns follow a normal distribution and that dependencies are constant over the business cycle. This approach underestimates both the frequency and the severity of a downturn and also underestimates the dependencies during a downturn. A measure of tail dependence helps investors stress-test their portfolios and adjust their positions to reduce their tail risk exposure.

Investors often allocate resources to different industries and select stocks within each industry based on some risk-return preference. This motivates us to analyze tail dependence within the 48 Fama and French (1997) industries.[1] Our analysis shows substantial cross-industry differences in diversification benefits of tail risks. We identify determinants of tail risk, and provide empirical evidence that historical tail dependence helps investors to predict which industries exhibit most tail dependence.

Our tail dependence measure is based on Hartmann et al. (2004), Poon et al. (2004) and Segoviano and Goodhart (2009). Roughly speaking, it is

[1] The seven smallest industries are too small, and we omit the industry with unclassified firms.

the number of simultaneous tail returns within a particular industry used in other studies. As such, it indicates the extent to which within-industry diversification does not reduce tail risk.

This measure has some advantages over other measures of tail risk. It is straightforward to calculate as it is constructed by setting a stock specific threshold equal to the stock's daily Value-at-Risk (VaR) and then simply counting the number of stocks with returns exceeding the VaR. We believe this measure is more intuitive than alternatives such as principal component decompositions or copulas. In addition, our measure looks solely at tail dependence, whereas other measures are also affected by regular non-tail dependencies.

Our work relates most directly to papers that focus on systemic risk in different industries.[2] Bühler and Prokopczuk (2010) compare systemic risk in twelve industries by estimating a copula that allows for tail dependence using the stock prices of the five most important firms within each industry. In their article, they define tail dependence as the probability that these five firms simultaneously have a tail return. Their results suggest that tail dependence is significantly larger in the banking industry. Kinlaw et al. (2012) compare systemic risk in ten industries with a principal components approach. They find that firms in finance, energy and technology have the most systemic risk.

Our research extends on the existing literature in a number of ways. First, we compare tail dependence across industries with a relatively simple and intuitive measure. Second, we assess more industries and use a larger sample of firms in each industry. This detects tail dependence that might be missed for more broadly defined industries. Third, we analyze the determinants of tail dependence in a panel regression. Fourth, we measure tail dependence in residual returns from the capital asset pricing model (CAPM). This measure is of particular relevance for an investor hedging his exposure to the market. Finally, we study time variation in tail dependence by using annual samples.

Though the industry ranking varies over time, we find that the banking and petroleum & natural gas (oil) industries exhibit the highest level of tail dependence. Utilities, financial trading and insurance show somewhat less tail dependence than banking and oil, but more than other industries. We document an overall increase in tail dependence, which implies that hedging tail risk has become more important. Fortunately, the shared exposure to the market return explains an increasing part of this tail dependence. This

[2] De Bandt and Hartmann (2000) and Kritzman et al. (2011) provide a more general overview of the literature on systemic risk.

makes hedging the market return a more effective strategy. We also find that this year's tail dependence is a significant predictor for the next year's tail dependence.

Our results on the determinants of tail dependence suggest that an investor targeting some VaR needs a more diversified portfolio for industries where firms have (*i*) a large market beta (i.e. electronic equipment, machinery), (*ii*) a large market cap (banks, oil) or (*iii*) less volatility (utils, financials). Alternatively, we could interpret our results as indicating which industries have a higher VaR because it is more costly to reduce tail risk through diversification.

3.2 Methodology

We measure tail dependence of a portfolio of selected stocks by the expected fraction of additional failing firms in the portfolio given that at least one firm fails (EAF):

$$EAF := E\left[\frac{\kappa-1}{n_0-1}\,\middle|\,\kappa \geq 1\right] = \frac{\sum\limits_{t\in\tau}\left([\#\text{selected firms that fail in period } t]-1\right)}{(n_0-1)\,|\tau\,|} \tag{3.1}$$

where κ represents the number of failing firms, n_0 is the number of firms considered in each period, τ is the set of periods where at least one selected firm fails, and $|\tau|$ is the number of periods in this set. We define firm i to be a failing firm when its daily stock return R_i is below some negative threshold level q_i. For a given percentage p, the threshold level is the $p\%$ lower quantile of daily stock returns of firm i. Thus, for $p = 2\%$, a firm's return is below its failure level for approximately five days per year. We measure EAF by computing the mean additional fraction of selected firms that fail in the periods where at least one selected firm failed and will often express it as a percentage rather than as a fraction.

Our measure is closely related to the fragility index proposed in Huang (1992), and applied in Hartmann et al. (2004), Poon et al. (2004), De Vries (2005) and Segoviano and Goodhart (2009), defined as the expected number of failing firms given that at least one firm is failing. EAF is a transformation of the fragility index such that it is always between zero and one.

We select the 20 largest firms in each industry. It is appropriate to focus on tail dependence of large stocks because failures of larger firms have a larger impact on the real economy. A measure that includes all stocks would be biased towards the characteristics of the smaller, less liquid, stocks. For each day we select the largest 20 firms in an industry as measured by market capitalization to mitigate potential distortions due to a varying number of firms and survivorship bias. This improves upon the common approach in the literature where a fixed set of firms is considered.[3] We use the market capitalization with a lag of one day because a firm's market capitalization based on the closing price is affected by an extreme negative return on that day, which may exclude some extreme negative returns.

To identify determinants of tail risk, we calculate EAF in two different ways, which differ in the returns used for identifying tail events. Our first measure, the baseline EAF, uses stock returns.

It turns out that market risk contributes to tail dependence. For this reason, in our second tail risk measure, EAF_{res}, we use abnormal returns for each firm i, instead of stock returns to determine tail events. More specifically, we use the expected number of simultaneous abnormal tail returns. The abnormal returns are defined as the residuals of the well-known CAPM regression

$$R_{i,t} = \alpha_i + \beta_{i,t} R_{M,t} + \varepsilon_{i,t}. \tag{3.2}$$

Here, $\beta_{i,t}$ captures the correlation of stock returns $R_{i,t}$ with the market return $R_{M,t}$ and α_i captures firm-specific omitted variables. The residuals $\varepsilon_{i,t}$ filter out tail events unrelated with the market. An abnormal tail return is a tail return below the stock specific $p\%$ quantile of abnormal returns. We use monthly values for β_i by estimating market sensitivity over a five year rolling window. Each time window provides the beta for each firm i for the next month. The difference $DIF = EAF - EAF_{\text{res}}$ measures the impact of the market return on tail dependence.

3.3 Data

We employ U.S. equity data and the value-weighted market index from the Center for Research for Security Prices (CRSP) over the period 1988-2011.[4] We classify the securities according to the 48 industries used in Fama and French (1997). We drop the industry with unclassified firms, plus seven industries that have less than 20 securities for at least one period.[5] This

[3] Our results are robust to using a constant sample of the twenty largest firms listed over the sample period.

[4] The returns for the period 1988-1992 are only used to estimate the betas for the first years of the sample 1993-2011.

[5] The neglected industries are agriculture, candy & soda, tobacco products, shipbuilding & railroad equipment, defense, coal, and shipping containers.

leaves 40 industries.

Panel A on the left in Table 3.1 presents descriptive statistics for the daily stock returns of these industries, sorted by their mean pair-wise correlation ρ_{20}. This correlation is largest in the banking industry (Banks), followed by the petroleum & natural gas (Oil), utilities (Util), financial trading (FinTr), and insurance industry (Insur). These five industries tend to have a smaller-than-average variance σ^2, indicating lower idiosyncratic risk. In each industry, the average return is small and positive. Skewness γ is present in many industries and is particularly negative in the petroleum & natural gas industry.

The cross-sectional mean $\bar{\beta}$ measures the correlation with the market and partly captures exposure to macro-economic shocks and indirect exposure due to feedback effects involving the real economy. The five industries with a high intra-industry correlation have a relatively low correlation with the market. Their $\bar{\beta}$ is comparable with non-cyclical industries such as food products (Food), precious metals (Gold) and beer & liquor (Beer). Comparing $\bar{\beta}_{20}$ and $\bar{\beta}$, we see that the twenty largest firms are often more sensitive to the market return than the average firm. This is intuitive as larger firms tend to be less locally oriented than their smaller counterparts.

The market sensitivity is much larger for the large financials (see $\bar{\beta}_{20}$). Still, the twenty largest banks, insurers and financial traders are less sensitive to the market return than many other industries, including the cyclical industries for business services (BusSv), computers (Comps), and electronic and electrical equipment (Chips and ElcEq). This suggests that the larger correlation ρ_{20} is not explained by a larger exposure to the market return.

The size of the twenty largest banks (MC_{20}) exceeds the industry average (MC) by a larger multiple than in any other industry. One explanation here is the larger number of firms in the banking industry. The time-varying sample of the twenty largest firms tends to be more stable over time for industries where the largest firms are less sensitive to the market return. For instance, only 33 firms in the non-cyclical beer & liquor industry are in the top 20 largest firms for at least one day, while this number is 112 for the cyclical business services industry.

3.4 Results

In this section, we present and discuss our results. First, we measure tail dependence for each industry and show that it provides additional information to investors over straightforward correlation measures. Subsequently,

Table 3.1: **Descriptive statistics for the period 1993-2011.**

ρ_{20} = time average of the mean pairwise correlation of the time-varying twenty largest firms, μ = mean daily return in hundreds of basis points (bp), σ^2 = variance daily log return in bp^2, γ = cross-sectional mean skewness daily log returns, $\bar{\beta}$ = cross-sectional mean of the firm-specific mean β over rolling estimation windows of 5 years, N = number of firms, N_{20} = number of firms in time-varying top 20, MC = cross-sectional mean of mean market cap for each firm, MC_{20} and $\bar{\beta}_{20}$ refer to the 20 firms with the largest maximal market cap, the five rightmost columns list EAF for different values of p.

	Panel A										Panel B				
											EAF				
Industry	ρ_{20}	μ	σ^2	γ	$\bar{\beta}$	$\bar{\beta}_{20}$	N	N_{20}	MC	MC_{20}	0.1%	0.5%	1%	5%	10%
Banks	0.52	7.6	12.4	0.19	0.6	1.2	1,561	66	0.5	25.9	7.0	9.8	10.0	15.7	19.5
Oil	0.51	12.8	26.0	-0.92	0.9	0.7	448	78	1.3	24.0	8.7	9.2	9.7	13.8	18.2
Util	0.47	5.7	6.7	0.08	0.6	0.5	256	66	1.3	8.5	3.8	6.6	7.8	13.4	17.0
FinTr	0.43	11.9	18.7	0.17	0.8	1.5	757	90	0.5	13.1	4.6	6.5	7.6	12.6	16.4
Insur	0.40	7.6	12.9	0.40	0.7	1.0	362	61	1.3	16.7	2.3	5.6	7.3	11.2	14.7
Chips	0.36	11.9	29.5	0.05	1.7	2.2	612	90	0.9	18.5	1.0	3.1	4.6	10.2	14.8
Chems	0.36	8.8	20.5	-0.41	1.0	1.1	182	72	0.9	6.6	2.7	4.0	4.8	9.0	13.0
Mach	0.36	10.5	21.3	0.43	1.1	1.3	322	75	1.2	16.3	3.0	4.8	5.5	9.8	13.7
Autos	0.33	6.3	19.0	0.12	1.1	1.3	149	70	0.8	4.5	1.9	3.9	4.4	8.7	12.2
Cnstr	0.33	10.4	27.3	0.21	1.1	1.3	138	75	0.2	1.2	1.9	3.8	5.2	8.8	12.8
Rtail	0.33	7.9	24.2	-0.08	1.1	0.9	640	67	0.9	19.1	1.1	3.2	4.2	8.7	12.6
Paper	0.32	6.6	15.2	0.54	0.9	1.1	108	59	0.7	3.2	1.6	3.8	4.7	8.6	12.6
Steel	0.31	6.9	18.0	0.03	1.3	1.5	139	74	0.5	3.1	2.5	4.4	5.1	9.3	13.4
Trans	0.31	7.5	20.7	0.47	1.0	1.0	252	75	0.6	5.5	2.8	3.5	4.7	8.3	12.1
Telcm	0.31	7.8	31.0	0.06	1.6	1.2	539	98	1.0	18.6	1.9	3.5	4.7	9.2	12.8
BusSv	0.31	10.7	35.0	0.04	1.5	2.2	2,028	112	0.4	19.7	1.1	2.5	3.7	8.3	12.7
Comps	0.29	12.0	34.6	-0.01	1.6	1.9	430	88	1.0	17.7	0.2	1.7	3.1	8.5	13.0
Drugs	0.28	10.2	31.5	0.06	1.5	0.9	620	70	1.2	31.8	0.7	3.0	3.9	8.0	12.0
Meals	0.27	8.5	23.7	-0.09	0.9	1.2	287	83	0.4	4.4	1.0	2.9	3.8	7.6	11.4
Books	0.26	7.6	21.6	0.12	1.0	1.1	115	70	0.6	2.7	1.2	3.4	4.3	7.9	11.5
LabEq	0.25	14.4	29.8	0.00	1.2	1.4	226	77	0.3	2.9	1.1	2.3	3.4	8.0	12.2
Hshld	0.23	7.5	21.4	0.60	1.0	1.0	162	65	1.2	9.0	1.7	3.1	3.8	7.1	10.8
Clths	0.23	8.7	25.5	0.02	0.8	1.3	116	65	0.3	1.6	0.5	1.9	2.6	6.6	10.8
BldMt	0.23	9.1	19.1	-0.25	0.9	1.0	174	76	0.3	2.0	1.5	2.9	3.6	7.1	11.1
MedEq	0.23	11.2	29.7	-0.25	1.1	0.9	443	76	0.4	7.5	0.5	2.1	2.9	6.9	10.8
Whlsl	0.21	10.2	28.7	-0.32	1.1	0.9	526	90	0.2	3.5	1.3	2.4	2.9	6.8	10.7
Food	0.21	9.2	19.5	-0.29	0.7	0.6	165	65	0.8	5.7	1.0	2.0	2.9	6.1	9.9
Aero	0.21	10.4	16.3	-0.19	0.8	1.0	51	50	2.0	4.9	1.8	3.3	3.9	6.9	10.2
ElcEq	0.21	11.5	32.9	0.19	1.5	1.8	307	111	0.2	1.8	1.3	2.3	3.2	7.0	11.2
Hlth	0.20	10.4	30.7	-0.26	1.1	0.9	355	113	0.2	2.4	0.8	2.4	3.0	7.2	11.0
Fun	0.20	11.9	39.2	-0.02	1.0	1.4	233	96	0.6	6.0	1.5	2.6	3.1	6.1	9.9
Gold	0.18	20.6	55.2	0.81	0.5	0.7	56	51	0.4	1.1	0.8	2.5	3.3	6.0	9.2
PerSv	0.17	9.9	27.8	0.10	1.1	1.1	177	95	0.2	1.4	0.3	1.8	2.3	5.7	9.5
Toys	0.13	7.9	31.4	0.16	1.0	1.4	112	77	0.2	0.8	0.7	1.4	2.1	5.2	8.5
RlEst	0.12	9.2	26.0	-0.12	0.7	1.1	98	83	0.2	0.7	1.6	2.3	2.9	5.3	8.5
Mines	0.10	9.3	31.0	0.27	1.0	1.4	52	52	0.3	0.8	0.4	1.4	1.8	4.2	8.1
Rubbr	0.07	9.8	23.4	0.55	0.9	1.3	96	68	0.2	0.7	0.3	1.0	1.2	4.1	7.5
Txtls	0.05	4.5	23.8	0.25	0.7	0.9	67	52	0.1	0.3	0.2	1.1	1.5	4.0	7.6
FabPr	0.04	11.4	21.7	-0.04	1.1	1.2	33	33	0.2	0.3	0.2	1.2	1.3	3.9	7.4
Beer	0.03	7.6	18.3	-0.21	0.6	0.7	33	33	1.0	1.6	0.3	0.7	1.0	3.7	7.3

we identify some determinants of this tail dependence. We then consider how tail dependence and the ranking of industries evolve over time. This highlights the need to continuously update tail risk mitigation strategies. Subsequently, we study out-of-sample predictions to analyze the forward-looking nature of our tail dependence measure. Finally, we focus on the five industries with highest tail dependence to further study the effects of the market return and time variation.

Panel B in Table 3.1 presents our tail dependence measure for each industry and several threshold levels. For failure levels above 0.1%, the ranking of the top five sectoral EAF from high to low is: banking, oil, utilities, financial trading, and insurance. The ranking of these five industries is the same if p is 0.5%. At this 0.5% VaR level, we find that on average 9.8% of the other 19 large banks fail on that particular day given that at least one bank is failing. For the oil industry, this number is 9.2%, for utilities 6.6%, for financial trading 6.5%, and for insurance 5.6%. For small failure probabilities (high loss thresholds) tail dependence in the banking and oil industry is substantially larger than in other industries, suggesting that the most extreme returns are more systemic and thus harder to reduce by diversification.

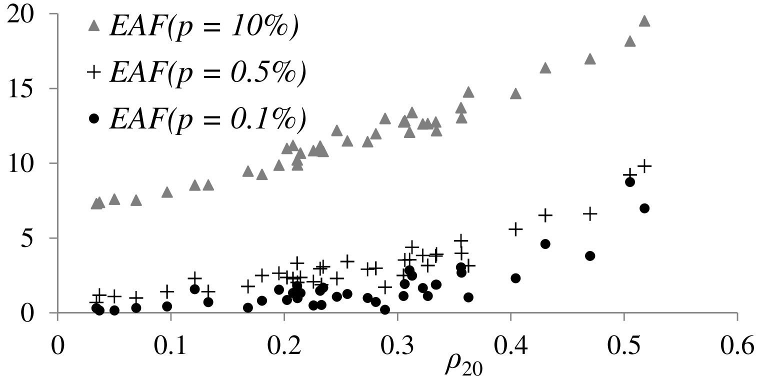
The ranking of tail dependence across industries is roughly the same as the ranking in mean correlation coefficients ρ_{20}. In other words, industries where individual stocks are highly correlated are also industries with larger tail dependence. The rank correlation between the two dependence measures exceeds 0.75. Intuitively, the relation between ρ_{20} and EAF is more linear for large values of p (small loss thresholds) as our tail measure is then less focused on the tail.

For small p, however, although the ranking of the top five industries with high tail dependence remains the same, the relation between mean correlation ρ_{20} and tail dependence EAF becomes nonlinear, as illustrated in Figure 3.1. At a high correlation, a small increase in correlation is associated with a large increase in tail dependence. This suggests that amplification effects may play a role. The mean pairwise correlation coefficient misses the nonlinear increase in tail dependence because it does not focus on the tail. Notably, the mean industry beta of the top 20 firms is insignificantly related to EAF in Table 3.1. Nevertheless, after controlling for other factors, we show that β does explain tail dependence.

An important step in mitigating tail risk is to identify some of the determinants of tail risk. We relate our tail dependence measure to the log of market capitalization, log volatility, and the market beta. Here, we define a tail event as the three smallest returns in each year for each stock ($p = 1.2\%$).

Figure 3.1: **Correlation and tail dependence**

EAF for the failure probabilities $p = 0.1\%, 0.5\%, 10\%$. Data is from Table 3.1.



We estimate a fixed-effects panel with annual periods (P-values in brackets, and $R^2 = 0.70$)

$$\log(EAF_i(t)) = \alpha_i + \gamma_t + \underset{(0.0\%)}{0.11} \log(MV_i(t)) - \underset{(0.0\%)}{0.84} \log(\sigma_i(t)) + \underset{(0.0\%)}{0.51} \hat{\beta}_i(t) \tag{3.3}$$

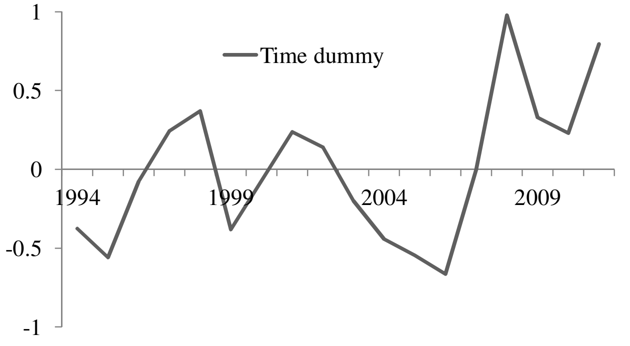
where $EAF_i(t)$ corresponds to industry i in year t, the variables are annual averages of the largest 20 firms in each industry, and $\hat{\beta}_i(t)$ is estimated in an instrumental variable panel regression up to a one year lag to account for estimation error in $\beta_i(t)$.

We find a significant positive coefficient for market cap. An explanation may be that a larger market cap makes firms less locally oriented, and so increases the overlap of tail exposures. We find a significant negative correlation for the stock volatility. This is intuitive, as a higher volatility implies more idiosyncratic risk, less dependence with other firms, and hence less joint tail risk. We also find a positive significant correlation for beta. This is reasonable, because a larger beta implies more shared dependence with the overall economy. Note that the significance of beta conflicts with our observation from Table 3.1 that the market beta and our tail dependence measure are unrelated. Apparently, the market beta becomes significantly related to tail dependence when controlling for market cap and volatility in annual samples. Figure 3.2 shows the time dummies γ_t in (3.3). The increasing pattern indicates that within industry tail dependence has increased since the financial crisis.

In order to determine whether the level of tail dependence provides in-

Figure 3.2: **Time variation in tail dependence**

Time dummies γ_t in (3.3).



formation to investors on future levels of tail dependence, we estimate the following fixed effects model in a first-difference GMM setup (Arellano and Bond (1991) robust P-value)

$$\log(EAF_i(t)) = \alpha_i + \gamma_t + \underset{(2.9\%)}{0.18} \log(EAF_i(t-1)) \tag{3.4}$$

The coefficient implies that $EAF_i(t)$ increases with a significant 0.2% if $EAF_i(t-1)$ increases with 1%. We thus find that EAF of the previous year helps investors in estimating current year's EAF.

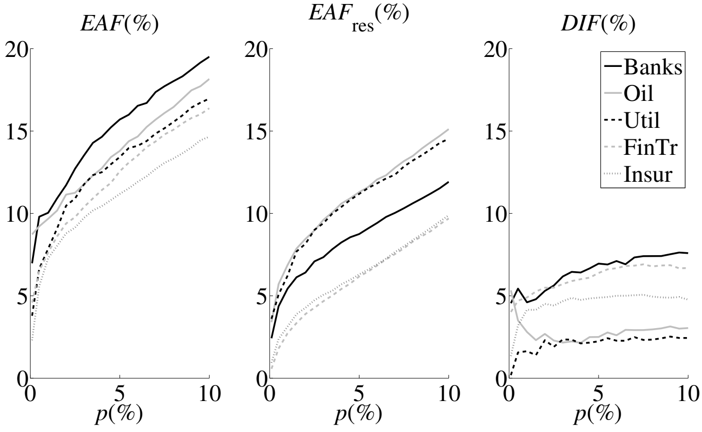
As an extra step towards identifying the determinants of tail dependence, we compare tail dependence using simple returns and abnormal returns from the CAPM (3.2). Because we are mostly interested in industries that exhibit high levels of tail dependence, we focus on the top five industries banking, oil, utilities, financial trading and insurance which have most within industry tail dependence.

The graph on the left hand in Figure 3.3 shows EAF as a function of the firm-specific probability of an extreme negative return, p. We determine the tail events by considering the whole sample period with p from 0.1% to 10% percent. As the failure probability p decreases, we focus on more extreme events and we zoom in on tail dependence. At the same time, the sampling variance increases when observed events are more scarce.

The graph in the middle of Figure 3.3 shows EAF_{res}, which is our tail dependence measure when using abnormal CAPM returns. In the right hand graph, the difference in the dependence measures, DIF, indicates that the market return explains a substantial part of the EAF in the left hand graph.

Figure 3.3: ***EAF***

EAF for simple returns (left), abnormal CAPM returns (middle), and the difference between the two measures (right). The sample period is 1993-2011.



The size of the correction differs per industry and is largest for the three financial industries, which indicates that aggregate economic shocks explain a large fraction of tail dependence within these industries.

To assess whether differences between the industries in the left hand graph in Figure 3.3 are significant, we performed a bootstrap exercise. Details are in the appendix. We find the following ordering in the case of simple returns

$$EAF^{(\text{Banks})} \gtrapprox EAF^{(\text{Oil})} > EAF^{(\text{Util})} \gtrapprox EAF^{(\text{FinTr})} \gtrapprox EAF^{(\text{Insur})}$$

The symbol $\gtrapprox$ denotes insignificantly larger than. The large tail dependence in the banking and oil industry suggests that tail risk of common risk factors (such as the market return, changes in the interest rate and changes in the oil price), or common heteroskedasticity plays a larger role in both industries. In unreported results, we find that daily changes in the oil price cannot explain $EAF^{(\text{Oil})}$.

The tail dependence is significantly lower for the abnormal returns than for the simple returns in all five industries. An investor who hedges his exposure to the market return therefore decreases his tail risk substantially. Differences remain, however, between the industries for the abnormal CAPM returns:

$$EAF_{\text{res}}^{(\text{Oil})} \gtrapprox EAF_{\text{res}}^{(\text{Util})} \gtrapprox EAF_{\text{res}}^{(\text{Banks})} > EAF_{\text{res}}^{(\text{Insur})} \gtrapprox EAF_{\text{res}}^{(\text{FinTr})}$$

Though not implied by the inequalities, the oil industry and the banking industry do differ significantly from each other. Thus we see the results in Figure 3.3 confirmed by our bootstrap analysis.

A similar bootstrap procedure on *DIF* indicates that the industry-specific decline based on abnormal returns is larger for the three financial industries:

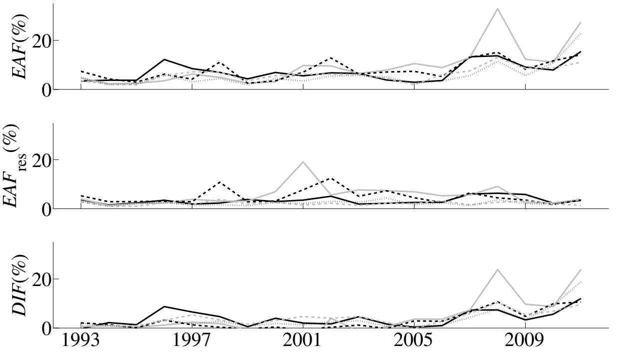
$$DIF^{(\text{FinTr})} \gtrapprox DIF^{(\text{Banks})} \gtrapprox DIF^{(\text{Insur})} \gtrapprox DIF^{(\text{Oil})} \gtrapprox DIF^{(\text{Util})}$$

This ordering is in line with the betas $\bar{\beta}_{20}$ in Table 3.1. Compared to the oil and utilities industry, tail dependence in the financial sector is highly associated with shocks in the market index, and is hence harder to diversify by investing in other industries.

To study how our measure of tail-dependence in the top five industries evolves over time, we use annual windows with 3 tail observations per stock in each year ($p = 1.2\%$). The top graph in Figure 3.4 confirms the results in Figure 3.2 which indicated that tail dependence varies over time. In addition, the order of the top five industries varies over time. The oil sector exhibits most tail dependence during the most recent half of our sample, which is in line with Kinlaw et al. (2012). The spikes in 2008 and 2011 suggest that shocks to firms in the oil industry are highly concentrated and occur on only a few days. The market faced a downturn at the same days as *DIF* shows similar spikes (see bottom graph).

Figure 3.4: **Tail dependence over time**

Tail dependence measures for annual *EAF* (top), EAF_{res} (middle), and *DIF* (bottom). Failure level $p = 1.2\%$, legend is in Figure 3.3.



In general, the increasing pattern of *DIF* in Figure 3.4 suggests that the increase in tail dependence over time is mainly attributable to a larger effect of the market return on tail dependence. That is, reducing within industry tail risk by hedging the market exposure has become more effective.

3.5 Conclusion

We measured downward tail dependence in the 48 Fama and French (1997) industries. Our measure for tail dependence is the average fraction of firms within an industry that experience an extreme negative return given that at least one firm in that industry experiences such a return on that day. Our analysis is based on a time-varying panel of the twenty largest U.S. firms in each industry.

We obtain the following results. Firstly, the banking industry and the oil industry exhibit the most tail dependence. The utility, financial trading and insurance industry exhibit somewhat less tail dependence, but in general more than the other industries. A top-down investor concerned about downside risks will find it harder to reduce his tail risk by diversification within these five industries.

Secondly, we find that (*i*) a large market beta, (*ii*) a large market cap, and (*iii*) a smaller volatility all increase within industry tail dependence. Nonetheless, an industry with a low market beta does not guarantee low tail dependence. Market risk, as estimated by the market beta, explains a relatively large part of the tail dependence in the banking, financial trading and insurance industry. This suggests that for these industries, it is more difficult to use market-wide diversification to reduce the downside risk.

Finally, tail dependence has increased over time, with the oil industry recently exhibiting the largest tail dependence. The market beta explains both observations. During the most recent years, hedging the market exposure was a more effective strategy to reduce within industry tail risk. Finally, our tail dependence measure is a valuable tool in predicting next year's tail dependence.

Appendix

3.A Appendix

Table 3.2 shows the P-values from our bootstrap exercise for the failure probability $p = 1\%$. The P-values below 5% and above 95% are significant at the 5% level of significance. The upper triangle refers to cross-industry EAF, the lower triangle to cross-industry EAF_{res}. The main diagonal compares EAF based on simple returns with EAF_{res} based on abnormal returns. Thus, it refers to the P-values for the null hypothesis $DIF = 0$.

Table 3.2: **Bootstrapped P-values for null hypothesis of $EAF^{(X)} = EAF^{(Y)}$ against the one-sided alternative $EAF^{(X)} > EAF^{(Y)}$.**

The upper triangle refers to EAF, the lower triangle to EAF_{res}, the main diagonal compares EAF with EAF_{res}. All results are based on 10,000 bootstraps of the (abnormal) returns of the time-varying sample of the largest 20 firms in each industry.

* indicates significant at the 5% level of confidence.

	Banks	Oil	Util	FinTr	Insur
Banks	0.000*	0.370	0.032*	0.001*	0.001*
Oil	0.981*	0.007*	0.045*	0.038*	0.004*
Util	0.916	0.307	0.034*	0.351	0.203
FinTr	0.000*	0.000*	0.000*	0.000*	0.358
Insur	0.000*	0.000*	0.000*	0.927	0.000*

The upper triangle shows that the banking industry has significantly more tail dependence than utilities, financial traders and insurers, but not significantly more than the oil industry.

Chapter 4

A welfare analysis on bank size, capital and asset dependence

Joint work with Chen Zhou.

4.1 Introduction

The financial crisis has forcefully shown that the transmission of adverse shocks on the financial system may have a severe effect on social welfare. Three fundamental characteristics of the financial system are crucial in this transmission mechanism. First, by their typically high leverage, the small capital ratio makes individual financial institutions highly sensitive to asset losses. Second, a similar asset balance of different institutions suggests that a single shock may hit multiple institutions simultaneously. Third, some financial institutions are extraordinarily large, and consequently their failure would impose a large shock to the system which is more severe than the failure of a group of small institutions. Thus, the stability of a financial system and the potential welfare loss once the system fails depend on three important bank characteristics: the capital structure, the asset portfolio composition and the bank size. All three characteristics are a consequence of banks' management decisions over time.

For each characteristic, banks tend to choose a level that is suboptimal for society. First, because equity financing is not tax deductible and banks' shareholders have limited liability in case of a default, bank owners prefer a lower level of capital than optimal for society. Second, after a major stress event, regulators tend to save banks ex-post to prevent a systemic banking crisis. In particular, large banks or a large number of smaller identical banks may trigger a major stress event. Banks anticipate such policies by growing large and holding similar asset portfolios ex-ante.

To deal with the negative social impact of such decisions, recent policy reforms impose higher capital ratios, stimulate specialization of banks, and implement a more careful supervision on banks that are "too-big-to-fail". In this study, we investigate the impact of capital structure, asset interdependence and size on social welfare in a theoretical model. In addition, we evaluate policy interventions directly or indirectly designed towards these characteristics.

We build an equilibrium model in which banks make management decisions on capital and size with the following considerations. First, banks determine their capital ratio by balancing the financing costs of debt and equity. Following the corporate finance literature, the tax shield is a key factor that favours debt financing, while a high level of debt corresponds to a high default probability. Second, large banks benefit from a more diversified portfolio, and thus a lower risk of assets. However, large banks may suffer from higher operational costs due to the complexity to manage such large banks. Third, similar to Ibragimov et al. (2011) banks invest in correlated assets in a myopic way. They do not take into account the adverse effect on social welfare of correlated asset portfolios.

To evaluate the impact on social welfare, we consider two welfare measures. First, we define a welfare measure as the total surplus generated by bank's loans. Second, we define social welfare by subtracting from the former welfare measure the potential cost of a systemic crisis. This cost depends on the frequency of a systemic crisis defined as the probability that some fraction of debt in the financial system is not repaid.

A key feature of our theoretical model is that the three characteristics – capital ratio, asset interdependence and size – are endogenously interrelated. Importantly, this interrelation affects systemic risk and social welfare in an ambiguous way. For example, large banks invest in a more diversified portfolio of assets. Consequently the asset dependence between two large banks tends to be higher which raises systemic risk. On the other hand, systemic risk may also be smaller, as large banks bear less individual risk due to their diversification benefits. Nonetheless, the smaller individual risk of large banks gives more room to leverage. This partly offsets the potential reduction in systemic risk due to diversification. This example demonstrates that a system consisting of a few large banks that are well diversified but also highly leveraged may or may not be more systemically risky. Indeed, an equilibrium analysis is non-trivial due to the complexity generated by the interaction between the three characteristics. Our theoretical model considers this interaction and provides a corresponding equilibrium analysis for policy evaluations.

Our main findings are as follows. First, without any regulation banks make a suboptimal decision on size and the capital ratio. Banks hold less capital than is socially optimal while the size may be larger or smaller than socially optimal. Second, if the bankruptcy costs are low, and financing costs and taxes are high, imposing a higher capital requirement results in larger banks. Nonetheless, imposing higher capital requirements always increases social welfare. Restricting bank size is not necessarily welfare improving in a crisis period where asset interdependence is high and investment opportunities are scarce.

The policy implication is that capital requirements are the preferred policy measure during economic booms as well as during crisis periods. In particular during economic booms where asset interdependence is low, capital requirements may have an additional mitigating effect on systemic risk through its indirect effect on a smaller size.

To the best of our knowledge, this paper is the first to address the joint impact of the three characteristics of banking on systemic risk and social welfare. The recent paper of De Nicolò et al. (2014) is most related to our approach. They calibrate a dynamic model to assess the effect of capital requirements on social welfare. They find that capital requirements affect both bank lending and social welfare with an inverted U-shape. Our model complements this paper because we include size as an additional endogenous factor, and allow for capital-size interaction effects. In addition, we calibrate our model by considering more severe shocks exhibiting heavy tails.

Our model on the choice of capital structure is related to the corporate finance literature on optimal capital structure. Starting with Modigliani and Miller (1958), the theoretical corporate finance literature has studied at length the determinants of an optimal capital structure, e.g., Kraus and Litzenberger (1973), Brennan and Schwartz (1978), Bradley et al. (1984), and Miao (2005) among many others. The main message of this literature is that taxes and bankruptcy costs are the key determinants of the capital structure.

Although our model is a static equilibrium model, this study can be compared with the literature on the interrelation between the optimal capital ratio and the business cycle, see, e.g., Angeloni and Faia (2013), and Repullo and Suarez (2013). Our static equilibrium model can be interpreted as the unconditional long-run effect. Alternatively, for a specific parameter choice we interpret our model as conditional on some state of the business cycle.

The asset dependence between banks models the systemic risk in the financial system. This is in line with the literature relating systemic risk to banks' common exposure, see Acharya (2009), Wagner (2010), and Ibragimov et al. (2011).[1] Our setup of asset dependence is in the spirit of the approach in Ibragimov et al. (2011) who also find that the similarity of financial intermediaries has an adverse impact on systemic risk and hence social welfare. This externality depends crucially on the distribution of the shocks. Our approach is different by endogenizing the capital-size decision in the presence of asset interdependence.

We do not explicitly model systemic risk generated by the liability side. However, the potential for a bank run decreases the value of the assets of all involved banks. Accordingly, the potential for a bank run is implicitly modelled through the asset interdependence of different banks. Modelling the bank run potential through an explicit effect of default probabilities on asset interdependence would unnecessarily complicate the model. A similar argument applies to systemic risk generated by the interbanking network. Furthermore, the latter mechanism is currently an insignificant transmission channel for a bank run in an advanced economy since banks can lend at negligible rates from central banks.

Lastly, this study also contributes to the emerging literature on the impact of bank size on banks' financing decisions and risk taking, and eventually systemic risk. On the one hand, Hughes and Mester (1993) find in an empirical study that large banks have lower funding costs on uninsured deposits. On the other hand, empirical evidence for the adverse effect of bank size on systemic risk is documented in Pais and Stork (2013), and Laeven et al. (2014). In addition, Davies and Tracey (2014) report that large banks do not benefit from scale economies after a correction for the potential bailout effect.

The paper proceeds as follows. Section 4.2 presents the model. In Section 4.3, we derive the optimal size and capital structure decision given the risk profile of the bank, and we consider the existence of a general equilibrium. Section 4.4 defines the social welfare function. Section 4.5 provides the setup on the asset dependence and the calibration of our model. Section 4.6 concludes. The proofs of the analytical results are postponed to the Appendix.

[1] Systemic risk is usually attributed to the following channels: direct linkages from mutual exposures, indirect linkages from common exposures, and interdependence from information contagion. We focus on the indirect linkage channel. Broad surveys on systemic risk are in De Bandt and Hartmann (2000), Santos (2001), and Galati and Moessner (2013).

4.2 The model

The model is a one period model in a risk neutral world[2] with a risk free rate r_f.[3] In this general equilibrium model, banks choose total assets x (referred to as size) and the equity to total assets ratio y (referred to as capital ratio). Consequently, a bank's equity and debt is $E = xy$ and $D = x - E = x(1-y)$, respectively.

To ensure the existence of banks, we adopt the standard assumption that investors in debt and equity cannot lend directly to entrepreneurs. This is consistent with the theory of financial intermediation in Diamond (1984). We assume that banks are perfectly competitive, as in Diamond and Dybvig (1983), and De Nicolò et al. (2014) among others.

Bank assets

The assets are loans to finance projects of entrepreneurs. A bank earns per unit of assets a stochastic return

$$R_A = \mu + \sigma X, \tag{4.1}$$

with mean $\mathbb{E}[R_A] = \mu$, standard deviation $\sigma(x, \mu)$, and X a random variable with zero mean and unit variance. We denote the probability density function and the distribution function of X by f and F, respectively. The mean asset return μ is endogenously determined by the competition between banks (see section 4.3.2 for details). The standard deviation $\sigma(x, \mu)$ is a function of the size x and the mean asset return μ. It decreases with size due to the potential diversification benefits, i.e., $\sigma_1(x, \mu) \leq 0$.[4] Charging a higher loan rate to entrepreneurs increases the probability on a loan restructuring. In addition, a higher rate may particularly deter risk-averse entrepreneurs, and less risky projects are possibly matched to a cheaper source of financing. Hence, we assume that the standard deviation $\sigma(x, \mu)$ is non-decreasing with the mean asset return, i.e., $\sigma_2(x, \mu) \geq 0$.

[2]In this setting, random variables are considered under a risk neutral probability measure. Risk neutral settings are also in Miao (2005), Titman and Tsyplakov (2007), Repullo and Suarez (2013), and Allen et al. (2014).

[3]Throughout the paper, we denote a net interest rate with a lowercase r, and the corresponding gross interest rate with an uppercase: $R = 1 + r$.

[4]The partial derivative to the ith argument of a function f is denoted by f_i ($i = 1, 2, \ldots$), while f_x denotes the total derivative to a variable x.

Operational cost

Bank operations have a cost function $c(x)$ that depends on size. We assume that the operational cost and the mean operational cost $\bar{c}(x) = c(x)/x$ are both convex in x with $\bar{c}(x)$ minimized at some $x_0 \geq 0$. Intuitively, at some point there are diseconomies of scale in operational costs by inefficiencies such as bureaucracy costs. This assumption is supported by the study in Davies and Tracey (2014) that shows that large banks do not benefit from scale economies. Empirical evidence for diseconomies of scale in banking is in Allen and Rai (1996).

Debt financing

We abstract from banks' maturity transformation function by assuming that debt has a fixed maturity of one period. Equivalently, debt is automatically rolled over at the end of the period. By assuming $y < 1$, we rule out the limiting case of no debt financing where $D = 0$.

A bank promises to pay at the end of the period the gross contracted interest rate R_c on debt outstanding. It can only fulfil this promise if the return on assets R_A is sufficient. Namely, a low realization of R_A results into default if total cost exceeds total revenue: $c(x) + R_c D > R_A x$. In case of a default, debt holders incur an additional deadweight bankruptcy cost $\eta > 0$ on each unit of debt to recover their debt outstanding.[5] In this risk neutral setting, risk aversion is another interpretation of this bankruptcy cost. Since holders of defaulting debt only receive $R_A x - c(x) - \eta D$, instead of the contracted $R_c D$, the expected gross return on debt is

$$\mathbb{E}[R_D] = \mathbb{E}\left[\min\left(\frac{R_A x - c(x)}{D}, R_c\right)\right] - \eta\, PD, \tag{4.2}$$

where the probability of default is

$$PD = \mathbb{P}(R_A x < c(x) + R_c D)\,. \tag{4.3}$$

Equity financing

The equity holders are the residual claimants who receive the after-tax profit at the end of the period. The end-of-period equity value before taxes (E_{BT})

[5] A similar feature is in Strebulaev (2007) and De Nicolò et al. (2014). Our bankruptcy cost differs from costs of illiquidity studied in Cifuentes et al. (2005) and Wagner (2011). They model illiquidity costs by fire sale losses which is not a deadweight loss.

equals revenues minus the sum of operational costs and costs of debt financing. Accordingly, the initial expectation of the end-of-period equity is

$$\mathbb{E}[E_{BT}] = \mathbb{E}\big[[R_A x - c(x) - R_c D]^+\big] \tag{4.4}$$

where $[x]^+ = \max(x, 0)$. The bank pays the corporate tax rate $\tau = 1 - \bar{\tau}$ on before-tax profits. Note that the cost of debt financing is tax deductible, while the cost of equity financing is not.

4.3 Equilibrium analysis

This section provides an equilibrium analysis at two different levels. At the bank level, we derive in section 4.3.1 the equilibrium borrowing and loan rate of a single bank with given size and capital ratio. In section 4.3.2, we obtain a general equilibrium result for banks in perfect competition.

4.3.1 Banking equilibrium

In the banking equilibrium, the risk neutrality of investors implies that equity holders and debt holders earn on average the risk free rate on their investments. From these returns, we determine for given size x and given capital ratio y, the following three equilibrium values: (i) the mean return on assets (μ^*), (ii) the probability of default (PD^*), and (iii) the contracted interest rate on debt (R_c^*).

Proposition 1. *If an equilibrium exists for the size-capital ratio pair (x, y), then a banking equilibrium is the triple (μ^*, PD^*, R_c^*) that satisfies the following three equations*

$$\mu^* = 1 + \bar{c}(x) + \frac{r_f}{\bar{\tau}} y + (r_f + \eta PD^*)(1 - y) \tag{4.5}$$

$$\mathbb{E}\big[\min\big(X, F^{-1}(PD^*)\big)\big] = -\frac{(1 + r_f/\bar{\tau})\, y}{\sigma(x, \mu^*)} \tag{4.6}$$

$$\mathbb{E}\left[\min\left(\frac{R_A - \bar{c}(x)}{1 - y}, R_c^*\right)\right] = R_f + \eta PD^* \tag{4.7}$$

where $R_A = \mu^* + \sigma(x, \mu^*)X$ as in (4.1), and, as before, $R_f = 1 + r_f$ is the gross risk free rate.

Note that the three equations in Proposition 1 are interdependent, or even more precisely μ^* and PD^* can be solved simultaneously from (4.5) and (4.6) and then R_c^* follows from (4.7).

Substituting the extremal values $PD^* = 0$ and $PD^* = 1$ in (4.5) gives that the expected gross return on assets μ^* satisfies

$$0 \leq \mu^* - \bar{c}(x) - R_f - r_\tau y \leq \eta\,(1-y) \tag{4.8}$$

where $r_\tau := r_f\left(\frac{1}{\bar{\tau}} - 1\right)$.

Since F^{-1} is an increasing function, equation (4.6) can be rewritten as

$$PD^* = g\left(\frac{\sigma(x,\mu^*)}{(1 + r_f/\bar{\tau})\,y}\right). \tag{4.9}$$

where g is some increasing function that depends on the given cdf F and the parameters r_f and $\bar{\tau}$.

An implicit expression for PD^* follows by substituting (4.5) into the RHS in (4.9). The obtained equation has PD^* as the single unknown value. In fact, both hand sides of the obtained equation are increasing in PD^*. This indicates that multiple banking equilibriums might exist, or none exists. The next proposition provides sufficient conditions for existence and uniqueness of the banking equilibrium.

Proposition 2.

Existence: *For all (x, y), a banking equilibrium (μ^*, PD^*, R_c^*) satisfying* (4.5)–(4.7) *exists if at least one of the following two conditions holds*

$$1 + \frac{r_f}{\bar{\tau}} \leq -\inf(X) \min_{\tilde{x},\mu} \sigma(\tilde{x},\mu) \tag{4.10}$$

$$\sigma_2 \equiv 0 \text{ for all } (x,y) \tag{4.11}$$

Uniqueness: *The equilibrium is unique if* (4.11) *holds.*

An equilibrium may not exist if the sufficient existence conditions do not hold. Debt holders prefer a small PD to maximize the payment on their debt. The equity holders in turn may prefer for any asset return μ a strictly higher PD to benefit from their limited liability. In such cases, (4.5) and (4.6) have no common solution (μ^*, PD^*). As a consequence, the equityholders and bondholders cannot agree upon an equilibrium (μ^*, PD^*, R_c^*) such that an equilibrium does not exist. The sufficient conditions in Proposition 2 exclude this case. We will impose the condition $\sigma_2 \equiv 0$ in the calibration in Section 4.5.

Consider the impact of the financing cost r_f and the operational cost $\bar{c}(x)$ on the endogenous probability of default PD^* if $\sigma_2 \equiv 0$ holds, i.e., σ

is invariant to the mean asset return μ. In this case (4.9) implies that PD^* does not depend on μ^*. Thus, increasing r_f has no feedback effect on PD^* through μ^*. Hence, we conclude from (4.9) that PD^* decreases with respect to the financing cost r_f.

The mean operational cost $\bar{c}(x)$ has no direct effect on PD^* (see (4.9)). Following the equilibrium equation (4.5), the asset return μ^* compensates any effect of the operational cost on PD^*. The random spread $R_A - \bar{c}(x)$ in the equilibrium equation (4.7) is then unaffected by changes in the operational cost. Therefore, PD^* (and R_c^* in (4.7)) is unaffected by the operational cost.

To study the impact of taxes and bankruptcy costs on the equilibrium, suppose that both are absent ($\tau = \eta = 0$). Then, equation (4.8) shows that $\mu^* = \bar{c}(x) + R_f$, which says that the required marginal revenue on assets equals its marginal cost. Having taxes and bankruptcy costs raises this marginal cost, and, consequently, the marginal revenue μ^* in equilibrium.

4.3.2 General equilibrium

In this section we obtain general equilibrium results where the size x and capital y are endogenously determined by the competitive setting. As in the banking equilibrium in section 4.3.1, entrepreneurs are price takers since they stick to the offered loan rate of the banks. Following Diamond (1984), we assume information asymmetry between entrepreneurs and banks. Namely, banks know the aggregate return distribution of projects, but not the return distribution of a specific project. By contrast, entrepreneurs do have private information on their project. An entrepreneur initiates a project provided the lowest offered rate is below the project's return in case the project is succesful. Consequently, the entrepreneur extracts all rents beyond the offered loan rate.

In the perfectly competitive banking industry, the entrepreneurs force banks to demand the lowest possible mean interest rate μ^c. Only banks that are able to demand this minimal rate survive in this perfectly competitive setting. The competitive equilibrium (x^c, y^c) corresponds to the pair (x, y) with the minimal equilibrium mean asset return μ^c. In this competitive equilibrium, efficient banks choose (i) the optimal size x^c where the marginal benefit of a smaller volatility ($\sigma_1 < 0$) equals the marginal cost of a higher mean operational cost ($\bar{c}' > 0$), and (ii) the optimal capital level y^c where the marginal benefit of a smaller expected bankruptcy cost equals the marginal cost of a smaller benefit from tax deductible debt financing.

The equilibrium values of the variables in the competitive equilibrium are denoted with a superscript c. We suppress asterisks for convenience. For instance:

$$(x^c, y^c) = \left\{ (x, y) \,\middle|\, (x, y) = \operatorname*{argmin}_{\tilde{x},\tilde{y}} \mu^*(\tilde{x}, \tilde{y}) \right\}$$

$$\mu^c = \mu^*(x^c, y^c) \qquad \mu_1^c = \frac{\partial}{\partial x}\mu^*(x, y)\bigg|_{(x=x^c, y=y^c)} = 0$$

$$PD^c = PD^*(x^c, y^c) \qquad PD_1^c = \frac{\partial}{\partial x}PD^*(x, y)\bigg|_{(x=x^c, y=y^c)}$$

The definition of the partial derivatives PD_2^c and PD_{12}^c, and the total derivatives PD_x^c and PD_y^c is similar to the definitions above. The derivatives of x^c and y^c are similarly defined. The same applies to the derivatives with respect to the model parameters σ_0, η, r_f and τ.

We consider the impact of size x and the capital ratio y on the competitive PD^c. Denote

$$x^c(y) = \operatorname*{argmin}_{x} \mu(x, y) \qquad y^c(x) = \operatorname*{argmin}_{y} \mu(x, y) \tag{4.12}$$

as the optimal (competitive) size x and the optimal (competitive) capital ratio y conditional on the other variable.

To avoid clutter, we make the following mild technical assumptions on the competitive equilibrium (x^c, y^c). The global minimum $\mu^c = \mu^*(x^c, y^c)$ is (i) unique, (ii) an interior point, and (iii) the Hessian matrix of μ^* is not zero at (x^c, y^c). This ensures that μ^c is a strict global minimum, and the correspondences $x^c(y)$ and $y^c(x)$ are properly defined functions in a neighborhood of (x^c, y^c). Note that we still allow for multiple local minimums.

We consider the following two conditions:

$$\frac{\mathrm{d}x^c(y^c)}{\mathrm{d}y} > 0 \qquad \frac{\mathrm{d}y^c(x^c)}{\mathrm{d}x} > 0 \tag{C}$$

The conditions in (C) state that a higher capital ratio forces banks to choose a larger size, and vice versa. In practice, the sign of the derivatives in (C) is ambiguous. Concerning the impact of capital on size ($\mathrm{d}x^c/\mathrm{d}y$), a higher capital ratio may result in a larger desire to enjoy diversification benefits by increasing size. On the other hand, the more expensive equity financing may make project investments too expensive, and hence reduce the bank size. Our condition in (C) puts a higher weight on the former effect.

For the impact of size on capital ($\mathrm{d}y^c/\mathrm{d}x$), a larger size may give banks an incentive to choose a lower capital ratio by the larger diversification benefits. On the other hand, a higher marginal operational cost from the larger size could also induce large banks to choose a more safe operational setup by choosing a higher capital ratio. The latter effect dominates in (C).

The following proposition gives the impact of x and y on PD in the competitive equilibrium.

Proposition 3.

(i) If $\sigma_1^c < 0$, the partial and total derivatives of PD and $\bar{c}$ at an interior competitive equilibrium (x^c, y^c) are as given in Table 4.1.

(ii) If $\sigma_1^c = 0$ then $\frac{\mathrm{d}y^c(x^c)}{\mathrm{d}x} = \frac{\mathrm{d}x^c(y^c)}{\mathrm{d}y} = 0$ and all derivatives in Table 4.1 are zero at (x^c, y^c) except $\frac{\partial}{\partial y} PD^c = \frac{\mathrm{d}}{\mathrm{d}y} PD^c < 0$.

(iii) The conditions in (C) are equivalent to

$$(1 - y^c) PD_{12}^c < PD_1^c. \tag{4.13}$$

<table>
<tr><th></th><th>$\frac{\partial}{\partial x}$</th><th>$\frac{\mathrm{d}}{\mathrm{d}x}$</th><th>$\frac{\partial}{\partial y}$</th><th>$\frac{\mathrm{d}}{\mathrm{d}y}$</th></tr>
<tr><td>PD^c</td><td>$-$</td><td>$-^*$</td><td>$-$</td><td>$-^*$</td></tr>
<tr><td>$\bar{c}^c$</td><td colspan="2">$+$</td><td>0</td><td>$+^{**}$</td></tr>
</table>

Table 4.1: **Derivatives of the default probability and the average cost function.**

The table reports the sign of the partial derivatives and the total derivatives of the default probability $PD^*(x, y)$ and the average cost function $\bar{c}(x, y)$ with respect to size x and capital ratio y at an interior competitive equilibrium that satisfies $\sigma_1^c < 0$. The competitive equilibrium (x^c, y^c) corresponds to the banking equilibrium (μ^*, PD^*, R_c^*) in (4.5)–(4.7) where the mean asset return μ^* is minimal. * if (C) holds. ** if and only if (C) holds.

We remark that the conditions in (C) and the equivalent conditions in (4.13) are not very stringent. Consider the decreasing function

$$h(z) = g\left(\frac{1}{(1 + r_f/\bar{\tau})\, z}\right),$$

with g as in (4.9). From $PD^* = h\left(y/\sigma(x, \mu^*)\right)$ and

$$\frac{\partial \sigma^c(x, \mu)}{\partial y} = \frac{\partial \sigma^c}{\partial \mu} \frac{\partial \mu^c}{\partial y} = 0,$$

which gives at the competitive equilibrium

$$PD_1^c = -\frac{\sigma_1^c y^c}{(\sigma^c)^2} h'\left(\frac{y^c}{\sigma^c}\right)$$
$$PD_{12}^c = -\frac{\sigma_1^c}{(\sigma^c)^2}\left[h'\left(\frac{y^c}{\sigma^c}\right) + \frac{y^c}{\sigma^c} h''\left(\frac{y^c}{\sigma^c}\right)\right]$$

For $\sigma_1^c < 0$, inequality (4.13) is then equivalent to

$$\frac{2y^c - 1}{(1-y^c)\,y^c}\sigma^c < \left.\frac{h''(z)}{h'(z)}\right|_{z=y^c/\sigma^c} = \left.\frac{\mathrm{d}}{\mathrm{d}z}\ln\left[h'(z)\right]\right|_{z=y^c/\sigma^c}$$

In other words, if $y^c < \frac{1}{2}$ and $\sigma_1^c < 0$ both hold, a function h that is concave or limited convex at the competitive value y^c/σ^c implies the conditions in (C).

4.3.3 A numerical example

Proposition 3 gives us the effect of size and capital on the probability of default under condition (C). If this condition does not hold, the total effects on PD^c depend on parameters that are relevant for policy intervention, such as σ_0, η, r_f and τ. A numerical example demonstrates how the sign of the total derivatives x_y^c, y_x^c, PD_x^c, and PD_y^c depends on the parameters if condition (C) does not hold. For analytical convenience, we assume that X follows a uniform distribution, similar to Acharya (2009), and Angeloni and Faia (2013). In contrast, the calibration study in Section 4.4 uses a t-distribution.

Assume that $X \sim U\left[-\sqrt{3}, \sqrt{3}\right]$, which satisfies the assumptions $\mathbb{E}[X] = 0$ and $\mathrm{Var}(X) = 1$. The mean operational cost is convex and given by $\bar{c}(x) = x^\gamma$ with $\gamma > 1$, and the standard deviation of a unit of assets is[6]

$$\sigma(x,\mu) = \frac{1}{\sqrt{3}}\left(\frac{\sigma_0}{x^\delta}\right)^2 \tag{4.14}$$

Without loss of generality, we assume $0 < \delta \leq \frac{1}{4}$ by the following reasoning. If assets are completely dependent, the standard deviation of total assets is linear in x. Then, the standard deviation per unit of assets is constant, which corresponds to $\delta = 0$ in (4.14). Since we assume that δ is strictly positive, we could choose a very small δ for this completely dependent case. At the other extreme, assets are completely independent. This means that the

[6] The scaling term $1/\sqrt{3}$ will cancel out later.

standard deviation of total assets is proportional to $\sqrt{x}$, which corresponds to $\delta = \frac{1}{4}$. The assumption $0 < \delta \leq \frac{1}{4}$ implies that the ratio $\zeta := \gamma/\delta$ exceeds 4.

Define the function

$$\phi(y) := \sqrt{y}\,(1-y)\left|3\sqrt{y} - \frac{1}{\sqrt{y}}\right|^{\zeta-1}$$

and the constants

$$\tilde{R}_f := \sqrt{1 + \frac{r_f}{\bar{\tau}}} \tag{4.15}$$

$$r_{\bar{\tau}} := r_f\left(\frac{1}{\bar{\tau}} - 1\right) \tag{4.16}$$

$$y_1 := \frac{1}{3} + \frac{2}{3}\sqrt{\frac{\zeta-1}{\zeta+2}} \tag{4.17}$$

$$\eta_a := \frac{r_{\bar{\tau}}}{1 - \frac{1}{2}\left(\frac{1}{\zeta}\phi(y_1)\,\tilde{R}_f^{\zeta}\sigma_0^{-\zeta} r_{\bar{\tau}}\right)^{\frac{1}{\zeta-1}}} \tag{4.18}$$

$$\eta_b := \frac{(2\sigma_0)^{\zeta}\zeta}{2\phi(y_1)\,\tilde{R}_f^{\zeta}} \tag{4.19}$$

$$\tilde{\sigma} := \left[\phi\left(\frac{\zeta}{\zeta+2}\right)\frac{\eta}{\zeta}\left|2 - \frac{2r_{\bar{\tau}}}{\eta}\right|^{1-\zeta}\right]^{1/\zeta}\tilde{R}_f \tag{4.20}$$

$$x_{\max} := \left([\eta - r_{\bar{\tau}}]\,\frac{\sigma_0}{\tilde{R}_f^2}\right)^{1/\delta} \tag{4.21}$$

The following proposition shows at the competitive equilibrium the relation between size and capital, and also the impact of size and capital on the probability of default. The restriction $\eta \notin [\eta_a, \eta_b]$ is only needed to exclude a boundary equilibrium with $y^c = 1$.

Proposition 4. *At the competitive equilibrium, the size x^c, the capital ratio y^c and the corresponding derivatives are as given in Table 4.2.*

We start with the interrelation between size and capital. Proposition 4 states that size and capital are positively related if the bankruptcy cost η is low. A low bankruptcy cost gives a low penalty on a higher default probability from debt financing. This results in low capital ratios and large marginal diversification benefits. That is, the effect of higher diversification

	$\eta < r_{\bar{T}}$	$\eta \geq r_{\bar{T}} \cap \eta \notin [\eta_a, \eta_b]$	
		$\sigma_0 < \tilde{\sigma}$	$\sigma_0 \geq \tilde{\sigma}$
x^c	$[0, x_{\max})$	$[x^c(y_1), x_{\max}]$	
x^c_y	$+$	$-$	
y^c	$\left[0, \frac{1}{3}\right)$	$\left[\frac{1}{3}, y_1\right]$	
y^c_x	$+$	$-$	
PD^c_x	$-$		
PD^c_y	$-$		$+$

Table 4.2: **Total derivatives and intervals at the competitive equilibrium.**

This table presents the intervals of the size x, and the capital ratio y. In addition, it presents the derivatives of x, y, and the default probability PD. The shocks follow a uniform distribution, and risk as in (4.14) which satisfies $\sigma_2(x, \mu) \equiv 0$. All statistics are at an interior competitive equilibrium (x^c, y^c). This corresponds to (x, y) with the banking equilibrium (μ^*, PD^*, R^*_c) in (4.5)–(4.7) with the smallest μ^*. The parameters η and σ_0 represent the bankruptcy cost and the uncertainty in the economy, respectively. The definitions of $r_{\bar{T}}$, y_1, η_a, η_b, $\tilde{\sigma}$, and $x_{\max}$ are in (4.16)–(4.21). A plus and a minus sign indicate a positive and a negative total derivative, respectively.

benefits dominates the effect of a higher marginal operational cost if and only if bankruptcy costs are low. Consequently, a higher capital ratio results in larger banks in case capital ratios are below $\frac{1}{3}$.

The previous result indicates an important trade-off when implementing a capital requirement when the actual levels of capital are low. In that case, a more stringent capital requirement results into smaller individual default probabilities, but larger banks. From the micro-prudential view of Table 4.2, since $PD^c_1 \leq 0$ and $PD^c_2 \leq 0$ (see Proposition 3(i)), the larger size reinforces the direct reduction in PD from the larger capital requirements. From a macro perspective, the upward effect on size counteracts the effect of the smaller individual default probabilities on systemic risk by a crowding effect on banks. Closed-form results are unavailable for this trade-off. Accordingly, we will elaborate on this trade-off in a calibration in Section 4.5.

Next, we consider the total effect on the probability of default. In this respect, we infer from Proposition 4 that large banks are preferred from a micro-prudential point of view. More specifically, an upward cap on size unambiguously increases the probability of default because $PD^c_x < 0$. Again, such a policy is only meaningful in a macro context which we discuss in Section 4.5.

The total effect of the capital ratio on the probability of default is ambiguous. If bankruptcy costs η and uncertainty σ_0 are both high, a higher capital requirement results in a higher default probability. The high bankruptcy cost results in a smaller marginal diversification benefit that is dominated by the effect of the operational costs. This explains the negative relation between size and capital ($x_y^c < 0$). Together with a high uncertainty, this results in a dominant upward effect of the smaller size on the probability of default. Hence, this case demonstrates that a larger capital requirement may fail to achieve the micro-prudential goal of a smaller individual probability of default.

The setup of this example case also enables us to derive the impact of the policy parameters at the competitive equilibrium: the uncertainty σ_0, the bankruptcy costs η, the financing costs r_f, and the tax rate τ.

Proposition 5. *Provided an interior equilibrium exists (a sufficient condition is $\eta \notin [\eta_a, \eta_b]$), the sensitivities of size, capital ratio and PD to the parameters σ_0, η, r_f and τ are given in Table 4.3.*

We discuss the impact of the parameters on size and capital in more detail. In most of the discussion we refer to the endogenous relation between size and capital. By Proposition 4, the competitive size and the competitive capital have a positive association if and only if bankruptcy costs are low.

(*i*) σ_0: When the uncertainty σ_0 in the economy is high, the diversification benefit received by large banks is low. Therefore, the optimal bank size is lower and closer to the size that minimizes the operational cost. Because capital protects banks against unexpected shocks, banks tend to hold more capital when the economy is more uncertain. However, this relation is reversed when bankruptcy costs are low. In that case, banks enjoy an initial low capital ratio due to the low cost of debt. Such highly leveraged banks enjoy the potential upside shocks due to the uncertainty in the economy. Therefore, with uncertainty increasing, banks may leverage more to exploit the potential upward shocks.

(*ii*) η: A higher bankruptcy cost η gives banks a stronger incentive to reduce their default probability by means of diversification. The positive relation with size reflects this diversification effect. In addition, a higher bankruptcy cost implies a higher marginal cost of debt. Banks are then more willing to hold more capital and prefer the diversification benefits of a larger size. However, such relations have to be considered in conjunction with the endogenous relation between size and capital.

<table>
<tr><th rowspan="2"></th><th rowspan="2">$\eta < r_{\bar{\tau}}$</th><th colspan="2">$r_{\bar{\tau}} \leq \eta < \zeta r_{\bar{\tau}}$</th><th colspan="2">$\eta \geq \zeta r_{\bar{\tau}}$</th></tr>
<tr><th>$\sigma_0 < \tilde{\sigma}$</th><th>$\sigma_0 \geq \tilde{\sigma}$</th><th>$\sigma_0 < \tilde{\sigma}$</th><th>$\sigma_0 \geq \tilde{\sigma}$</th></tr>
<tr><td>$x^c_{\sigma_0}$</td><td colspan="5">−</td></tr>
<tr><td>x^c_{η}</td><td>+</td><td colspan="2">?</td><td colspan="2">+</td></tr>
<tr><td>$x^c_{r_f}$</td><td>?</td><td colspan="4">+</td></tr>
<tr><td>x^c_{τ}</td><td>?</td><td colspan="4">+</td></tr>
<tr><td>$y^c_{\sigma_0}$</td><td>−</td><td colspan="4">+</td></tr>
<tr><td>y^c_{η}</td><td colspan="3">+</td><td colspan="2">−</td></tr>
<tr><td>$y^c_{r_f}$</td><td colspan="5">−</td></tr>
<tr><td>y^c_{τ}</td><td colspan="5">−</td></tr>
<tr><td>$PD^c_{\sigma_0}$</td><td>+</td><td>?</td><td>+</td><td>?</td><td>+</td></tr>
<tr><td>PD^c_{η}</td><td colspan="2">−</td><td colspan="2">?</td><td>−</td></tr>
<tr><td>$PD^c_{r_f}$</td><td colspan="2">?</td><td>−</td><td>?</td><td>−</td></tr>
<tr><td>PD^c_{τ}</td><td colspan="2">?</td><td>−</td><td>?</td><td>−</td></tr>
</table>

Table 4.3: **Comparative statics at the competitive equilibrium.** The table presents comparative statics in the competitive equilibrium for size x, the capital ratio y, and the default probability PD with respect to the uncertainty σ_0, the bankruptcy cost η, the risk free rate r_f and the tax rate τ. The shocks follow a uniform distribution, and risk is constant in return ($\sigma_2(x, \mu) \equiv 0$). All statistics are at the competitive equilibrium (x^c, y^c) which corresponds to (x, y) with the the banking equilibrium (μ^*, PD^*, R^*_c) in (4.5)–(4.7) with the smallest μ^*. The parameter $\zeta := \gamma/\delta$ follows from the average cost function $\bar{c}(x) = x^\gamma$ and the standard deviation of each unit of assets $\sigma(x, \mu) = \sigma_0 x^{-2\delta}$. The definitions of $r_{\bar{\tau}}$, η_a, η_b, and $\hat{\sigma}$ are in (4.16)–(4.20). A plus and a minus sign indicate a positive and a negative total derivative, respectively.

Therefore, we observe for a low bankruptcy cost η a positive impact of this bankruptcy cost on the capital ratio, whereas for a high bankruptcy cost the endogenous size effect dominates such that the impact of the bankruptcy cost on capital is negative.

(iii) r_f: A high risk free rate r_f corresponds to a high expected return on debt and equity. This makes debt relatively more attractive due to its tax deductibility. As a consequence, the capital ratio decreases when the risk free rate increases. Since a higher risk free rate results in more expensive financing, the desired diversification is also higher, which leads to larger banks. However, this result is only observed for a high bankruptcy cost. Otherwise, the endogenous positive relation between size and capital makes the relation between the risk free rate and size

ambiguous.

(*iv*) The explanation for the effect of the tax rate is very similar to the explanation with the risk free rate. Tax deductibility and costs of financing are again the relevant channels.

The impact of the policy parameters on the probability of default PD are mostly in line with the impact on size and the capital ratio. Namely, a high uncertainty in the economy, a low bankruptcy cost, a low financing cost and a low tax rate correspond to a high PD.

4.4 Social welfare

We study the effect of size, the capital ratio, and the policy parameters on social welfare for a financial system consisting of a number of banks. Social welfare is defined here as the total surplus generated by the banking system discounted by a penalty for the associated systemic risk. We first define measures for the two components total surplus and systemic risk. Then, we calibrate the model to obtain the welfare effects of different policies on size and the capital ratio.

4.4.1 Total surplus

By risk neutrality for the net return on equity and debt ($\mathbb{E}[r_E] = \mathbb{E}[r_D] = r_f$),[7] and the tax effect of equity ($\mathbb{E}[r_E] = \bar{\tau}\left(\mathbb{E}[E_{BT}] - E\right)/E$), the operations of each bank generate the surplus

$$(\mathbb{E}[E_{BT}] - E - r_f E) + (\mathbb{E}[D_{\text{end}}] - D - r_f D) = \left(\frac{1}{\bar{\tau}} - 1\right) r_f E + 0. \quad (4.22)$$

where $\mathbb{E}[D_{\text{end}}]$ and $\mathbb{E}[E_{BT}]$ are the end-of-period values of debt and before-tax equity, respectively. The left hand side (LHS) contains two end-of-period values for the expected surplus: the before-tax surplus of the equity holders, and the surplus of the debt holders.

Debt does not generate any surplus, while the taxation of equity generates a surplus. The right hand side (RHS) of (4.22) represents this surplus. The surplus is zero if taxes are absent ($\bar{\tau} = 1 - \tau = 1$). Given the size of equity, an increase in the corporate tax rate τ (a decrease in $\bar{\tau}$) increases the surplus since an increase in τ leads to a higher asset return μ^* on bank

[7] The expectations of the return distributions are of course under the risk neutral measure.

loans to entrepreneurs. Otherwise, banks cannot pay both the higher tax and the required return of equity holders. The higher borrowing rate for entrepreneurs (implied by the higher μ^*) results in a smaller number of initiated projects, and may thus not be socially optimal. We model this trade-off as follows.

The total amount of accepted loans depends on the mean return on assets μ^* as

$$Q(\mu^*) = a - b\left(\mu^* - 1\right), \tag{4.23}$$

where a and b are some positive constants. Since Q is the aggregate amount of assets in the economy, the number of banks equals $n = Q(\mu^*)/x$.[8] We consider equilibria in which banks make identical management decisions on size and capital. The banks differ ex-post as they are exposed to bank-specific shocks that are correlated in a way that we will specify in section 4.4.3.[9]

Under a fixed number of banks (n) the equilibrium μ^* is the minimal mean asset return under the additional constraint $x = Q(\mu^*(x, y))/n$. Thus, size would no longer be part of a bank's decision making. Consequently, our main result that policies on capital are most important remains unaffected. Importantly, a fixed n violates the free entry assumption in our long-run model with perfect competition. Therefore, we focus on the case with n endogenous.

Entrepreneurs only initiate a project if the expected return exceeds the offered loan rate. They earn on average a consumer surplus from the project:

$$CS = \int_{\mu^*}^{1+a/b} Q(\tilde{\mu})d\tilde{\mu} = \frac{1}{2b}Q\left(\mu^*\right)^2. \tag{4.24}$$

The financiers do not earn a surplus by the risk neutrality of the equity holders and debt holders. The government surplus follows from multiplying the surplus in (4.22) by the number of banks,

$$GS = Q(\mu^*)r_{\mathcal{T}}y \tag{4.25}$$

where as before $r_{\mathcal{T}} = \left(\frac{1}{\mathcal{T}} - 1\right) r_f$, $E = xy$, and $n = Q(\mu^*)/x$. Figure 4.1 illustrates the concepts graphically. The total surplus on loans, TS, is the sum of (4.24) and (4.25):

$$TS = Q(\mu^*)\left(\frac{1}{2b}Q(\mu^*) + r_{\mathcal{T}}y\right) \tag{4.26}$$

[8] We avoid clutter by allowing for non-integer n. This does not affect our main results.

[9] Ex-ante symmetric setups are also in Ibragimov et al. (2011), Repullo and Suarez (2013), and De Nicolò et al. (2014), among others.

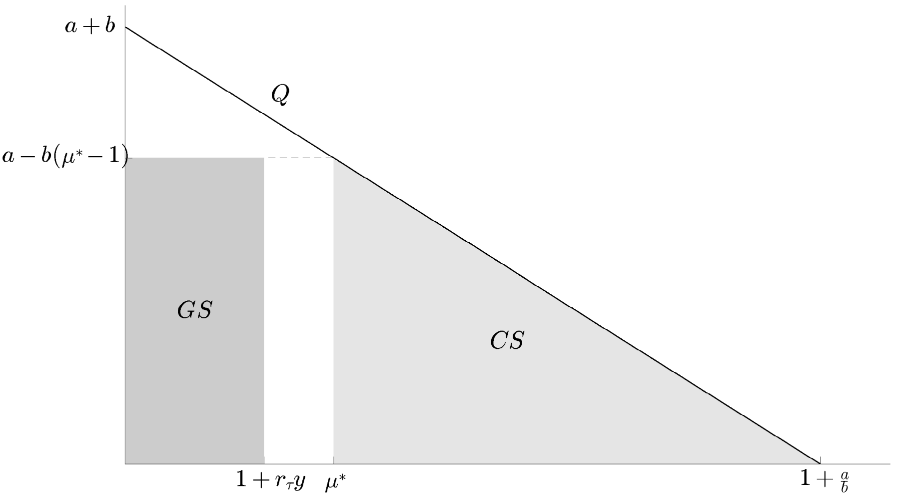


Figure 4.1: **Consumer surplus and government surplus.**
The figure shows the consumer surplus CS from (4.24) and the government surplus GS from (4.25). The parameter $r_{\bar{\tau}}$ and the total amount of accepted loans Q are defined in (4.16) and (4.23), respectively. The surfaces CS and GS are nonoverlapping by (4.8) in Proposition 1.

We have four straightforward observations. First, perfect competition among banks erodes the surplus of banks such that the entrepreneurs (CS) and the government (GS) share the total surplus on bank loans.

Second, the size x is absent in (4.26). In other words, for given capital ratio, the size has only an indirect effect on TS through μ^*. It follows from $Q'(\mu^*) < 0$ in (4.26) that $\partial TS(\mu^*, y)/\partial\mu^* < 0$. This indicates that maximizing TS conditional on y is the same as minimizing μ^* conditional on y. Therefore, the optimal size function $x^{TS}(y)$ for total surplus is the same function as $x^c(y)$ of the competitive equilibrium in Proposition 3.

Third, since the competitive equilibrium minimizes the mean asset return μ^* (see section 4.3.2) such that $\mu_x^c = \mu_y^c = 0$,

$$\left.\frac{\partial Q\left(\mu^*(x,y)\right)}{\partial x}\right|_{(x,y)=(x^c,y^c)} = \left.\frac{\partial Q\left(\mu^*(x,y)\right)}{\partial y}\right|_{(x,y)=(x^c,y^c)} = 0.$$

Consequently, we obtain from (4.26)

$$\left.\frac{\partial TS\left(x,y\right)}{\partial y}\right|_{(x,y)=(x^c,y^c)} = Q\left(\mu^c\right) r_{\mathcal{T}} > 0.$$

Thus, the total surplus TS is strictly increasing in the capital ratio y at the competitive equilibrium. That is, the competitive equilibrium does not maximize the total surplus.

Fourth, equation (4.26) does *not* imply that full equity financing ($y = 1$) maximizes the total surplus. In that case, the required return μ exceeds the competitive return μ^c, which may lead to a lower TS.

4.4.2 Systemic risk

The total surplus analysis neglects the cost of replacing defaulting banks. In particular, it neglects the nonlinear effect of a systemic crisis, i.e., having multiple defaults simultaneously. We define such a systemic crisis as the event where the amount of unpaid debt exceeds some fraction δ_S of total assets. This is similar to the approach in Ibragimov et al. (2011). [10]

Let $PS(x, y)$ represent the probability of a systemic crisis given the size x and capital ratio y of the individual banks

$$PS(x,y) = \mathbb{P}\left(\frac{\text{unpaid debt}}{\text{total assets in the economy}} > \delta_S\right) \tag{4.27}$$

Define $\lambda > 0$ as the loss due to a systemic crisis as a fraction of the total surplus on banking operations (TS). We define social welfare as

$$SW(x,y) = (1 - \lambda PS(x,y))\, TS(x,y). \tag{4.28}$$

Thus, a social planner needs to balance the total surplus with the probability of a systemic crisis. The risk preference against a systemic crisis is more pronounced if the parameter λ is high.

The bank size and capital ratio that maximize the total surplus does not correspond to the social optimum, i.e., it does not maximize social welfare. Since the probability on a systemic crisis increases in the default probability, the PD that maximizes total surplus TS is too high for maximizing social welfare. Hence, it is beneficial to have larger banks by their lower PD. However, having larger banks implies that a smaller number of banks suffices

[10] Recall that banks are ex-ante identical, but differ ex-post by the bank-specific shocks. Thus, some banks may default whilst others may not.

to finance a given proportion of loans to the entrepreneurs. Consequently, a smaller number of defaults can cause a systemic crisis when banks are larger. These two counterbalancing effects determine the optimal bank size that maximizes social welfare.

4.4.3 The distribution of banks' asset values

Recall from (4.1) that we model the gross return per unit of assets of a single bank i by the random variable $R_A = \mu^* + Y_i$, where μ^* is the equilibrium mean asset return conditional on size x and capital ratio y, while $Y_i = \sigma(x, \mu^*)X_i$ is random noise with mean zero. For a banking system consisting of n banks, we model the joint distribution of $(Y_1, \ldots, Y_n)$ as follows. We assume a symmetric banking system where each of the n banks has a market share $1/n$. Suppose there exists an infinite number of entrepreneurs uniformly spread on a circle with unit circumference. The entrepreneurs t and u are parameterized by $t, u \in [0,1]$ with distance

$$d(t,u) = \min\left(|u-t|, 1-|u-t|\right). \tag{4.29}$$

The investment project of entrepreneur t is exposed to the shock $Z(t) \sim N(0,1)$. The dependence between the shocks of entrepreneurs t and u is given by[11]

$$\operatorname{Cov}\left(Z(t), Z(u)\right) = \rho^{d(t,u)}.$$

Thus, the returns of two projects are at least $\sqrt{\rho}$ and more correlated if the projects are closer to each other.

In practice, asset backed securities are collateralized by assets having a similar risk profile, such as mortgages within the same region. This motivates us to assume that banks prefer adjacent projects. Indeed, such preferences minimize the distance cost. Bank i invests in the projects of the entrepreneurs located on the interval $[(i-1)/n, i/n]$. The shock distribution per unit of assets of bank i is given by

$$Y_i = \hat{\sigma}\sqrt{W} n \int_{t=(i-1)/n}^{i/n} Z(t)dt, \tag{4.30}$$

[11] Ibragimov et al. (2011) derive this result in a discrete setting with a large number m of entrepreneurs. The shocks to projects of two adjacent entrepreneurs follow an AR(1) process $x_t = \rho_m x_{t-1} + w$ with $w \sim N\left(0, 1-\rho_m^2\right)$. We present the limiting case where $m \to \infty$. Notice that $Z(t)$ is not a Wiener process because of $\rho_m < 1$ and the circular condition $Z(0) = Z(1)$. Indeed, $\operatorname{Cov}\left(Z(t), Z(u)\right) \neq \min(t,u)$ which would be true for a Wiener process.

where $W \sim IG(\frac{\nu}{2}, \frac{\nu}{2})$ follows an inverse gamma distribution with $\nu > 2$, and represents a common stochastic scaling factor that affects all Y_i simultaneously. The factor $\hat{\sigma}$ is a deterministic scaling factor for the standard deviation of the common shock. An extreme outcome of W increases the likelihood that multiple banks face an extreme outcome simultaneously.

Using well-known properties of the t-distribution (see, e.g., McNeil et al. (2010)), we have that the total shock Y_i to bank i follows a scaled Student $t(\nu)$-distribution with[12]

$$\mathrm{Var}\left(Y_i\right) = \frac{2\nu\hat{\sigma}^2 n^2}{\nu - 2} \frac{\rho^{1/n} - 1 - \frac{1}{n}\ln\left(\rho\right)}{\ln^2\left(\rho\right)} \tag{4.31}$$

Consider banks 1 and i that invest in projects on the intervals $[0, 1/n]$ and $[(i-1)/n, i/n]$, respectively. The covariance of their shocks Y_1 and Y_i is

$$\mathrm{Cov}\left(Y_1, Y_i\right) = \frac{\nu}{\nu - 2}\hat{\sigma}^2 \mathrm{Cov}\left(Z_1, Z_i\right) \tag{4.32}$$

where the covariance of the bank-specific shocks (excluding the common shock W) is

$$\mathrm{Cov}\left(Z_1, Z_{1+i}\right) = \begin{cases} \rho^{(i-1)/n} n^2 \left(\frac{\rho^{1/n}-1}{\ln(\rho)}\right)^2 & \frac{i+1}{n} \leq \frac{1}{2} \\ \frac{1}{\ln^2(\rho^s)} f(i, n, \rho) & \frac{i}{n} \leq \frac{1}{2} \leq \frac{i+1}{n} \\ \frac{n^2}{\ln^2(\rho)} f\left(n - i, n, \rho\right) & \frac{i-1}{n} \leq \frac{1}{2} \leq \frac{i}{n} \\ \rho^{1-(i+1)/n} n^2 \left(\frac{\rho^{1/n}-1}{\ln(\rho)}\right)^2 & \frac{1}{2} \leq \frac{i-1}{n} \end{cases}$$

with

$$f(i, n, \rho) = \rho^{\frac{1}{2}} \ln\left(\rho\right) \left(\frac{2\left(i+1\right)}{n} - 1\right) + \sqrt[n]{\rho}^{\,i-1} - 2\sqrt[n]{\rho}^{\,i} + \rho^{1-(i+1)/n}$$

It is straightforward to generalize this result to an arbitrary pair of banks by normalizing the project locations.

4.5 Calibration

Analytical results are unavailable for the impact of the size and the capital ratio on social welfare. Therefore, we investigate this impact with a calibration study. First, we discuss the implementation, then we present the numerical results.

[12]The derivations of (4.31) and (4.32) are in the appendix.

4.5.1 Implementation

We assume that the operational cost function has the form $c(x) = c_0 + c_1 x + c_2 x^2$ with $c_0, c_2 \geq 0$. This implies a convex average cost function

$$\bar{c}(x) = \frac{c_0}{x} + c_1 + c_2 x.$$

As discussed in section 4.4.2, we assume that a systemic crisis occurs if the expected total loss on debt is at least $\delta_S Q(\mu^*)$, where $\delta_S \in [0,1]$ is a threshold parameter. Since each default results in a dead weight loss of $D\eta$, a systemic crisis occurs if the number of defaults exceeds $n_D := \delta_S Q/(D\eta)$:

$$PS(x,y) = \mathbb{P}\left(\#\text{defaults} \geq \frac{\delta_S Q(\mu^*(x,y))}{(1-y)x\eta}\right) \tag{4.33}$$

If the threshold $\delta_S Q/(D\eta)$ is non-integer, we linearly interpolate the probability on a systemic crisis $PS(x,y)$ by weighting the probabilities based on the thresholds equal to the two nearest integers.

Table 4.4 presents the baseline parameter values. The rationale behind the chosen values is as follows. The riskfree interest rate is set at 0.1 such that our model represents a period of about 5 years. Our tax rate should be read as the joint effect of the corporate tax rate and the tax on interest income. Based on the estimates in Graham (1999), Strebulaev (2007) chooses the marginal rate on interest income 22.9 percentage points higher than the marginal rate on dividends. In our model, dividends are implicit in r_E, which is a total return that includes dividend distributions. Using a corporate tax rate of 0.35, we choose the effective corporate tax parameter equal to $\tau = 1 - (1-0.35)(1-0.229) \approx 0.5$.

r_f	τ	η	ν	$\hat{\sigma}$	c_0	c_1	c_2	ρ	a	b	δ_S	λ
0.1	0.5	0.1	4	0.05	10^{-3}	0	10^{-3}	0.5	25	25	0.1	1

Table 4.4: **Baseline parameters.**

This table reports baseline parameters of the risk free rate r_f, the tax rate τ, the bankruptcy cost η, the tail index ν of the shock distribution, the uncertainty $\hat{\sigma}$, the cost function parameters c_0, c_1, and c_2, the asset dependence ρ, the parameters a and b in the asset demand function $Q(\mu^*) = a - b(\mu^* - 1)$, the threshold parameter δ_S for a systemic crisis, and the fraction λ of total surplus lost in case of a systemic crisis.

The bankruptcy cost parameter is set at $\eta = 0.1$ to match the value in De Nicolò et al. (2014). In addition, it satisfies $\eta = r_f\,(1/\bar{\tau} - 1)$ which is the border case in Propositions 4 and 5. The tail index ν is set at 4 to match the estimates from the stock indices in Poon et al. (2004). This estimate is also the middle of the three tail indices studied in Ibragimov et al. (2011). The scaling factor of the volatility $\hat{\sigma}$ is half of the riskfree rate, which gives reasonable default probabilities and matches the volatility-risk free rate ratio in De Nicolò et al. (2014). We set $c_0 = c_2$ to set the average cost minimizing size at one.

With the chosen correlation $\rho = 0.5$, the asset portfolios of two banks with maximal distance, i.e., opposite to each other on the unit circle, still have a high correlation between 0.85 and 0.96 if the number of banks is between 5 and 20. This high correlation represents the increased dependence during crisis events. The values for the parameters $a = 25$ and $b = 25$ of the demand function imply that the number of banks Q/x is less than 25 if all banks minimize operational costs at $x = 1$. The number of banks equals 20 if the excess return $\mu^* - R_f$ on assets equals the risk free rate r_f. We set the systemic crisis threshold δ_S at 0.10. That is, 10% of the banks in distress represents a systemic crisis. By setting $\lambda = 1$, the surplus on banking operations vanishes in case of a systemic crisis.

4.5.2 Numerical results: three optima

We run the optimization procedure on a two-dimensional grid with intervals of 0.005. The first dimension corresponds to size and runs from 0.8 to 1.2, which is centered around the average cost minimizing size of one. The second dimension represents the capital ratio y and runs from 5% to 40%.[13]

Table 4.5 reports the optimal results in the baseline setup. First, we consider the size in each optimum. The competitive size is 5% larger than the size of one that minimizes the average cost. In the total surplus optimum, the optimal size is close to one because larger banks crowd out other banks in financing projects. Having fewer banks would be suboptimal from a social point of view. The social welfare optimum takes systemic risk into account. This reduces the optimal size even further because more banks corresponds to a more stable system. Although the socially optimal size is smaller than in the total surplus optimum, the number of banks is also lower. This stems from a stronger preference for profitable projects to increase the mean return on assets. The higher return reduces default probabilities, and hence systemic risk.

Next, we consider the capital ratio. The capital ratio of the total surplus optimum is substantially higher than the capital ratio in the competitive

[13] Our computations would extend the grid if the optimum is at a boundary point at the initial grid.

equilibrium. Since equity is not tax deductible, the total surplus is higher for higher capital ratios. This result coincides with the direct impact of y on TS in (4.26). By taking into account the additional effect of the probability of a systemic crisis, the capital ratio for maximal social welfare increases further in order to reduce default probabilities. Notably, a higher capital ratio in the three optima is associated with a smaller size.

Considering individual and systemic default probabilities, moving from the optimum of CE to TS and from TS to SW shows that the probability of default PD and systemic risk PS are both decreasing. Indeed, this ordering corresponds to the different weight each equilibrium attaches to the stability of the system.

The net contracted interest rate r_c is close to the riskfree rate r_f of 10%.[14] This indicates that our results would not be materially affected by imposing deposit insurance which would bring r_c only marginally down to r_f.[15]

The welfare measures TS and SW in the two bottom rows of Table 4.5 show that the competitive equilibrium has a lower social welfare that stems from systemic risk effects. However, by comparing the TS and SW optimum we find that maximizing total surplus increases capital ratios sufficiently such that systemic risk is substantially reduced, even without taking systemic risk explicitly into account.

4.5.3 Policy implications

We consider policies that limit size below the size of the competitive equilibrium, 1.05. Similarly, we study capital ratio requirements requiring a ratio above the capital ratio of the competitive equilibrium (10.0%). We consider the size and the capital ratio in the social welfare optimum, respectively 1.00 and 37.5%, as the other bound for such policies.

The results are shown in Figure 4.2. The top plots present for each of the three optima the capital ratio that maximizes social welfare and the corresponding level of social welfare. The size varies between the competitive equilibrium size and the size for optimal social welfare. The horizontal lines in the figure suggest that a policy which restricts size is an ineffective policy. More specifically, the optimal capital ratio for banks and society do both not change under different size restrictions. This result is in line with Proposition 4 although the model assumptions differ from the conditions in this

[14] Recall that the corresponding horizon of our long-run model exceeds 5 years.

[15] In our model, more risk leads to a higher contracted interest rate on deposits without affecting the expected return on deposits. This corresponds to an insurance scheme with a fair insurance premium.

		CE	TS	SW
x	size	1.050	1.005	1.000
$y(\%)$	capital ratio	10	24	37.5
n	number of banks	21.1	21.7	21.5
$PD(\%)$	probability of default	3.8	0.22	0.04
$PS(\%)$	probability of systemic crisis	7.8	0.46	0.08
$r_A(\%)$	net mean asset return	11.5	12.6	14.0
$r_c(\%)$	net contracted interest rate	10.0	10.1	10.1
TS	total surplus on bank operations	10.0	10.1	10.1
SW	social welfare on bank operations	9.2	10.0	10.1

Table 4.5: **The three optimums in the baseline model.**
This table reports several variables at the three different optimums. CE is the competitive equilibrium ($\min_{x,y} \mu^*(x,y)$), TS is the total surplus optimum ($\max_{x,y} TS$), and SW is the social welfare optimum ($\max_{x,y} SW$). The parameter values are in Table 4.4.

proposition. The condition $\eta = r_{\bar{T}}$ leads to the result that the competitive capital ratio is insensitive to size. The level of social welfare appears to be insensitive to size when optimizing social welfare.

By varying the imposed capital ratio y in each of the three optima, the bottom plots in Figure 4.2 show the size that maximizes social welfare and the corresponding level of social welfare. The optimal size is shown to be sensitive to the capital ratio. As such, policies that restrict the capital ratio do affect the optimal size. For given capital ratio y, the optimal size is the same in the competitive equilibrium and the optimum maximizing total surplus. This is in line with our result that $x^c(y) = x^{TS}(y)$. Therefore, differences in the optimal competitive size and the optimal size for social welfare are explained by differences in systemic risk rather than total surplus.

With a low capital requirement, banks prefer a larger size than socially optimal (left bottom plot). The intuition is that the default probability is high under a low capital requirement. Then, it is important for individual banks to utilize the diversification benefits. By contrast, it is for the social optimum necessary to have a large number of banks to prevent a systemic crisis. This leads to smaller banks in the social optimum. As the capital requirement becomes more restrictive, i.e., a higher capital ratio closer to the social optimum, the difference between the size of the two optima shrinks. Nonetheless, the social welfare cost due to the size difference is always small (right bottom plot). This is in line with the observation that policies on size have small effects on social welfare. In any case, social welfare improves by

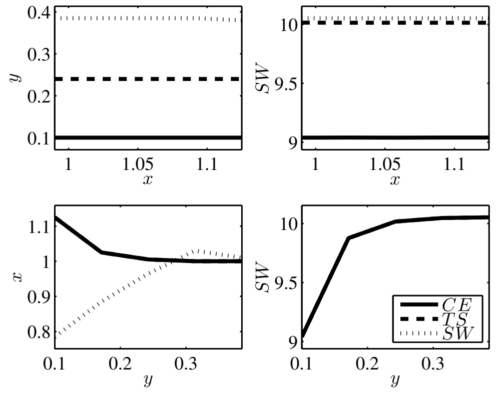


Figure 4.2: **Size and capital restrictions.**

The figure plots the optimal size x, the optimal capital ratio y and social welfare SW as a function of size and capital. CE is the competitive equilibrium ($\min_{x,y} \mu^*(x, y)$), TS is the total surplus optimum ($\max_{x,y} TS$), and SW is the social welfare optimum ($\max_{x,y} SW$). The solid line and dashed line coincide in the two bottom plots with a restriction on capital since $x^c(y) = x^{TS}(y)$. Parameter values are in Table 4.4.

a stricter capital requirement.

We highlight the following two observations from Figure 4.2. First, a policy that raises capital ratios is more effective than a policy that restricts size. Second, it is not necessary to raise the capital requirement to the socially optimal level. In our calibration, a policy with a capital ratio at 18% instead of the social optimal 37.5% suffices to raise social welfare close to its optimal value.

4.5.4 Sensitivity analysis

We check the robustness of our results by calibrating our model with different parameter choices. This serves simultaneously as an investigation on the impact of alternative policies.

The impact of asset dependence

It is generally acknowledged that correlations are higher during crisis periods. Instead of the baseline choice $\rho = 0.5$, we repeat the exercise using a low correlation ($\rho = 0.2$) and a high correlation ($\rho = 0.8$). Table 4.6 shows that the correlation has only a clear impact on the size in the competitive

	$\rho = 0.2$			$\rho = 0.8$		
	CE	TS	SW	CE	TS	SW
x	1.125	1.005	0.990	1.015	1.000	1.000
$y(\%)$	10	24	38.5	10	24	36
n	19.7	21.7	21.7	21.8	21.8	21.5
$PD(\%)$	3.6	0.21	0.04	3.9	0.22	0.05
$PS(\%)$	9.6	0.54	0.08	6.1	0.36	0.08
$r_A(\%)$	11.5	12.6	14.1	11.5	12.6	13.8
$r_c(\%)$	10.6	10.0	10.0	10.6	10.1	10.0
TS	10.0	10.1	10.1	10.0	10.1	10.1
SW	9.1	10.0	10.1	9.4	10.0	10.1

Table 4.6: **Sensitivity to asset dependence.**
This table reports several variables at the three different optimums. Except for the asset dependence ρ, the parameter values are as in Table 4.4. The definition of CE, TS, SW, and the variables in the rows is in Table 4.5.

equilibrium. The difference between the competitive size and the socially optimal size is most pronounced in a prosperous economy where correlations are low. The left plot in Figure 4.3 shows this effect for a wider range of ρ.

The left panel in Figure 4.3 indicates that a policy that restricts size is potentially helpful for small ρ. The more substantial diversification benefits with a small ρ provokes banks to diversify more by growing large. In doing so, they neglect the externality on the smaller number of banks in the economy. Consequently, if the correlation ρ is low, the loss in social welfare at the competitive equilibrium is higher, and the level of social welfare is lower (right plot). Nonetheless, a size restriction leads under $\rho = 0.2$ again to a suboptimal capital ratio and hence a suboptimal social welfare (not plotted).

The middle plot suggests a general opportunity for policy interventions on the capital ratio. That is, the state of the economy as measured by the asset dependence ρ has hardly any effect on the optimal capital ratio and the welfare gain of a policy that restricts the capital ratio. Our qualitative conclusions regarding policies that restrict size or capital ratios remain unchanged.

The impact of demand elasticity

We consider the effect of a bad economy where the amount of project investments is highly sensitive to the offered interest rate. In this setting, the elasticity of demand b is 100 rather than $b = 25$. Consequently, the amount

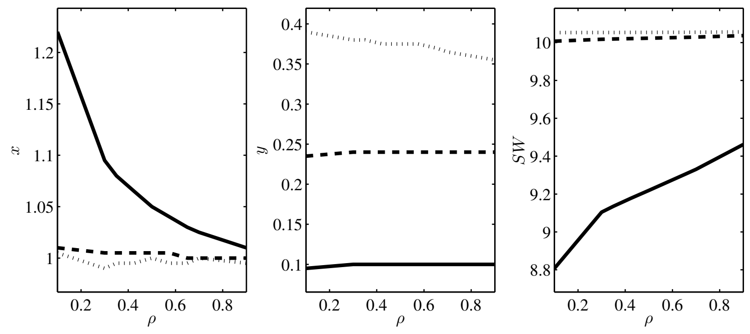


Figure 4.3: **Sensitivity plots of the asset dependence.**
The figure presents the optimal size x, the optimal capital ratio y, and social welfare SW for several values of the asset dependence parameter ρ. The solid line is for the competitive equilibrium ($\min_{x,y} \mu^*(x, y)$), the dashed line is for maximal total surplus ($\max_{x,y} TS$), and the dotted line is for maximal social welfare ($\max_{x,y} SW$). Baseline parameter values are in Table 4.4.

of investments Q decreases dramatically. Figure 4.4 shows the impact of policies on size and capital. Compared to Figure 4.2, the lower amount of investments reduces social welfare substantially (right plots), and results in a smaller number of banks (not plotted).

The qualitative result that the socially optimal capital ratio is insensitive to the size remains unchanged (top left plot). Nevertheless, with the high demand elasticity the socially optimal capital ratio is closer to the competitive equilibrium capital ratio. This is because by the higher demand elasticity a deviation from the competitive capital ratio leads to a higher penalty in terms of written loans Q. Still, social welfare of the competitive equilibrium is again about 10% lower than the social optimum (top right plot).

Another robust result is that compared to size restrictions, capital requirements are more effective in raising social welfare. For low capital ratios, the probability of a systemic crisis is much higher due to the higher default probabilities. With a less strict policy on the capital ratio, it turns out to be socially optimal to reduce systemic risk by increasing diversification benefits through larger banks (bottom left plot), at the expense of a smaller number of banks. In other words, a small number of large banks is socially optimal for small capital ratios.

By requiring a slightly higher capital ratio at about 12%, the optimal size is aligned for both the competitive equilibrium and the social optimum. For

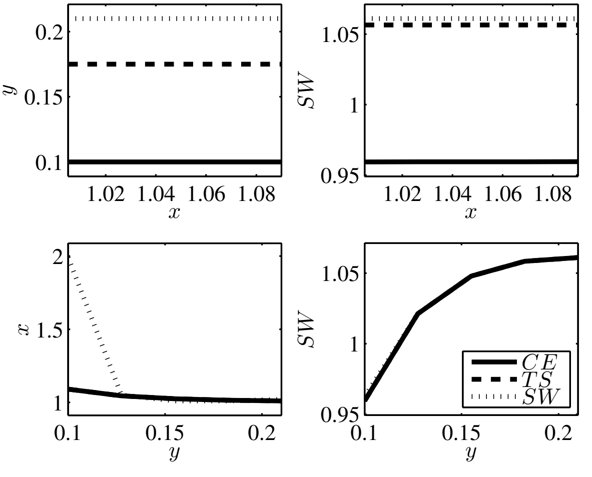


Figure 4.4: **Size and capital restrictions for a high demand elasticity.** The figure plots the optimal size x, the optimal capital ratio y and social welfare SW as a function of size and capital. CE is the competitive equilibrium ($\min_{x,y} \mu^*(x,y)$), TS is the total surplus optimum ($\max_{x,y} TS$), and SW is the social welfare optimum ($\max_{x,y} SW$). The solid line and the dashed line coincide in the two bottom plots with a restriction on capital since $x^c(y) = x^{TS}(y)$. Parameter values are in Table 4.4, except for $b = 100$.

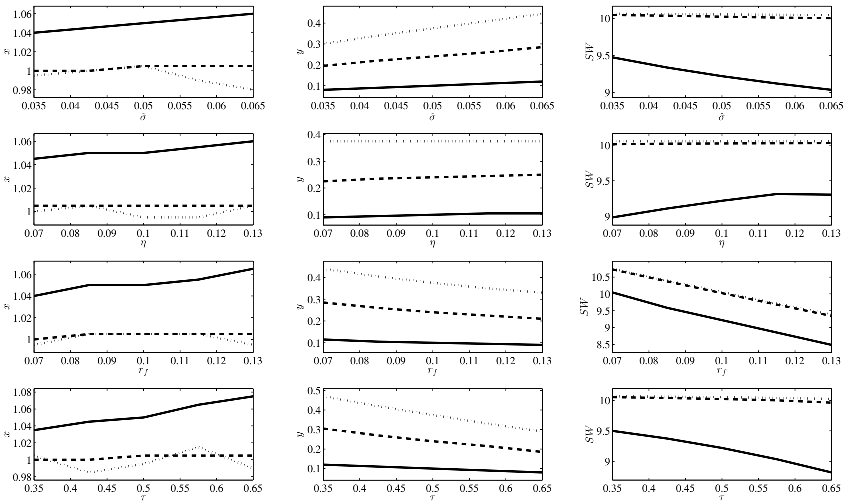


Figure 4.5: **Sensitivity plots.**

This figure presents the optimal size x, the optimal capital ratio y, and social welfare SW by varying the uncertainty $\hat{\sigma}$, the bankruptcy cost η, the risk free rate r_f, or the tax rate τ. The solid line is for the competitive equilibrium ($\min_{x,y} \mu^*(x,y)$), the dashed line is for maximal total surplus ($\max_{x,y} TS$), and the dotted line is for maximal social welfare ($\max_{x,y} SW$). Baseline parameter values are in Table 4.4.

such capital ratios, social welfare is also raised to a comparable level (bottom right plot). Again, it is not necessary to impose a capital requirement close to the social optimum. A capital requirement of 15% brings social welfare already close to the social optimum.

The impact of other parameters

Figure 4.5 reports the impact of the parameters $\hat{\sigma}$, η, r_f and τ at the optima. The discussion focuses on the competitive equilibrium (solid line) and the social optimum (dotted line). First, the result that banks tend to prefer a larger size and a lower capital ratio than socially optimal is robust for all parameters. Thus, policies on size and capital have at least some welfare enhancing effect.

Second, focusing on the uncertainty parameter $\hat{\sigma}$ (the top row), the competitive size increases with the uncertainty which is in contrast to Proposition

5. The reason is that the fat-tailed shocks considered here have a different impact on the probability of default than the uniform shocks considered in the proposition. The welfare loss at the competitive equilibrium is maximal for a high uncertainty (right plot), because banks overdiversify to overcome the high uncertainty. The policy implications in an economy with a high uncertainty are similar to the low correlation case (Figure 4.3). That is, a general need for capital policies regardless of the business cycle.

Third, a low bankruptcy cost η corresponds to the highest welfare loss in the competitive equilibrium. For any bankruptcy cost, raising capital requirements improves social welfare. Hence, bankruptcy costs do not play a key role in policy discussions.

Fourth, the impact of the risk-free rate r_f is similar to the analytical results in Table 4.3: When r_f is low, competitive banks are smaller and hold more capital. A low r_f moves the focus from minimizing the financing costs towards minimizing the average operational cost. In addition, a low r_f reduces the tax benefit of debt financing relative to equity financing. As a consequence, banks prefer a smaller size and hold somewhat more capital. Society prefers even smaller banks in order to reduce the probability on a systemic crisis. In general, the welfare loss at the competitive equilibrium does not depend on r_f since welfare decreases in both optimums with a similar magnitude. In our risk-neutral setting, this financing cost r_f is the only parameter that materially affects the optimal level of social welfare. The intuition is that a higher financing cost for banks results in a higher required asset return, which results in a drop in the number of financed projects and hence a lower social welfare. The difference between the socially optimal capital ratio and the competitive capital ratio is maximal when r_f is low. This indicates that when monetary policy is loosening, it is important to tighten prudential policy.

Lastly, concerning the tax rate τ, a high tax leads to a large welfare loss. The higher marginal tax rate amplifies the effect of the lower capital ratio on the surplus in (4.22), and thus social welfare. Therefore, a policy on the capital ratio is most effective when taxes are high.

4.5.5 Summary

Overall, we draw the following robust conclusions from our calibration.

(i) Size preferences of banks and society are more aligned than capital preferences. Consequently, policy should target at capital ratios. A capital policy may endogenously lead to smaller bank sizes which makes a policy aiming at smaller banks obsolete.

(*ii*) Optimal social welfare is approximately obtained when maximizing total surplus. In conjunction with the fact that $x^c(y) = x^{TS}(y)$, a capital requirement is effective in optimizing social welfare. It is not necessary to raise capital requirements close to the social optimum. In our calibration, it suffices to choose a capital requirement close to the average of the capital ratios of the competitive equilibrium and the social optimum.

(*iii*) Policy interventions are particularly helpful if the correlation ρ is low, the uncertainty $\hat{\sigma}$ is high, the bankruptcy cost (or risk aversion) η is low, and taxes τ are high. This corresponds to periods with major innovations and stringent tax regimes.

(*iv*) The difference between the socially optimal capital ratio and the competitive capital ratio is maximal when r_f is low. Thus, a less stringent monetary policy should be accompanied with a more stringent prudential policy.

(*v*) Although the optimal level of social welfare is constant in most of the parameters, the optimal capital requirement to attain this level may change substantially when the parameters change. This suggests a contingent policy on capital requirements.

4.6 Conclusion

We have studied the joint effect of bank size, capital structure, and asset dependence on social welfare. Our analysis takes into account the interaction between size and capital. Namely, higher capital requirements simultaneously result in larger banks if bankruptcy costs are low, and financing costs and taxes are high. We studied policies on size, capital as well as other policy parameters.

Without any regulation, banks take a suboptimal size-capital decision. Banks tend to hold insufficient capital while the size may be too low or too high (e.g., if the demand elasticity is low). Thus, a policy imposing a higher capital ratio always helps to increase social welfare. By contrast, a restriction on bank size has a minor impact particularly in a bad economy where correlations are high.

Our main conclusion is that capital requirements are more effective than policy measures on size. Capital requirements may have an additional channel to limit systemic risk through its effect on size. This effect is relevant in economies where dependencies are low, uncertainty is high, bankruptcy costs (or risk aversion) is low, and taxes are high.

Our study contributes to the policy debate in at least the following two related directions. First, our results support the countercyclical capital buffer in Basel III. Building up capital during a good economy (low asset dependence) is more effective for social welfare. Instead, a strict capital requirement in bad times with high asset dependence is not particularly effective.

Second, our result illustrates the joint welfare effect of monetary policy and prudential policy. The financing cost is the only parameter that materially affects the optimal level of social welfare. It indicates that a less stringent monetary policy should be accompanied with a more stringent prudential policy.

Appendix

Appendix

4.A Proofs

Proof of Proposition 1
Risk neutrality ($\mathbb{E}[R_D] = R_f$), together with $D = x(1-y)$ and (4.2) implies (4.7). Applying (4.7) to (4.4) yields

$$\begin{aligned}\mathbb{E}[E_{BT}] &= \mathbb{E}\big[[R_A x - c(x) - R_c D]^+\big] \\ &= x\mathbb{E}[[R_A - \bar{c}(x) - (1-y)\,R_c] - \min\left(R_A - \bar{c}(x) - (1-y)\,R_c, 0\right)] \\ &= x\mathbb{E}[R_A - \bar{c}(x) - \min\left(R_A - \bar{c}(x), (1-y)\,R_c\right)] \\ &= x\left[\mu - \bar{c}(x) - (R_f + \eta PD)\,(1-y)\right].\end{aligned}$$

Consequently, the expected after-tax return on the bank's equity equals

$$\begin{aligned}\mathbb{E}[r_E] &= \frac{\bar{\tau}\left(\mathbb{E}[E_{BT}] - E\right)}{E} \\ &= \frac{\bar{\tau}}{y}\left(\mu - \bar{c}(x) - 1 - (r_f + \eta PD)\,(1-y)\right)\end{aligned} \tag{4.34}$$

We now obtain equation (4.5) from the risk neutrality condition $\mathbb{E}[r_E] = r_f$.

Next, we prove the remaining equilibrium equation (4.6). Define the default threshold of X

$$h(\mu, R_c; x) = \frac{\bar{c}(x) + R_c(1-y) - \mu}{\sigma(x,\mu)} \tag{4.35}$$

such that a default corresponds to the event $\{X < h\}$. In the following derivation, we suppress the fixed argument x for convenience.

By using $R_A = \mu^* + \sigma^* X$ in the banking equilibrium, we rewrite (4.7) and (4.5) as respectively

$$\mathbb{E}[\min\left(X, h\right)] = \frac{\bar{c}(x) - \mu^* + (1-y) + (r_f + \eta\mathbb{P}(X < h))\,(1-y)}{\sigma(x,\mu^*)} \tag{4.36}$$

$$0 = \frac{\bar{c}(x) - \mu^* + 1 + \frac{r_f}{\bar{\tau}}y + (r_f + \eta\mathbb{P}(X < h))(1-y)}{\sigma(x, \mu^*)} \tag{4.37}$$

Subtracting (4.37) from (4.36) results in

$$\mathbb{E}[\min(X, h)] = -\frac{y}{\sigma(x, \mu^*)}\left(1 + \frac{r_f}{\bar{\tau}}\right) \tag{4.38}$$

Equation (4.6) follows by noting that the equilibrium mean asset return μ^* solves the equilibrium equation (4.38), and $F^{-1}(PD)$ is another expression for the default threshold h.

□

Proof of Proposition 2

We find conditions for existence and uniqueness of the banking equilibrium (R_c^*, μ^*, PD^*). Following the proof of Proposition 1, h refers to the default threshold for X. Given μ^* and x, the function $h : \mathbb{R} \to \mathbb{R}$ in (4.35) is a bijection of the gross contracted interest rate on debt (R_c) to the default threshold h. As such, we find conditions for existence and uniqueness of the equilibrium (h^*, μ^*).

Existence

By (4.38), an equilibrium necessarily satisfies

$$-\frac{1 + r_f/\bar{\tau}}{\sigma(x, \mu^*)}y \leq h(\mu^*) \leq \sup(X) \tag{4.39}$$

By $\sigma_2 \geq 0$, it suffices for existence to have

$$-\frac{1 + r_f/\bar{\tau}}{\sigma(x, \mu^l)}y \geq \inf(X) \tag{4.40}$$

Hence, the condition in (4.10) is sufficient since $y \leq 1$.

Next, we derive the sufficient condition for existence in (4.11). Since $\sigma_2 \equiv 0$, we have $\tilde{\sigma}(x) := \sigma(x, \mu(x, y))$ such that $\tilde{\sigma}(x)$ is not affected by R_c^* through μ. Substituting (4.1) and (4.5) into (4.7), we get

$$\mathbb{E}\Big[\min\left(\tilde{\sigma}(x)X + \frac{r_f}{\bar{\tau}}y, (1-y)(R_c^* - r_f - \eta PD) - 1\right)\Big] = -y \tag{4.41}$$

For $R_c^* = 0$, the right hand side (RHS) exceeds the left hand side (LHS) since

$$\text{LHS}|_{R_c^*=0} < -(1-y)(r_f + \eta PD) - 1 < -1 < -y = \text{RHS}|_{R_c^*=0}$$

while for $R_c^* \to \infty$ the LHS exceeds the RHS. The continuity in R_c^* of the LHS of (4.41) implies that (4.41) has a solution in R_c^*. This implies the sufficient condition in (4.11).

Uniqueness
Under (4.11), the RHS of (4.6) is constant which pins down PD^* uniquely. Then, μ^* and R_c^* are also unique by (4.5) and (4.7), respectively.

□

Proof of Proposition 3
The competitive equilibrium is the special case of the banking equilibrium where the mean asset return μ^* is minimal. The corresponding variables are denoted with a superscript c.

(*i*) We find from (4.9), $\mu_1^c = \mu_2^c = 0$, $\sigma_1(x,\mu) < 0$ and $\sigma(x,\mu) > 0$, the partial derivatives of $PD^*(x,y)$ at the competitive equilibrium:

$$PD_1^c = \frac{\partial}{\partial x} PD^*(x,y)\bigg|_{(x,y)=(x^c,y^c)} \tag{4.42}$$

$$= \frac{\partial}{\partial x} g\left(\frac{\sigma(x,\mu^*(x,y))}{(1+r_f/\bar{\tau})\,y}\right)\bigg|_{(x,y)=(x^c,y^c)} \tag{4.43}$$

$$= (g')^c \frac{\sigma_1(x,\mu^c) + \sigma_2(x,\mu^c)\mu_1^c}{(1+r_f/\bar{\tau})y^c} < 0 \tag{4.44}$$

$$PD_2^c = (g')^c \frac{\partial}{\partial y} \frac{\sigma(x^c,\mu^*(x,y))}{(1+r_f/\bar{\tau})y}\bigg|_{y=y^c}$$

$$= -(g')^c \frac{\sigma(x^c,\mu^c)/y^c + \sigma_2(x^c,\mu^c)\mu_2^c}{(1+r_f/\bar{\tau})y^c} < 0 \tag{4.45}$$

The inequalities in (4.44) and (4.45)) imply under (C)

$$PD_x^c = PD_1^c + \frac{\mathrm{d}x^c(y^c)}{\mathrm{d}y} PD_2^c < 0$$

$$PD_y^c = \frac{\mathrm{d}y^c(x^c)}{\mathrm{d}x} PD_1^c + PD_2^c < 0$$

For the operational cost function $\bar{c}(x)$, it immediately follows that $\frac{\partial}{\partial y}\bar{c}(x) = 0$. Taking the partial derivative of (4.5) to x and using $\mu_1^c = 0$ and (4.44) gives $\bar{c}'(x^c) > 0$. Thus, (C) is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}y}\bar{c}(x^c(y)) = \bar{c}'(x^c)\frac{\mathrm{d}x^c(y^c)}{\mathrm{d}y} > 0$$

(ii) The claim on the derivatives in case of $\sigma_1^c = 0$ is straightforward from the representation of PD in (4.6) in Proposition 1.

(iii) Because $\mu^c = \mu^*(x^c, y^c)$ has a positive definite Hessian matrix at (x^c, y^c):

$$\mu_{11}^c \mu_{22}^c - (\mu_{12}^c)^2 > 0 \qquad \mu_{11}^c > 0 \qquad \mu_{22}^c > 0 \tag{4.46}$$

Let $y^c(x)$ as in (4.12), which means $\mu_2^*(x, y^c(x)) = 0$ for all x. Differentiating to x,

$$\frac{\mathrm{d}}{\mathrm{d}x}\mu_2^*(x, y^c(x)) = \mu_{12}^*(x, y^c(x)) + \mu_{22}^*(x, y^c(x))\frac{\mathrm{d}y^c}{\mathrm{d}x} = 0$$

Evaluating this expression at $x = x^c$,

$$\frac{\mathrm{d}y^c(x^c)}{\mathrm{d}x} = -\frac{\mu_{12}^c}{\mu_{22}^c}. \tag{4.47}$$

Similarly, $\mu_1^*(x^c(y), y) \equiv 0$ leads to

$$\frac{\mathrm{d}x^c(y^c)}{\mathrm{d}y} = -\frac{\mu_{12}^c}{\mu_{11}^c} \tag{4.48}$$

The cross partial derivative of $\mu(x, y)$ at μ^c follows from (4.5):

$$\mu_{12}^c = [(1 - y^c)\, PD_{12}^c - PD_1^c]\, \eta \tag{4.49}$$

By (4.49) and $\eta > 0$, the inequality in (4.13) is equivalent to $\mu_{12}^c < 0$. By (4.46) – (4.48), this is equivalent to condition (C).

□

Proof of Proposition 4 and 5

We combine the proofs of Proposition 4 and 5 because of the similar structure of the derivations. Existence and uniqueness of the banking equilibrium follows from $\sigma_2(x, \mu) \equiv 0$ in Proposition 2. From the cdf of X,

$$F(x) = \frac{1}{2\sqrt{3}}\left(x + \sqrt{3}\right) \qquad -\sqrt{3} \le x \le \sqrt{3},$$

we define for $PD^* \in [0, 1]$

$$F^{-1}(PD^*) = \sqrt{3}\,(2PD^* - 1)\,.$$

Using $\mathbb{P}\big(X < F^{-1}(PD^*)\big) = PD^*$,

$$\begin{aligned}\mathbb{E}\big[\min\left(X, F^{-1}(PD^*)\right)\big] &= F^{-1}(PD^*)(1-PD^*) + \frac{F^{-1}(PD^*)-\sqrt{3}}{2}PD^* \\ &= F^{-1}(PD^*)\left(1-\frac{PD^*}{2}\right) - \frac{\sqrt{3}}{2}PD^* \\ &= \frac{\sqrt{3}}{2}\left[(2PD^*-1)(2-PD^*)-PD^*\right] \\ &= -\sqrt{3}\,(PD^*-1)^2\end{aligned}$$

Substitution of (4.6), (4.14), and $PD^* \leq 1$ yields for the banking equilibrium of a given (size, capital)-pair (x, y):

$$PD^*(x,y) = 1 - \sqrt{\frac{(1+r_f/\bar{\tau})\,y}{\sigma_0^2 x^{-2\delta}}} = 1 - w(x)\sqrt{y} \tag{4.50}$$

where we have defined for brevity the positive and increasing function

$$w(x) := \sqrt{1+\frac{r_f}{\bar{\tau}}}\frac{x^\delta}{\sigma_0} \tag{4.51}$$

with, as before, $r_f \geq 0$, $\bar{\tau} \in (0,1]$, $\sigma_0 > 0$ and $0 < \delta \leq \frac{1}{4}$. The partial derivatives up to the second order of PD^* in (4.50) are

$$PD_1^*(x,y) = -w'(x)\sqrt{y} \leq 0 \tag{4.52}$$

$$PD_2^*(x,y) = -\frac{w(x)}{2\sqrt{y}} \leq 0 \tag{4.53}$$

$$PD_{11}^*(x,y) = -w''(x)\sqrt{y} \tag{4.54}$$

$$PD_{12}^*(x,y) = -\frac{w'(x)}{2\sqrt{y}} \tag{4.55}$$

$$PD_{22}^*(x,y) = \frac{w(x)}{4y\sqrt{y}} \tag{4.56}$$

Differentiating (4.5), and substituting (4.52)-(4.56)

$$\mu_1^*(x,y) = \bar{c}'(x) - \eta w'(x)\sqrt{y}\,(1-y) \tag{4.57}$$

$$\mu_2^*(x,y) = -\eta\left[\frac{1}{2\sqrt{y}}w(x) - \frac{3}{2}\sqrt{y}w(x) + 1\right] + r_f\left(\frac{1}{\bar{\tau}} - 1\right) \tag{4.58}$$

$$\mu_{11}^*(x,y) = \bar{c}''(x) - \eta w''(x)\sqrt{y}\,(1-y) \tag{4.59}$$

$$\mu_{12}^*(x,y) = \frac{1}{2}\eta w'(x)\frac{1}{\sqrt{y}}(3y-1) \tag{4.60}$$

$$\mu_{22}^*(x,y) = \eta w(x)\left(\frac{1+3y}{4y\sqrt{y}}\right) \tag{4.61}$$

The results above are for the banking equilibrium of any given (x, y). Next, we apply these results to the competitive equilibrium (x^c, y^c), which is the special case where the mean asset return $\mu^*(x, y)$ is minimal. As before, we denote the corresponding variables with a superscript c. The first and second order derivative of $w(x)$ at x^c are denoted by w_1^c and w_{11}^c, respectively. We adopt a similar notation for $\bar{c}^c$, the average cost function $\bar{c}(x)$ at x^c.

The inequality (4.13) in Proposition 3 holds if and only if $y^c < \frac{1}{3}$. To see this, use (4.52) and (4.55) to find for $y^c < \frac{1}{3}$,

$$(1-y^c)\,PD_{12}^c = \left(1-\frac{1}{y^c}\right)\frac{w_1^c\sqrt{y^c}}{2} < -w_1^c\sqrt{y^c} = PD_1^c$$

Thus, by Proposition 3(iii), the capital ratio increases in the size $(y_x^c > 0)$ and the size increases in the capital ratio $(x_y^c > 0)$ if and only if $y^c < \frac{1}{3}$.

In line with our technical assumption, the choice $\bar{c}(x) = x^\gamma$ with $\gamma > 1$ leads to properly defined functions $x^c(y)$ and $y^c(x)$ near the competitive equilibrium. More specifically, (i) μ^* is convex in x along $(x, y^c(x))$, and (ii) μ^* is convex in y along $(x^c(y), y)$. Statement (i) follows from isolating $\eta\sqrt{y}(1-y)$ in (4.57) with $\mu_1^c = 0$, and substituting this result into (4.59) with the imposed restriction $\delta < \gamma$:

$$\mu_{11}^c = \gamma\,(\gamma-1)\,x^{\gamma-2} - \gamma\,(\delta-1)\,x^{\gamma-2} > 0$$

Statement (ii), the convexity of μ^* in y at $(x^c(y), y)$, is straightforward from μ_{22}^* in (4.61). This confirms that $x^c(y)$ and $y^c(x)$ lead to the unique minimum of μ^* conditional on y and x, respectively.

Next, we find the unique pair (x^c, y^c) of the competitive equilibrium in terms of the parameters. The size $x^c(y)$ minimizes μ given the capital ratio y. From $\mu_1(x^c(y), y) = 0$, (4.15), (4.51) and (4.57), we obtain

$$x^c(y) = \left(\frac{\eta\tilde{R}_f\delta}{\gamma\sigma_0}\sqrt{y}\,(1-y)\right)^{1/(\gamma-\delta)}. \tag{4.62}$$

The function $x^c(y)$ has an inverted U-shape for the optimal size x as a function of the capital ratio y. This optimal size is zero if $y = 0$ or $y = 1$, and maximal at $y = \frac{1}{3}$, i.e.,

$$0 \leq x^c(y) \leq \left(\frac{2\eta \tilde{R}_f \delta}{3\sqrt{3}\gamma\sigma_0} \right)^{1/(\gamma-\delta)} := x_{\max}.$$

The capital ratio $y^c(x)$ minimizes μ^* given the size x. From $\mu_2(x, y^c(x)) = 0$, (4.15), (4.16), (4.51) and (4.58), we obtain

$$0 = \frac{1}{2\sigma_0}\eta \tilde{R}_f x^\delta \left(\frac{1}{\sqrt{y^c(x)}} - 3\sqrt{y^c(x)} \right) + \eta - r_\tau \tag{4.63}$$

which gives

$$x^\delta = \frac{2\left(\eta - r_\tau\right)\sigma_0}{\eta \tilde{R}_f \left(3\sqrt{y^c(x)} - \frac{1}{\sqrt{y^c(x)}} \right)} \qquad y^c(x) \neq \frac{1}{3} \tag{4.64}$$

Equation (4.64) defines $y^c(x)$ implicitly. The denominator always increases with $y^c(x)$ and switches sign at $y^c(x) = \frac{1}{3}$, which implies $\lim_{x\to\infty} y^c(x) = \frac{1}{3}$. This corresponds to $\mu_{12} = 0$ in (4.60), and $x^c_y = y^c_x = 0$. As a consequence, the sign of the numerator in (4.64) has an important effect. It indicates if the size and the capital ratio move in the same direction. This has important implications for the effect of the parameters. Based on the sign of the numerator of (4.64), we distinguish three cases:

(i) $\eta > r_\tau$:
The positive numerator in (4.64) and $x^\delta \geq 0$ imply $y^c(x) > \frac{1}{3}$, and $y^c(x)$ decreases in x. Capital ratios $y^c(x) \leq \frac{1}{3}$ are excluded because no $x \geq 0$ corresponds to such $y^c(x)$. Thus, $y^c(x)$ is a decreasing function with $y^c\left(\left(\frac{(\eta - r_\tau)\sigma_0}{\eta\tilde{R}_f}\right)^{1/\delta}\right) = 1$ and asymptotic lower bound $\lim_{x\to\infty} y^c(x) = \frac{1}{3}$. From $y^c > \frac{1}{3}$, we know that $y^c_x < 0$ and $x^c_y < 0$ hold.

(ii) $\eta = r_\tau$
The terms with $\sqrt{y^c(x)}$ in (4.63) sum to zero. As a consequence, $\{(x, y^c(x)) : x \geq 0\}$ is a horizontal line with $y^c(x) \equiv \frac{1}{3}$ and $y^c_x \equiv 0$.

(iii) $\eta < r_\tau$:
The negative numerator in (4.64) and $x^\delta \geq 0$ give $y^c(x) < \frac{1}{3}$ and $y^c(x)$ increases in x. Here, capital ratios $y^c(x) \geq \frac{1}{3}$ are excluded as no $x \geq 0$ corresponds to such $y^c(x)$. Now, y^c is an increasing function with asymptotic bounds $\lim_{x\downarrow 0} y^c(x) = 0$ and $\lim_{x\to\infty} y^c(x) = \frac{1}{3}$. From $y^c < \frac{1}{3}$, we obtain $y^c_x > 0$ and $x^c_y > 0$.

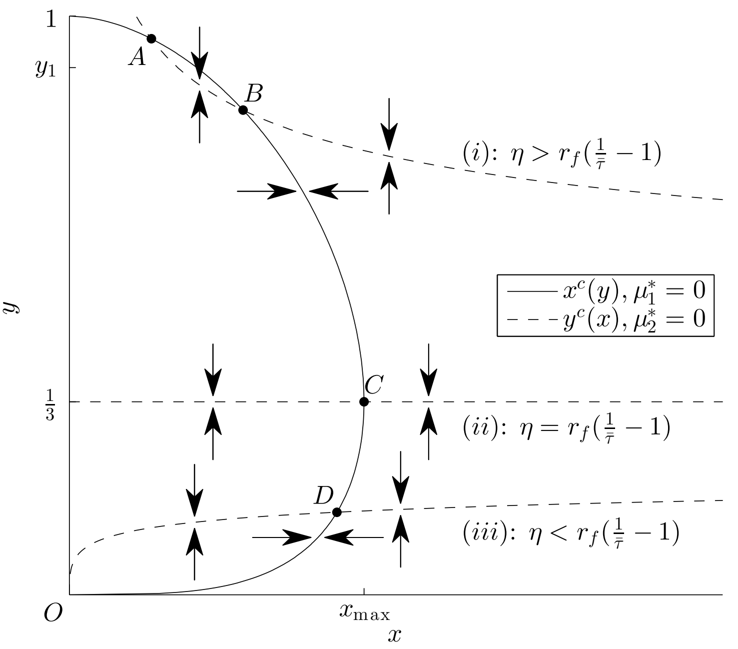


Figure 4.6:

Phase diagram of $x^c(y)$ and the three possible shapes for $y^c(x)$.

The figure shows the optimal size curve $x^c(y)$ and three possible shapes for $y^c(x)$. The parameters y_1 and $x_{\max}$ are defined in (4.17) and (4.21). The curves are defined in (4.62) and (4.64). The shape of $y^c(x)$ is determined by the bankruptcy cost η, the risk free rate r_f, and the tax rate $\tau = 1 - \bar{\tau}$. Horizontal and vertical arrows indicate changes towards the optimal size $x^c(y)$ and the optimal capital ratio $y^c(x)$, respectively.

The phase diagram in Figure 4.6 shows an example for each of the three possible cases. The competitive equilibrium (x^c, y^c) is determined by the cutting points of $x^c(y)$ and $y^c(x)$. In case (i), there are at most two intersection points (A and B in Figure 4.6), while there exists a unique intersection point in cases (ii) and (iii), the points C and D, respectively. We show this formally by studying the analytical properties of the cutting points of the curves $x^c(y)$ and $y^c(x)$, in respectively (4.62) and (4.64).

It is most instructive to consider the curves in $(\sqrt{x}, y)$-space. Rewriting the equilibrium equations (4.62) and (4.64) produces

$$\sqrt{x} = \left(\frac{\eta \tilde{R}_f \delta}{\gamma \sigma_0} \sqrt{y} \, (1-y) \right)^{\frac{1}{2(\gamma-\delta)}} \tag{4.65}$$

$$\sqrt{x} = \left| \frac{2\sigma_0 \left[\eta - r_\tau \right]}{\eta \tilde{R}_f \left(3\sqrt{y} - \frac{1}{\sqrt{y}} \right)} \right|^{\frac{1}{2\delta}}. \tag{4.66}$$

We first show that the RHS of (4.65) is concave in y. After some algebra, the second order derivative to y is

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2} \left(\sqrt{y} \, (1-y) \right)^a = a \left(\sqrt{y} \, (1-y) \right)^a \frac{n_a(y)}{4 \, (y-1)^2 \, y^2}$$

where $a := \frac{1}{2(\gamma-\delta)} \in \left(0, \frac{2}{3}\right)$ and

$$n_a(y) := 3 \, (3a-2) \, y^2 - 2 \, (3a-2) \, y + a - 2.$$

The denominator is nonnegative, while the numerator $n_a(y)$ has an extremum at $y = \frac{1}{3}$. At $y \in \left\{0, \frac{1}{3}, 1\right\}$, the numerator is given by respectively

$$\begin{aligned} n_a(0) &= a - 2 < 0 \\ n_a\left(\frac{1}{3}\right) &= \frac{1}{3}(3a-2) - \frac{2}{3}(3a-2) + a - 2 = -\frac{4}{3} < 0 \\ n_a(1) &= 3 \, (3a-2) - 2 \, (3a-2) + a - 2 = 4a - 4 < 0 \end{aligned}$$

This implies that the numerator is negative for all $y \in (0,1)$, and so $\sqrt{x^c(y)}$ is a strictly concave function at the interval $(0,1)$.

Next, we show that the RHS of (4.66) is convex in y, notice that

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2} \left| 3\sqrt{y} - \frac{1}{\sqrt{y}} \right|^b = b \left| \frac{3y-1}{\sqrt{y}} \right|^b \frac{n_b(y)}{4y^2 \, (3y-1)^2} \tag{4.67}$$

where $b := -\frac{1}{2\delta} < -2$ and

$$n_b(y) = 9 \, (b-2) \, y^2 + 6 \, (b-2) \, y + 2 + b.$$

The numerator term $n_b(y)$ in (4.67) is decreasing if $y \in (0,1)$ with maximum $n_b(0) = 2 + b < 0$. The last denominator term in (4.67) is positive on $(0,1] \backslash \{\frac{1}{3}\}$. The second order derivative of the RHS in (4.66) is equal to

(4.67) scaled by a certain positive constant if $\eta \neq r_\mathcal{T}$. This means that the second order derivative of (4.66) is positive for each $y \in [0,1] \setminus \left\{\frac{1}{3}\right\}$. Therefore, the curve $(\sqrt{x}, y^c(x))$ is convex in y if $\eta \neq r_\mathcal{T}$.

Since a convex line and a concave line share at most two points of intersection, it follows that the curves $\left(\sqrt{x^c(y)}, y\right)$ and $(\sqrt{x}, y^c(x))$ have at most two intersection points if $\eta \neq r_\mathcal{T}$. For $\eta = r_\mathcal{T}$ (case (ii)), this follows from $y^c(x) \equiv \frac{1}{3}$. A monotonic transformation from $(\sqrt{x}, y)$-space back to (x, y)-space will, of course, not change the number of intersection points. For the three different cases, we have

(i) $\eta > r_\mathcal{T}$: In this case, $y^c(x) > \frac{1}{3}$, and the curves have either zero, one, or two points of intersection. The arrows point in the direction of a lower μ^*. In case of no intersection point, it is straightforward from the direction of the arrows that the optimum μ^c (the minimal μ^*) is the boundary point $(x^c, y^c) = (0, 1)$. In case of a single intersection point, any deviation from this point results into an increase in μ^*. Thus, this single intersection point is the minimizer of μ^*. Two intersection points produce an additional unstable equilibrium. The arrows in Figure 4.6 suggest that the equilibrium A in the Northwest with the smaller size x and the higher capital ratio y is the unstable equilibrium. Moving towards the Southeast equilibrium B decreases μ^*. Hence, the unique minimum of μ^* is at B.

(ii) $\eta = r_\mathcal{T}$: The pair $\left(x^c\left(\frac{1}{3}\right), \frac{1}{3}\right)$ is the unique point of intersection C in Figure 4.6.

(iii) $\eta < r_\mathcal{T}$, we have $y^c(x) < \frac{1}{3}$. From $\frac{d}{\mathrm{d}y} x^c(0) \to \infty$, $\lim_{x\to\infty} y^c(x) = \frac{1}{3}$, and the intersection point $x^c(0) = y^c(0) = 0$, we obtain that a unique interior competitive equilibrium (x^c, y^c) exists that satisfies (4.62) and (4.64), simultaneously. The arrows in Figure 4.6 indicate that this interior solution D is the minimizer of μ^*.

Hence, the competitive equilibrium is unique in each of the three cases.

Next, we express the optimal solution (x^c, y^c) in terms of the parameters. Substitution of (4.62) in (4.64) and using the positive sign of x^δ gives

$$\left|3\sqrt{y^c} - \frac{1}{\sqrt{y^c}}\right| \left(\frac{\eta \tilde{R}_f \delta}{\gamma \sigma_0} \sqrt{y^c}\,(1 - y^c)\right)^{\delta/(\gamma-\delta)} = \frac{\sigma_0}{\tilde{R}_f} \left|2 - \frac{2 r_\mathcal{T}}{\eta}\right| \tag{4.68}$$

Define $\zeta := \gamma/\delta > 4$ which gives for any $y^c \in (0,1)\setminus\{\frac{1}{3}\}$ an important expression in terms of the model parameters:

$$\sqrt{y^c}\,(1-y^c)\left|3\sqrt{y^c}-\frac{1}{\sqrt{y^c}}\right|^{\zeta-1}=\frac{\zeta}{\eta}\frac{\sigma_0^\zeta}{\left(1+\frac{r_f}{\bar\tau}\right)^{\zeta/2}}\left|2-\frac{2r_\mathcal{T}}{\eta}\right|^{\zeta-1} \tag{4.69}$$

where we used (4.15). Define for the LHS and RHS of (4.69), respectively:

$$\phi(y;\zeta):=\sqrt{y}\,(1-y)\left|3\sqrt{y}-\frac{1}{\sqrt{y}}\right|^{\zeta-1} \tag{4.70}$$

$$\psi\left(\eta;r_f,\sigma_0,\bar\tau,\zeta\right):=\frac{\zeta}{\eta}\frac{\sigma_0^\zeta}{\left(1+\frac{r_f}{\bar\tau}\right)^{\zeta/2}}\left|2-\frac{2r_\mathcal{T}}{\eta}\right|^{\zeta-1} \tag{4.71}$$

The left panel in Figure 4.7 sketches the shape of ϕ as a function of y. The competitive capital ratio must satisfy $\phi(y^c)=\psi$. As discussed before, $y^c-\frac{1}{3}$ has the same sign as $\eta-r_\mathcal{T}$. Thus, the left panel in Figure 4.7 provides at most two solutions for the equation $\phi(y)=\psi$. When two solutions exist, which may occur if $\eta>r_\mathcal{T}$, we have shown that the smallest solution on $\left(\frac{1}{3},1\right)$ corresponds to the global minimum (x^c,y^c) of $\mu^*(x,y)$. To find an upper bound for this solution, we derive an explicit expression for the point $y_1\in\left(\frac{1}{3},1\right)$ that satisfies $\phi'(y_1)=0$.

To formally study the shape and derivative of ϕ, it is convenient to consider a monotonic transform of ϕ:

$$\begin{aligned}\tilde\phi(y)&:=\ln\left(\phi(y;\zeta)\right)\\&=\ln\left(\sqrt{y}\right)+\ln\left(1-y\right)+(\zeta-1)\ln\left|3\sqrt{y}-\frac{1}{\sqrt{y}}\right|\end{aligned} \tag{4.72}$$

The derivatives of $\tilde\phi(y)$ and $\phi(y)$ have the same sign if $\phi(y)>0$, i.e. if $y\neq\frac{1}{3}$. Hence, y_1 is also a zero of $\tilde\phi'$. The latter derivative is

$$\tilde\phi'(y)=\frac{3(\zeta+2)y^2-2(\zeta+2)y+2-\zeta}{2y(y-1)(3y-1)}$$

For $y^c>\frac{1}{3}$, $\hat\phi'(y^c)>0$ is equivalent to

$$h(y^c):=3\left(\zeta+2\right)\left(y^c\right)^2-2\left(\zeta+2\right)y^c+2-\zeta<0$$

Similarly, if $y^c<\frac{1}{3}$ then $\tilde\phi'$ is positive if and only if $h(y^c)>0$. The zeros of h and so ϕ' are at

$$y_0=\tfrac{1}{3}-\tfrac{2}{3}\sqrt{\tfrac{\zeta-1}{\zeta+2}}<0 \qquad y_1=\tfrac{1}{3}+\tfrac{2}{3}\sqrt{\tfrac{\zeta-1}{\zeta+2}}<1 \tag{4.73}$$

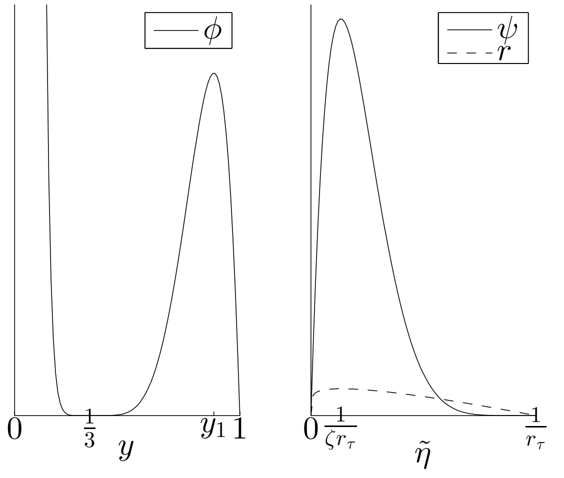


Figure 4.7: **The functions $\phi(y)$, $\psi(\tilde{\eta})$, and $r(\tilde{\eta})$.**
The figure shows the functions $\phi(y)$, $\psi(\tilde{\eta})$, and $r(\tilde{\eta})$ which are defined in respectively (4.70), (4.71), and (4.74).

The function h satisfies $h(0) = 2 - \zeta < 0$ and $h(1) = 4$. It has a global minimum $h(\frac{1}{3}) = \frac{4}{3}(1-\zeta) < 0$. Considering $y^c \in [0,1]$, the function h is only positive at the interval $(y_1, 1]$. Hence, the function ϕ decreases on the interval $I_0 := \left(0, \frac{1}{3}\right) \cup (y_1, 1]$, and increases on $I_1 := \left(\frac{1}{3}, y_1\right)$. The extremal points of the continuous ϕ are

$$\begin{aligned} &\phi(0;\zeta) \to \infty && \phi(y_1;\zeta) = \sqrt{y_1}\,(1-y_1)\left(3\sqrt{y_1} - \frac{1}{\sqrt{y_1}}\right)^{\zeta-1} > 0 \\ &\phi\left(\tfrac{1}{3};\zeta\right) = 0 && \phi(1;\zeta) = 0 \end{aligned}$$

The preceding results imply the following results on the competitive equilibrium (x^c, y^c) with minimal $\mu^*(x^c, y^c) = \mu^c$:

(*i*) $\eta \leq r_\tau$:
a unique equilibrium with $y^c \leq \frac{1}{3}$ and $x^c \leq x_{\max}$

(*ii*) $\eta > r_\tau$ and $\phi(y_1) \geq \psi(\eta; r_f, \sigma_0, \bar{\tau}, \zeta)$:
a unique equilibrium with $y^c \in \left(\frac{1}{3}, y_1\right]$ and $x^c < x_{\max}$

(*iii*) $\eta > r_\tau$ and $\phi(y_1) < \psi(\eta; r_f, \sigma_0, \bar{\tau}, \zeta)$:
a boundary solution $(x^c, y^c) = (0, 1)$

where $x_{\max} = \left(\frac{2\eta\tilde{R}_f}{3\sqrt{3}\sigma_0\zeta}\right)^{1/(\gamma-\delta)}$ and $y_1 = \frac{1}{3} + \frac{2}{3}\sqrt{\frac{\zeta-1}{2+\zeta}}$.

To guarantee an interior solution in case (iii), we derive a sufficient condition which ensures $\phi\left(y_1\right) \geq \psi\left(\eta; r_f, \sigma_0, \bar{\tau}, \zeta\right)$. Define $\tilde{\eta} := \frac{1}{\eta}$, and

$$a := 2\left(\frac{\zeta \sigma_0^{\zeta}}{(1 + r_f/\bar{\tau})^{\zeta/2}}\right)^{\frac{1}{\zeta-1}} > 0$$

$$r(\tilde{\eta}) := \psi^{\frac{1}{\zeta-1}} = a\tilde{\eta}^{\frac{1}{\zeta-1}}\left(1 - r_\tau \tilde{\eta}\right) \tag{4.74}$$

The right panel in Figure 4.7 sketches the shape of ψ and r as a function of $\tilde{\eta}$. As we are considering case (iii), it follows that $\tilde{\eta} < 1/r_\tau$ holds. The condition $\phi\left(y_1\right) \geq \psi$ is thus equivalent to

$$\phi\left(y_1\right)^{\frac{1}{\zeta-1}} \geq r(\tilde{\eta}). \tag{4.75}$$

Notice that $r(0) = r\left(\frac{1}{r_\tau}\right) = 0$, and

$$r'(\tilde{\eta}) = a\tilde{\eta}^{\frac{1}{\zeta-1}}\left(\frac{1}{\zeta-1}\left(\frac{1}{\tilde{\eta}} - r_\tau\right) - r_\tau\right) = \frac{a}{\zeta-1}\tilde{\eta}^{\frac{2-\zeta}{\zeta-1}}\left(1 - \zeta r_\tau \tilde{\eta}\right)$$

$$r'\left(\frac{1}{\zeta r_\tau}\right) = 0$$

$$r''(\tilde{\eta}) = -\frac{a}{\zeta-1}\tilde{\eta}^{\frac{2-\zeta}{\zeta-1}}\left(\frac{\zeta-2}{\zeta-1}\frac{1}{\tilde{\eta}} + \zeta r_\tau\right) < 0$$

Thus, the function r attains its maximum on $[0, \frac{1}{r_\tau}]$ at $\frac{1}{\zeta r_\tau}$. Therefore, (4.75) holds with equality for some $\tilde{\eta}$ if and only if

$$\max_{\tilde{\eta}} r(\tilde{\eta}) = r\left(\frac{1}{\zeta r_\tau}\right) \geq \phi\left(y_1\right)^{\frac{1}{\zeta-1}}. \tag{4.76}$$

If condition (4.76) holds strictly then two distinct $\tilde{\eta}$ at $\left(0, \frac{1}{r_\tau}\right)$ solve $r(\tilde{\eta}) = \phi(y_1)^{\frac{1}{\zeta-1}}$. One solution $\tilde{\eta}_0 \in \left(0, \frac{1}{\zeta r_\tau}\right)$ and another solution $\tilde{\eta}_1 \in \left(\frac{1}{\zeta r_\tau}, \frac{1}{r_\tau}\right)$. In addition, $r(\tilde{\eta}) > \phi(y_1)^{\frac{1}{\zeta-1}}$ is equivalent to $\tilde{\eta} \in (\tilde{\eta}_0, \tilde{\eta}_1)$ because $r'' < 0$. Then, no equilibrium exists on the interval $(\tilde{\eta}_0, \tilde{\eta}_1)$ between the two solutions. For such $\tilde{\eta}$, we obtain the boundary solution $(x^c, y^c) = (0, 1)$. Thus, an interior equilibrium exists if and only if $\tilde{\eta} = \frac{1}{\eta} \in [0, \tilde{\eta}_0] \cup [\tilde{\eta}_1, 1/r_\tau]$.

We first study a lower bound for $\tilde{\eta}_0$. After that, we provide an upper bound for $\tilde{\eta}_1$. Together, the bounds give a sufficient condition for the existence of an interior equilibrium (x^c, y^c).

An upper bound on r in (4.74) is given by $r_0^U(\tilde{\eta}) := a\tilde{\eta}^{\frac{1}{\zeta-1}}$, which produces the following lower bound of $\tilde{\eta}_0$ on $\left(0, \frac{1}{\zeta r_\tau}\right)$

$$\tilde{\eta}_0^L = \frac{\phi(y_1)}{a^{\zeta-1}}$$

Indeed, $r(\tilde{\eta}) < r^U(\tilde{\eta}) < \phi(y_1)^{\frac{1}{\zeta-1}}$ for all $\tilde{\eta} \in (0, \tilde{\eta}_0^L)$ such that (4.75) is satisfied.[16]

For an upper bound on $\tilde{\eta}_1$, notice that $\tilde{\eta}^{\frac{1}{\zeta-1}} \leq (1/r_\tau)^{\frac{1}{\zeta-1}}$ since $\tilde{\eta} \leq 1/r_\tau$ holds in the relevant case (iii). Therefore, an upper bound on $r(\tilde{\eta})$ at the interval $[0, 1/r_\tau]$ is given by

$$\tilde{r}_1^U(\tilde{\eta}) := a\,(1/r_\tau)^{\frac{1}{\zeta-1}}\,(1 - r_\tau\tilde{\eta})$$

Letting $\tilde{r}_1^U\left(\tilde{\eta}_1^U\right) = \phi(y_1)^{\frac{1}{\zeta-1}}$ implies the following upper bound on $\tilde{\eta}_1$

$$\tilde{\eta}_1^U = \frac{1}{r_\tau}\left(1 - \frac{1}{a}\left[r_\tau\,\phi(y_1)\right]^{\frac{1}{\zeta-1}}\right)$$

The upper bound $\tilde{\eta}_1^U$ is an upper bound for the solution $\tilde{\eta}_1$ of the original equation $r(\tilde{\eta}) = \phi(y_1)^{\frac{1}{\zeta-1}}$. Since this solution is above $\frac{1}{\zeta r_\tau}$, the upper bound $\tilde{\eta}_1^U$ is on $\left(\frac{1}{\zeta r_\tau}, \frac{1}{r_\tau}\right)$.[17]

All $\tilde{\eta} \leq \tilde{\eta}_0^L$ and all $\tilde{\eta} \geq \tilde{\eta}_1^U$ satisfy (4.75). Provided (4.76) holds, a sufficient condition for $\phi(y_1) \geq \psi$ is $\tilde{\eta} \notin [\tilde{\eta}_0, \tilde{\eta}_1] \subset [\tilde{\eta}_0^L, \tilde{\eta}_1^U]$. In terms of $\eta = 1/\tilde{\eta}$, it follows that a sufficient condition is $\eta \notin [\eta_a, \eta_b]$ where

$$\begin{aligned}
\eta_a &:= \frac{1}{\tilde{\eta}_1^U} \\
&= \frac{r_\tau}{1 - \phi(y_1)^{\frac{1}{\zeta-1}}\,\frac{r_\tau^{\frac{1}{\zeta-1}}}{a}} \\
&= \frac{r_\tau}{1 - \frac{1}{2}\left(\frac{1}{\zeta}\phi(y_1)\,\tilde{R}_f^\zeta\sigma_0^{-\zeta}r_\tau\right)^{\frac{1}{\zeta-1}}} \\
\eta_b &:= \frac{1}{\tilde{\eta}_0^L} = \frac{(2\sigma_0)^\zeta\zeta}{2\phi(y_1)\,(1 + r_f/\bar{\tau})^{\zeta/2}}
\end{aligned}$$

[16] By $r'(0) \to \infty$, the solution $\tilde{\eta}_0$ could be small such that is corresponds to a very high η. Bounds $\tilde{\eta}_0^L \leq 1$ have no practical interpretation since $\eta = 1/\tilde{\eta}$ pertains to the bankruptcy cost on each *unit* of debt.

[17] One can show that linearizing r around $\tilde{\eta} = 1/r_\tau$ produces the same upper bound $\tilde{\eta}_1^U$.

This proves that $\eta \notin [\eta_a, \eta_b]$ is sufficient to guarantee the existence of an interior equilibrium.

The bound η_a exceeds $r_\tau = r_f\left(\frac{1}{\bar{\tau}} - 1\right)$, which means that the corresponding equilibrium capital ratio $y^c \in \left(\frac{1}{3}, 1\right]$ where the restriction $\phi(y^c) \geq \psi$ may bind (case (*iii*)). Using (4.69), the bound η_b exceeds ζr_τ. This proves Proposition 4 except for the sign of the derivative of PD. We find the latter jointly with the comparative statics in Proposition 5.

To study the marginal effect of the parameters σ_0^ζ, η, r_f, and $\bar{\tau}$ on y^c, use $\tilde{\phi}(y^c) := \ln(\phi(y^c;\zeta))$ from (4.72), and

$$\tilde{\psi}(\eta; r_f, \sigma_0, \bar{\tau}, \zeta) \quad := \quad \ln(\psi(\eta; r_f, \sigma_0, \bar{\tau}, \zeta))$$

where ϕ and ψ are as in (4.70) and (4.71), respectively. The identity $\tilde{\phi}(y^c) = \tilde{\psi}$ from (4.69) implies for any parameter $p \in \{\sigma_0, \eta, r_f, \bar{\tau}\}$

$$\frac{\mathrm{d}y^c}{\mathrm{d}p} = \left(\frac{\mathrm{d}\tilde{\phi}(y^c)}{\mathrm{d}y}\right)^{-1} \frac{\mathrm{d}\tilde{\psi}}{\mathrm{d}p} = \frac{\tilde{\psi}_p}{\tilde{\phi}_y^c} \tag{4.77}$$

As argued after (4.73), $\tilde{\phi}_y(y^c)) < 0$ is equivalent to $y^c < \frac{1}{3}$, which is the same as $\eta < r_\tau = r_f(1/\bar{\tau} - 1)$. Therefore,

$$\tilde{\phi}_y(y^c) \begin{cases} < 0 & \text{if } \eta < r_\tau \\ > 0 & \text{if } \eta > r_\tau \end{cases} \tag{4.78}$$

We assess the effect of the parameters on $\tilde{\psi}$, and derive the total effect on y^c by combining (4.77) and (4.78). We assume an interior minimum of $\mu^*(x, y)$, or equivalently $\phi(y_1) \geq \psi$ when $\eta > r_\tau$.

(*i*) σ_0: Because $\tilde{\psi}_{\sigma_0} > 0$, we find from (4.77) that the capital ratio y^c increases with the volatility scalar σ_0 if and only if $\eta > r_\tau$.

(*ii*) η: Differentiation to η yields

$$\begin{aligned} \frac{\mathrm{d}\tilde{\psi}}{\mathrm{d}\eta} &= \frac{\mathrm{d}}{\mathrm{d}\eta}\left[-\ln(\eta) + (\zeta - 1)\ln\left|2 - \frac{2r_\tau}{\eta}\right|\right] \\ &= \frac{1}{\eta}\frac{\eta - \zeta r_\tau}{r_\tau - \eta} \end{aligned}$$

Therefore, $\tilde{\psi}_\eta > 0$ if and only if one of the following two conditions holds:
(*a*) $r_\tau < \eta < \zeta r_\tau$

(b) $\zeta r_\tau < \eta < r_\tau$
We exclude case (b) since $\zeta \geq 4$. Hence, $\psi_\eta > 0$ if and only if $\eta \in (r_\tau, \zeta r_\tau)$. This indicates $y^c_\eta = \tilde{\psi}_\eta / \tilde{\phi}_y > 0$ if and only if $\eta \in [0, \zeta r_\tau)$.

(iii) r_f: Differentiating $\tilde{\psi}$ to r_f,

$$\begin{aligned}\frac{\mathrm{d}\tilde{\psi}}{\mathrm{d}r_f} &= \frac{\mathrm{d}}{\mathrm{d}r_f}\left[-\frac{\zeta}{2}\ln\left(1+\frac{r_f}{\bar{\tau}}\right)+(\zeta-1)\ln\left|2-\frac{2r_\tau}{\eta}\right|\right] \\ &= -\frac{1}{2}\frac{\zeta}{r_f+\bar{\tau}}-\frac{(\zeta-1)\left(1/\bar{\tau}-1\right)}{\eta-r_\tau}\end{aligned}$$

The denominator of the second term changes sign at $\eta = r_\tau$. We have $\tilde{\psi}_{r_f} < 0$ if $\eta > r_\tau$. From $\tilde{\psi}_{r_f} > 0$ for $\eta = 0$ and $\frac{\mathrm{d}}{\mathrm{d}\eta}\tilde{\psi}_{r_f} > 0$ for $\eta \in [0, r_\tau)$, it follows that $\tilde{\psi}_{r_f} > 0$ for all $\eta \in [0, r_\tau)$. Now (4.77) gives that the capital ratio decreases with r_f: $y^c_{r_f} = \tilde{\psi}_{r_f}/\tilde{\phi}_y < 0$ for all η.

(iv) τ: We study the effect of the complement $\bar{\tau} = 1 - \tau$. Differentiating $\tilde{\psi}$ to $\frac{1}{\bar{\tau}}$,

$$\begin{aligned}\frac{\mathrm{d}\tilde{\psi}}{\mathrm{d}\left(1/\bar{\tau}\right)} &= \frac{\mathrm{d}}{\mathrm{d}\left(1/\bar{\tau}\right)}\left[-\frac{\zeta}{2}\ln\left(1+\frac{r_f}{\bar{\tau}}\right)+(\zeta-1)\ln\left|2-\frac{2r_\tau}{\eta}\right|\right] \\ &= -\frac{1}{2}\frac{\zeta r_f \bar{\tau}}{\bar{\tau}+r_f}-\frac{(\zeta-1)\,r_f}{\eta-r_\tau}\end{aligned}$$

The denominator of the second term changes sign at $\eta = r_\tau$. Along similar lines as $\tilde{\psi}_{r_f}$, it can be shown that $\tilde{\psi}_{1/\bar{\tau}} < 0$ if $\eta > r_\tau$, and $\tilde{\psi}_{1/\bar{\tau}} > 0$ if $\eta < r_\tau$. This gives that for all η that the capital ratio decreases with $\frac{1}{\bar{\tau}}$, which gives $y^c_\tau < 0$.

Collecting the previous results gives the sensitivity results of y^c.

We do a similar sensitivity analysis on x^c. Using $\zeta = \gamma/\delta$, we find from equation (4.62)

$$(\zeta-1)\ln\left(x^c(y)^\delta\right) = \ln\left(\frac{\eta\tilde{R}_f}{\zeta\sigma_0}\sqrt{y}\,(1-y)\right)$$

with derivatives

$$(\zeta-1)\frac{\mathrm{d}}{\mathrm{d}y}\ln\left(x^c(y)^\delta\right) = \frac{1}{2}\frac{1-3y}{y\,(1-y)}$$

$$
\begin{aligned}
(\zeta-1)\frac{\mathrm{d}}{\mathrm{d}\sigma_0}\ln\left(x^c(y)^\delta\right) &= -\frac{1}{\sigma_0}+\frac{1}{2}\frac{1-3y}{y\,(1-y)}y_{\sigma_0}\\
(\zeta-1)\frac{\mathrm{d}}{\mathrm{d}\eta}\ln\left(x^c(y)^\delta\right) &= \frac{1}{\eta}+\frac{1}{2}\frac{1-3y}{y\,(1-y)}y_{\eta}\\
(\zeta-1)\frac{\mathrm{d}}{\mathrm{d}r_f}\ln\left(x^c(y)^\delta\right) &= \frac{1}{\tilde{R}_f}\frac{\mathrm{d}\tilde{R}_f}{\mathrm{d}r_f}+\frac{1}{2}\frac{1-3y}{y\,(1-y)}y_{r_f}\\
&= \frac{1}{2\left(\bar{\tau}+r_f\right)}+\frac{1}{2}\frac{1-3y}{y\,(1-y)}y_{r_f}\\
(\zeta-1)\frac{\mathrm{d}}{\mathrm{d}\bar{\tau}}\ln\left(x^c(y)^\delta\right) &= \frac{1}{\tilde{R}_f}\frac{\mathrm{d}\tilde{R}_f}{\mathrm{d}\bar{\tau}}+\frac{1}{2}\frac{1-3y}{y\,(1-y)}y_{\bar{\tau}}\\
&= -\frac{r_f}{2\bar{\tau}\left(\bar{\tau}+r_f\right)}+\frac{1}{2}\frac{1-3y}{y\,(1-y)}y_{\bar{\tau}}
\end{aligned}
$$

where we used $\tilde{R}_f := \sqrt{1+\frac{r_f}{\bar{\tau}}}$. Combining this with the results on y^c gives the sensitivity results of x^c.

Next, we consider the effects on PD^c. Regardless of $\bar{c}(x)$, the sign of the total derivative of PD^c to x is always negative for the chosen functional form of $w(x)$:

$$
\begin{aligned}
PD^c_x &= PD^c_1+\frac{\mathrm{d}y^c(x)}{\mathrm{d}x}PD^c_2\\
&= -w'(x^c)\sqrt{y^c}+\frac{\frac{1}{2}\eta w'(x^c)\left(3y^c-1\right)/\sqrt{y^c}}{\eta w(x^c)\left(\frac{1+3y^c}{4y^c}\right)/\sqrt{y^c}}\frac{w(x^c)}{2\sqrt{y^c}}\\
&= -\frac{2}{3y+1}w^c_1\sqrt{y^c}<0
\end{aligned}
$$

where we used (4.47), (4.52), (4.53), (4.60), and (4.61). In other words, a possible positive indirect effect through capital is dominated by the direct effect of size on PD.

The total derivative of PD to the capital ratio y does depend on $\bar{c}$ through ζ. Notice first from (4.50), (4.51), and (4.62) that in any banking equilibrium $(x^c(y), y)$:

$$
\begin{aligned}
PD^*(x^c(y),y) &= 1-\frac{\tilde{R}_f}{\sigma_0}\sqrt{y}x^c(y)^\delta \qquad (4.79)\\
&= 1-\frac{\tilde{R}_f}{\sigma_0}\sqrt{y}\left(\frac{\eta\tilde{R}_f}{\zeta\sigma_0}\sqrt{y}\,(1-y)\right)^{1/(\zeta-1)}
\end{aligned}
$$

$$= 1 - \left(\frac{\tilde{R}_f}{\sigma_0}\sqrt{y}\right)^{\zeta/(\zeta-1)} \left(\frac{\eta}{\zeta}(1-y)\right)^{1/(\zeta-1)}$$

After some rewriting,

$$(\zeta - 1)\ln(1 - PD^*) = \ln(\eta) - \ln(\zeta) + \zeta \ln\left(\sqrt{1 + \frac{r_f}{\bar{\tau}}}\right) - \zeta \ln(\sigma_0) + \frac{\zeta}{2}\ln(y) + \ln(1-y)$$

with derivatives

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}y}(\zeta - 1)\ln(1 - PD^*) &= \frac{\zeta}{2y} - \frac{1}{1-y} \\
\frac{\mathrm{d}}{\mathrm{d}\sigma_0}(\zeta - 1)\ln(1 - PD^*) &= -\frac{\zeta}{\sigma_0} + \left(\frac{\zeta}{2y} - \frac{1}{1-y}\right) y_{\sigma_0} \\
\frac{\mathrm{d}}{\mathrm{d}\eta}(\zeta - 1)\ln(1 - PD^*) &= \frac{1}{\eta} + \left(\frac{\zeta}{2y} - \frac{1}{1-y}\right) y_{\eta} \\
\frac{\mathrm{d}}{\mathrm{d}r_f}(\zeta - 1)\ln(1 - PD^*) &= \frac{\zeta}{2(\bar{\tau} + r_f)} + \left(\frac{\zeta}{2y} - \frac{1}{1-y}\right) y_{r_f} \\
\frac{\mathrm{d}}{\mathrm{d}\bar{\tau}}(\zeta - 1)\ln(1 - PD^*) &= -\frac{\zeta r_f}{2\bar{\tau}(\bar{\tau} + r_f)} + \left(\frac{\zeta}{2y} - \frac{1}{1-y}\right) y_{\bar{\tau}}
\end{aligned}$$

Since $\frac{\mathrm{d}}{\mathrm{d}p}(\zeta - 1)\ln(1 - PD^*)$ and PD^*_p have opposite sign for any parameter p, and $\frac{\zeta}{2+\zeta} > \frac{1}{3}$, we obtain at the competitive equilibrium

(*i*) $PD^c_y < 0$ if and only if $y^c \in \left[0, \frac{\zeta}{2+\zeta}\right]$.

(*ii*) $PD^c_{\sigma_0} > 0$ if (*a*) $\eta \leq r_{\mathcal{T}}$, or (*b*) $y^c \in \left[\frac{\zeta}{2+\zeta}, y_1\right]$

(*iii*) $PD^c_{\eta} < 0$ if (*a*) $y^c \in \left[0, \frac{\zeta}{2+\zeta}\right]$ and $\eta < \zeta r_{\mathcal{T}}$, or (*b*) $y^c \in \left[\frac{\zeta}{2+\zeta}, y_1\right]$ and $\eta > \zeta r_{\mathcal{T}}$.

(*iv*) $PD^c_{r_f} < 0$ if $y^c \in \left[\frac{\zeta}{2+\zeta}, y_1\right]$

(*v*) $PD^c_{\bar{\tau}} > 0$ if $y^c \in \left[\frac{\zeta}{2+\zeta}, y_1\right]$

The sign of the partial derivatives is ambiguous outside these intervals. Using (4.69) and $\frac{\zeta}{2+\zeta} < y_1$, the condition $y^c \in \left[0, \frac{\zeta}{2+\zeta}\right]$ is equivalent to $\phi\left(\frac{\zeta}{\zeta+2}\right) \geq \psi$. Rewriting the latter inequality gives $\sigma_0 \leq \tilde{\sigma}$ with $\tilde{\sigma}$ as in (4.20). □

Proof of equation (4.31)**:** Using (4.29) and (4.30), the shock $Y_i \sim c\,(s)\,t(\nu)$ of a bank with market share $s := 1/n$ follows a scaled univariate Student $t(\nu)$-distribution with scaling factor c and standard deviation σ_Y:

$$c\,(s) = \hat{\sigma}\frac{1}{s}\left(\int\limits_{t=(i-1)s}^{is}\int\limits_{u=(i-1)s}^{is}\rho^{d(t,u)}\,\mathrm{d}t\,\mathrm{d}u\right)^{1/2}$$
$$= \frac{\hat{\sigma}}{\ln\left(\rho^s\right)}\sqrt{2\left(\rho^s - 1 - \ln\left(\rho^s\right)\right)} \tag{4.80}$$
$$\sigma_Y(s) = \sqrt{\frac{\nu}{\nu-2}}c(s) \tag{4.81}$$

The standard deviation σ_Y is a function of size x and the mean asset return μ^* since $s(x,\mu^*) = x/Q(\mu^*)$. We thus write $\sigma_Y(s)$ and $\sigma_Y(x,\mu)$ interchangeably to refer to the standard deviation of Y. At the competitive equilibrium, the capital ratio has no effect on the market share $s^c = s(x^c, y^c)$ since $\frac{\partial s^c}{\partial y} = \frac{\partial s^c}{\partial \mu}\frac{\partial \mu^c}{\partial y} = 0$.

The bank-specific shock for a bank that invests in projects on the interval $[0,s]$ with $s \le \frac{1}{2}$ is

$$\begin{aligned}
\mathrm{Var}\left(Z\right) &= \frac{1}{s^2}\int_{u=0}^{s}\int_{t=0}^{s}\mathrm{Cov}\left(Z\left(t\right), Z\left(u\right)\right)\mathrm{d}t\,\mathrm{d}u \\
&= \frac{2}{s^2}\int_{u=0}^{s}\int_{t=0}^{u}\rho^{\min(u-t,1-|u-t|)}\mathrm{d}t\,\mathrm{d}u \\
&= \frac{2}{s^2}\int_{u=0}^{s}\int_{t=0}^{u}\rho^{u-t}\mathrm{d}t\,\mathrm{d}u \\
&= \frac{2\left(\rho^s - 1 - \ln\left(\rho^s\right)\right)}{\ln^2\left(\rho^s\right)}
\end{aligned}$$

The total shock $Y = \frac{1}{s}\int_{t=0}^{s} Y\left(t\right)\mathrm{d}t = \frac{\hat{\sigma}\sqrt{W}}{s}\int_{t=0}^{s} Z\left(t\right)\mathrm{d}t$ that includes the macro shock is a scaled univariate Student $t(\nu)$-distribution with

$$\begin{aligned}
\mathrm{Var}\left(Y\right) &= \hat{\sigma}^2\mathrm{Var}\left(\frac{\sqrt{W}}{s}\int_{t=0}^{s} Z\left(t\right)\mathrm{d}t\right) \\
&= \hat{\sigma}^2\mathbb{E}[\mathrm{W}]\mathrm{Var}\left(Z\right) \\
&= \frac{2\nu\hat{\sigma}^2}{\nu-2}\frac{\rho^s - 1 - \ln\left(\rho^s\right)}{\ln^2\left(\rho^s\right)}
\end{aligned}$$

$$= \frac{2\nu\hat{\sigma}^2 n^2}{\nu - 2} \frac{\rho^s - 1 - \frac{1}{n}\ln(\rho)}{\ln^2(\rho)}$$

□

Proof of equation (4.32)**:** It is more convenient to derive the shocks for banks 1 and $1+i$. The covariance of the shocks, including the macro shock, is

$$\begin{aligned} \mathrm{Cov}(Y_1, Y_{1+i}) &= \mathrm{Cov}\left(\hat{\sigma}\sqrt{W} Z_1, \hat{\sigma}\sqrt{W} Z_{1+i}\right) \\ &= \hat{\sigma}^2 \mathbb{E}[W]\,\mathrm{Cov}(Z_1, Z_{1+i}) \\ &= \frac{\nu}{\nu - 2}\hat{\sigma}^2 \mathrm{Cov}(Z_1, Z_{1+i}) \end{aligned}$$

To find the covariance of the bank-specific shocks Z_1 and Z_{1+i}, notice that banks 1 and $1+i$ invest in projects on the intervals $[0, 1/n]$ and $[i/n, (1+i)/n]$, respectively. Letting $s = 1/n$, the covariance of the bank-specific shocks, excluding the macro-shock W, is

$$\begin{aligned} &\mathrm{Cov}(Z_1, Z_{1+i}) \\ &= \frac{1}{s^2}\mathrm{Cov}\left(\int_{t=0}^{s} Z(t)\,\mathrm{d}t, \int_{u=is}^{(i+1)s} Z(u)\,\mathrm{d}u\right) \\ &= \frac{1}{s^2}\int_{u=is}^{(i+1)s}\int_{t=0}^{s} \mathrm{Cov}(Z(t), Z(u))\,\mathrm{d}t\,\mathrm{d}u \\ &= \frac{1}{s^2}\int_{u=is}^{(i+1)s}\int_{t=0}^{s} \rho^{\min(|u-t|, 1-|u-t|)}\mathrm{d}t\,\mathrm{d}u \\ &= \tfrac{1}{s^2}\int_{u=is}^{(i+1)s}\int_{t=\min\left(\max\left(u-\frac{1}{2},0\right),s\right)}^{s} \rho^{u-t}\mathrm{d}t\,\mathrm{d}u \\ &\quad + \tfrac{1}{s^2}\int_{u=is}^{(i+1)s}\int_{t=0}^{\min\left(\max\left(u-\frac{1}{2},0\right),s\right)} \rho^{1-(u-t)}\mathrm{d}t\,\mathrm{d}u \end{aligned} \tag{4.82}$$

where we used

$$\min_{t<u}(|u-t|, 1-|u-t|) = \begin{cases} u - t & t \geq u - \frac{1}{2} \\ 1 - (u-t) & t < u - \frac{1}{2} \end{cases}$$

We use (4.82) to work out three different cases.

(*i*) If $(i+1)s \leq \frac{1}{2}$, then all projects of banks 1 and $1+i$ are on $\left[0, \frac{1}{2}\right]$ such that the shortest distance between projects t and u is $u - t$:

$$\mathrm{Cov}(Z_1, Z_{1+i}) = \frac{1 - \rho^{-s}}{s^2 \ln(\rho)}\int_{u=is}^{(i+1)s} \rho^u \mathrm{d}u = \frac{\rho^{(i-1)s}(1-\rho^s)^2}{\ln^2(\rho^s)}$$

(ii) If $is \geq \frac{1}{2}+s$, each project t of bank 1 is on $[0,s]$ while each project u of bank i is on $[s+\frac{1}{2},1]$. The shortest distance is then $t+1-u$:

$$\begin{aligned}
&\mathrm{Cov}\left(Z_1, Z_{1+i}\right)\\
&= \frac{1}{s^2}\frac{\rho\left(\rho^s-1\right)}{\ln(\rho)}\int_{u=is}^{(i+1)s}\rho^{-u}\mathrm{d}u\\
&= \frac{\rho^{1-(i+1)s}\left(1-\rho^s\right)^2}{\ln^2(\rho^s)}
\end{aligned}$$

(iii) In the intermediate case, the projects of bank $1+i$ cover either $\frac{1}{2}$ or $\frac{1}{2}+s$. If the projects cover $\frac{1}{2}$,

$$is \leq \frac{1}{2} < (i+1)\,s \leq \frac{1}{2}+s \tag{4.83}$$

Starting from (4.82),

$$\begin{aligned}
&s^2\mathrm{Cov}\left(Z_1, Z_{1+i}\right)\\
&= \int_{u=is}^{(i+1)s}\int_{t=0}^{\min\left(\max\left(u-\frac{1}{2},0\right),s\right)}\rho^{t+(1-u)}\mathrm{d}t\,\mathrm{d}u\\
&\quad+\int_{u=is}^{(i+1)s}\int_{t=\min\left(\max\left(u-\frac{1}{2},0\right),s\right)}^{s}\rho^{u-t}\mathrm{d}i\,\mathrm{d}u\\
&= \int_{u=is}^{1/2}\int_{t=0}^{0}\rho^{t+(1-u)}\mathrm{d}t\,\mathrm{d}u+\int_{u=\frac{1}{2}}^{(i+1)s}\int_{t=0}^{u-\frac{1}{2}}\rho^{t+(1-u)}\mathrm{d}t\,\mathrm{d}u\\
&\quad+\int_{u=is}^{1/2}\int_{t=0}^{s}\rho^{u-t}\mathrm{d}t\,\mathrm{d}u+\int_{u=\frac{1}{2}}^{(i+1)s}\int_{t=u-\frac{1}{2}}^{s}\rho^{u-t}\mathrm{d}i\,\mathrm{d}u\\
&= \rho\int_{u=\frac{1}{2}}^{(i+1)s}\rho^{-u}\int_{t=0}^{u-\frac{1}{2}}\rho^{t}\mathrm{d}t\,\mathrm{d}u+\int_{u=is}^{1/2}\rho^{u}\int_{t=0}^{s}\rho^{-t}\mathrm{d}t\,\mathrm{d}u\\
&\quad+\int_{u=\frac{1}{2}}^{(i+1)s}\rho^{u}\int_{t=u-\frac{1}{2}}^{s}\rho^{-t}\mathrm{d}t\,\mathrm{d}u\\
&= \rho\int_{u=\frac{1}{2}}^{(i+1)s}\rho^{-j}\left(\frac{\rho^{u-\frac{1}{2}}-1}{\ln\left(\rho\right)}\right)\mathrm{d}u+\frac{\left(\rho^{\frac{1}{2}}-\rho^{is}\right)\left(1-\rho^{-s}\right)}{\ln^2\left(\rho\right)}\\
&\quad+\int_{u=\frac{1}{2}}^{(i+1)s}\rho^{u}\left(\frac{\rho^{\frac{1}{2}-u}-\rho^{-s}}{\ln\left(\rho\right)}\right)\mathrm{d}u\\
&= \frac{1}{\ln^2\left(\rho\right)}\left[\rho\int_{u=\frac{1}{2}}^{(i+1)s}\left(\frac{\rho^{-\frac{1}{2}}-\rho^{-u}}{\ln\left(\rho\right)}\right)\mathrm{d}u+\left(\rho^{\frac{1}{2}}-\rho^{is}\right)\left(1-\rho^{-s}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{\ln^2(\rho)}\left[\int_{u=\frac{1}{2}}^{(i+1)s}\left(\frac{\rho^{\frac{1}{2}}-\rho^{u-s}}{\ln(\rho)}\right)\mathrm{d}u\right]\\
&=\frac{1}{\ln^2(\rho)}\left[\rho^{\frac{1}{2}}\ln(\rho)\left(2s\left(1+i\right)-1\right)+\left(\rho^{\frac{1}{2}}-\rho^{is}\right)\left(1-\rho^{-s}\right)\right]\\
&+\frac{1}{\ln^2(\rho)}\left[-\left(\rho^{\frac{1}{2}}-\rho^{1-(i+1)s}\right)-\left(\rho^{is}-\rho^{\frac{1}{2}-s}\right)\right]\\
&=\frac{1}{\ln^2(\rho)}\left[\rho^{\frac{1}{2}}\ln(\rho)\left(2s\left(1+i\right)-1\right)+\rho^{(i-1)s}-2\rho^{is}+\rho^{1-(i+1]s}\right]
\end{aligned}
\tag{4.84}
$$

If the projects of bank $1+i$ cover $\frac{1}{2}+s$,

$$\frac{1}{2}\leq is<\frac{1}{2}+s\leq\left(i+1\right)s.$$

Let $\tilde{i}s=1-is$,

$$\frac{1}{2}\leq 1-\tilde{i}s<\frac{1}{2}+s\leq 1-\tilde{i}s+s$$

such that

$$\tilde{i}s\leq\frac{1}{2}<(\tilde{i}+1)s\leq\frac{1}{2}+s.$$

This is similar to (4.83). Hence, the expression for the covariance of Z_1 and Z_{1+i} follows by substituting $i=\frac{1}{s}-\tilde{i}$ for i in (4.84), and $s=1/n$.

In all cases, the correlation decreases by a factor ρ^{is} when i is moving away from $[0,s]$. Further, $i=1$ and $i=\frac{1}{s}-1$ produce the same result if $s\leq\frac{1}{4}$, as required by the symmetry of the intervals $[s,2s]$ and $[1-s,1]$ with respect to $[0,s]$.

□

4.B Estimation procedure

We adopt the following procedure to find for a given pair (x,y) the corresponding equilibrium (PD^*,μ^*,R_c^*). The scaled Student $t(\nu)$-distribution $Y=\sigma(s,y)X$ with density g and cdf G has a default threshold equal to $G^{-1}(PD)$. Rewrite (4.6) as

$$\mathbb{E}\left[\min\left(Y,G^{-1}\left(PD\right)\right)\right]=-y\left(1+\frac{r_f}{\bar{\tau}}\right)\tag{4.85}$$

where we use that X has variance one. The shock Y has an unbounded negative tail which implies that the sufficient existence and uniqueness conditions of an equilibrium in Proposition 1 are both satisfied. Indeed, the left hand side in (4.85) tends to $-\infty$ if $PD \to 0$, increases in PD, and it tends to $\mathbb{E}[Y] = 0$ if $PD \to 1$, while the right hand side in (4.85) is constant in PD. The left hand side in (4.85) is equal to

$$\mathbb{P}\big(Y < G^{-1}(PD)\big)\,\mathbb{E}\left[Y \,\middle|\, Y < G^{-1}(PD)\right] + \mathbb{P}\big(Y \geq G^{-1}(PD)\big)\; G^{-1}(PD)$$

Substitution in (4.85) leads to

$$PD\,\mathbb{E}\left[Y \,\middle|\, Y < G^{-1}(PD)\right] + (1 - PD)\; G^{-1}(PD) = -y\left(1 + \frac{r_f}{\bar{\tau}}\right) \tag{4.86}$$

Since Y has standard deviation $\sigma(s,\mu)$, it follows that $Y = \sigma(s,\mu)\sqrt{\frac{\nu-2}{\nu}}T$ with $T \sim t(\nu)$ a standard, i.e., unscaled, Student $t(\nu)$-distribution which has variance $\nu/(\nu-2)$. Let f_T and F_T represent the density function and cdf of T, respectively. Rewriting (4.86) in terms of T,

$$PD\,\mathbb{E}\left[T \,\middle|\, T < F_T^{-1}(PD)\right] + (1 - PD)\; F_T^{-1}(PD) = -\frac{y}{\sigma(s,\mu)}\sqrt{\frac{\nu}{\nu-2}}\left(1 + \frac{r_f}{\bar{\tau}}\right) \tag{4.87}$$

For the conditional loss of T (see equation (2.27) in McNeil et al. (2010))

$$\mathbb{E}\left[T \,\middle|\, T < F_T^{-1}(PD)\right] = -\frac{f_T\left(F_T^{-1}(PD)\right)}{PD}\left(\frac{\nu + \left(F_T^{-1}(PD)\right)^2}{\nu - 1}\right) \tag{4.88}$$

Substituting (4.88) in (4.87) indicates that we are looking for the solution PD^* of

$$\begin{aligned} &- f_T\left(F_T^{-1}(PD)\right)\left(\frac{\nu + \left(F_T^{-1}(PD)\right)^2}{\nu - 1}\right) + (1 - PD)\; F_T^{-1}(PD) \\ &= -\frac{y}{\sigma(s,\mu)}\sqrt{\frac{\nu}{\nu-2}}\left(1 + \frac{r_f}{\bar{\tau}}\right) \end{aligned} \tag{4.89}$$

We adopt an iterative procedure to find for given x and y the equilibrium $(PD^*(x,y), \mu^*(x,y), R_c^*(x,y))$:

(i) Initialize μ at the mean of its bounds in (4.8)

(ii) Compute σ in (4.31) using $s = x/Q(x,\mu)$

(iii) Find the unique PD that solves (4.89) (*iv*) Compute μ in (4.5)

(*v*) Stop if some stopping criterium is satisfied, otherwise go to ii.

After convergence, we find the contracted interest rate $R_c^*(x,y)$ as follows. By (4.3),

$$PD^*(x,y) = F_T\left(\sqrt{\frac{\nu}{\nu-2}}\frac{\bar{c}(x) + (1-y)\,R_c^*(x,y) - \mu^*}{\sigma(s,\mu^*)}\right)$$

from which we obtain $R_c^*(x,y)$

$$R_c^*(x,y) = \frac{1}{1-y}\left(\mu^* - \bar{c}(x) + \sigma\left(s,\mu^*\right)\sqrt{\frac{\nu-2}{\nu}}F_T^{-1}\left(PD^*\right)\right)$$

The concavity of μ in x and y ensures that we do not need to compute μ for each possible (x,y) to find the minimum of μ. Instead, we run over a numerical grid of (x,y) by moving in a direction where μ decreases.

The social welfare $SW(x,y)$ in (4.28) follows from the total surplus in (4.26), and PS in (4.33). The probability PS is based on the expressions in (4.31), (4.32) and (4.33).[18]

[18] The Matlab package `mvtcdf` evaluates the multivariate Student t-distribution in Matlab. This package however turned out to be slow, or inaccurate in computing the joint default probability of a high dimensional Student t-distribution. We resolved this by implementing Algorithm 2 in Han and Wu (2010) with the following four modifications. First, our indicator function is $\sum_i 1(X_i < 0) \geq n_D$ since we define a systemic crisis by a minimal number n_D of defaults instead of a simultaneous default of all banks. Second, rather than directly setting $\mu^* = \Sigma^{-1}C$, we iterate beforehand over a set of candidates for μ^* to test numerically on the variance of each candidate. Each candidate is a multiple of the all ones vector, because our setting is completely symmetric. Third, to take the different definition of μ^* into account, we substitute $A^T\mu^*$ from Lemma 2 in Han and Wu (2010) for $A^{-1}C$. Fourth, as follows from the proof of Lemma 2, our $W^{(i)}$ has mean $\sqrt{Y/\nu}A^{-1}C$ without a minus sign.

Numerical tests on small dimensional problems confirm that our algorithm is accurate. Compared to the `mvtcdf` package, it produces for high dimensions, i.e., a large number of banks, faster and more stable estimates for the systemic default probability PS.

Chapter 5

Risk allocation and the Shapley value

5.1 Introduction

A large literature on systemic risk measures has evolved since the recent financial crisis.[1] Attributing systemic risk to individual institutions is akin to a problem already addressed by game theorists. Noble Laureate Lloyd Shapley developed in Shapley (1953) an allocation methodology where the player's worth (value) equals the average of his marginal contribution to the value created by all possible subsets of other players. This results in a fair allocation of the total value in the sense that the value created jointly by two players is split equally between the players.

Several papers have implemented the Shapley value to allocate systemic risk to banks, see, e.g., Tarashev et al. (2013), Drehmann and Tarashev (2013), and Cao (2014). Nonetheless, both the widely used allocation measures MES and $\Delta CoVaR$ do not fit the Shapley value allocation. Adjustments of the two measures to satisfy the Shapley axioms lead to more ($O(2^n)$) computations and a less transparent allocation.

The contribution of this paper is that we provide a general class of simple risk allocation measures that are Shapley values, and demonstrate that two measures in the literature fit into this class.

5.2 Shapley value

Since we apply the Shapley value to risk measures of banks, we refer to banks as players in game theory. Let $N = \{1, \ldots, n\}$ denote the set of all banks, and let B_i denote a binary variable which is one if bank i fails,

I am very grateful to Gerard van der Laan for comments on an earlier version.

[1] The survey in Bisias et al. (2012) discusses 31 systemic risk measures.

and zero otherwise. The total number of failing banks in the subset S is $\kappa_S = \sum_{i \in S} B_i$, and $\kappa := \kappa_N$ is the total number of failing banks. The characteristic function v is a function defined on each $S \subseteq N$ and satisfies $v(\emptyset) = 0$.

The Shapley value assigned to bank i is given by

$$\phi_i(v) = \sum_{\{S: i \notin S\}} \frac{|S|!(n - |S| - 1)!}{n!} \left(v(S \cup i) - v(S)\right)$$

That is, a bank's portion of the total value $v(N)$ equals the average of this bank's marginal contribution to all possible subsets of other banks. For $v(S) = a\, w(S) + b\, z(S)$ $(a, b \geq 0)$, it follows immediately that the Shapley value is linear in the characteristic functions w and z:

$$\phi(v) = a\, \phi(w) + b\, \phi(z) \qquad S \subseteq N$$

The Shapley value is fair in the sense that it satisfies several axiomatizations.[2] For instance, Myerson (1980) shows that the Shapley value is the unique allocation that satisfies the following two axioms:

(i) *Efficiency:* The allocated values sum to the total value:

$$\sum_{i \in N} \phi_i(v) = v(N)$$

(ii) *Balanced contributions*: For any $S \subseteq N$ and any pair of banks $i, j \in S$, the loss that bank i incurs when removing bank j equals the loss that bank j incurs when removing bank i:

$$\phi_i(v|_S) - \phi_i(v|_{S \setminus j}) = \phi_j(v|_S) - \phi_j(v|_{S \setminus i}) \tag{5.1}$$

where $v|_M$ is the restriction of v to the subsets of M.

The following proposition enables us to obtain the Shapley value in $O(n)$ computations for a special class of characteristic functions.

Proposition 1. *Suppose the characteristic function v has the functional form $v(S) = \kappa_S f(\kappa_S)$ with f finite for any κ_S.*

(i) The Shapley value assigned to bank i equals $\phi_i(v) = B_i f(\kappa)$.

[2] see Winter (2002) for an overview.

(ii) *The weighted Shapley value of Kalai and Samet (1987) is not linear in f as in (i), except for the special case $w_1 = \ldots = w_n \neq 0$.*

Proof. (i) We show that the allocation $\phi_i(v) = B_i f(\kappa)$ satisfies the efficiency axiom as well as the balanced contributions axiom. It is easy to see that $\phi(v)$ satisfies the efficiency axiom because

$$\sum_{i \in N} \phi_i(v) = \sum_{i \in N} B_i f(\kappa) = \kappa f(\kappa) = v(N)$$

For any distinct $i, j \in S$ with $S \subseteq N$, we have $\phi_i(v|_S) = B_i f(\kappa_S)$ and $\phi_i(v|_{S \setminus j}) = B_i f(\kappa_S - B_j)$ such that the balanced contributions axiom is necessarily satisfied if

$$B_i f(\kappa_S) - B_i f(\kappa_S - B_j) = B_j f(\kappa_S) - B_j f(\kappa_S - B_i)$$

Obviously, this equation holds for any of the four possibilities of the pair (B_i, B_j). This implies that $\phi(v)$ satisfies the balanced contributions axiom, as required.

(ii) For an arbitrary nonzero weight vector $w \in \mathbb{R}^n$, the balanced contributions axiom (5.1) changes for weighted Shapley values into (see Kalai and Samet (1987)):

$$w_j \left[\phi_i(v_S) - \phi_i(v|_{S \setminus j})\right] = w_i \left[\phi_j(v_S) - \phi_j(v|_{S \setminus i})\right] \tag{5.2}$$

Write the allocation in the desired linear form in f: $\phi_i(v|_S) = \gamma_i^S w_i B_i f(\kappa_S)$ with γ_i^S unknown. Using $v(S) = \kappa_S f(\kappa_S)$ and the efficiency axiom on the subgame (S, v_S) implies

$$\sum_{i \in S} \gamma_i^S w_i B_i = \kappa_S$$

Therefore, we need that one of the following equations holds true for all $S \subseteq N$:

$$\gamma_i^S = \frac{1}{w_i} \tag{5.3}$$

$$\gamma_i^S = \frac{\kappa_S}{\sum_{k \in S} w_k B_k} \tag{5.4}$$

The balanced contributions axiom (5.2) is satisfied for a pair (i, j) with $(B_i, B_j) = (1, 1)$ if for all $S \subseteq N$

$$w_j w_i \left[\gamma_i^S f(\kappa_S) - \gamma_i^{S \setminus j} f(\kappa_S - 1)\right] = w_i w_j \left[\gamma_j^S f(\kappa_S) - \gamma_j^{S \setminus i} f(\kappa_S - 1)\right]$$

Since in general $w_i, w_j \neq 0$:

$$\left[\gamma_i^S - \gamma_j^S\right] f(\kappa_S) = \left[\gamma_i^{S\setminus j} - \gamma_j^{S\setminus i}\right] f(\kappa_S - 1) \qquad \text{for all } S \subseteq N$$

The latter equation is only satisfied for any function f if $\gamma_i^S = \gamma_j^S$ and $\gamma_i^{S\setminus j} = \gamma_j^{S\setminus i}$. Thus, γ_i^S does neither depend on i, nor on S. This implies that (5.3) and (5.4) are both inappropriate, except for the special case $w_1 = \ldots = w_n \neq 0$ that corresponds to Proposition 1(i). □

Proposition 1 implies that the Shapley value assigned to bank i equals the fraction B_i/κ of the total value $v(N) = \kappa f(\kappa)$. By the linearity of the Shapley value, and the linearity of the expectations operator, the proposition is easily extended to characteristic functions of the type $v(S) = \mathbb{E}[\kappa_S f(\kappa_S)]$ with κ_S a random variable.

5.3 Systemic risk measures

The two most common measures to allocate systemic risk are the marginal expected shortfall (*MES*) from Acharya et al. (2012) and $\Delta CoVaR$ from Adrian and Brunnermeier (2011). We briefly discuss both measures. Let R_i denote the loss of bank i, and R the weighted loss of a portfolio of banks in the banking network:

$$R = \sum_{i \in N} w_i R_i$$

The *MES* of bank $i \in N$ is the marginal expected shortfall with respect to the weight w_i given that R is above some typically high threshold q:

$$MES_i = \frac{\partial}{\partial w_i} \mathbb{E}\left[R \mid R \geq q\right] = \mathbb{E}\left[R_i \mid R \geq q\right]$$

For a given probability p, let $VaR_i(p)$ denote the $p\%$ quantile of bank i. The measure $\Delta CoVaR_i$ of bank i is the difference in some given quantile of R, between (i) R_i is at its $p\%$ quantile return, and (ii) R_i is at its median:

$$\Delta CoVaR_i = CoVaR_i^{R_i = VaR_i(p)} - CoVaR_i^{R_i = VaR_i(50\%)}$$

where $CoVaR_i$ satisfies for given p and x

$$p = \mathbb{P}\left(R \geq CoVaR_i^{R_i = VaR_i(p)} \mid R_i = x\right).$$

Benoit et al. (2013) discuss several properties of both measures.

Using the MES or $\Delta CoVaR$ as a characteristic function does not result in a useful Shapley value (see Drehmann and Tarashev (2013), and Cao (2014)). More specifically, the marginal contribution of a bank to a subset of banks does not depend on the subset. Using an interbank network, Drehmann and Tarashev (2013) show that such a measure allocates the risk generated by an interbank link solely to the lender. By contrast, a risk measure based on the Shapley value treats both involved banks of an interbank link symmetrically and, as a consequence, splits the associated risk equally between the lender and the borrower. In addition, the construction of the Shapley value from marginal contributions mitigates the impact of model risk by (i) errors in the total loss R by e.g. neglecting banks, and (ii) idiosyncratic losses that do not reflect systemic risk.

To overcome the disadvantages above, a related characteristic function is available for both measures. Adding a conditional event extends both measures to subsets of banks. The characteristic function for the expected shortfall measure MES is (see Drehmann and Tarashev (2013))

$$v^{MES}(S) = \mathbb{E}\left[\sum_{i\in S} R_i \,\middle|\, \sum_{i\in S} R_i \geq q\right], \qquad \text{for all } S \subseteq N.$$

We get the following characteristic function for $\Delta CoVaR$:[3]

$$v^{\Delta CoVaR}(S) = CoVaR_S^{R_i = VaR_i(p)} - CoVaR_S^{R_i = VaR_i(50\%)_i}, \qquad \text{for all } S \subseteq N$$

where $CoVaR_S^{R_i = VaR_i(p)}$ follows from $R = \sum_{i\in N} R_i$ and

$$\mathbb{P}\left(R \geq CoVaR_S^{R_i = x_i} \mid R_i = x_i \text{ for all } i \in S\right) = p.$$

Using the adjusted characteristic functions, the marginal contribution of a bank depends on the subset S of banks, i.e., $v(S \cup i) - v(S)$ depends on S. The risk allocation according to the Shapley value is obtained when we allocate the risk to bank i by averaging marginal contributions of all subsets S that do not contain bank i.

The downside of this procedure is that the characteristic function must be computed for each subset of banks. Since the number of subsets grows exponentially in the number of banks (n), the number of evaluations also grows exponentially $(O(2^n))$. In addition, the intuition behind the allocated risk is less straightforward than by using the original risk measures.

We show that the following two allocation measures directly generate a Shapley value:

[3] A similar extension of $\Delta CoVaR$ is in Cao (2014).

(i) The conditional probability of at least one extra bank failure, given that bank $i \in N$ fails

$$\begin{aligned} PAO_i &= \mathbb{P}\left(\exists j \in N \setminus i \;\text{ such that } B_j = 1 \mid B_i = 1\right) \\ &= \frac{\mathbb{P}\Big(B_i\left[1 - \prod_{j\in N\setminus i}(1 - B_j)\right] = 1\Big)}{\mathbb{P}(B_i = 1)} \\ &= \frac{\mathbb{P}(B_i 1\left(\kappa > 1\right) = 1)}{\mathbb{P}(B_i = 1)}, \end{aligned}$$

where $1(x)$ is the indicator function which is one if x is true and zero otherwise.

(ii) The systemic impact index defined as the expected number of failures, given that bank $i \in N$ fails

$$\begin{aligned} SII_i &= \mathbb{E}\left[\kappa \mid B_i = 1\right] \\ &= \frac{\mathbb{E}[B_i \kappa]}{\mathbb{P}(B_i = 1)}. \end{aligned}$$

The previously discussed *MES* and $\Delta CoVaR$ use returns or, alternatively, another additive performance measure. The measures *PAO* and *SII* simply consider extreme events without quantifying the magnitude of an extreme shock to an individual bank. In other words, the latter two measures are linkage measures by focusing on interbank linkages, and refrain from assessing the magnitude on individual banks.

Segoviano and Goodhart (2009) define the linkage measures *PAO* and *SII* as the banking stability index (*BSI*) and the probability of cascading events (*PCE*), respectively. Zhou (2010) implements these two measures in an extreme value theory setting to test if size is a valid proxy for systemic importance.

Under a VaR approach, each failure probability $\mathbb{P}(B_i = 1)$ equals some fixed probability p. Then, the characteristic functions

$$\begin{aligned} v^{PAO}(S) &= \frac{1}{p}\mathbb{E}[\kappa_S 1\left(\kappa_S > 1\right)], && \text{for all } S \subseteq N \\ v^{SII}(S) &= \frac{1}{p}\mathbb{E}\left[\kappa_S^2\right], && \text{for all } S \subseteq N. \end{aligned}$$

generate the Shapley allocations PAO_i and SII_i, respectively. Namely, Proposition 1 shows that

$$\phi_i(v^{PAO}) = \frac{1}{p}\mathbb{E}\left[\frac{B_i}{\kappa}\kappa 1\left(\kappa > 1\right)\right] = \frac{1}{p}\mathbb{P}(B_i 1\left(\kappa > 1\right) = 1) = PAO_i$$

$$\phi_i(v^{SII}) = \frac{1}{p}\mathbb{E}\left[\frac{B_i}{\kappa}\kappa^2\right] = \frac{1}{p}\mathbb{E}[B_i\kappa] = SII_i$$

The total risk (value) of the banking system is

$$v^{PAO}(N) = \frac{1}{p}\mathbb{E}[\kappa 1\,(\kappa > 1)]$$
$$v^{SII}(N) = \frac{1}{p}\mathbb{E}\left[\kappa^2\right]$$

Thus PAO corresponds to total risk that linearly increases in the number of failures, while *SII* corresponds to a quadratic increase in the number of failures.

5.4 Conclusion

Under a VaR approach where each bank faces a failure probability p, we have shown that the bank-specific measures PAO_i and SII_i are Shapley value allocations of a more complicated characteristic function. Since a separate evaluation of each possible marginal contribution is redundant, the computation of both measures grows only linearly in the number of banks. The total risk of a banking system with PAO and *SII* equals the linear $\mathbb{E}[\kappa 1\,(\kappa > 1)]\,/p$ and the quadratic $\mathbb{E}\left[\kappa^2\right]/p$, respectively. The well-known risk measures *MES* and $\Delta CoVaR$ do not correspond to a Shapley value with the number of computations growing linearly in the number of banks.

More general, Proposition 1 shows that any risk measure that solely depends on the total number of failures κ is the Shapley value allocation of some underlying characteristic function that solely varies in κ. This obviates the need to iterate over all $O(2^n)$ possible marginal contributions.

Chapter 6

A new evaluation method for the multivariate normal distribution

6.1 Introduction

Consider a Gaussian random vector Z in $\mathbb{R}^d$, and the probability that Z is in a region $T = (a_1, b_1) \times \ldots \times (a_d, b_d)$. Without loss of generality, Z has mean $\mathbf{0}$, covariance matrix Σ with ones at the diagonal, and each lower bound a_i is finite. The upper bounds b_i are allowed to be infinite. We evaluate the multivariate integral

$$\mathbb{P}(Z \in T) = \frac{1}{(2\pi|\mathbf{\Sigma}|)^{d/2}} \int_{x_1=a_1}^{b_1} \cdots \int_{x_d=a_d}^{b_d} \exp\left(-\frac{\mathbf{x}'\mathbf{\Sigma}^{-1}\mathbf{x}}{2}\right) \mathrm{d}x_1 \ldots \mathrm{d}x_d \tag{6.1}$$

Genz and Bretz (2009) provide an extensive overview of methods to evaluate this integral. The method of Genz (1992) is the standard method in the software packages R and Matlab. This method transforms the integration region in (6.1) in three steps to a unit hypercube. The resulting integral is numerically integrated by means of a simulation procedure. Nonetheless, this method is not for all parameters in (6.1) the optimal method in terms of accuracy or speed, see, e.g., Fayed and Atiya (2014) for the bivariate case. In particular, the focus on a small absolute error of the method in Genz (1992) may give an inaccurate estimate for the ratio $p_1/p_2 = \mathbb{P}(Z \in T_1)/P(Z \in T_2)$ if p_1 and p_2 are both small. The estimated ratio $(p_1 + \varepsilon_1)/(p_2 + \varepsilon_2)$ is then close to the error ratio $\varepsilon_1/\varepsilon_2$.

Our method bounds the probability in (6.1) by combining a change of measure with a Taylor expansion. We truncate the expansion, and bound the

I am very grateful to Ludolf Meester for extensive discussions and comments on earlier versions. I thank Jan Brinkhuis and John Geweke for useful suggestions.

remainder by using a transformation to spherical coordinates. The relative error of the bounds declines in the true probability. This ensures an estimate of the probability ratio $(p_1 + \varepsilon_1)/(p_2 + \varepsilon_2)$ close to the true ratio p_1/p_2, even if both p_1 and p_2 are small.

We are not the first to use a Taylor expansion. For instance, Moran (1983) evaluates (6.1) by performing a Taylor expansion with an implicit change of measure that affects the covariance matrix. A sufficient convergence condition for the series expansion is that the sum of the absolute partial correlations is for each variable smaller than one. Hashorva and Hüsler (2003) use Jensen's inequality, and the smallest and largest eigenvalue of $\mathbf{\Sigma}$ to derive analytically simple lower and upper bounds for (6.1). While a Taylor series expansion is absent, they implicitly apply a change of measure by changing the mean and covariance matrix. We find that their measure provides suboptimal bounds.

Our method is neither the first to evaluate (6.1) with an importance sampling approach. For example, Glasserman and Li (2005) compute loss quantiles of a credit portfolio by applying importance sampling. The loss distribution is a weighted summation of Bernoulli variables indicating defaults. They perform a Monte Carlo simulation using importance sampling to change the mean of the multivariate normal distribution. Phinikettos and Gandy (2011) employ Monte Carlo simulation using an importance sampler with a fixed mean and a flexible, though identical, variance for each variable. Using some additional techniques to reduce the variance simulation, their method tends to be more accurate than the method of Genz (1992) if the eigenvalues of the covariance matrix are more dispersed.

Nomura (2014) numerically integrates over a spherical coordinate system. Rather than adopting importance sampling, the integration region is transformed such that the first principle axis of the probability distribution is along the first axis. The method still faces an exponential increase in computation time when the dimension increases, even with a dimension reduction technique. This is similar to the combinatorial explosion that our Taylor series expansion faces.

To the best of our knowledge, there is no existing study that combines a Taylor series expansion with an importance sampling approach. In addition, our bounds are for at least some specific cases more accurate than any other bound in the literature.

Section 6.2 describes our method. We give convergence criteria in Section 6.3. Error bounds are in Section 6.4. Section 6.5 discusses possible importance samplers. Section 6.6 illustrates the accuracy of our method with some examples. Conclusions are in Section 6.7.

6.2 The method

6.2.1 The series expansion

We apply a change of measure such that $Z \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ is transformed into $Z \sim N(\mu, \boldsymbol{\Omega})$ under the new measure. Here, the vector μ and the matrix $\boldsymbol{\Omega}$ are arbitrary, except that $\boldsymbol{\Omega}$ is positive definite and symmetric. The new measure enables a more efficient calculation to estimate the probability in (6.1). We have

$$\begin{aligned} \mathbb{P}(\mathbf{a} \leq Z \leq \mathbf{b}) &= \mathbb{E}[1_{\mathbf{a} \leq Z \leq \mathbf{b}}] && (6.2) \\ &= \tilde{\mathbb{E}}[L(Z) 1_{\mathbf{a} \leq Z \leq \mathbf{b}}] && (6.3) \end{aligned}$$

where $\tilde{\mathbb{E}}$ indicates an expectation under the new probability measure, and $L(\mathbf{z})$ is the likelihood ratio of the original density relative to the new density:

$$L(\mathbf{z}) = \sqrt{\frac{|\boldsymbol{\Omega}|}{|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}\mathbf{z}'\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Omega}^{-1}\right)\mathbf{z} - \mu'\boldsymbol{\Omega}^{-1}\mathbf{z} + \frac{1}{2}\mu'\boldsymbol{\Omega}^{-1}\mu\right) \qquad (6.4)$$

Substitution of (6.4) into (6.3) yields

$$\begin{aligned} &\mathbb{P}(\mathbf{a} \leq Z \leq \mathbf{b}) \\ &= \sqrt{\frac{|\boldsymbol{\Omega}|}{|\boldsymbol{\Sigma}|}} e^{\frac{1}{2}\mu'\boldsymbol{\Omega}^{-1}\mu} \tilde{\mathbb{E}}\left[\exp\left(-\frac{1}{2}Z'\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Omega}^{-1}\right)Z - \mu'\boldsymbol{\Omega}^{-1}Z\right) 1_{\mathbf{a} \leq Z \leq \mathbf{b}}\right] \\ &= c_0 \tilde{\mathbb{E}}\left[\exp\left(-\frac{1}{2}Z'\mathbf{A}Z - \mathbf{w}'Z\right) 1_{\mathbf{a} \leq Z \leq \mathbf{b}}\right] && (6.5) \\ &= c_0 e^{\frac{1}{2}\nu'\mathbf{A}\nu} \tilde{\mathbb{E}}\left[\exp\left(-\frac{1}{2}(Z-\nu)'\mathbf{A}(Z-\nu)\right) 1_{\mathbf{a} \leq Z \leq \mathbf{b}}\right] && (6.6) \end{aligned}$$

where

$$c_0 = \sqrt{\frac{|\boldsymbol{\Omega}|}{|\boldsymbol{\Sigma}|}} e^{\frac{1}{2}\mu'\boldsymbol{\Omega}^{-1}\mu} \qquad \mathbf{A} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Omega}^{-1} \qquad \mathbf{w} = \boldsymbol{\Omega}^{-1}\mu$$

and ν solves

$$\mathbf{A}\nu = -\mathbf{w}. \qquad (6.7)$$

Equivalently,

$$\left(\mathbf{I}-\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\right)\nu=\mu \tag{6.8}$$

If $\mathbf{A}$ is singular and $\mathbf{w}\notin \mathrm{Im}(\mathbf{A})$ then (6.7) has no solution and, as a consequence, the representation in (6.6) does not exist. Therefore, we impose that $\mathbf{A}$ is nonsingular. Equivalently, $\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}$ has no eigenvalue equal to 1. Under this condition,

$$\nu=\left(\mathbf{I}-\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\right)^{-1}\mu \tag{6.9}$$

Different choices of μ and $\boldsymbol{\Omega}$ lead to different expressions in (6.6). It will turn out that a diagonal structure of $\boldsymbol{\Omega}$ ensures that the integration region is rectangular under the new measure. Therefore, we impose that $\boldsymbol{\Omega}$ is diagonal with positive diagonal entries. Let $\mathbf{D}$ be the diagonal matrix with positive entries for which $\boldsymbol{\Omega}=\mathbf{D}^2$.

Moran (1983) chooses $\mu=\mathbf{0}$ and for $\boldsymbol{\Omega}$ a diagonal matrix with $\boldsymbol{\Omega}_{ii}=1/(\boldsymbol{\Sigma}^{-1})_{ii}$. In this case $\nu=\mathbf{0}$ always solves (6.7) regardless if $\mathbf{I}-\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}$ is nonsingular. Using the same $\boldsymbol{\Omega}$, Hashorva and Hüsler (2003) derive bounds from (6.5) for the case $\mathbf{b}=\infty\mathbf{1}$. To determine μ, they first obtain $\hat{\mathbf{x}}=\operatorname{argmin}_{\mathbf{x}\geq\mathbf{a}}\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}$, and derive for μ: If $\hat{x}_i=a_i$, then $\nu_i=a_i$ which gives $\mu_i\leq a_i$, otherwise $\mu_i=\nu_i=\hat{x}_i>a_i$.

Let $\tilde{\mathbf{x}}=\mathbf{D}^{-1}(\mathbf{x}-\nu)$ for any vector $\mathbf{x}$. Under the new probability measure, we find

$$\tilde{Z}=\mathbf{D}^{-1}(Z-\nu)\sim N(\mathbf{D}^{-1}(\mu-\nu),\mathbf{I})=N(\tilde{\mu},\mathbf{I}).$$

where using (6.8) and (6.9),

$$\begin{aligned}\tilde{\mu}&=\mathbf{D}^{-1}(\mu-\nu)\\&=-\mathbf{D}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\nu\\&=-\mathbf{D}\boldsymbol{\Sigma}^{-1}\nu &(6.10)\\&=-\mathbf{D}\boldsymbol{\Sigma}^{-1}\left(\mathbf{I}-\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\right)^{-1}\mu\\&=\left(\mathbf{I}-\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\right)^{-1}\mathbf{D}^{-1}\mu &(6.11)\end{aligned}$$

Rewriting (6.6), and using $\mathbf{D}_{ii}>0$

$$\begin{aligned}&\mathbb{P}(\mathbf{a}\leq Z\leq\mathbf{b})\\&=c_1\tilde{\mathbb{E}}\left[\exp\left(-\frac{1}{2}(Z-\nu)'\mathbf{D}^{-1}\mathbf{DADD}^{-1}(Z-\nu)\right)1_{\mathbf{D}^{-1}(\mathbf{a}-\nu)\leq\tilde{Z}\leq\mathbf{D}^{-1}(\mathbf{b}-\nu)}\right]\\&=c_1\tilde{\mathbb{E}}\left[\exp\left(\frac{1}{2}\tilde{Z}'\mathbf{B}\tilde{Z}\right)1_{\tilde{\mathbf{a}}\leq\tilde{Z}\leq\tilde{\mathbf{b}}}\right] &(6.12)\end{aligned}$$

$$= c_1 \sum_{k=0}^{\infty} \frac{1}{k!} \tilde{\mathbb{E}} \left[\left(\frac{1}{2} \left(\tilde{Z} 1_{\tilde{\mathbf{a}} \le \tilde{Z} \le \tilde{\mathbf{b}}} \right)' \mathbf{B} \left(\tilde{Z} 1_{\tilde{\mathbf{a}} \le \tilde{Z} \le \tilde{\mathbf{b}}} \right) \right)^k 1_{\tilde{\mathbf{a}} \le \tilde{Z} \le \tilde{\mathbf{b}}} \right] \tag{6.13}$$

$$= c_1 \left(\prod_{j=1}^{d} \left(\Phi(\tilde{b}_j - \tilde{\mu}_j) - \Phi(\tilde{a}_j - \tilde{\mu}_j) \right) + \sum_{k=1}^{\infty} \frac{1}{k!} \tilde{\mathbb{E}} \left[\left(\frac{U' \mathbf{B} U}{2} \right)^k \right] \right) \tag{6.14}$$

where $c_1 = c_0 e^{\frac{1}{2}\nu' \mathbf{A} \nu}$, $\mathbf{B} = -\mathbf{DAD} = \mathbf{I} - \mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{D}$, $\tilde{\mathbf{a}} = \mathbf{D}^{-1}(\mathbf{a} - \nu)$, $\tilde{\mathbf{b}} = \mathbf{D}^{-1}(\mathbf{b} - \nu)$, $U = \tilde{Z} 1_{\tilde{\mathbf{a}} \le \tilde{Z} \le \tilde{\mathbf{b}}}$, and $\Phi(x) = \mathbb{P}(Z_0 \le x)$ is the cumulative distribution function (cdf) of the univariate standard normal distribution $Z_0 \sim N(0, 1)$.

6.2.2 Restrictions on the new measure

The left panel in Figure 6.1 illustrates an example where $Z \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ is a bivariate normal distribution with correlation parameter $\rho = 0.7$. The right panel shows the same distribution under the new measure. More specifically, it represents (6.12) by plotting iso-probability curves of $\tilde{Z} \sim N(\tilde{\mu}, \mathbf{I})$ and iso-value curves of $\exp(\tilde{Z}'\mathbf{B}\tilde{Z}/2)$. The latter is a weighting function which is one at the origin and, by the identity $\tilde{\mathbf{z}}'\mathbf{B}\tilde{\mathbf{z}} = (-\tilde{\mathbf{z}})'\mathbf{B}(-\tilde{\mathbf{z}})$, symmetric with respect to the origin. For the specific case in Figure 6.1, the weight from $\exp(\tilde{Z}'\mathbf{B}\tilde{Z}/2)$ is maximal at the Northwest corner and the Southeast corner. This is because the one-dimensional eigenspace of the maximal eigenvalue of $\mathbf{B}$ points from the origin towards these directions. Under the original measure in (6.2), i.e., the left panel in Figure 6.1, a weighting function is absent, thus the weighting function is implicitly uniform.

Define $\hat{\mathbf{x}} = \mathbf{D}^{-1}\mathbf{x}$ such that $\tilde{\mathbf{x}} = \mathbf{D}^{-1}(\mathbf{x} - \nu) = \hat{\mathbf{x}} - \hat{\nu}$ for any vector $\mathbf{x}$. This gives the following d-dimensional vectors

$$\tilde{\mathbf{a}} = \hat{\mathbf{a}} - \hat{\nu} \qquad \tilde{\mathbf{b}} = \hat{\mathbf{b}} - \hat{\nu} \qquad \tilde{\mu} = -\mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{D}\hat{\nu} \qquad \tilde{\nu} = \mathbf{0}$$

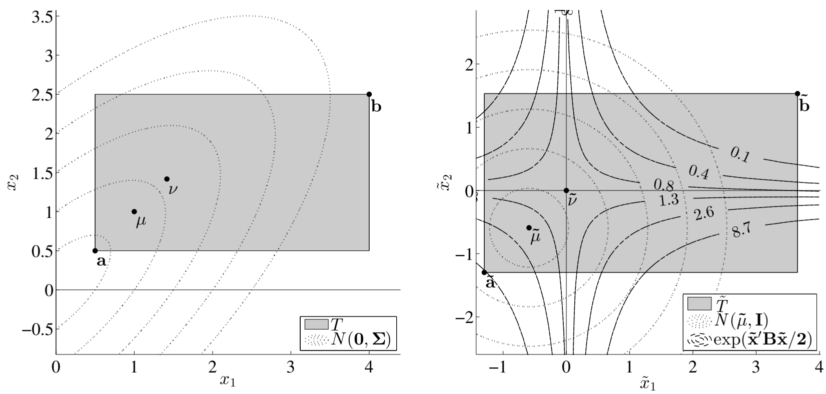
where we used (6.10) for $\tilde{\mu}$.

The new measure is completely determined by the set $\left\{\tilde{\mu}, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}\right\}$ as $\tilde{\nu} = 0$ always holds (see for instance the right panel in Figure 6.1). Both the vector μ and the diagonal matrix $\mathbf{D}$ with positive diagonal elements can be chosen arbitrarily, which gives a total of $d + d = 2d$ free parameters. Therefore, we can choose at most $2d$ parameters from a total of $3d$ parameters in the set $\left\{\tilde{\mu}, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}\right\}$. For instance, if one needs specific values for each of the elements in the vectors $\tilde{\mu}$ and $\tilde{\mathbf{a}}$, then the available $2d$ degrees of freedom of μ and $\mathbf{D}$ are used by pinning down $\tilde{\mu}$ and $\tilde{\mathbf{a}}$. This means that both $\tilde{\mathbf{b}}$ and $\mathbf{B}$ follow from the implied values of μ and $\mathbf{D}$.

Not all values are possible for the vectors in $\left\{\tilde{\mu}, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}\right\}$. For instance, the non-negativity of $\mathbf{D}^{-1}$ implies that the ranking of $\hat{\mu}_i$, $\hat{a}_i$, and $\hat{b}_i$ corresponds to the ranking of a_i, μ_i and b_i with $i = \{1, \ldots, d\}$. After subtracting $\tilde{\nu}_i$ from $\hat{\mu}_i$, $\hat{a}_i$, and $\hat{b}_i$, the ranking of $\tilde{\mu}_i$, $\tilde{a}_i$, and $\tilde{b}_i$ is of course still the same. Hence, we cannot freely choose this ranking. An additional restriction $\tilde{b}_i = \infty$ applies to all i with $b_i = \infty$.

Figure 6.1: **Change of measure.**

From bivariate normal $Z \sim N(\mathbf{0}, \mathbf{\Sigma})$ with correlation parameter $\rho = 0.7$ and zero means to $\mu = (1, 1)'$ and $\mathbf{\Omega} = 0.5\mathbf{I}$. Left panel: Integration region T, and iso-probability curves of $Z \sim N(\mathbf{0}, \mathbf{\Sigma})$. Right panel: Integration region $\tilde{T}$, iso-probability curves of $\tilde{Z} \sim N(\tilde{\mu}, \mathbf{I})$, and iso-value curves of $\exp(\tilde{Z}'\mathbf{B}\tilde{Z}/2)$.



We truncate the infinite series in (6.14) after n terms, and estimate the sum of the remaining terms. The series may diverge however for certain matrices $\mathbf{B}$, and hence for certain $\mathbf{\Omega} = \mathbf{D}^2$. Therefore, we derive necessary and sufficient convergence conditions in Section 6.3. This restricts the available $\mathbf{D}$, and so restricts the length $|\tilde{b}_i - \tilde{a}_i| = (\mathbf{D}_{ii})^{-1}|b_i - a_i|$ of the integration region along each dimension $i \in \{1, \ldots, d\}$ under the new probability measure. Next, we describe a procedure similar to Moran (1983) to obtain the moments of $U'\mathbf{B}U/2$ in (6.14). From now on, we simply write $\mathbb{E}$ to refer to $\tilde{\mathbb{E}}$.

6.2.3 The summants

We discuss the computation of the elements $(U'\mathbf{B}U/2)^k$ of the series in (6.14). The expansion of $(U'\mathbf{B}U)^k$ is a summation over d^{2k} different paths of expectations. Each path contains k elements of the matrix $\mathbf{B}$. The symmetry of **B** reduces the maximal number of different expectations over all paths to $(d(d+1)/2)^k$. In addition, the expectation of a certain path only depends on the number of times each nonzero element in **B** appears in the path. In other words, the ordering of appearance is irrelevant such that a path is characterized by a $d \times d$ upper triangular matrix **M** with $\Sigma_{i,j} m_{ij} = k$ where m_{ij} represents the number of times element b_{ij} is present in the path. For instance, if $k = 2$

$$\begin{aligned} &\mathbb{E}[U_1 b_{12} U_2 U_1 b_{13} U_3] && =\mathbb{E}[U_2 b_{12} U_1 U_1 b_{13} U_3] && =\mathbb{E}[U_1 b_{12} U_2 U_3 b_{13} U_1] \\ =&\mathbb{E}[U_2 b_{12} U_1 U_3 b_{13} U_1] && =\mathbb{E}[U_1 b_{13} U_3 U_1 b_{12} U_2] && =\mathbb{E}[U_3 b_{13} U_1 U_1 b_{12} U_2] \\ =&\mathbb{E}[U_1 b_{13} U_3 U_2 b_{12} U_1] && =\mathbb{E}[U_3 b_{13} U_1 U_2 b_{12} U_1] && =b_{12} b_{13} \mathbb{E}\left[U_1^2 U_2 U_3\right] \end{aligned}$$

All 8 paths are characterized by the $d \times d$ upper triangular matrix **M** with elements $m_{12} = m_{13} = 1$, and $m_{ij} = 0$ else.

More general, consider **M** with $k_0 = \sum_{i=1}^{d} m_{ii}$ selected diagonal elements of **B**. By the symmetry of **B**, we have for any path (a multiple of) 2^{k-k_0} paths with an identical representation **M**. Accordingly, we reduce the number of identical paths by a factor 2^{k-k_0} by using the matrix $\bar{\mathbf{B}}$ defined as the sum of the upper triangular matrix of **B** without the diagonal, and half the diagonal of **B**. The elements of the diagonal of $\bar{\mathbf{B}}$ are thus $\bar{b}_{ii} = b_{ii}/2$. This compensates for the double counting of paths that contain $U_i b_{ii} U_i$, and implies $\mathbf{B} = \bar{\mathbf{B}} + \bar{\mathbf{B}}'$. Let the function $c(\mathbf{M})$ represent the number of paths that share a given **M**. This number equals the multinomial coefficient $(k!/(m_{11}!m_{12}!\dots m_{dd}!))$. Using some straightforward manipulations,

$$\begin{aligned} \frac{1}{k!}\mathbb{E}\left[\left(\frac{U'\mathbf{B}U}{2}\right)^k\right] &= \frac{1}{k!}\mathbb{E}\left[\left(U'\bar{\mathbf{B}}U\right)^k\right] \\ &= \frac{1}{k!}\sum_{\sum_{i\le j} m_{ij}=k} c(\mathbf{M})\bar{b}_{11}^{m_{11}}\bar{b}_{12}^{m_{12}}\dots\bar{b}_{dd}^{m_{dd}}\mathbb{E}\left[U_1^{m_1}\dots U_d^{m_d}\right] \\ &= \sum_{\sum_{i\le j} m_{ij}=k} \frac{\bar{b}_{11}^{m_{11}}\bar{b}_{12}^{m_{12}}\dots\bar{b}_{dd}^{m_{dd}}}{m_{11}!m_{12}!\dots m_{dd}!}\mathbb{E}[U_1^{m_1}]\dots\mathbb{E}\left[U_d^{m_d}\right] \quad (6.15) \end{aligned}$$

The number (m_i) of times each U_i appears in a particular combination **M** is a simple summation over the elements corresponding to element i: $m_i = \sum_{j\ge i} m_{ij} + \sum_{j\le i} m_{ji}$. Indeed, this summation counts m_{ii} twice to account for the double appearance of U_i in $U_i \bar{b}_{ii} U_i$.

Equation (6.15) sums over a product of univariate moments. The moments of a truncated univariate standard normal distribution $V = Z_0 1_{l\le Z_0\le u}$

with $Z_0 \sim N(0,1)$ follow from the well-known recursion

$$\mathbb{E}\big[V^i\big] = (i-1)\,\mathbb{E}\big[V^{i-2}\big] + l^{i-1}\varphi\,(l) - u^{i-1}\varphi\,(u)$$

with initial conditions

$$\mathbb{E}\big[V^0\big] = \Phi(u) - \Phi(l) \qquad \mathbb{E}\big[V^1\big] = \int_l^u \xi\varphi\,(x)\,\mathrm{d}x = \varphi\,(l) - \varphi(u) \tag{6.16}$$

where the functions φ and Φ denote for a standard normal distribution the probability density function and the cdf, respectively.

Rather than the moments of V, we need the moments of each $U_i = \tilde{Z}1_{\tilde{\mathbf{a}}\leq\tilde{Z}\leq\tilde{\mathbf{b}}}$ in (6.14) with $\tilde{Z} \sim N(\tilde{\mu}_i, 1)$. Equivalently, we need the moments of the univariate random variables $U_i = (Z_0 + \tilde{\mu}_i)1_{l_i\leq Z_0\leq u_i}$ with $l_i = \tilde{a}_i - \tilde{\mu}_i$ and $u_i = \tilde{b}_i - \tilde{\mu}_i$. The moments of U_i follow by applying (6.15) and (6.16) on

$$\mathbb{E}\Big[U_i^k\Big] = \mathbb{E}\Big[(Z_0 + \tilde{\mu}_i)^k 1_{l_i\leq Z_0\leq u_i}\Big] = \sum_{j=0}^{k} \binom{k}{j} \tilde{\mu}_i^{k-j}\,\mathbb{E}\big[V^j\big] \tag{6.17}$$

where $k = 0, 1, 2, \ldots$.

6.3 Convergence conditions

The following theorem gives necessary and sufficient conditions for convergence of the series in (6.15), and thus (6.14).

Theorem 1. *Define* $X = \frac{1}{2}Z'\mathbf{B}Z1_{\mathbf{a}\leq Z\leq\mathbf{b}}$ *with*

- $Z \sim N(\mu, \mathbf{I})$
- $\mathbf{B} = \mathbf{I} - \mathbf{D}\mathbf{\Sigma}^{-1}\mathbf{D}$ *with both the matrices* $\mathbf{D}$ *and* $\mathbf{\Sigma}$ *positive definite*
- *the vector* $\mathbf{a}$ *satisfies* $|a_i| < \infty$

Further, let $I = \{i : b_i = \infty\}$, $\hat{d} = |I|$, $\hat{\mathbf{B}} = \mathbf{B}_{II}$, *and* $\hat{\lambda}(\hat{\mathbf{s}}) = \hat{\mathbf{s}}'\hat{\mathbf{B}}\hat{\mathbf{s}}$.

The series expansion $\sum_{k=0}^{\infty}\mathbb{E}\big[X^k/k!\big]$ *of* $\mathbb{E}[\exp(X)]$ *converges if one of the following conditions is true*

(i) $\mu_i \notin [a_i, b_i]$ *for some* $i \notin I$

(ii) the set $\mathcal{S} = \{\hat{\mathbf{s}} \in \mathbb{R}^{\hat{d}} : \hat{\lambda}(\hat{s}) \leq -1, \|\hat{\mathbf{s}}\| = 1, \hat{\mathbf{s}} \geq 0\}$ *is empty*

(iii) the set I *is empty*

(iv) the eigenvalues of $\hat{\mathbf{B}}$ *are larger than* -1

(v) the dominant eigenvalue of $\hat{\mathbf{B}}$ *is smaller than 1*

(vi) the absolute row sum $\sum_{j=1}^{\hat{d}} |\hat{\mathbf{B}}_{ij}|$ *is for each* i *smaller than 1*

(vii) any of the conditions (iv)-(vi) *above holds with* $\mathbf{B}$ *in place of* $\hat{\mathbf{B}} = \mathbf{B}_{II}$

The series expansion $\sum_{k=0}^{\infty} \mathbb{E}\big[X^k/k!\big]$ *of* $\mathbb{E}[\exp(X)]$ *diverges if the set* $\mathcal{S}$ *has nonzero measure.*

The crucial point of Theorem 1 is that a convergent series is available for any $\mathbf{\Sigma}$. For instance, $\mathbf{D} = \delta\mathbf{I}$ with

$$0 < \delta < \sqrt{2 \min_i \lambda_i(\mathbf{\Sigma})} = \sqrt{\frac{2}{\max_i \lambda_i(\mathbf{\Sigma}^{-1})}}$$

ensures

$$\min_i \lambda_i(\mathbf{B}) = \min_i \lambda_i(\mathbf{I} - \delta^2\mathbf{\Sigma}^{-1}) > \min_i \lambda_i \left(\mathbf{I} - \frac{2\mathbf{\Sigma}^{-1}}{\max_i \lambda_i(\mathbf{\Sigma}^{-1})}\right) = -1$$

which produces by $1(iv)$ and (vii) in Theorem 1 a convergent series $\mathbb{E}\big[X^k/k!\big]$.

Moran (1983) provides condition (vi) as a sufficient convergence condition. The absolute value of each eigenvalue is necessarily smaller than one if this sufficient condition is satisfied. Thus, our condition (iv) is less strict. The following example illustrates that the sufficient conditions in Theorem $1(iii)$-(vii) are not necessary for convergence.

Example Let $\mathbf{a} = 0$, $\mathbf{b} = \infty$, and define the positive definite matrix

$$\Sigma = \begin{pmatrix} 1 & 0.864 & 0.964 & -0.966 \\ & 1 & 0.938 & -0.933 \\ & & 1 & -0.983 \\ & & & 1 \end{pmatrix}$$

where we neglect the symmetric lower triangular part for convenience. Following Moran (1983), choose $\mathbf{\Omega} = \mathbf{D}^2 = (\mathrm{diag}(\mathbf{\Sigma}^{-1}))^{-1}$. This gives

$$\mathbf{B} = \begin{pmatrix} 0 & -0.549 & 0.495 & -0.510 \\ & 0 & 0.501 & -0.404 \\ & & 0 & -0.394 \\ & & & 0 \end{pmatrix}$$

In this case, $\mathbf{B} = \mathbf{I} - \mathbf{D}\Sigma^{-1}\mathbf{D}$ has zero diagonal and the off-diagonal elements contain the partial correlation coefficients. This matrix $\mathbf{B}$ has an eigenvalue -1.14 which is smaller than -1. Thus, the matrix does not satisfy any of the sufficient conditions (*iv*)-(*vii*) in Theorem 1.

The dominant eigenvector of $\mathbf{B}$ that corresponds to the eigenvalue -1.14 is given by $(0.610, 0.590, -0.405, 0.340)$. This vector has a negative entry which means that this eigenvector is not in $\mathcal{S}$. The unit vectors $\mathbf{s}$ in the positive orthant do satisfy $\lambda(\mathbf{s}) > -1$ since the vector $\mathbf{s}^* = (0.609, 0.568, 0, 0.554)$ solves

$$\min_{\{\mathbf{s}:\|\mathbf{s}\|=1, \mathbf{s}>0\}} \lambda(\mathbf{s})$$

and admits $\lambda(\mathbf{s}^*) = \mathbf{s}^{*\prime}\mathbf{B}\mathbf{s}^* = -0.978 > -1$. This implies that the set $\mathcal{S}$ in Theorem 1(*ii*) is empty such that the series $\mathbb{E}\left[(Z'\mathbf{B}Z/2)^k\right]/k!$ converges for any $Z \sim N(\mu, \mathbf{I})$).

6.4 Error bounds

A truncation of the expansion in (6.14) after n terms leads to a truncation error equal to the remainder $c_1 \sum_{k=n+1}^{\infty} \frac{1}{k!}\tilde{\mathbb{E}}[U'\mathbf{B}U/2]^k$. In this section we find upper and lower bounds for this remainder. First, we introduce some notation.

Let X be a random variable with distribution function F. Unless indicated otherwise, we assume $\mathbb{P}(X < 0) > 0$, and $\mathbb{P}(X > 0) > 0$. Denote the positive and negative parts of X by $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$, respectively. Let μ_n represent the n-th moment of X, and define the moment ratio $m_n = \mu_n/\mu_{n-1}$. Let μ_n^+, μ_n^-, m_n^+ and m_n^- refer to the moments and moment ratios of the positive parts and the negative parts, respectively. Notice that $\mu_n = \mu_n^+ + (-1)^n\mu_n^-$, and, if n even, $\mu_n > 0$.

For $\mu_n > 0$, let $X^{(n)}$ represent the random variable with cumulative distribution function $F^{(n)}$ that satisfies

$$\mathrm{d}F^{(n)}(x) = \begin{cases} \frac{1}{\mu_n} x^n \mathrm{d}F(x) & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{6.18}$$

This defines a distribution function since $\mathrm{d}F^{(n)}(x) \geq 0$ and $\int \mathrm{d}F^{(n)}(x) = 1$. It follows that the i-th moment of $X^{(n)}$ equals

$$\mu_i^{(n)} = \frac{\mu_{n+i}}{\mu_n} \tag{6.19}$$

We use the operator $\leq_{\mathrm{st}}$ and the term dominance to refer to first order stochastic dominance.

The basic idea is to estimate the remainder of $\exp(X) = \exp\left(U'\mathbf{B}U/2\right)$ in (6.14) after n terms by bounding

$$\sum_{i=n+1}^{\infty} \frac{\mathbb{E}\left[X^i\right]}{i!} = \sum_{i=n+1}^{\infty} \frac{\mathbb{E}\left[(X^-)^i\right] + \mathbb{E}\left[(X^+)^i\right]}{i!}$$
$$= \mu_n^- s_n(X_n^-) + \mu_n^+ s_n(X_n^+) \tag{6.20}$$

where

$$s_n(X_n^-) = \frac{1}{\mu_n^-} \sum_{i=n+1}^{\infty} \frac{\mu_i^-}{i!} \qquad s_n(X_n^+) = \frac{1}{\mu_n^+} \sum_{i=n+1}^{\infty} \frac{\mu_i^+}{i!}$$

We bound each of the four elements in the right-hand side of (6.20). Bounds on μ_n^- and μ_n^+ are in section 6.4.1, whilst section 6.4.2 contains bounds on $s_n(X^-)$ and $s_n(X^+)$. They jointly provide a lower and upper bound on the remainder $\sum_{i=n+1}^{\infty} \mathbb{E}\left[X^i\right]/i!$.

6.4.1 Bounds on μ_n^- and μ_n^+

This section starts with five lemmas. We end this section with Theorem 7 that summarizes the algorithm to bound μ_n^- and μ_n^+.

Lemma 2. *(i) For nonnegative X and $n \geq 2$, or for arbitrary X and even $n \geq 2$*

$$\mu_n \mu_{n-2} - \mu_{n-1}^2 \geq 0$$

(ii) For nonnegative X with $\mu_1 > 0$

$$0 < m_1 \leq m_2 \leq m_3 \leq \dots$$

The inequalities are all strict if and only if X is nondegenerate.

The next lemma provides preliminary bounds on μ_i^s and m_i^s ($s \in \{+, -\}$).

Lemma 3. *For $i \geq 1$, even $n \geq 2$ and $s \in \{+, -\}$, we have*

$$\mu_i^s, m_i^s \geq 0 \qquad m_1^s = \mu_1^s \qquad \underline{x}^s \leq m_{i+1}^s \leq \bar{x}^s \tag{6.21}$$

$$\mu_n^s \leq \mu_n \tag{6.22}$$

$$\mu_{n-1}^+ \geq \mu_{n-1} \qquad \frac{1}{m_n^+} \geq \frac{1}{m_n} \qquad m_{n-1}^+ \geq m_{n-1} \tag{6.23}$$

$$\mu_{n-1}^- \geq -\mu_{n-1} \qquad \frac{1}{m_n^-} \geq -\frac{1}{m_n} \qquad m_{n-1}^- \geq -m_{n-1} \tag{6.24}$$

where the support $[\underline{x}^s, \bar{x}^s]$ *of* X^s $(s \in \{+,-\})$ *on the positive half-line* $(0,\infty)$ *is given by*

$$\underline{x}^s = \inf\left(x \,|\, \mathbb{P}(X^s \le x) > \mathbb{P}(X^s = 0)\right) \qquad \bar{x}^s = \min\left(x \,|\, \mathbb{P}(X^s \le x) = 1\right)$$

(*i*) *The nonnegativity constraints on* μ_i^s *and* m_i^s *hold with equality if and only if* $X^s \equiv 0$.

(*ii*) *The bounds* $\underline{x}^s \le m_{i+1}^s \le \bar{x}^s$ *hold with equality if and only if* X^s *is degenerate on* $(0,\infty)$.

(*iii*) *The bounds on* μ_i^+ *and* m_i^+ *in* (6.22)–(6.24) *hold with equality if and only if* $X^- \equiv 0$.

(*iv*) *The bounds on* μ_i^- *and* m_i^- *in* (6.22)–(6.24) *hold with equality if and only if* $X^+ \equiv 0$.

The next lemma enables us to bound μ_i^s in terms of the bounds on m_i^s. The required moments and upper bounds will be available in our application of the lemma.

Lemma 4. *Consider* X *with known moments* μ_{n-2}, μ_{n-1}, μ_n *(*n *even). Upper and lower bounds on* m_{n-1}^s *and* m_n^s *are known (*$s \in \{+,-\}$*) and denoted by lower and upper bars, respectively. Let*

$$a_n = \mu_n - \frac{\mu_{n-1}^2}{\mu_{n-2}} \qquad b_n = \frac{\mu_n}{\mu_{n-2}} \tag{6.25}$$

$$u_n^+(x) = \frac{a_n}{b_n/x - 2m_{n-1} + x} \qquad u_n^-(x) = \frac{a_n}{b_n/x + 2m_{n-1} + x} \tag{6.26}$$

$$v_n^s(c,d) = \begin{cases} u_n^s(c) & \sqrt{b_n} < c \\ u_n^s(\sqrt{b_n}) & c \le \sqrt{b_n} \le d \\ u_n^s(d) & \sqrt{b_n} > d \end{cases}$$

The moments of X^s *are bounded by*

$$\begin{aligned} \mu_{n-2}^s &\le \frac{u_n^s(\underline{m}_{n-1}^s)}{\underline{m}_{n-1}^s} \\ \mu_{n-1}^s &\le \min\left(v_n^s\left(\underline{m}_n^s, \bar{m}_n^s\right), v_n^s\left(\underline{m}_{n-1}^s, \bar{m}_{n-1}^s\right)\right) \\ \mu_n^s &\le \bar{m}_n^s u_n^s(\bar{m}_n^s) \end{aligned}$$

The bounds on μ_{n-2}^s *and* μ_n^s *are (i) strictly tighter if the bounds* $\underline{m}_{n-1}^s$ *and* $\bar{m}_n^s$ *are tighter, and (ii) strict if and only if* $|X|$ *is nondegenerate.*

To implement possible improvements in the bounds, the next lemma applies $\mu_i^s = m_i^s \mu_{i-1}^s$ and $\mu_i = \mu_i^+ + (-1)^i \mu_i^-$ on the bounds in Lemma 4.

Lemma 5. *Let n even, $s \in \{+,-\}$, then for $i = 2, 3, 4, \ldots$*

$$\mu_i^s \geq \max\left(\frac{\left(\underline{\mu}_{i-1}^s\right)^2}{\bar{\mu}_{i-2}^s}, \frac{\left(\underline{\mu}_{i+1}^s\right)^2}{\bar{\mu}_{i+2}^s}, \underline{m}_i^s \underline{\mu}_{i-1}^s, \frac{\underline{\mu}_{i+1}^s}{\bar{m}_{i+1}^s}\right)$$

$$\mu_i^s \leq \min\left(\sqrt{\bar{\mu}_{i-1}^s \bar{\mu}_{i+1}^s}, \bar{m}_i^s \bar{\mu}_{i-1}^s, \frac{\bar{\mu}_{i+1}^s}{\underline{m}_{i+1}^s}\right)$$

$$\mu_{n-1}^+ \geq \underline{\mu}_{n-1}^- + \mu_{n-1} \qquad \mu_{n-1}^+ \leq \bar{\mu}_{n-1}^- + \mu_{n-1}$$

$$\mu_{n-1}^- \geq \underline{\mu}_{n-1}^+ - \mu_{n-1} \qquad \mu_{n-1}^- \leq \bar{\mu}_{n-1}^+ - \mu_{n-1}$$

$$\mu_n^+ \geq \mu_n - \bar{\mu}_n^- \qquad \mu_n^+ \leq \mu_n - \underline{\mu}_n^-$$

$$\mu_n^- \geq \mu_n - \bar{\mu}_n^+ \qquad \mu_n^- \leq \mu_n - \underline{\mu}_n^+$$

Lemma 6 bounds m_i^s in terms of the bounds on μ_i^s.

Lemma 6. *Consider a random variable X with known moments μ_i and known bounds $\underline{\mu}_i^s$ and $\bar{\mu}_i^s$ on μ_i^s ($s \in \{+,-\}$). The following inequalities hold*

$$\frac{\underline{\mu}_i^s}{\bar{\mu}_{i-1}^s} \leq m_i^s \leq \frac{\bar{\mu}_i^s}{\underline{\mu}_{i-1}^s} \qquad \sqrt{\frac{\underline{\mu}_i^s}{\bar{\mu}_{i-2}^s}} \leq m_i^s \leq \sqrt{\frac{\bar{\mu}_{i+1}^s}{\underline{\mu}_{i-1}^s}} \tag{6.27}$$

$$\frac{\mu_i - \bar{\mu}_i^-}{\bar{\mu}_{i-1}^+} \leq m_i^+ \leq \frac{\mu_i - \underline{\mu}_i^-}{\underline{\mu}_{i-1}^+} \qquad \frac{\mu_i - \bar{\mu}_i^+}{\bar{\mu}_{i-1}^-} \leq m_i^- \leq \frac{\mu_i - \underline{\mu}_i^+}{\underline{\mu}_{i-1}^-} \qquad \textit{even } i \tag{6.28}$$

$$\frac{\mu_i + \underline{\mu}_i^-}{\bar{\mu}_{i-1}^+} \leq m_i^+ \leq \frac{\mu_i + \bar{\mu}_i^-}{\underline{\mu}_{i-1}^+} \qquad \frac{\underline{\mu}_i^+ - \mu_i}{\bar{\mu}_{i-1}^-} \leq m_i^- \leq \frac{\bar{\mu}_i^+ - \mu_i}{\underline{\mu}_{i-1}^-} \qquad \textit{odd } i \tag{6.29}$$

The bounds on m_i^s are strict if and only if the bounds on μ_i^s are strict. The identity $\sqrt{\frac{\underline{\mu}_i^s}{\bar{\mu}_{i-2}^s}} = m_i^s = \sqrt{\frac{\bar{\mu}_{i+1}^s}{\underline{\mu}_{i-1}^s}}$ holds if and only if X^s is degenerate on $(0, \infty)$.

In addition to Lemma 3-6, we check for improvements in the bounds by employing Lemma 2(ii) on the obtained bounds ($s \in \{+,-\}$): $m_{i-1}^s \leq m_i^s \leq m_{i+1}^s$. Subsequently, we return to Lemma 4 and repeat the loop until none of the bounds has changed during a loop. Theorem 7 summarizes the algorithm that enables us to find bounds on the moments μ_n^- and μ_n^+, and the corresponding moment ratios $m_n^- = \mu_n^-/\mu_{n-1}^-$ and $m_n^+ = \mu_n^+/\mu_{n-1}^+$.

Theorem 7. *Assume that the required moments μ_i are known. The index j denotes an index that depends on i.*

(i) Initialize bounds on m_i^- and m_i^+ (Lemma 3).

(ii) Bound μ_i^- and μ_i^+ in terms of the bounds on m_j^- and m_j^+ (Lemma 4).

(iii) Bound μ_i^- and μ_i^+ in terms of the bounds on μ_j^-, μ_j^+, m_j^-, and m_j^+ (Lemma 5).

(iv) Bound m_i^- and m_i^+ in terms of the bounds on μ_j^- and μ_j^+ (Lemma 6).

(v) Bound m_i^- and m_i^+ in terms of the bounds on m_j^- and m_j^+ (Lemma 2(ii)).

(vi) Stop if none of the bounds has changed since step (ii), otherwise go to step (ii).

6.4.2 Bounds on $s_n(X^-)$ and $s_n(X^+)$

Define for $n = 0, 1, 2, \ldots$

$$p_n(x) = \sum_{i=0}^{n} \frac{x^i}{i!} \quad r_n(x) = e^x - p_n(x) \quad t_n(x) = \begin{cases} x^{-n} r_n(x) & x \neq 0 \\ 0 & x = 0 \end{cases} \tag{6.30}$$

and for a random variable X with $\mathbb{E}[X^n] \neq 0$ the scaled remainder

$$s_n(X) = \frac{\mathbb{E}[r_n(X)]}{\mathbb{E}[X^n]}. \tag{6.31}$$

Define for $t < 1$ and $\lambda \neq 0$:

$$\begin{aligned} c(t; d, \xi, u) &= \mathbb{E}\left[Z_{t,\xi}^{d-1} 1_{0 \le Z_{t,\xi} \le u\sqrt{1-t}}\right] \qquad Z_{t,\xi} \sim N\left(\frac{\xi}{\sqrt{1-t}}, 1\right) \\ m(t; d, \xi, u) &= \frac{c(t; d, \xi, u)}{c(0; d, \xi, u)(1-t)^{d/2}} \exp\left(\frac{\xi^2 t}{2-2t}\right) \\ \hat{\mu}_i &= \left(\frac{\lambda}{2}\right)^i \frac{c(0; d+2i, \xi, u)}{c(0; d, \xi, u)} \\ X_{\xi,u} &= \frac{1}{2} R^2 \end{aligned} \tag{6.32}$$

$$s_n\left(\lambda X_{\xi,u}\right)=\begin{cases}\frac{1}{\lambda^n\Gamma(d/2+n)}\left(\frac{\Gamma(d/2)}{(1-\lambda)^{d/2}}-\sum_{i=0}^{n}\frac{\lambda^i\Gamma(d/2+i)}{i!}\right) & \text{if } \xi=0\\ =\frac{1}{n!}\left[{}_2F_1\left(1,n+\frac{d}{2};n+1;\lambda\right)-1\right] & \text{and } u=\infty\\ \frac{1}{\hat{\mu}_n}\left(m(\lambda;d,\xi,u)-\sum_{i=0}^{n}\frac{\hat{\mu}_i}{i!}\right) & \text{else}\end{cases} \tag{6.33}$$

where the random variable R has the density

$$f_R(r;d,\xi,u)=\frac{r^{d-1}}{c(0;d,\xi,u)\sqrt{2\pi}}\exp\left(-\frac{(r-\xi)^2}{2}\right)1_{r\le u}. \tag{6.34}$$

A recursive expression for $c(t;d,\xi,u)$ is in (6.17).

Theorem 8. *Suppose $\mu\le 0$ and $\mathbf{a}\le\mathbf{0}\le\mathbf{b}$ with $\mathbf{a}$ finite. Let $X=\frac{1}{2}Z'\mathbf{B}Z1_{\mathbf{a}\le Z\le\mathbf{b}}$ with $Z\sim N(\mu,\mathbf{I})$. The scaled remainder after n terms of X^- and X^+ is bounded by respectively*

$$\min\left(s_n\left(\underline{\lambda}X_{0,\infty}\right),s_n\left(\underline{\lambda}X_{\|\mu\|,\bar{u}}\right)\right)\le s_n(-X^-)<0 \tag{6.35}$$

$$t_n(\underline{m}_{n+1}^+)\le s_n(X^+)\le\max\left(s_n\left(\bar{\lambda}X_{0,\infty}\right),s_n\left(\bar{\lambda}X_{\|\mu\|,\bar{u}}\right)\right) \tag{6.36}$$

where t_n and s_n are as in (6.30) *and* (6.33)*, and*

(*i*) $\underline{\lambda}$ *is a lower bound on* $\min_{\|\mathbf{s}\|=1}\{\lambda(\mathbf{s}):\lambda(\mathbf{s})\le 0, s_i\ge 0 \text{ if } a_i=0\}$*, i.e.,* $\underline{\lambda}\ge\lambda_{\min}$

$\bar{\lambda}$ *is an upper bound on* $\max_{\|\mathbf{s}\|=1}\{\lambda(\mathbf{s}):\lambda(\mathbf{s})\ge 0, s_i\le 0 \text{ if } b_i=0\}$*, i.e.,* $\bar{\lambda}\le\lambda_{\max}$

where $\lambda(\mathbf{s})=\mathbf{s}'\mathbf{B}\mathbf{s}$ and $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalue of $\mathbf{B}$*, respectively.*

(*ii*) $\underline{m}_{n+1}^+$ *is a lower bound on* $\mathbb{E}\left[(X^+)^{n+1}\right]/\mathbb{E}\left[(X^+)^n\right]$

(*iii*) $\bar{u}$ *is an upper bound on* $\sup_{\mu'\mathbf{x}>0,\mathbf{x}\ge\mathbf{a}}\|\mathbf{x}\|$*, i.e.,*

$$\bar{u}\le\max_j\sqrt{\left(\frac{1}{\mu_j}\sum_{i\ne j}a_i\mu_i\right)^2+\sum_{i\ne j}a_i^2} \tag{6.37}$$

(*iv*) $s_n(\lambda X_{0,\infty})$ *is strictly convex and strictly increasing in λ with*

$$\lim_{\lambda\to 0}s_n(\lambda X_{0,\infty})=0.$$

Compared to Theorem 8, tighter bounds on $s_n(X^-)$ and $s_n(X^+)$ are available by (i) considering $X_{\mathbf{s}} + X_{-\mathbf{s}}$ along a line, instead of $X_{\mathbf{s}}$ and $X_{-\mathbf{s}}$ along separate half-lines, or (ii) splitting the d-dimensional unit ball S_d in the proof of Theorem 8 in smaller subsets for $\mathbf{s}$ than only the two subsets based on the sign of $\lambda(\mathbf{s})$. Both adjustments could improve our method at the expense of even more complicated bounds than in Theorem 8.

We employ Theorem 8 under the new measure, i.e., with tildes: $-\infty\mathbf{1} < \tilde{\mathbf{a}} \leq \mathbf{0} \leq \tilde{\mathbf{b}}$, $\tilde{\mu}$, and $\tilde{Z} \sim N(\tilde{\mu}, \mathbf{I})$.

6.5 The covariance matrix and the location vector

Our method has some degrees of freedom in choosing the covariance matrix $\mathbf{\Omega}$ and the location vector μ. We propose some possible choices for $\mathbf{\Omega}$ and μ. The discussion follows an intuitive though general approach. Specific choices are preferable for specific parameters in (6.1).

6.5.1 The covariance matrix $\mathbf{\Omega}$

The matrix $\mathbf{\Omega}$ should at least ensure convergence of the series in (6.14). Moran (1983) selects $\mathbf{\Omega}$ such that $\mathbf{B}$ has a zero diagonal. Unfortunately, this choice may result into a divergent series if the dominant eigenvalue of $\mathbf{B}$ exceeds one (see Theorem 1). Each of the three proposed choices for $\mathbf{\Omega}$ resolves this issue. When feasible, a zero diagonal on $\mathbf{B}$ is still interesting from a computational point of view in the summation (6.15).

Without loss of generality, we assume that the variance of each Z_i in the original problem, i.e., the diagonal elements of $\mathbf{\Sigma}$, are identical. Therefore, we solely consider multiples of the identity matrix: $\mathbf{\Omega}(\omega) := \omega\mathbf{I}$ such that $\mathbf{B} = \mathbf{I} - \omega\mathbf{\Sigma}^{-1}$, and $\lambda(\mathbf{B}) = 1 - \omega\lambda(\mathbf{\Sigma}^{-1})$. This gives

$$\lambda_{\min}(\mathbf{B}) = 1 - \omega\lambda_{\max}(\mathbf{\Sigma}^{-1}) \qquad \lambda_{\max}(\mathbf{B}) = 1 - \omega\lambda_{\min}(\mathbf{\Sigma}^{-1}) \tag{6.38}$$

We propose the following three choices for ω:

(i) By Theorem 8(iv), the absolute value of the scaled remainder increases in the absolute value of λ. That is, a small dominant eigenvalue leads to a small scaled remainder. To minimize the dominant eigenvalue, choose $\omega > 0$ such that $-\lambda_{\min}(\mathbf{B}) = \lambda_{\max}(\mathbf{B})$. Equivalently, by (6.38)

$$-1 + \omega\lambda_{\max}(\mathbf{\Sigma}^{-1}) = 1 - \omega\lambda_{\min}(\mathbf{\Sigma}^{-1}),$$

which gives

$$\omega = \frac{2}{\lambda_{\min}(\mathbf{\Sigma}^{-1}) + \lambda_{\max}(\mathbf{\Sigma}^{-1})} \tag{6.39}$$

This ω minimizes the dominant eigenvalue of $\mathbf{B}$ among all $\mathbf{\Omega}(\omega)$. Lower ω correspond to higher $|\lambda_{\min}(\mathbf{B})|$, whilst higher ω correspond to higher $|\lambda_{\max}(\mathbf{B})|$. Convergence of the expansion in (6.13) follows from Theorem 1(*iv*) and (*vii*), and the positive definiteness of $\mathbf{\Sigma}^{-1}$:

$$\lambda_{\min}(\mathbf{B}) = 1 - \omega\lambda_{\max}(\mathbf{\Sigma}^{-1}) = 1 - \frac{2}{\lambda_{\min}(\mathbf{\Sigma}^{-1})/\lambda_{\max}(\mathbf{\Sigma}^{-1}) + 1} > -1$$

Notice that $|\lambda_{\min}(\mathbf{B})|$ is close to zero if the condition number of $\mathbf{\Sigma}^{-1}$, i.e., the ratio $\lambda_{\max}(\mathbf{\Sigma}^{-1})/\lambda_{\min}(\mathbf{\Sigma}^{-1})$ is close to one. Equivalently, the condition number of $\mathbf{\Sigma}$ is ideally close to one.

(*ii*) Method (*i*) does not take into account the convexity of s_n in λ, see Theorem 8(*iv*). In other words, it underestimates the adverse effect of a large positive $\lambda_{\max}(\mathbf{B})$ on s_n. The following more complicated approach resolves this issue by minimizing the scaled remainder, or the remainder of the upper bound of unbounded unit vectors. Motivated by the bounds on nontruncated directions $X_{0,\infty}$ in (6.35)–(6.36) and (6.38), we find $\omega > 0$ that minimizes

$$\min_{\omega>0} \big|\, |s_n(\lambda_{\max}(\mathbf{B})X_{0,\infty})| - |s_n(\lambda_{\min}(\mathbf{B})X_{0,\infty})| \,\big|$$

Equivalently, using (6.33)

$$\min_{\omega>0} \left| \left| {}_2F_1\begin{pmatrix} 1, n+\frac{d}{2}; n+1; \\ \lambda_{\max}(\mathbf{B}) \end{pmatrix} \right| - \left| {}_2F_1\begin{pmatrix} 1, n+\frac{d}{2}; n+1; \\ \lambda_{\min}(\mathbf{B}) \end{pmatrix} \right| \right|$$

Uniqueness of this optimization problem follows from the convexity of the hypergeometric function ${}_2F_1\left(1, n+\frac{d}{2}; n+1; z\right)$ in z.

(*iii*) Another approach is to equalize the bounds on the remainders r_n of X^+ and X^-:

$$\min_{\omega>0} \big|\, |r_n(\lambda_{\max}(\mathbf{B})X_{0,\infty})| - |r_n(\lambda_{\min}(\mathbf{B})X_{0,\infty})| \,\big|$$

where, by (6.31) and (6.33)

$$r_n(\lambda X_{0,\infty}) = \frac{1}{(1-\lambda)^{d/2}} - \frac{1}{\Gamma(d/2)} \sum\nolimits_{i=0}^{n} \frac{\lambda^i \Gamma\left(d/2+i\right)}{i!}$$

Uniqueness of ω follows from the convexity of

$$|r_n(\lambda X_{0,\infty})| = |s_n(\lambda X_{0,\infty})| |\mathbb{E}[\lambda X_{0,\infty}]^n|$$

in λ, which in turn follows from the convexity in λ of both components on the right-hand side.

6.5.2 The central vector μ

Recall from (6.14),

$$\mathbb{P}(\mathbf{a} \leq Z \leq \mathbf{b}) = c_1 \left(\prod_{j=1}^{d} \left(\Phi(\tilde{b}_j - \tilde{\mu}_j) - \Phi(\tilde{a}_j - \tilde{\mu}_j) \right) + \sum_{k=1}^{\infty} \frac{1}{k!} \mathbb{E}\left[\left(\frac{1}{2} U' \mathbf{B} U \right)^k \right] \right)$$

where

$$c_1 = \sqrt{\frac{|\boldsymbol{\Omega}|}{|\boldsymbol{\Sigma}|}} e^{\frac{1}{2}\mu'\boldsymbol{\Omega}^{-1}\mu + \frac{1}{2}\nu'(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Omega}^{-1})\nu} \qquad \mathbf{B} = \mathbf{I} - \mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{D}$$

$$\nu = -\left(\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1} - \mathbf{I}\right)^{-1} \mu \qquad \tilde{\mu} = \mathbf{D}^{-1}(\mu - \nu)$$

$$\tilde{\mathbf{a}} = \mathbf{D}^{-1}(\mathbf{a} - \nu) \qquad \tilde{\mathbf{b}} = \mathbf{D}^{-1}(\mathbf{b} - \nu)$$

$$U = \tilde{Z} 1_{\tilde{\mathbf{a}} \leq \tilde{Z} \leq \tilde{\mathbf{b}}} \qquad \tilde{Z} \sim N(\tilde{\mu}, \mathbf{I})$$

To compute the summants of $\frac{1}{k!}\mathbb{E}\left[\left(\frac{1}{2}U'\mathbf{B}U\right)^k\right]$ in (6.14), we derived (6.15):

$$\frac{1}{k!}\mathbb{E}\left[\left(\frac{1}{2}U'\mathbf{B}U\right)^k\right] = \sum_{\sum_{i \leq j} m_{ij} = k} \frac{\bar{b}_{11}^{m_{11}} \bar{b}_{12}^{m_{12}} \ldots \bar{b}_{d,d}^{m_{d,d}}}{m_{11}! m_{12}! \ldots m_{d,d}!} \mathbb{E}[U_1^{m_1}] \ldots \mathbb{E}\left[U_d^{m_d}\right]$$

where $\bar{b}_{ii} = b_{ii}/2$, $\bar{b}_{ij} = b_{ij}$, $(i \neq j)$, and $m_i = \sum_{j \geq i} m_{ij} + \sum_{j \leq i} m_{ji}$. For a small power k, single powers $m_i = 1$ are relatively abundant in the summation. As we prefer to truncate the series after a small number of elements, we choose the vector μ such that $\mathbb{E}[U_i] = 0$ for $i = 1, \ldots, d$. This ensures $\mathbb{E}[U'\mathbf{B}U] = 0$ provided $\mathbf{B}$ has a zero diagonal.

Lemma 9. *The unique $\mu = \mathbf{D}\mathbf{B}\hat{\nu}$ that satisfies $\mathbb{E}[U_i] = 0$ corresponds to the unique fixed-point $\mathbf{f}(\hat{\nu}) = \hat{\nu}$ with*

$$f_i(\hat{\nu}) = \frac{\phi_i\left(\hat{\mathbf{a}} - \mathbf{B}\hat{\nu}\right) - \phi_i\left(\hat{\mathbf{b}} - \mathbf{B}\hat{\nu}\right)}{\Phi_i\left(\hat{\mathbf{b}} - \mathbf{B}\hat{\nu}\right) - \Phi_i\left(\hat{\mathbf{a}} - \mathbf{B}\hat{\nu}\right)} + \mathbf{e}_i\mathbf{B}\hat{\nu} \qquad i = 1, \ldots, d \tag{6.40}$$

For the sequence $\hat{\nu}^{(k)} = \mathbf{f}(\hat{\nu}^{(k-1)})$, the distance $\|\hat{\nu}^{(k)} - \hat{\nu}^*\|$ to the fixed point $\hat{\nu}^*$ decreases each step with at least the factor $\max_i |\lambda_i(\mathbf{B})|$.

Numerical problems may arise if both terms in the denominator of f_i in (6.40) are close to zero or close to one. Accordingly, we initialize the sequence with $\hat{\nu}^{(0)} = \mathbf{B}^{-1}\hat{\mathbf{a}}$ to ensure that the second term in the denominator of f_i in (6.40) is one half.

Numerical experiments confirm that the sequence $\hat{\nu}^{(k)} = \mathbf{f}(\hat{\nu}^{(k-1)})$ converges fast towards the fixed-point $\hat{\nu}^* = \mathbf{f}(\hat{\nu}^*)$ that corresponds to $\mathbb{E}[U_i] = 0$ $(i = 1, \ldots, d)$. The location vector μ follows from $\mu = \mathbf{DB}\hat{\nu}$.

Using this vector μ, the summation in (6.15) ideally neglects summants involving at least one unit power $\mathbb{E}[U_i] = 0$. This could speed up the computation time of our algorithm further, as particularly this summation is computationally the most involving. However, the error bounds should take into account the very small, though existent, approximation error in μ from the fixed-point algorithm. This error translates into a small though nonzero $|\mathbb{E}[U_i]|$ which leads to $\mathbb{E}[U'\mathbf{B}U] \neq 0$, even if $\mathbf{B}$ has a zero diagonal.

Powers $m > 1$ of $\mathbb{E}[U_i^m] = 0$ $(i = 1, \ldots, d)$ are more abundant in (6.15) when the power k of $\mathbb{E}\big[(U'\mathbf{B}U)^k\big]$ is higher. It is straightforward to extend the analysis to the condition $\mathbb{E}[U_i^m] = 0$ for some odd integer $m > 1$. The benefit of $\mathbb{E}[U_i^m] = 0$ with a higher power m is a potentially smaller $\mathbb{E}\big[(U'\mathbf{B}U)^k\big]$ for a higher k which is ideally close to the number of truncation terms n. Though interesting, we keep the analysis simple here by choosing $m = 1$, since it is computationally more expensive to compute μ for $\mathbb{E}[U_i^m] = 0$ with a higher m.

6.5.3 Comparison for equicorrelated distributions

We compare several μ and $\mathbf{\Omega}$ for the special case that $\mathbf{\Sigma}$ corresponds to an equicorrelated distribution. In doing this, we consider the d-dimensional equicorrelated multivariate normal distribution Z with zero mean vector, correlations ρ and unit standard deviations. It is well-known that $\rho \in \left(-\frac{1}{d-1}, 1\right)$ is equivalent to a positive definite covariance matrix $\mathbf{\Sigma}$. Using some chosen pair $(\mu, \mathbf{\Omega})$, we estimate the probability $\mathbb{P}(a\mathbf{1} \leq Z \leq \infty\mathbf{1})$ for given $a \geq 0$ and integer $d \geq 2$.

Using the results in section 6.5.1 and section 6.5.2, we derive expressions for the covariance matrix $\mathbf{\Omega} = \omega\mathbf{I}$, the location vector μ, and the corresponding vectors $\tilde{\mathbf{a}}$, ν and $\tilde{\mu}$ and parameters $\underline{\lambda}$ and $\bar{\lambda}$. We compare our results with the implicit choices in Moran (1983) and Hashorva and Hüsler (2003).

Indeed, fast simulation procedures are available for $\mathbb{P}(a\mathbf{1} \leq Z \leq \infty)$ of an equicorrelated normal distribution Z because of the simple representation $Z_i = \sqrt{\rho}Z_0 + \sqrt{1-\rho}\varepsilon_i$ where $i = 1, \ldots, d$, $Z_0 \sim N(0,1)$ and independent $\varepsilon_i \sim N(0,1)$. Nevertheless, we can derive interesting analytical expressions for such distributions. For instance, we obtain convergence criteria in terms of d and ρ for different $\mathbf{\Omega}$. We stress that our method is wider applicable, i.e., it is not restricted to the equicorrelated case.

First, we derive some general expressions for equicorrelated normal distributions. Subsequently, we invoke Theorem 1 for convergence criteria and bounds on the parameters under different setups.

Equicorrelated distributions

Define $\mathbf{M}$ as a $d \times d$ matrix with each diagonal element equal to m_{dg}, and each off-diagonal element equal to m_{off}. Since

$$\lambda(\mathbf{M}) = \lambda\left(m_{\mathrm{off}}\mathbf{1}\mathbf{1}' + (m_{\mathrm{dg}} - m_{\mathrm{off}})\,\mathbf{I}\right) = \lambda\left(m_{\mathrm{off}}\mathbf{1}\mathbf{1}'\right) + m_{\mathrm{dg}} - m_{\mathrm{off}}$$

we find two distinct eigenvalues

$$\lambda_{\mathbf{1}}(\mathbf{M}) = m_{\mathrm{dg}} + (d-1)m_{\mathrm{off}} \qquad \lambda_{\mathbf{1}\perp}(\mathbf{M}) = m_{\mathrm{dg}} - m_{\mathrm{off}} \tag{6.41}$$

with multiplicity 1 and $d-1$, respectively. The corresponding eigenspaces are $\mathbf{1}$ and the space orthogonal to $\mathbf{1}$. It can be verified that the condition number of $\mathbf{M}$ is

$$\left|\frac{\lambda_{\max}(\mathbf{M})}{\lambda_{\min}(\mathbf{M})}\right| = 1 + d\left|\frac{m_{\mathrm{off}}}{m_{\mathrm{dg}} - m_{\mathrm{off}}}\right| \tag{6.42}$$

Further, $\mathbf{M}^{-1}$ has the following diagonal elements $m_{\mathrm{dg}}^{\mathrm{inv}}$ and off-diagonal elements $m_{\mathrm{off}}^{\mathrm{inv}}$:

$$m_{\mathrm{dg}}^{\mathrm{inv}} = \frac{m_{\mathrm{dg}} + m_{\mathrm{off}}(d-2)}{(m_{\mathrm{dg}} - m_{\mathrm{off}})(m_{\mathrm{dg}} + m_{\mathrm{off}}(d-1))} = \frac{(d-1)\lambda_{\mathbf{1}}(\mathbf{M}) + \lambda_{\mathbf{1}\perp}(\mathbf{M})}{d\lambda_{\mathbf{1}}(\mathbf{M})\lambda_{\mathbf{1}\perp}(\mathbf{M})} \tag{6.43}$$

$$m_{\mathrm{off}}^{\mathrm{inv}} = -\frac{m_{\mathrm{off}}}{(m_{\mathrm{dg}} - m_{\mathrm{off}})(m_{\mathrm{dg}} + m_{\mathrm{off}}(d-1))} = \frac{\lambda_{\mathbf{1}\perp}(\mathbf{M}) - \lambda_{\mathbf{1}}(\mathbf{M})}{d\lambda_{\mathbf{1}}(\mathbf{M})\lambda_{\mathbf{1}\perp}(\mathbf{M})} \tag{6.44}$$

Consider $\mathbf{M} = \mathbf{\Sigma}$. Substituting $m_{\mathrm{dg}} = 1$ and $m_{\mathrm{off}} = \rho$ in (6.43) and (6.44) implies that the inverse covariance matrix $\mathbf{\Sigma}^{-1}$ has positive diagonal elements s_{dg}, and off-diagonal elements s_{off}:

$$s_{\mathrm{dg}} = \frac{1 + (d-2)\rho}{(1-\rho)\,(1 + (d-1)\rho)} \qquad s_{\mathrm{off}} = -\frac{\rho}{(1-\rho)\,(1 + (d-1)\rho)} \tag{6.45}$$

Notice that s_{off} and $\rho \in \left(-\frac{1}{d-1}, 1\right)$ have opposite signs. Equation (6.41) gives the two distinct positive eigenvalues of $\mathbf{\Sigma}^{-1}$

$$\lambda_{\mathbf{1}}(\mathbf{\Sigma}^{-1}) = s_{\text{dg}} + (d-1)s_{\text{off}} = \frac{1}{1+(d-1)\rho} \tag{6.46}$$

$$\lambda_{\mathbf{1}\perp}(\mathbf{\Sigma}^{-1}) = s_{\text{dg}} - s_{\text{off}} = \frac{1}{1-\rho} \tag{6.47}$$

Each considered $\mathbf{\Omega}$ is of the type $\mathbf{\Omega} = \omega\mathbf{I}$, which means

$$\mathbf{B} = \mathbf{I} - \omega\mathbf{\Sigma}^{-1} \qquad \nu = \left(\mathbf{I} - \mathbf{\Omega}\mathbf{\Sigma}^{-1}\right)^{-1}\mu = \mathbf{B}^{-1}\mu \tag{6.48}$$

$$\tilde{\mathbf{a}} = \frac{1}{\sqrt{\omega}}\left(\mathbf{a} - \mathbf{B}^{-1}\mu\right) \qquad \tilde{\mu} = \frac{1}{\sqrt{\omega}}(\mu - \nu) = \frac{1}{\sqrt{\omega}}\left(\mathbf{I} - \mathbf{B}^{-1}\right)\mu \tag{6.49}$$

We may find from (6.46), (6.47), and $\lambda(\mathbf{B}) = 1 - \omega\lambda(\mathbf{\Sigma}^{-1})$:

$$\lambda_{\mathbf{1}}(\mathbf{B}) = 1 - \frac{\omega}{1+(d-1)\rho} \qquad \lambda_{\mathbf{1}\perp}(\mathbf{B}) = 1 - \frac{\omega}{1-\rho} \tag{6.50}$$

Applying (6.43), (6.44), and (6.50) on $\mathbf{B}$ gives for $\mathbf{B}^{-1}$ the diagonal elements $b_{\text{dg}}^{\text{inv}}$ and off-diagonal elements $b_{\text{off}}^{\text{inv}}$:

$$b_{\text{dg}}^{\text{inv}} = \frac{(d-1)\lambda_{\mathbf{1}}(\mathbf{B}) + \lambda_{\mathbf{1}\perp}(\mathbf{B})}{d\lambda_{\mathbf{1}}(\mathbf{B})\lambda_{\mathbf{1}\perp}(\mathbf{B})} = \frac{(1+(d-1)\rho)(1-\rho) - \omega}{(1+(d-1)\rho-\omega)(1-\rho-\omega)}$$

$$b_{\text{off}}^{\text{inv}} = \frac{\lambda_{\mathbf{1}\perp}(\mathbf{B}) - \lambda_{\mathbf{1}}(\mathbf{B})}{d\lambda_{\mathbf{1}}(\mathbf{B})\lambda_{\mathbf{1}\perp}(\mathbf{B})} = -\frac{\rho\omega}{(1+(d-1)\rho-\omega)(1-\rho-\omega)}$$

The probability measure of Moran (1983)

Let the superscript (M) refer to variables of the measure of Moran (1983). Recall that Moran (1983) uses $\mu^{(M)} = \mathbf{0}$ and for $\mathbf{\Omega}^{(M)} = \left(\mathbf{D}^{(M)}\right)^2$ a diagonal matrix with $\mathbf{\Omega}_{ii}^{(M)} = 1/(\mathbf{\Sigma}^{-1})_{ii}$. Substituting $\omega = 1/s_{\text{dg}}$ in the expressions of the equicorrelated case leads to

$$\mathbf{D}^{(M)} = \frac{1}{\sqrt{s_{\text{dg}}}}\mathbf{I} \qquad \mathbf{B}_{ij}^{(M)} = \frac{\rho 1_{i\neq j}}{1+(d-2)\rho}$$

$$\nu^{(M)} = \tilde{\mu}^{(M)} = \mathbf{0} \qquad \tilde{\mathbf{a}}^{(M)} = a\sqrt{s_{\text{dg}}}\mathbf{1}$$

$$\lambda_{\mathbf{1}}(\mathbf{B}^{(M)}) = -\frac{(d-1)s_{\text{off}}}{s_{\text{dg}}} = \frac{(d-1)\rho}{1+(d-2)\rho} \quad \lambda_{\mathbf{1}\perp}(\mathbf{B}^{(M)}) = \frac{s_{\text{off}}}{s_{\text{dg}}} = \frac{-\rho}{1+(d-2)\rho}$$

Under the measure of Moran (1983), each boundary in $a\mathbf{1}$ is thus scaled by $\sqrt{s_{\text{dg}}}$ to correct for the standardization of $\mathbf{\Sigma}^{-1}$. The benefit of this standardization is a zero diagonal of $\mathbf{B}^{(M)}$.

While $\lambda_{\mathbf{1}\perp}(\mathbf{B}^{(M)}) > -1$ holds for any $\rho \in \left(-\frac{1}{d-1}, 1\right)$,[1] we have $\lambda_{\mathbf{1}}(\mathbf{B}^{(M)}) \leq -1$ if $\rho \leq 1/(3-2d)$. For such ρ, the matrix $\mathbf{\Omega}^{(M)}$ leads to a divergent series expansion by the sufficient divergence condition in Theorem 1. In addition, the error bounds we derived in Theorem 8 are unavailable when $\mu^{(M)} = \mathbf{0}$ and $a > 0$, because $\tilde{a}_1^{(M)} > 0$ violates the assumption of nonpositive lower bounds.[2] In summary, the pair $(\mu^{(M)}, \mathbf{\Omega}^{(M)})$ implicitly used in Moran (1983) is inappropriate if $\rho \leq 1/(3-2d)$ or $a > 0$ holds.

The probability measure of Hashorva and Hüsler (2003)

With a completely different approach, Hashorva and Hüsler (2003) derive bounds for $\mathbb{P}(Z \geq \mathbf{a})$ from Jensen's inequality, and the smallest and largest eigenvalue of $\mathbf{\Sigma}$. Similar to the method of Moran (1983), the implicit choice $\mathbf{\Omega}^{(H)} = \mathbf{\Omega}^{(M)}$ implies that the convergence condition is that both $\rho > 1/(3-2d)$ and $a \leq 0$ hold. Further, $\mathbf{D}^{(H)} = \mathbf{D}^{(M)}$ and $\mathbf{B}^{(H)} = \mathbf{B}^{(M)}$, $\lambda_{\mathbf{1}}(\mathbf{B}^{(H)}) = \lambda_{\mathbf{1}}(\mathbf{B}^{(M)})$, and $\lambda_{\mathbf{1}\perp}(\mathbf{B}^{(H)}) = \lambda_{\mathbf{1}\perp}(\mathbf{B}^{(M)})$. The vector $\nu^{(H)}$ is determined by the location where the probability mass is maximal. Since $\mathbf{\Sigma^{-1}}$ is symmetric and positive definite:

$$\nu^{(H)} = \underset{\mathbf{x} \geq \mathbf{a}}{\operatorname{argmin}}\, \mathbf{x}'\mathbf{\Sigma}^{-1}\mathbf{x} = \left(\underset{x \geq a}{\operatorname{argmin}}\, x^2 \mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}\right)\mathbf{1} = \max(a, 0)\mathbf{1}$$

The case $a \leq 0$ coincides with $\mu^{(M)} = \mathbf{0}$ in Moran (1983): $\mu^{(H)} = \nu^{(H)} = \mathbf{0}$. Now, we consider $a > 0$ for which $\nu^{(H)} = \hat{\mathbf{x}} = \mathbf{a}$. Using $\mu = \mathbf{B}\nu$ from (6.48) and $\lambda_{\mathbf{1}}(\mathbf{B}^{(M)}) < 1$,

$$\mu^{(H)} = \mathbf{B}^{(M)}\nu^{(H)} = \mathbf{B}^{(M)}\mathbf{a} = a\lambda_{\mathbf{1}}(\mathbf{B}^{(M)})\mathbf{1} < a\mathbf{1}$$

Hence,

$$\mu^{(H)} = \frac{(d-1)\rho a}{1+(d-2)\rho}\mathbf{1} \qquad \nu^{(H)} = a\mathbf{1}$$

$$\tilde{\mu}^{(H)} = \frac{a}{\sqrt{s_{\text{dg}}}}\frac{\rho-1}{1+(d-2)\rho}\mathbf{1} < \mathbf{0} \qquad \tilde{\mathbf{a}}^{(H)} = \frac{1}{\sqrt{s_{\text{dg}}}}\left(\mathbf{a} - \nu^{(H)}\right) = \mathbf{0}$$

For the two dimensional case in Figure 6.1 on p.130, we now have $\tilde{\mathbf{a}}^{(H)} = \tilde{\nu}^{(H)} = \mathbf{0}$ and $\tilde{\mathbf{b}}^{(H)} = \infty\mathbf{1}$. As $\tilde{\mu}^{(H)} < \tilde{\mathbf{a}}^{(H)}$, the right panel in Figure 6.1

[1] Recall that a positive definite equicorrelated $\mathbf{\Sigma}$ corresponds to $\rho \in \left(-\frac{1}{d-1}, 1\right)$.

[2] Because $Y := \lambda X_{0,\infty}$ with $X_{0,\infty}$ as in (6.32) is dominated by the conditional distribution $Y_a := Y|_{Y \geq a}$ with $a > 0$, it follows that the scaled remainder of Y cannot be an upper bound for the scaled remainder of Y_a.

indicates that we can restrict the analysis to directions $\mathbf{s} \geq 0$ starting from $\tilde{\mu}^{(H)}$. Using $\mathbf{a} = a\mathbf{1}$, $\mathbf{b} = \infty\mathbf{1}$, $\lambda(\mathbf{s}) = \mathbf{s}'\mathbf{B}\mathbf{s}$, and $\text{sign}(\mathbf{B}_{ij}^{(M)}) = \text{sign}(\rho 1_{i\neq j})$, we have the following sharp bounds on $\lambda(\mathbf{s})$

$$\underline{\lambda}^{(H)} = \min_{\|\mathbf{s}\|=1} \{\lambda(\mathbf{s}) : \lambda(\mathbf{s}) \leq 0, \mathbf{s} \geq 0\} = \lambda_{\min}(\mathbf{B}^{(M)}) 1_{\rho<0} = \lambda_{\mathbf{1}}(\mathbf{B}^{(M)}) 1_{\rho<0}$$
$$\bar{\lambda}^{(H)} = \max_{\|\mathbf{s}\|=1} \{\lambda(\mathbf{s}) : \lambda(\mathbf{s}) \geq 0, \mathbf{s} \geq 0\} = \lambda_{\max}(\mathbf{B}^{(M)}) 1_{\rho>0} = \lambda_{\mathbf{1}}(\mathbf{B}^{(M)}) 1_{\rho>0}$$

where we use $\lambda(\mathbf{e}_i) = 0$ for any vector $\mathbf{e}_i$ with a 1 at the ith entry and zeros elsewhere, since $\mathbf{B}^{(M)}$ has zero diagonal.

Each element in $U = \tilde{Z} 1_{\tilde{\mathbf{a}}^{(H)} \leq \tilde{Z} \leq \tilde{\mathbf{b}}^{(H)}} = \tilde{Z} 1_{\tilde{Z} \geq \mathbf{0}}$ with $\tilde{Z} \sim N(\tilde{\mu}^{(H)}, \mathbf{I})$ follows an independent standard normal distribution, first increased by $\tilde{\mu}_i^{(H)} < 0$ and then all values below $\tilde{a}_i^{(H)} = 0$ are set at zero. This implies that $X = \frac{1}{2} U'\mathbf{B}U = \frac{1}{2} d(d-1) \mathbf{B}_{12}^{(M)} \sum_{i<j} U_i U_j$ is either nonpositive ($\mathbf{B}_{12}^{(H)} \leq 0$ if $\rho \leq 0$) or nonnegative ($\mathbf{B}_{12}^{(H)} \geq 0$ if $\rho \geq 0$). The latter case with $\lambda(\mathbf{s}) > 0$ if $\mathbf{s} > 0$ is unfavorable from a computational view as this case leads to non-alternating summants in (6.15) and so larger remainders that are in general harder to bound.

Notice that the bound with $\bar{u}$ in (6.37) of Theorem 8 is redundant here because these bounds correspond to bounded directions $\mathbf{s}$ from $\tilde{\mu}^{(H)}$, whilst $\tilde{\mu}^{(H)} < \tilde{\mathbf{a}}^{(H)} < \tilde{\mathbf{b}}^{(H)} = \infty\mathbf{1}$ excludes such bounded directions.[3]

A new probability measure

Consider $\mathbf{\Omega} = \omega\mathbf{I}$ with ω from (6.39), (6.46), and (6.47) to ensure a minimal absolute eigenvalue:

$$\omega = \frac{2}{(d-2)s_{\text{off}} + 2s_{\text{dg}}} = \frac{2(1-\rho)(1+(d-1)\rho)}{2+(d-2)\rho} \tag{6.51}$$

By construction, both eigenvalues of $\mathbf{B}$ are equal in absolute value:

$$\lambda_{\mathbf{1}}(\mathbf{B}) = 1 - \omega\lambda_{\mathbf{1}}(\mathbf{\Sigma}^{-1}) = 1 - \frac{2(1-\rho)}{2+(d-2)\rho} = \frac{d\rho}{2+(d-2)\rho}$$
$$\lambda_{\mathbf{1}\perp}(\mathbf{B}) = 1 - \omega\lambda_{\mathbf{1}\perp}(\mathbf{\Sigma}^{-1}) = 1 - \frac{2(1+(d-1)\rho)}{2+(d-2)\rho} = -\frac{d\rho}{2+(d-2)\rho}$$

[3]For the intuition, imagine again vectors satisfying $\tilde{\mu}^{(H)} < \tilde{\mathbf{a}}^{(H)} < \tilde{\mathbf{b}}^{(H)} = \infty\mathbf{1}$ for the two-dimensional case in Figure 6.1 on p.130.

It is easy to see that $\lambda_{\mathbf{1}}(\mathbf{B}), \lambda_{\mathbf{1}\perp}(\mathbf{B}) > -1$ if $0 \leq \rho < 1$, which ensures convergence of the series in (6.14) for $\rho \geq 0$ (see Theorem 1). If $\rho < 0$, our algorithm converges if $d\rho/(2+(d-2)\rho) > -1$, or equivalently $\rho > -1/(d-1)$. This coincides with the condition for a positive definite equicorrelated $\mathbf{\Sigma}$. Hence, our method converges for any equicorrelated matrix when using $\mathbf{B}$. This benefit has the cost that $\mathbf{B}$ has a nonzero diagonal, which slows down computations of the summation in (6.15).

Concerning μ, the condition $\mathbb{E}\left[\tilde{Z}_i 1_{\tilde{a}_i \leq \tilde{Z}_i}\right] = 0$ with $\tilde{Z}_i \sim N(\tilde{\mu}_i, 1)$ implies that we need $\tilde{a}_i, \tilde{\mu}_i < 0$ for $i = 1, \ldots, d$. Since $\tilde{a}_i < 0$, $\tilde{b}_i = \infty$, and $\lambda(\mathbf{s}) = \mathbf{s}'\mathbf{B}\mathbf{s}$,

$$\underline{\lambda} = \min_{\|\mathbf{s}\|=1} \{\lambda(\mathbf{s}) : \lambda(\mathbf{s}) \leq 0\} = \lambda_{\min}(\mathbf{B}) = -\left|\frac{\rho d}{2 + (d-2)\rho}\right|$$

$$\bar{\lambda} = \max_{\|\mathbf{s}\|=1} \{\lambda(\mathbf{s}) : \lambda(\mathbf{s}) \geq 0\} = \lambda_{\max}(\mathbf{B}) = \left|\frac{\rho d}{2 + (d-2)\rho}\right|$$

By $\tilde{\mathbf{a}}, \tilde{\mu} < 0$ and symmetry of the variables, we find for $\bar{u}$ in Theorem 8(iii):[4]

$$\bar{u} = \sup_{\tilde{\mu}'\mathbf{x} > 0, \mathbf{x} \geq \tilde{\mathbf{a}}} \|\mathbf{x}\| = \|\tilde{\mathbf{a}}\|$$

The following lemma provides our μ that ensures $\mathbb{E}[U] = \mathbf{0}$, and indicates whether $\mu_1 > a$ holds.

Lemma 10. *The location vector μ equals $\mu_1 \mathbf{1}$ with $\mu_1 = a - \sqrt{\omega} y$, ω from* (6.51)*, and y the unique solution of*

$$-\frac{a}{\sqrt{\omega}} + y = \frac{\rho d}{2(\rho - 1)} \frac{\phi(y)}{1 - \Phi(y)} \tag{6.52}$$

A vector μ interior in the integration region T ($\mu_1 > a$) is equivalent to

$$a < \rho d \sqrt{\frac{1 + (d-1)\rho}{\pi (1-\rho)(2 + (d-2)\rho)}}$$

6.6 Numerical examples

This section first discusses numerical examples with equicorrelated distributions. Then, we consider random draws from a larger class of correlation matrices.

[4] Figure 6.1 is illustrative here as $\bar{u}$ is the maximal distance from the origin to the boundary of $\tilde{T}$ among all vectors $\mathbf{x}$ that satisfy $\tilde{\mu}'\mathbf{x} > 0$.

Consider the d-dimensional equicorrelated multivariate normal distribution $Z \sim N(\mathbf{0}, \mathbf{\Sigma})$ with correlation parameter $\rho \in \left(-\frac{1}{d-1}, 1\right)$ and unit standard deviations. We are interested in estimating the probability $p := \mathbb{P}(Z_1 \geq a, \ldots, Z_d \geq a)$ for given $a \geq 0$ and $d = 2, 3, \ldots$.

We obtain the location vector μ^* from Lemma 10, and the covariance matrix $\mathbf{\Omega}^* = \omega \mathbf{I}$ from ω in (6.51). This pair results in a matrix $\mathbf{B}$ with the smallest dominant eigenvalue, convergence of our method, and $\mathbb{E}[U] = \mathbf{0}$. Table 6.1 compares the performance of four different setups: *(i)* $(\mu^{(H)}, \mathbf{\Omega}^{(H)})$, *(ii)* $(\mu^*, \mathbf{\Omega}^{(H)})$, *(iii)* $(\mu^*, \mathbf{\Omega}^*)$, and *(iv)* the simulation method of Genz (1992).

For the setups *(i)*–*(iii)*, we compute the probability p as the average of the lower bound and the upper bound. This minimizes the absolute error bound ε at half the difference between the two bounds. The method *(iv)* of Genz (1992) only provides an error estimate $\hat{\varepsilon}$, not an error bound ε. Nonetheless, the error estimates of Genz (1992) are more conservative than our error bounds. The error bounds ε of the three setups *(i)*–*(iii)* are on average comparable for $n = 4$ in each of the four panels in Table 6.1, while the bounds with $n = 8$ terms are at least as accurate as the method of Genz (1992).

Due to space limitations, we do not report the bounds of Hashorva and Hüsler (2003) in Table 6.1. Although these bounds require a computation time of no more than 2.0 ms, the bounds are not competitive with our bounds. More specifically, the method of Hashorva and Hüsler (2003) gives error bounds of 0.0702 and 0.1813 for the case in Panel A. For the case in Panel B, the interval $[1.77 \cdot 10^{-3}, 3.12 \cdot 10^{-3}]$ using the method of Hashorva and Hüsler (2003) is also inaccurate compared to our method. This also holds for the cases in Panels C and D, where the intervals of Hashorva and Hüsler (2003) are $[5.28 \cdot 10^{-2}, 0.101]$ and $[1.01 \cdot 10^{-4}, 1.29 \cdot 10^{-4}]$, respectively.

The columns labeled r report the ratio of the best error estimate and the error bound: $r = (p_n - p_{14}) / \varepsilon_n$. A small $|r|$ indicates that the upper and lower bound of p are equally accurate. As a consequence, the true error is very small while the error bound is very conservative. A value of r close to one means that the estimate for p is too large, and the lower bound is more accurate. The upper bound is more accurate if r is close to minus one.

Our values of $|r|$ are not very close to one, thereby indicating that the accuracy of the lower bound and upper bound are of the same order of magnitude. Still, r tends to be positive and most estimates for p monotonically decrease in n. This indicates that the subsequent downward revisions in the upper bound exceed the upward revisions in the lower bound. In other words, the lower bound tends to be more accurate than the upper bound.

Table 6.1: **Effect of n, μ and $\mathbf{\Omega}$ for equicorrelated distributions.**

Panel A: $d=5$, $\rho=0.2$, $a=0.1$, $b=\infty$. Genz (1992): $p=0.1002862$, $\hat{\varepsilon}=9.371\cdot10^{-5}$, $\hat{r}=-0.47$, time(s) = 0.02.

	$(\mu^{(H)},\mathbf{\Omega}^{(H)})$				$(\mu^{*},\mathbf{\Omega}^{(H)})$				$(\mu^{*},\mathbf{\Omega}^{*})$			
n	p	ε	r	time(s)	p	ε	r	time(s)	p	ε	r	time(s)
4	0.1009143	$2.612\cdot10^{-3}$	0.22	0.02	0.1004242	$1.319\cdot10^{-4}$	0.72	0.04	0.1004678	$2.017\cdot10^{-4}$	0.68	0.02
6	0.1004262	$5.152\cdot10^{-4}$	0.19	0.01	0.1003344	$7.903\cdot10^{-6}$	0.58	0.03	0.1003340	$8.079\cdot10^{-6}$	0.53	0.02
8	0.1003459	$1.043\cdot10^{-4}$	0.15	0.01	0.1003300	$5.914\cdot10^{-7}$	0.43	0.04	0.1003300	$4.371\cdot10^{-7}$	0.51	0.12
10	0.1003325	$2.162\cdot10^{-5}$	0.12	0.03	0.1003298	$7.255\cdot10^{-8}$	0.42	0.05	0.1003298	$2.553\cdot10^{-8}$	0.54	0.80
12	0.1003302	$4.582\cdot10^{-6}$	0.08	0.11	0.1003298	$9.717\cdot10^{-9}$	0.36	0.14	0.1003298	$1.543\cdot10^{-9}$	0.58	4.03
14	0.1003298	$9.891\cdot10^{-7}$		0.35	0.1003298	$1.382\cdot10^{-9}$		0.38	0.1003298	$9.972\cdot10^{-11}$		17.96

Panel B: $d=5$, $\rho=0.2$, $a=1$, $b=\infty$. Genz (1992): $p=1.9380104\cdot10^{-3}$, $\hat{\varepsilon}(\cdot10^{-3})=5.942\cdot10^{-3}$, $\hat{r}=-0.26$, time(s) = 0.02

n	$p\ (\cdot10^{-3})$	$\varepsilon\ (\cdot10^{-3})$	r	time(s)	$p\ (\cdot10^{-3})$	$\varepsilon\ (\cdot10^{-3})$	r	time(s)	$p\ (\cdot10^{-3})$	$\varepsilon\ (\cdot10^{-3})$	r	time(s)
4	1.9520833	$1.780\cdot10^{-2}$	0.70	0.01	1.9397726	$3.320\cdot10^{-4}$	0.63	0.02	1.9399039	$7.880\cdot10^{-4}$	0.43	0.01
6	1.9408602	$1.872\cdot10^{-3}$	0.69	0.01	1.9395731	$1.664\cdot10^{-5}$	0.61	0.02	1.9395709	$2.378\cdot10^{-5}$	0.33	0.02
8	1.9397108	$2.157\cdot10^{-4}$	0.68	0.02	1.9395637	$1.157\cdot10^{-6}$	0.63	0.03	1.9395632	$8.563\cdot10^{-7}$	0.32	0.12
10	1.9395811	$2.674\cdot10^{-5}$	0.67	0.03	1.9395630	$9.599\cdot10^{-8}$	0.62	0.05	1.9395630	$3.397\cdot10^{-8}$	0.37	0.76
12	1.9395653	$3.521\cdot10^{-6}$	0.58	0.12	1.9395630	$8.914\cdot10^{-9}$	0.56	0.13	1.9395629	$1.363\cdot10^{-9}$	0.51	3.97
14	1.9395633	$4.874\cdot10^{-7}$		0.35	1.9395629	$9.005\cdot10^{-10}$		0.38	1.9395629	$6.524\cdot10^{-11}$		18.52

Panel C: $d=4$, $\rho=-0.1$, $a=0.1$, $b=\infty$. Genz (1992): $p=0.057753885$, $\hat{\varepsilon}=6.235\cdot10^{-5}$, $\hat{r}=0.12$, time(s) = 0.02

n	p	ε	r	time(s)	p	ε	r	time(s)	p	ε	r	time(s)
8	0.057755788	$1.081\cdot10^{-5}$	0.87	0.01	0.057746368	$1.237\cdot10^{-9}$	0.58	0.03	0.057746367	$1.231\cdot10^{-10}$	0.57	0.02
10	0.057747757	$1.548\cdot10^{-6}$	0.90	0.01	0.057746367	$7.107\cdot10^{-11}$	0.58	0.03	0.057746367	$1.752\cdot10^{-12}$	0.63	0.06
14	0.057746397	$3.224\cdot10^{-8}$	0.91	0.01	0.057746367	$2.955\cdot10^{-13}$	0.61	0.04	0.057746367	$4.454\cdot10^{-16}$	0.79	0.41
18	0.057746367	$6.803\cdot10^{-10}$		0.02	0.057746367	$1.467\cdot10^{-15}$		0.05	0.057746367	$1.513\cdot10^{-19}$		7.49

Panel D: $d=4$, $\rho=-0.1$, $a=1$, $b=\infty$. Genz (1992): $p=1.019637138\cdot10^{-4}$, $\hat{\varepsilon}(10^{-4})=1.308\cdot10^{-3}$, $\hat{r}=0.37$, time(s) = 0.02

n	$p\ (\cdot10^{-4})$	$\varepsilon\ (\cdot10^{-4})$	r	time(s)	$p\ (\cdot10^{-4})$	$\varepsilon\ (\cdot10^{-4})$	r	time(s)	$p\ (\cdot10^{-4})$	$\varepsilon\ (\cdot10^{-4})$	r	time(s)
8	1.019156347	$8.435\cdot10^{-8}$	0.09	0.01	1.019156339	$2.364\cdot10^{-10}$	-0.28	0.02	1.019156340	$3.164\cdot10^{-11}$	0.72	0.02
10	1.019156340	$3.153\cdot10^{-9}$	0.16	0.01	1.019156339	$5.284\cdot10^{-12}$	-0.25	0.02	1.019156339	$2.095\cdot10^{-13}$	0.73	0.04
14	1.019156339	$6.006\cdot10^{-12}$	0.26	0.01	1.019156339	$3.992\cdot10^{-15}$	-0.17	0.03	1.019156339	$1.236\cdot10^{-17}$	0.00	0.40
18	1.019156339	$1.544\cdot10^{-14}$		0.02	1.019156339	$4.181\cdot10^{-18}$		0.04	1.019156339	$9.542\cdot10^{-22}$		2.51

We may exploit this for an improvement of our estimates for p.

The columns labeled time(s) in Table 6.1 report the computation time in seconds.[5] Differences in computation time are small for low n. The computation time increases when the number of terms n increases, particularly with the covariance matrix $\mathbf{\Omega}^*$ since the corresponding $\mathbf{B}^*$ has a nonzero diagonal. Nonetheless, for $n = 8$ our method is still comparable with the method of Genz (1992) in terms of computation time. This is remarkable because in contrast to Genz (1992) we do not use a compiled version of our code. Another potential improvement is the adoption of a parallel computing package. For the equicorrelated case, additional computational benefits are possible in (6.15) where each U_i follows the same distribution, and $\bar{B}_{i,j}$ only depends on whether $\bar{b}_{i,j}$ is a diagonal element of $\bar{\mathbf{B}}$ (see (6.45) and (6.48)).

Consider the results for $d = 5$ and $\rho = 0.2$ with $a = -0.1$ and $a = 1$ in Panel A and B, respectively. It follows from Section 6.5 that our method converges for each of the three considered setups. Among these three setups, the convergence speed per term of the expansion is highest with the pair $(\mu^*, \mathbf{\Omega}^*)$ by the following two reasons. First, the vector μ^* ensures that $\mathbb{E}[U] \approx \mathbf{0}$, which reduces $\mathbb{E}\big[(U\mathbf{B}U)^k\big]$ and thereby the truncation error. Second, the matrix $\mathbf{\Omega}^*$ corresponds to a matrix $\mathbf{B}^*$ with smaller absolute eigenvalues. Namely, $\underline{\lambda}^* = \bar{\lambda}^* = 1/2.6 \approx 0.385$ compared to $\underline{\lambda}^{(H)} = 0$ and $\bar{\lambda}^{(H)} = 0.5$ for $\mathbf{\Omega}^{(H)}$. However, the matrix $\mathbf{B}^*$ has nonzeros on the diagonal, which slows the computation speed severely down. The results suggest that $(\mu^*, \mathbf{\Omega}^{(H)})$ is the optimal setup for equicorrelated distributions in dealing with convergence speed and computation time per term.

A similar picture emerges from the results with $d = 4$ and $\rho = -0.1$ using either $a = -0.1$ (Panel C) or $a = 1$ (Panel D). The estimates are more accurate than in Panels A and B. This follows from the condition number (6.42) and the corresponding smaller eigenvalues of $\mathbf{B}$ (see p.141). Here, $-\underline{\lambda}^* = \bar{\lambda}^* = 2/9$ while $\underline{\lambda}^{(H)} = -0.375$ and $\bar{\lambda}^{(H)} = 0$.

Figure 6.2 shows that the relative error bound ε/p declines in the threshold value a for $n = 8$. That is, while a increases, the error bound decreases more than the corresponding probability decreases. This decline in the *relative* error contrasts with standard algorithms such as Genz (1992) that focus on a small *absolute* error. Hence, our method is particularly helpful in estimating tail probabilities. Another remarkable pattern is the almost linear decrease in the log error bound ε when the number of series terms n increases. This is illustrated in Figure 6.3 for the case $a = 1$. This feature of the error bound has the potential to enable us to preselect a specific n for

[5] The computations times are obtained with an i5-3230 CPU with 6 GB RAM.

Figure 6.2: **Effect of a change in** a

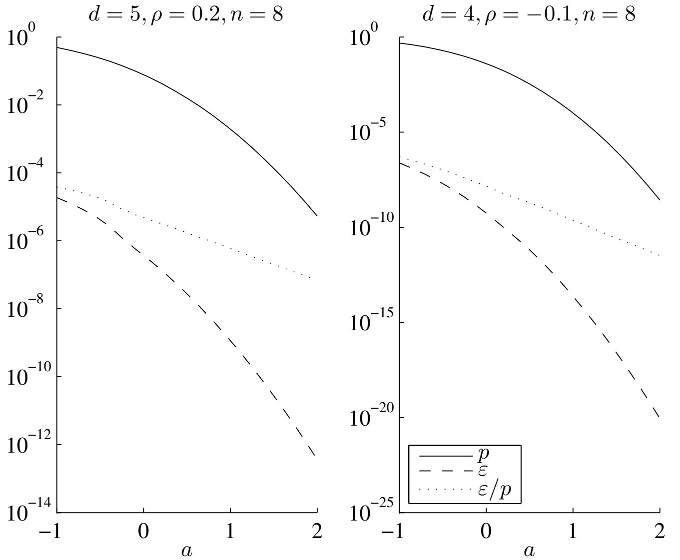
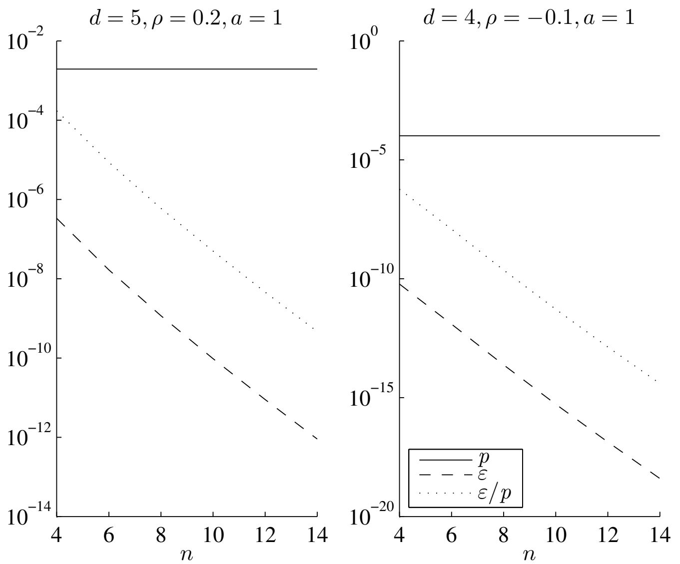


Figure 6.3: **Effect of a change in** n



a desired ε.

To study arbitrary non-equicorrelated covariance matrices $\mathbf{\Sigma}$, we generated 10,000 correlation matrices of dimension $d = 4$ by sampling d independent eigenvalues from a uniform distribution on $[0, 1]$, and sampling d independent eigenvectors from a uniform distribution on the d-dimensional unit sphere. Following Genz (1993), we generate the lower bound $\mathbf{a} = a\mathbf{1}$ from a uniform distribution on $[-\sqrt{d}, 0]$. We calculate p as the mean of the lower bound and the upper bound after $n = 8$ terms, and the error bound equals half the difference between the two bounds. We use μ^* and $\mathbf{\Omega}^*$ to ensure convergence.

The top left plot in Figure 6.4 contains the relative frequencies by error bound ε. It suggests that most, but far from all, Taylor series indicate convergence after a truncation of 8 terms. For instance, while half of the error bounds is below $9.0 \cdot 10^{-4}$, a substantial 7.1% of the computed probabilities exceeds 1 (top right plot). The middle plots indicate a similar widespread pattern for the relative error bound. Fortunately, this pattern is predictable to some extent. The left bottom plot in Figure 6.4 shows that a large error bound ε is associated with a high dominant eigenvalue $\max|\lambda|$. This is intuitive as a high $|\lambda|$ indicates a large remainder. Interestingly, Phinikettos and Gandy (2011) provide a simulation method that is relatively efficient for covariance matrices with *high* singular values which in general correspond to high condition numbers. In contrast, our method performs best with a small dominant eigenvalue, thus a small condition number.

The relation between the relative error bound ε/p and the dominant eigenvalue $|\lambda|$ is somewhat less clear (not shown). Indeed, the dominant eigenvalue mainly affects ε, not p. As a consequence, there is no clear mapping from the probability estimate p to the relative error bound ε/p, as indicated by the right bottom plot in Figure 6.4.

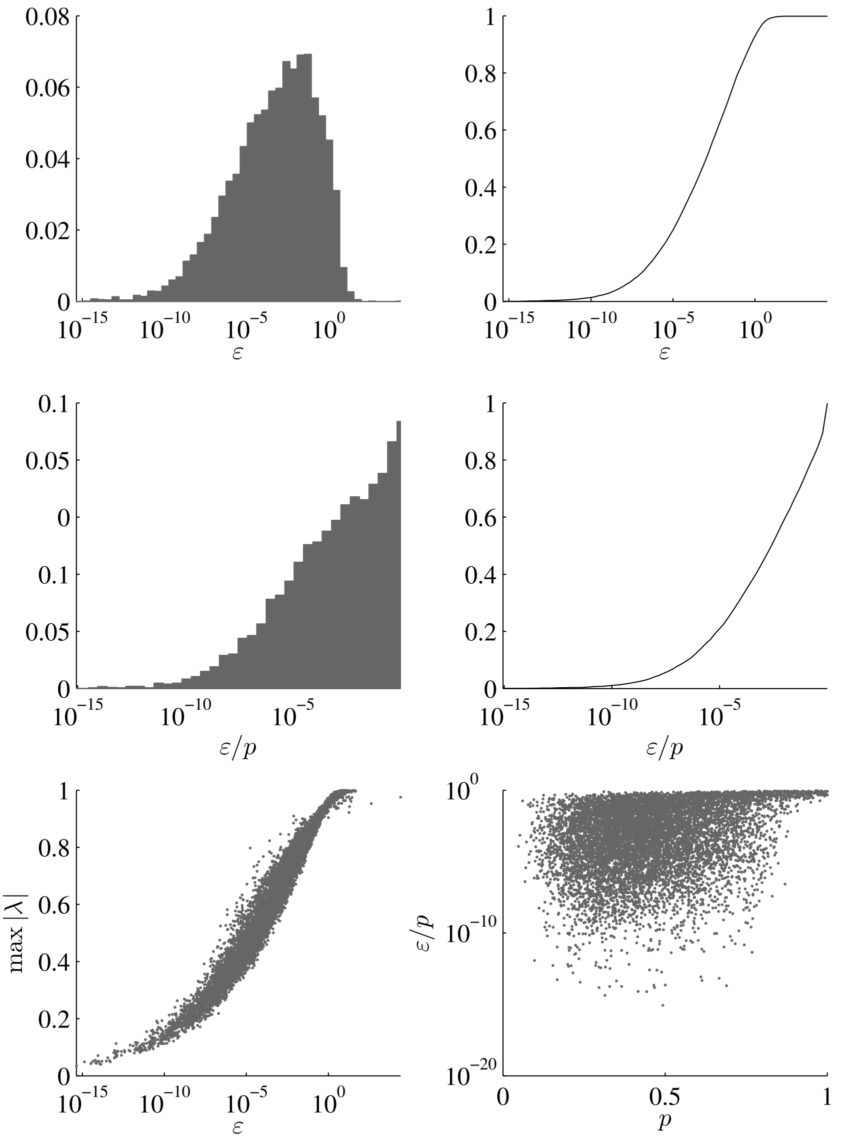
6.7 Conclusion

This paper develops an analytical method to bound the probability of a multivariate normal distribution on a rectangular region. The method combines a change of measure with a Taylor series expansion. The change of measure is similar to an importance sampling approach. We derived error bounds by exploiting that a multivariate normal distribution is an infinite mixture of random variables that are closely related to the gamma distribution. The bounds are tightest if the condition number of the covariance matrix is small.

Up to five dimensions, the bounds might be more precise than simula-

Figure 6.4: **Simulation results.**

10, 000 random covariance matrices with $d = 4$, and truncation after $n = 8$ terms. Relative frequency and cdf of error bound ε (top plots), relative frequency and cdf of relative error bound ε/p (middle), scatter plot of error bound and maximal absolute eigenvalue $\max|\lambda(\mathbf{B})|$ (bottom left), and scatter plot of relative error bound relative to estimated probability (bottom right).



tion methods. The intuition is that simulation methods tend to focus on a small absolute error. In addition, simulation methods only provide an error estimate. For a given distribution, our relative bound increases in the size of the integration region. This makes the method very well suited to estimate tail probabilities.

Additional improvements are possible with a compiled version of our code, a parallel computing package, or an even more accurate analysis of the error bounds. Another interesting question is how to deal with the exponentially increasing number of combinations when the dimension increases. Here, a numerical integration method to estimate the remainder term is potentially helpful. The numerical integration could be substituted by a faster simulation method, at the expense of an error estimate instead of an error bound.

Appendix

6.A Appendix

6.A.1 Proofs Section 6.3

We start with two lemmas that are helpful in proving Theorem 1.

Lemma 11. *Consider the noncentral χ^2-distribution X with $k \geq 2$ degrees of freedom and noncentrality parameter $\xi \geq 0$. The shape of the tail of the density function g of $Y = \lambda X/2$ ($\lambda > 0$) is independent of ξ and satisfies*

$$\lim_{x \to \infty} \frac{g(x)}{\exp(-x/\lambda_0)} = 0 \qquad \text{if and only if} \qquad \lambda > \lambda_0$$

Proof of Lemma 11

It is well known that the density function $f_{k,\xi}$ of X is a Poisson-weighted average of central χ^2-distributions:

$$f_{k,\xi}(x) = \sum_{i=0}^{\infty} \frac{e^{-\xi/2}(\xi/2)^i}{i!} f_{k+2i}(x)$$

where f_j is the density of the central χ^2-distribution with j degrees of freedom:

$$f_j(x) = \begin{cases} \frac{x^{(j/2)-1}}{2^{j/2}\Gamma(j/2)} e^{-x/2} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

An upper bound on the density of $Y = \lambda X/2$ at $Y = y$ follows from

$$\begin{aligned} g(y) &= f_{k,\xi}\left(\frac{2y}{\lambda}\right) \\ &= \frac{2}{\lambda} \sum_{i=0}^{\infty} \frac{e^{-\xi/2}(\xi/2)^i}{i!} \frac{1}{2^{k/2+i}\Gamma\left(k/2+i\right)} \left(\frac{2y}{\lambda}\right)^{\frac{k}{2}+i-1} \exp\left(-\frac{y}{\lambda}\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{2e^{-\xi/2}}{\lambda 2^{k/2}} \left(\frac{2y}{\lambda}\right)^{\frac{k}{2}-1} \exp\left(-\frac{y}{\lambda}\right) \sum_{i=0}^{\infty} \frac{(\xi/2)^i}{i!\Gamma\left(k/2+i\right)} \frac{1}{2^i} \left(\frac{2y}{\lambda}\right)^i \\
&\leq \frac{e^{-\xi/2}}{\lambda} \left(\frac{y}{\lambda}\right)^{\frac{k}{2}-1} \exp\left(-\frac{y}{\lambda}\right) \sum_{i=0}^{\infty} \frac{1}{(i!)^2} \left(\frac{\xi y}{2\lambda}\right)^i \\
&= \frac{e^{-\xi/2}}{\lambda} \left(\frac{y}{\lambda}\right)^{\frac{k}{2}-1} \exp\left(-\frac{y}{\lambda}\right) I_0\left(2\sqrt{\frac{\xi y}{2\lambda}}\right) \\
&\to \frac{e^{-\xi/2}}{\lambda} \left(\frac{y}{\lambda}\right)^{\frac{k}{2}-1} \exp\left(-\frac{y}{\lambda}\right) \frac{1}{\sqrt{\pi}} \left(\frac{2\lambda}{\xi y}\right)^{1/4} e^{\sqrt{\frac{2\xi y}{\lambda}}} \text{ as } y \to \infty \qquad (6.53)
\end{aligned}
$$

where I_0 is the modified Bessel function of the first kind with the well-known property $I_0(z) \to \sqrt{\frac{2}{\pi z}} e^z$ as $z \to \infty$. An asymptotic lower bound follows along similar lines from

$$
\begin{aligned}
g(y) &= \frac{2e^{-\xi/2}}{\lambda 2^{k/2}\Gamma\left(k/2\right)} \left(\frac{2y}{\lambda}\right)^{\frac{k}{2}-1} \exp\left(-\frac{y}{\lambda}\right) \sum_{i=0}^{\infty} \frac{(\xi/2)^i}{i!\Gamma\left(k/2+i\right)} \frac{1}{2^i} \left(\frac{2y}{\lambda}\right)^i \\
&\geq \frac{e^{-\xi/2}}{\lambda\Gamma\left(k/2\right)} \frac{\lambda}{y} \exp\left(-\frac{y}{\lambda}\right) \left(\frac{\xi}{2}\right)^{-k/2} \sum_{i=0}^{\infty} \frac{1}{((k/2+i)!)^2} \left(\frac{\xi y}{2\lambda}\right)^{k/2+i} \qquad (6.54)
\end{aligned}
$$

The summation term in (6.54) is of course strictly dominated by $\exp(\frac{\xi y}{2\lambda})$. Because $\exp(-y/\lambda)$ is the dominant term that drives both the asymptotic upper bound (6.53) and the lower bound (6.54) to zero,

$$
\lim_{y\to\infty} \frac{g(y)}{\exp(-y/\lambda_0)} = 0 \qquad \text{if and only if} \qquad \lambda > \lambda_0
$$

which does not depend on ξ. □

Lemma 12. *Consider a random variable X with a probability density function f that has a tail similar to a gamma distribution with parameters $d > 0$ and $\lambda > 0$:*

$$
f(x) \propto |x|^{d-1} \exp(-\lambda|x|) \qquad |x| \to \infty
$$

The series expansion $\sum_{k=0}^{\infty} \mathbb{E}\left[X^k\right]/k!$ of $\mathbb{E}[X]$ converges if and only if $\lambda > 1$.

Proof of Lemma 12
Let $m_k := \mathbb{E}\left[X^k\right]/k!$. For $k = 0, 1, 2, \ldots$

$$
m_k = \frac{1}{k!} \lim_{z\to\infty} \int_{x=-z}^{z} x^k f(x) \,\mathrm{d}x \qquad (6.55)
$$

$$= \lim_{z\to\infty} \frac{z^{k+1}}{(k+1)!} \left[(-1)^k \int_{x=-z}^{0} w_k(x) f(x)\,\mathrm{d}x + \int_{x=0}^{z} w_k(x) f(x)\,\mathrm{d}x \right] \tag{6.56}$$

where $w_k(x) := (k+1)(|x|/z)^k/z$ is the weight of the strictly positive $f(x)$ in the integral. The total weight of both integrals in (6.56) equals one:

$$\lim_{z\to\infty} \int_{x=-z}^{0} w_k(x)\,\mathrm{d}x = \lim_{z\to\infty} \int_{x=0}^{z} w_k(x)\,\mathrm{d}x = 1$$

For an arbitrary $\varepsilon \in (0,1]$ and k large, the relative weight of $w_k(x)$ in the subset $[0,(1-\varepsilon)z]$ of $[0,z]$ converges to zero if $z \to \infty$:

$$\lim_{k,z\to\infty} \frac{\int_{x=0}^{(1-\varepsilon)z} w_k(x)\mathrm{d}x}{\int_{x=0}^{z} w_k(x)\mathrm{d}x} = \lim_{k,z\to\infty} \frac{\int_{x=0}^{(1-\varepsilon)z} \frac{k+1}{z} \left(\frac{x}{z}\right)^k \mathrm{d}x}{\int_{x=0}^{z} \frac{k+1}{z} \left(\frac{x}{z}\right)^k \mathrm{d}x} = \lim_{k\to\infty} (1-\varepsilon)^{k+1} = 0$$

Since $\left|\int_{x=0}^{z} w_k(x)\mathrm{d}x\right| < \infty$, we have for all $y \in [0,\infty)$,

$$\lim_{k\to\infty} \int_{x=-y}^{0} w_k(x) f(x)\,\mathrm{d}x = \lim_{k\to\infty} \int_{x=0}^{y} w_k(x) f(x)\,\mathrm{d}x = 0.$$

We find for (6.56) with finite y

$$\lim_{k\to\infty} \frac{y^{k+1}}{(k+1)!} \left[(-1)^k \int_{x=-y}^{0} w_k(x) f(x)\,\mathrm{d}x + \int_{x=0}^{y} w_k(x) f(x)\,\mathrm{d}x \right] = 0.$$

Thus, $\lim_{k\to\infty} m_k$ in (6.55) is completely determined by $f(x)$ with $|x| > y$:

$$m_k \to \frac{1}{k!} \lim_{y\to\infty} \left[\int_{x=-\infty}^{-y} x^k\, f(x)\,\mathrm{d}x + \int_{x=y}^{\infty} x^k\, f(x)\,\mathrm{d}x \right]$$

In other words, the gamma-shaped tail of f fully determines the limit of the sequence m_k. Next, we identify the gamma distributions with a convergent sequence m_k.

Consider the moments of a gamma distribution Y with parameters $d > 0$ and $\lambda > 0$:

$$\mathbb{E}\left[Y^k\right] = \frac{\Gamma(d+k)}{\Gamma(d)\lambda^k}$$

By Stirling's formula,

$$\frac{\mathbb{E}\left[Y^k\right]}{k!} = \frac{\Gamma(d+k)}{k!\,\Gamma(d)\lambda^k}$$

$$
\begin{aligned}
&\propto \frac{\sqrt{2\pi(d+k-1)}}{\Gamma(d)\lambda^k\sqrt{2\pi k}} \left(\frac{d+k-1}{e}\right)^{d+k-1} \left(\frac{e}{k}\right)^k \\
&\propto \frac{1}{\Gamma(d)\lambda^k} \left(\frac{d+k-1}{e}\right)^{d-1} \left(1+\frac{d-1}{k}\right)^{k+1/2} \\
&\propto \frac{1}{\Gamma(d)\lambda^k} \left(\frac{d+k-1}{e}\right)^{d-1} e^{d-1} \\
&\propto \frac{1}{\Gamma(d)\lambda^k} (d+k-1)^{d-1}.
\end{aligned}
$$

Therefore, $\mathbb{E}\left[Y^k\right]/k! \to 0$ if and only if $\lambda > 1$.

Since the tails determine the asymptotics of the sequence m_k of X, the sequence m_k converges to zero ($m_k \to 0$) if and only if $\lambda > 1$.

□

Proof of Theorem 1

Diagonalize the symmetric matrix $\mathbf{B} = \mathbf{I} - \mathbf{D}\mathbf{\Sigma}^{-1}\mathbf{D}$ as $\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$ with $\mathbf{\Lambda}_{jj} = \lambda_j$ ($j = 1, \ldots, d$). The matrix $\mathbf{D}\mathbf{\Sigma}^{-1}\mathbf{D}$ is positive definite, which implies $\lambda_j < 1$. It follows from $Z \sim N(\mu, \mathbf{I})$ that $Z \sim \mu + SZ_0$ with S the direction of Z uniform on the unit d-sphere, and Z_0, the distance to μ, has density

$$
h(z) = \frac{2^{1-d/2}}{\Gamma(d/2)} z^{d-1} e^{-z^2/2}
$$

The factor z^{d-1} is the Jacobian in a spherical representation. This gives for $\mathbb{E}[\exp(X)]$ with $X = \frac{1}{2}Z'\mathbf{B}Z 1_{\mathbf{a}\le Z\le \mathbf{b}}$ the spherical representation

$$
\mathbb{E}[\exp(X)] = \frac{1}{S_d}\int_{\|\mathbf{s}\|=1} \mathbb{E}\left[\exp\left(\frac{1}{2}(\mu+\mathbf{s}Z_0)'\mathbf{B}(\mu+\mathbf{s}Z_0)\right) 1_{\mathbf{a}\le\mu+\mathbf{s}Z_0\le\mathbf{b}}\right] \mathrm{d}\mathbf{s} \tag{6.57}
$$

where $S_d = \int_{\|\mathbf{s}\|=1} \mathrm{d}\mathbf{s}$ is the surface of the unit d-sphere. As follows from the proof of Lemma 12, it suffices for studying the convergence of (6.57) to consider the tail of each $X_{\mathbf{s}} := \frac{1}{2}(\mu+\mathbf{s}Z_0)'\mathbf{B}(\mu+\mathbf{s}Z_0)$ in the mixture (6.57).

A truncated $X_{\mathbf{s}}$ has finite moments and, hence, a convergent Taylor series expansion for $\exp(X_{\mathbf{s}})$. That is, only non-truncated $X_{\mathbf{s}}$ can result in a divergent series expansion of $\exp(X)$. The truncation of the (univariate) distribution $X_{\mathbf{s}}$ depends on the indicator function in (6.57). Since $\mathbf{a} > -\infty$ and $I = \{i : b_i = \infty\}$, we have a non-truncated distribution $X_{\mathbf{s}}$ (thus $\lim_{z\to\infty} 1_{\mathbf{a}\le\mu+\mathbf{s}z\le\mathbf{b}} = 1$) if and only if each of the following three conditions holds:

(a) $s_I \geq 0$

(b) $s_i > 0$ if $i \in I$ and $\mu_i < a_i$

(c) $s_i = 0$ and $\mu_i \in [a_i, b_i]$ if $i \notin I$

A violation of any of the necessary divergence conditions (a)–(c) implies that the series $\exp(X_{\mathbf{s}})$ converges. That is, the violation of (c) gives the sufficient convergence condition (i) in Theorem 1.

Next, we study the elements in the set I to further identify distributions $X_{\mathbf{s}}$ with a divergent series expansion. We refer to a variable x in this reduced dimension as $\hat{x}$. Notice that $\hat{\mathbf{B}} = \mathbf{B}_{II}$ with eigenvalue decomposition $\hat{\mathbf{V}}\hat{\mathbf{\Lambda}}\hat{\mathbf{V}}'$ (with $\hat{\mathbf{V}}' = \hat{\mathbf{V}}^{-1}$) has in general $\hat{\mathbf{V}} \neq \mathbf{V}_{II}$, $\hat{\mathbf{\Lambda}} \neq \mathbf{\Lambda}_{II}$, and $\hat{\lambda}_i < 1$. Since $\hat{\mathbf{b}} = \mathbf{b}_I = \infty\mathbf{1}$,

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big(\hat{X}\Big)\Big] &= \mathbb{E}\left[\exp\left(\frac{1}{2}\hat{Z}'\hat{\mathbf{B}}\hat{Z}\right)1_{\hat{\mathbf{a}}\leq\hat{Z}}\right] \\
&= \mathbb{E}\left[\exp\left(\frac{1}{2}\hat{Z}'\hat{\mathbf{V}}\hat{\mathbf{\Lambda}}\hat{\mathbf{V}}'\hat{Z})\right)1_{\hat{\mathbf{a}}\leq\hat{Z}}\right] \\
&= \mathbb{E}\left[\exp\left(\frac{1}{2}\hat{W}'\hat{\mathbf{\Lambda}}\hat{W}\right)1_{\hat{\mathbf{a}}\leq\hat{\mathbf{V}}^{-1}\hat{W}}\right]
\end{aligned}$$

where $\hat{W} = \hat{\mathbf{V}}'\hat{Z} \sim N(\hat{\xi}, \mathbf{I})$ with $\hat{\xi} = \hat{\mathbf{V}}'\hat{\mu}$. This means that we consider a sum of $\hat{d}$ independent noncentral χ^2-distributions with ξ containing the noncentrality parameters. Lemma 11 indicates that the noncentrality parameters in $\hat{\xi}$ have no effect on the tail shape. Therefore, we neglect $\hat{\xi}$ and so $\hat{\mu}$ by considering the spherical representation (6.57) in the reduced dimension with $\hat{X}_{\hat{\mathbf{s}}} := \frac{1}{2}Z_0^2\hat{\lambda}(\hat{\mathbf{s}})$ and $\hat{\lambda}(\hat{\mathbf{s}}) := \hat{\mathbf{s}}'\hat{\mathbf{B}}\hat{\mathbf{s}}$ along each direction $\hat{\mathbf{s}}$. Motivated by the divergence condition (a), we only consider nonnegative $\hat{\mathbf{s}}$. For large Z_0,

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big(\hat{X}\Big)\Big] &\sim \hat{c}_2\int_{||\hat{\mathbf{s}}||=1,\hat{\mathbf{s}}\geq 0}\mathbb{E}\left[\exp\left(\frac{1}{2}(\hat{\mathbf{s}}Z_0)'\mathbf{B}(\hat{\mathbf{s}}Z_0)\right)1_{\hat{\mathbf{a}}\leq\mu+\hat{\mathbf{s}}Z_0}\right]\mathrm{d}\hat{\mathbf{s}} \\
&= \hat{c}_2\int_{||\hat{\mathbf{s}}||=1,\hat{\mathbf{s}}\geq 0}\mathbb{E}\Big[\exp\Big(\hat{X}_{\hat{\mathbf{s}}}\Big)1_{\hat{\mathbf{a}}\leq\mu+\hat{\mathbf{s}}Z_0}\Big]\mathrm{d}\hat{\mathbf{s}}
\end{aligned}$$

where $\hat{c}_2$ is a proportionality constant, and $\sim$ means that the approximation is valid for large Z_0, i.e., large $\left|\hat{X}_{\hat{\mathbf{s}}}\right|$. Each $\hat{X}_{\hat{\mathbf{s}}} = \frac{1}{2}Z_0^2\hat{\lambda}(\hat{\mathbf{s}})$ has probability density function

$$h_{\hat{\mathbf{s}}}(z)\mathrm{d}z = \begin{cases} \frac{h_{\hat{\mathbf{s}}}\left(\sqrt{2x/\hat{\lambda}(\hat{\mathbf{s}})}\right)}{\sqrt{2\hat{\lambda}(\hat{\mathbf{s}})x}}\mathrm{d}x \propto x^{\frac{\hat{d}}{2}-1}e^{-\frac{x}{\hat{\lambda}(\hat{\mathbf{s}})}}\mathrm{d}x & \text{if } \mathrm{sign}(x) = \mathrm{sign}(\lambda(\hat{\mathbf{s}})) \\ 0 & \text{otherwise} \end{cases} \tag{6.58}$$

where we used the implied derivative $\mathrm{d}z = \mathrm{d}x/\sqrt{2\hat{\lambda}(\hat{\mathbf{s}})x}$ of $Z_0 = \sqrt{2\hat{X}_{\hat{\mathbf{s}}}/\hat{\lambda}(\hat{\mathbf{s}})}$. The density function in (6.58) implies that $\left|\hat{X}_{\hat{\mathbf{s}}}\right|$ follows a Gamma$(\hat{d}/2, |\hat{\lambda}(\hat{\mathbf{s}})|)$-distribution. Using $\hat{\lambda}_i < 1$, all directions $\hat{\mathbf{s}}$ satisfy $\hat{\lambda}(\hat{\mathbf{s}}) < 1$. Therefore, directions with $\hat{\lambda}(\hat{\mathbf{s}}) > -1$ satisfy $|\hat{\lambda}(\hat{\mathbf{s}})| < 1$, and thus have by Lemma 12 a convergent Taylor series in (6.57).

Collecting previous results implies that only $\hat{\mathbf{s}} \geq 0$ with $\hat{\lambda}(\hat{\mathbf{s}}) \leq -1$ have a divergent series expansion. As a consequence, emptiness of the set $\mathcal{S}$ in (*ii*) of Theorem 1 is a sufficient condition for convergence. It is straightforward that the set $\mathcal{S}$ is empty if one of the convergence conditions (*iii*)-(*vii*) holds. On the other hand, it suffices for divergence if the set $\mathcal{S}$ has a positive measure.

□

6.A.2 Proofs Section 6.4.1

Proof of Lemma 2

(*i*) First consider a strictly positive X. Then $Y = \log X$ is well-defined. Define $m(\alpha) = \mathbb{E}[X^\alpha]$, then $\log m(\alpha)$ is easily seen to be the cumulant generating function of Y. As these functions are known to be strictly convex for non-degenerate random variables, $2\log m((\alpha+\beta)/2) < \log m(\alpha) + \log m(\beta)$ and setting $\alpha = n$, $\beta = n-2$ then proves $\mu_{n-2}\mu_n > \mu_{n-1}^2$.

For general nonnegative X the inequality can be shown to be true by conditioning, since $\mathbb{E}[X^n] = \mathbb{E}\left[X^n \mid X > 0\right]\mathbb{P}(X > 0)$. The only condition, $\mathbb{P}(X > 0) > 0$, follows from the assumed non-degeneracy.

For X unrestricted in sign, note that $\mu_n \leq \mathbb{E}[|X|^n]$, with equality for even n. The strict inequality $\mu_n^2 < \mathbb{E}[|X|^n]^2$ holds for odd n and X with nonzero probability mass both left and right of zero. These facts imply that $\mu_n\mu_{n-2} > \mu_{n-1}^2$ for any nondegenerate X and even n.

It is straightforward that equality holds for the degenerate case.

(*ii*) Follows from $\mu_i > 0$ and (*i*). □

Proof of Lemma 3

We prove strict bounds for X with both X^- and X^+ nondegenerate, since the equality cases of (i)-(iv) are straightforward.

The constraints in (6.21) are straightforward. By definition, $m_1^s = \mu_1^s/\mu_0^s = \mu_1^s$. The first column of inequalities in (6.22)-(6.24) follows from $\mu_i = \mu_i^+ + (-1)^i \mu_i^-$. For a nonnegative random variable X with cdf F and support on $[\underline{x}, \bar{x}]$

$$\underline{x} < m_{i+1} = \frac{\int_{\underline{x}}^{\bar{x}} x x^i \mathrm{d}F(x)}{\int_{\underline{x}}^{\bar{x}} x^i \mathrm{d}F(x)} < \bar{x}$$

The second and third column in (6.23) and (6.24) follow from

$$\begin{aligned}
\frac{1}{m_n^+} &= \frac{\mu_{n-1}^+}{\mu_n^+} > \frac{\mu_{n-1}^+ - \mu_{n-1}^-}{\mu_n^+ + \mu_n^-} = \frac{1}{m_n} \\
m_{n-1}^+ &= \frac{\mu_{n-1}^+}{\mu_{n-2}^+} > \frac{\mu_{n-1}^+ - \mu_{n-1}^-}{\mu_{n-2}^+ + \mu_{n-2}^-} = m_{n-1} \\
\frac{1}{m_n^-} &= \frac{\mu_{n-1}^-}{\mu_n^-} > \frac{\mu_{n-1}^- - \mu_{n-1}^+}{\mu_n^+ + \mu_n^-} = -\frac{1}{m_n} \\
m_{n-1}^- &= \frac{\mu_{n-1}^-}{\mu_{n-2}^-} > \frac{\mu_{n-1}^- - \mu_{n-1}^+}{\mu_{n-2}^+ + \mu_{n-2}^-} = -m_{n-1}
\end{aligned}$$

□

Proof of Lemma 4

We first obtain strict bounds for X with both X^- and X^+ both nondegenerate. The moments μ_{n-2}, μ_{n-1} and μ_n are known and satisfy $\mu_i = \mu_i^+ + (-1)^i \mu_i^-$ $(i = n-2, n-1, n)$. Combining this with $0 < m_{n-1}^+ < m_n^+$ (apply Lemma 2 to X^+)

$$\begin{aligned}
\mu_{n-2}^- &= \mu_{n-2} - \frac{\mu_{n-1}^+}{\mu_{n-1}^+}\mu_{n-2}^+ = \mu_{n-2} - \frac{\mu_{n-1}^+}{m_{n-1}^+} < \mu_{n-2} - \frac{\mu_{n-1}^+}{m_n^+} \\
\mu_{n-1}^- &= \mu_{n-1}^+ - \mu_{n-1} \\
\mu_n^- &= \mu_n - m_n^+ \mu_{n-1}^+
\end{aligned}$$

It is most convenient to derive bounds in terms of the reciprocal of μ_i^+, μ_i^-, m_i^+ and m_i^-:

$$m_n^- > m_{n-1}^- \Leftrightarrow \mu_{n-2}^- \mu_n^- > (\mu_{n-1}^-)^2$$

$$
\begin{aligned}
&\Rightarrow \left(\mu_{n-2} - \frac{\mu_{n-1}^{+}}{m_n^{+}}\right)\mu_n^{-} > \left(\mu_{n-1}^{-}\right)^2 \\
&\Leftrightarrow \left(\mu_{n-2} - \frac{\mu_{n-1}^{+}}{m_n^{+}}\right)\left(\mu_n - m_n^{+}\mu_{n-1}^{+}\right) > \left(\mu_{n-1}^{+} - \mu_{n-1}\right)^2 \\
&\Leftrightarrow \mu_{n-2}\mu_n - m_n^{+}\mu_{n-1}^{+}\mu_{n-2} - \frac{\mu_{n-1}^{+}\mu_n}{m_n^{+}} + \left(\mu_{n-1}^{+}\right)^2 \\
&\quad > (\mu_{n-1}^{+})^2 - 2\mu_{n-1}^{+}\mu_{n-1} + \mu_{n-1}^2 \\
&\Leftrightarrow \frac{1}{\mu_{n-1}^{+}} > \frac{\frac{\mu_n}{m_n^{+}} - 2\mu_{n-1} + \mu_{n-2}m_n^{+}}{\mu_{n-2}\mu_n - \mu_{n-1}^2}
\end{aligned} \tag{6.59}
$$

Starting from

$$
\mu_{n-2}^{-} = \mu_{n-2} - \frac{\mu_{n-1}^{+}}{m_{n-1}^{+}} \qquad \mu_{n-1}^{-} = \mu_{n-1}^{+} - \mu_{n-1} \qquad \mu_n^{-} < \mu_n - m_{n-1}^{+}\mu_{n-1}^{+}
$$

gives along similar lines (replace m_n^{+} by m_{n-1}^{+} in (6.59).)

$$
\frac{1}{\mu_{n-1}^{+}} > \frac{\frac{\mu_n}{m_{n-1}^{+}} - 2\mu_{n-1} + \mu_{n-2}m_{n-1}^{+}}{\mu_{n-2}\mu_n - \mu_{n-1}^2} \tag{6.60}
$$

Equations (6.59) and (6.60) imply, respectively

$$
\frac{1}{\mu_n^{+}} = \frac{1}{\mu_{n-1}^{+}m_n^{+}} > \frac{\frac{\mu_n}{(m_n^{+})^2} - \frac{2\mu_{n-1}}{m_n^{+}} + \mu_{n-2}}{\mu_{n-2}\mu_n - \mu_{n-1}^2} \tag{6.61}
$$

$$
\frac{1}{\mu_{n-2}^{+}} = \frac{m_{n-1}^{+}}{\mu_{n-1}^{+}} > \frac{\mu_n - 2\mu_{n-1}m_{n-1}^{+} + \mu_{n-2}\left(m_{n-1}^{+}\right)^2}{\mu_{n-2}\mu_n - \mu_{n-1}^2} \tag{6.62}
$$

Since $-X = X^{-} - X^{+}$, we can switch all $-$ and $+$ symbols and use $E[(-X)^i] = (-1)^i\mu_{n-1}$ to obtain from (6.59)–(6.62)

$$
\frac{1}{\mu_{n-1}^{-}} > \frac{\frac{\mu_n}{m_n^{-}} + 2\mu_{n-1} + \mu_{n-2}m_n^{-}}{\mu_{n-2}\mu_n - \mu_{n-1}^2} \tag{6.63}
$$

$$
\frac{1}{\mu_{n-1}^{-}} > \frac{\frac{\mu_n}{m_{n-1}^{-}} + 2\mu_{n-1} + \mu_{n-2}m_{n-1}^{-}}{\mu_{n-2}\mu_n - \mu_{n-1}^2} \tag{6.64}
$$

$$
\frac{1}{\mu_n^{-}} > \frac{\frac{\mu_n}{(m_n^{-})^2} + \frac{2\mu_{n-1}}{m_n^{-}} + \mu_{n-2}}{\mu_{n-2}\mu_n - \mu_{n-1}^2} \tag{6.65}
$$

$$\frac{1}{\mu_{n-2}^-} > \frac{\mu_n + 2\mu_{n-1}m_{n-1}^- + \mu_{n-2}\left(m_{n-1}^-\right)^2}{\mu_{n-2}\mu_n - \mu_{n-1}^2} \tag{6.66}$$

The bounds in (6.59)–(6.66) depend on the known moments μ_{n-2}, μ_{n-1} and μ_n as well as on the unknown moment ratios m_{n-1}^s and m_n^s ($s \in \{+,-\}$). We first bound (6.59) and (6.63) in terms of the known moments. Define

$$g^+(x) = \frac{1}{a_n}\left(b_n x - 2m_{n-1} + \frac{1}{x}\right) \qquad g^-(x) = \frac{1}{a_n}\left(b_n x + 2m_{n-1} + \frac{1}{x}\right)$$

where a_n and b_n are as in (6.25), and we suppress the dependence of g^+ and g^- on n for brevity. The bounds in (6.59) and (6.63) are $1/\mu_{n-1}^s > g^s(1/m_n^s)$. The bounds on m_n^s satisfy from Lemma 3

$$\max\left(\frac{1}{m_n}, 0\right) \le \frac{1}{\bar{m}_n^+} \le \frac{1}{m_n^+} \le \frac{1}{\underline{m}_n^+} \quad \max\left(-\frac{1}{m_n}, 0\right) \le \frac{1}{\bar{m}_n^-} \le \frac{1}{m_n^-} \le \frac{1}{\underline{m}_n^-} \tag{6.67}$$

Thus, we need a lower bound on g^s at the positive but unknown value $1/m_n^s \in [1/\bar{m}_n^s, 1/\underline{m}_n^s]$. For the first derivative of g^s,

$$\frac{\mathrm{d}g^s}{\mathrm{d}x} = \frac{1}{a_n}\left(b_n - \frac{1}{x^2}\right)$$

By Lemma 2(i), $a_n > 0$ such that g^s decreases on $(0, 1/\sqrt{b_n})$ and increases on $(1/\sqrt{b_n}, \infty)$. Therefore, the minimizing x of $g^s(x)$ on $[1/\bar{m}_n^s, 1/\underline{m}_n^s]$ is as close as possible to $1/\sqrt{b_n}$. Equivalently, the minimizing x of $g^s(1/x)$ for $x \in [\underline{m}_n^s, \bar{m}_n^s]$ is as close as possible to $x = \sqrt{b_n}$:

$$\frac{1}{\mu_{n-1}^s} > w_n^s\left(\underline{m}_n^s, \bar{m}_n^s\right) \qquad\qquad s \in \{+,-\} \tag{6.68}$$

where

$$w_n^s(c,d) = \begin{cases} g^s(1/c) & \sqrt{b_n} < c \\ g^s(1/\sqrt{b_n}) & c \le \sqrt{b_n} \le d \\ g^s(1/d) & \sqrt{b_n} > d \end{cases}$$

The bounds in (6.68) are positive because $m_n > m_{n-1}$ implies $\sqrt{b_n} > |m_{n-1}|$ such that the global minimum of g^s admits $g^s(1/\sqrt{b_n}) > 0$. Thus, the reciprocal of (6.68) gives upper bounds for $\mu_{n-1}^s < v_n^s\left(\underline{m}_n^s, \bar{m}_n^s\right) := 1/w_n^s\left(\underline{m}_n^s, \bar{m}_n^s\right)$. A similar derivation starting from (6.60) and (6.64) yields $\mu_{n-1}^s < v_n^s\left(\underline{m}_{n-1}^s, \bar{m}_{n-1}^s\right)$.

Letting $\bar{\mu}^s_{n-1} = \min\left(v^s_n\left(\underline{m}^s_n, \bar{m}^s_n\right), v^s_n\left(\underline{m}^s_{n-1}, \bar{m}^s_{n-1}\right)\right)$ represent an upper bound on μ^s_{n-1}, an upper bound on μ^s_n is available from $\mu^s_n \le \bar{m}^s_n \bar{\mu}^s_{n-1}$. Nonetheless, if $\sqrt{b_n} < \bar{m}^s_n$ we can obtain tighter bounds from (6.61) and (6.65). More precisely, we prove that in our range of interest the bounds in (6.61) and (6.65) increase monotonically in the unknown parameters m^s_n. Accordingly, the upper bounds on m^+_n and m^-_n give the tightest lower bounds on (6.61) and (6.65), respectively. Let

$$h(x) = x\, g^+(x) = \frac{1}{a_n}\left(b_n x^2 - 2m_{n-1}x + 1\right)$$

The bounds in (6.61) and (6.65) are $1/\mu^+_n > h(1/m^+_n)$ and $1/\mu^-_n > h(-1/m^-_n)$, respectively. For the first derivative of h (note $m_{n-1} = b_n/m_n$),

$$\frac{\mathrm{d}h}{\mathrm{d}x} = \frac{2b_n}{a_n}\left(x - \frac{1}{m_n}\right)$$

The function h decreases on $(-\infty, 1/m_n)$ and increases on $(1/m_n, \infty)$. By (6.67) we have $1/m_n \le 1/\bar{m}^+_n \le 1/m^+_n$, which means $h\left(1/\bar{m}^+_n\right) \le h\left(1/m^+_n\right)$. Thus, we can safely use $\bar{m}^+_n$ without violating the lower bound $h(1/m^+_n)$ on $1/\mu^+_n$ in (6.61). We also assumed $-1/m^-_n \le -1/\bar{m}^-_n < 0 < 1/m_n$, which ensures $h(-1/\bar{m}^-_n) \le h(-1/m^-_n)$. Hence, the upper bound $\bar{m}^-_n$ on m^-_n cannot violate the lower bound on $1/\mu^-_n$ in (6.65). Obviously, a strictly tighter bound $\bar{m}^s_n$ on m^s_n produces a strictly tighter bound on $1/\mu^s_n$.

The minimum of h, and so each lower bound, is positive:

$$h\left(\frac{1}{m_n}\right) = \frac{1}{a_n}\left(\frac{b_n}{m_n^2} - \frac{2m_{n-1}}{m_n} + 1\right) = \frac{1}{a_n}\left(1 - \frac{m_{n-1}}{m_n}\right) > 0$$

Inverting the positive lower bounds gives

$$\begin{aligned} \mu^+_n &< 1/h(1/\bar{m}^+_n) = \bar{m}^+_n u^+_n(\bar{m}^+_n) \\ \mu^-_n &< 1/h(-1/\bar{m}^-_n) = -m^-_n u^+_n(-m^-_n) = m^-_n u^-(m^-_n). \end{aligned}$$

with the functions $u^+_n(.)$ and $u^-_n(.)$ as in (6.26). The upper bounds on μ^s_{n-2} follow from (6.62), (6.66), and a similar analysis on $k(x) := x^2 h(1/x) = x/u^+_n(x)$. The function k has a minimum $k(m_{n-1}) = (b_n - m^2_{n-1})/a_n > 0$, decreases on $[-\bar{m}^-_{n-1}, -\underline{m}^-_{n-1}]$, and increases on $[\underline{m}^+_{n-1}, \bar{m}^+_{n-1}]$. Since $-\underline{m}^-_{n-1} < m_{n-1} < \underline{m}^+_{n-1}$, the upper bounds on μ^s_{n-2} are determined by $\underline{m}^s_{n-1}$.

The case with equality signs follows by noting that $b_n = m_{n-1}$ if and only if $|X|$ degenerates at $(0, \infty)$.

□

Proof of Lemma 5
Follows from $\mu_i^s = m_i^s \mu_{i-1}^s$ and $\mu_i = \mu_i^+ + (-1)^i \mu_i^-$ and Lemma 2.

□

Proof of Lemma 6
We prove the case with strict bounds, the case with equality signs is trivial. The two leftmost inequalities in (6.27) are straightforward. The two rightmost inequalities in (6.27) follow immediately from Lemma 2(*ii*):

$$m_{i-1}^s < \sqrt{m_{i-1}^s m_i^s} = \sqrt{\frac{\mu_i^s}{\mu_{i-2}^s}} < m_i^s$$

Since $\mu_i = \mu_i^+ + (-1)^i \mu_i^-$ and $\mu_i^s = m_i^s \mu_{i-1}^s$

$$(-1)^i m_i^+ = \frac{(-1)^i \mu_i - \mu_i^-}{\mu_{i-1}^+} \qquad (-1)^i m_i^- = \frac{\mu_i - \mu_i^+}{\mu_{i-1}^-}$$

By carefully constructing the most conservative bounds from the implied sign of the numerators above,

$$\text{If } \underline{\mu}_i^- < (-1)^i \mu_i\text{:} \quad (-1)^i m_i^+ < \frac{(-1)^i \mu_i - \underline{\mu}_i^-}{\underline{\mu}_{i-1}^+} \quad \text{else } (-1)^i m_i^+ < \frac{(-1)^i \mu_i - \underline{\mu}_i^-}{\bar{\mu}_{i-1}^+}$$

$$\text{If } \bar{\mu}_i^- < (-1)^i \mu_i\text{:} \quad (-1)^i m_i^+ > \frac{(-1)^i \mu_i - \bar{\mu}_i^-}{\bar{\mu}_{i-1}^+} \quad \text{else } (-1)^i m_i^+ > \frac{(-1)^i \mu_i - \bar{\mu}_i^-}{\underline{\mu}_{i-1}^+}$$

$$\text{If } \underline{\mu}_i^+ < \mu_i\text{:} \quad (-1)^i m_i^- < \frac{\mu_i - \underline{\mu}_i^+}{\underline{\mu}_{i-1}^-} \quad \text{else } (-1)^i m_i^- < \frac{\mu_i - \underline{\mu}_i^+}{\bar{\mu}_{i-1}^-}$$

$$\text{If } \bar{\mu}_i^+ < \mu_i\text{:} \quad (-1)^i m_i^- > \frac{\mu_i - \bar{\mu}_i^+}{\bar{\mu}_{i-1}^-} \quad \text{else } (-1)^i m_i^- > \frac{\mu_i - \bar{\mu}_i^+}{\underline{\mu}_{i-1}^-}$$

By Lemma 3, $\mu_i^+ < \mu_i$ and $\mu_i^- < \mu_i$ for even i such that (6.28) contains the if-cases above. For odd i, we have by Lemma $\mu_i^+ > \mu_i$ and $\mu_i^- > -\mu_i$ such that (6.29) contains the else-cases above.

□

6.A.3 Proofs Section 6.4.2

We derive in some lemmas the required building blocks for the bounds in Theorem 8. We extensively use the definitions in (6.30)–(6.31).

Lemma 13. *The function t_n has the following properties* ($n = 0, 1, 2, \ldots$)

$$t_n(x) = \sum_{k=1}^{\infty} \frac{x^k}{(n+k)!} \qquad \operatorname{sign}(t_n(x)) = \operatorname{sign}(x)$$

$$t_n > -\frac{1}{n!} \qquad t_n' > 0 \qquad t_n'' > 0$$

Proof of Lemma 13

The summation representation of t_n follows from the Taylor series of e^x in the definition of $r_n(x)$. Consider the function

$$t_0(x; u) := e^{ux} - 1 = \sum_{k=1}^{\infty} \frac{(ux)^k}{k!}.$$

Define $t_n(x; u)$ recursively by

$$t_n(x; u) = \int_{v=0}^{u} t_{n-1}(x; v)\, \mathrm{d}v \qquad n = 1, 2, \ldots \tag{6.69}$$

It follows that $t_n(x; u) = \sum_{k=1}^{\infty} \frac{u^{n+k} x^k}{(n+k)!}$ such that $t_n(x; 1) = t_n(x)$. If $x_0 < x_1$ then $t_0(x_0; u) < t_0(x_1; u)$ for all $u > 0$ and, by induction on n in (6.69), $t_n(x_0; u) < t_n(x_1; u)$ for all $u > 0$. This shows $\frac{d}{dx} t_n(x; u) > 0$, and in particular $t_n' > 0$. The sign of t_n now follows from $t_n(0) = 0$. Further,

$$\lim_{x \to -\infty} t_n(x) = \lim_{x \to -\infty} x^{-n} e^x - \sum_{k=0}^{n} \frac{x^{k-n}}{k!} = -\frac{1}{n!}$$

For the second derivative, notice that by the convexity of t_0,

$$t_0(x_0 + \delta; u) - t_0(x_0; u) < t_0(x_1 + \delta; u) - t_0(x_1; u) \qquad \delta, u > 0 \qquad x_0 < x_1$$

Induction on n through (6.69) implies

$$t_n(x_0 + \delta; u) - t_n(x_0; u) < t_n(x_1 + \delta; u) - t_n(x_1; u)$$

which implies an increasing slope of t_n, i.e., $t_n'' > 0$.

□

Lemma 14 contains several properties of s_n from (6.31).

Lemma 14. *Suppose $\mu_n \neq 0$ where n is an arbitrary integer. Then,*

(*i*) $s_n(X) = \mathbb{E}\big[t_n(X^{(n)})\big]$ *with the cdf of $X^{(n)}$ as defined in* (6.18).

(*ii*) *If* $X \geq 0$, $s_n(\lambda X)$ *is strictly convex and strictly increasing in* λ *with*

$$\lim_{\lambda \to 0} s_n(\lambda X) = 0$$

(*iii*) *If* $X \geq 0$,

$$-\frac{1}{n!} < s_n(-X) < 0 < t_n(m_{n+1}) \leq s_n(X)$$

where as before, $m_i = \mu_i/\mu_{i-1}$. *The inequalities are all strict if and only if* X *is nondegenerate.*

(*iv*) *If* $0 \leq X^{(n)} \leq_{\mathrm{st}} Y^{(n)}$,

$$-\frac{1}{n!} < s_n(-Y) \leq s_n(-X) < 0 < s_n(X) \leq s_n(Y)$$

The inequalities are all strict if and only if X *is strictly dominated by* Y.

(*v*) *If* $X \equiv x$,

$$s_n(X) = t_n(x)$$

(*vi*) *If* $X = BY$ *with* B *an independent Bernoulli random variable with* $\mathbb{P}(B = 1) > 0$,

$$s_n(X) = s_n(Y)$$

Proof of Lemma 14

(*i*) Using subsequently (6.31), (6.19), and Lemma 13 produces

$$s_n(X) = \frac{1}{\mu_n} \sum_{i=1}^{\infty} \frac{\mu_{n+i}}{(n+i)!} = \sum_{i=1}^{\infty} \frac{\mu_i^{(n)}}{(n+i)!} = \mathbb{E}\Big[t_n(X^{(n)})\Big]$$

(*ii*) Similar to (*i*),

$$s_n(\lambda X) = \frac{\mathbb{E}[r_n(\lambda X)]}{\mathbb{E}[(\lambda X)^n]} = \mathbb{E}\left[\sum_{i=1}^{\infty} \frac{(\lambda X^{(n)})^i}{(n+i)!}\right] = \mathbb{E}\Big[t_n(\lambda X^{(n)})\Big]$$

which leads to

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} s_n(\lambda X) = \mathbb{E}\Big[X^{(n)} t_n'(\lambda X^{(n)})\Big]$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} s_n(\lambda X) = \mathbb{E}\left[\left(X^{(n)}\right)^2 t_n''(\lambda X^{(n)})\right]$$

The required result follows from $\mathbb{P}\big(X^{(n)} \neq 0\big) > 0$ and $t_n'' > 0$ in Lemma 13.

(*iii*) From Lemma 13, we have $-1/n! < t_n(x) < 0$ if $x < 0$. Substituting the result of Lemma 14(*i*),

$$-\frac{1}{n!} < s_n(-X) = \mathbb{E}\left[t_n(-X^{(n)})\right] < 0$$

Lemma 2(*ii*) indicates that the moment ratios m_i of $X \geq 0$ is a positive and nondecreasing series. Since $t_n(x) > 0$ for $x > 0$ (see Lemma 13)

$$\begin{aligned} 0 < t_n(m_{n+j+1}) &= \sum_{k=1}^{\infty} \frac{m_{n+j+1}^k}{(n+k)!} \\ &\leq \sum_{k=1}^{\infty} \frac{1}{(n+k)!} \prod_{l=n+j+1}^{n+j+k} m_l = \frac{1}{\mu_{n+j}} \sum_{k=1}^{\infty} \frac{\mu_{n+j+k}}{(n+k)!} = s_n\left(X^{(j)}\right) \end{aligned}$$

Substituting $j = 0$ gives Lemma 14(*iii*). The inequalities are strict unless X is degenerate.

(*iv*) Rewriting s_n in terms of t_n (see (*i*)), the result follows from the lower bound $-1/n!$ on t_n, $t_n' > 0$, and $t_n(0) = 0$ (see Lemma 13). The inequalities are strict unless $X \stackrel{d}{=} Y$, because then also $X^{(n)} \stackrel{d}{=} Y^{(n)}$.

(*v*) Follows immediately from the definitions in (6.30) and (6.31).

(*vi*) By the law of total expectations,

$$\begin{aligned} s_n(X) &= \frac{\mathbb{E}\left[r_n(X) \mid B=0\right] \mathbb{P}(B=0) + \mathbb{E}\left[r_n(X) \mid B=1\right] \mathbb{P}(B=1)}{\mathbb{E}\left[X^n \mid B=0\right] \mathbb{P}(B=0) + \mathbb{E}\left[X^n \mid B=1\right] \mathbb{P}(B=1)} \\ &= \frac{0 + \mathbb{E}[r_n(Y)]\, \mathbb{P}(B=1)}{0 + \mathbb{E}[Y^n]\, \mathbb{P}(B=1)} = s_n(Y) \end{aligned}$$

□

Below, we consider the random variables $X \geq 0$ and $Y \geq 0$ with density functions f and g, and positive support on $S_X = (0, \bar{x}]$ and $S_Y = (0, \bar{y}]$, respectively. The density functions are also allowed to be nonzero at 0.

The operator $X \leq_{\text{MLR}} Y$ indicates $\bar{x} \leq \bar{y}$, and a nondecreasing likelihood ratio on the shared support $S_{XY} = \{S_X \cap S_Y\} \backslash \{0\} = (0, \min(\bar{x}, \bar{y})]$:[6]

$$\frac{g(x_1)}{f(x_1)} \leq \frac{g(x_2)}{f(x_2)} \qquad x_1 < x_2 \qquad x_1, x_2 \in S_{XY}$$

The pair X and Y is then said to satisfy the monotone likelihood ratio (MLR) property. We denote by $X <_{\text{MLR}} Y$ the case where (i) $X \leq_{\text{MLR}} Y$ and (ii) $\bar{x} < \bar{y}$ or $g(x_1)/f(x_1) < g(x_2)/f(x_2)$ holds for all $x_1 < x_2$ with x_1 and x_2 in some dense subset of S_{XY}.

Notice that no condition is imposed on the likelihood ratio at 0. Lemma 15 contains some useful properties of the MLR property. $X|_{X>0}$ refers to the distribution of X conditional on $X > 0$. The corresponding cdf is

$$\tilde{F}(x) = \mathbb{P}\big(X|_{X>0} \leq x\big) = \frac{\mathbb{P}(0 < X \leq x)}{\mathbb{P}(X > 0)} = \frac{F(x) - F(0)}{1 - F(0)}$$

Lemma 15. *Let the monotonic function $h : \mathbb{R}^+ \to \mathbb{R}^+$ be continuously differentiable with $h(0) = 0$ and $h' > 0$.*

If $X \leq_{\text{MLR}} Y$, then for $n = 0, 1, 2, \ldots$

- (i) $h(X^{(j)}) \leq_{\text{MLR}} h(Y^{(j)})$ *for all* $j \in \mathbb{R}$
- (ii) $X|_{X>0} \leq_{\text{MLR}} Y|_{Y>0}$
- (iii) $X \leq_{\text{st}} Y$
- (iv) $-1/n! < s_n(-h(Y)) \leq s_n(-h(X)) < 0 < s_n(h(X)) \leq s_n(h(Y))$

The inequalities $\leq$, $\leq_{\text{st}}$ and $\leq_{\text{MLR}}$ are all strict if and only if (a) $\bar{x} < \bar{y}$ or (b) $X <_{MLR} Y$.

Proof of Lemma 15

Denote the density of the distributions X, $\tilde{X} := X|_{X>0}$, Y, and $\tilde{Y} := Y|_{Y>0}$ by f, g, $\tilde{f}$ and $\tilde{g}$, respectively.

- (i) First, notice that $X \leq_{\text{MLR}} Y$ and $X^{(j)} \leq_{\text{MLR}} Y^{(j)}$ are equivalent because the support remains the same under the transformation (6.18), and

$$\frac{g^{(j)}(x_1)}{f^{(j)}(x_1)} = \frac{c x_1^j g(x_1)}{x_1^j f(x_1)} \leq \frac{c x_2^j g(x_2)}{x_2^j f(x_2)} = \frac{g^{(j)}(x_2)}{f^{(j)}(x_2)} \quad x_1 < x_2 \quad x_1, x_2 \in S_{XY}$$

[6] More general results follow from the formal, but less intuitive, definition of the MLR property, $f(x_2)g(x_1) \leq f(x_1)g(x_2)$, or allowing for a discontinuous support of f and g.

where the proportionality constant $c = \frac{\mathbb{E}[X^{(j)}]\mathbb{E}[Y]}{\mathbb{E}[Y^{(j)}]\mathbb{E}[X]}$ corrects for the different moments after applying (6.18).

Second, we show that the monotonic transformation h does not affect the MLR property. Define $\hat{X} = h(X)$ and $\hat{Y} = h(Y)$ with densities $\hat{f}$ and $\hat{g}$, respectively. The support of $\hat{X}$ is still a subset of the support of $\hat{Y}$ since $S_X \subset S_Y$ and h does not change the support. Because $h(0) = 0$, the support of X and Y still contains arbitrarily small positive values. By the inverse function theorem,

$$\begin{aligned}\hat{f}(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}\big(X \leq h^{-1}(x)\big) = \frac{\mathrm{d}}{\mathrm{d}x}F(h^{-1}(x)) = \\ &= \frac{\mathrm{d}h^{-1}(x)}{\mathrm{d}x}f(h^{-1}(x)) = \frac{f(h^{-1}(x))}{h'(x)}\end{aligned}$$

Because $h^{-1}(x)$ is positive and increasing in x, and $X \leq_{\mathrm{MLR}} Y$, we find again an increasing likelihood ratio:

$$\frac{\hat{g}(x)}{\hat{f}(x)} = \frac{h'(x)}{f(h^{-1}(x))}\frac{g(h^{-1}(x))}{h'(x)} = \frac{g(h^{-1}(x))}{f(h^{-1}(x))}$$

This proves $\hat{X} \leq_{\mathrm{MLR}} \hat{Y}$.

(*ii*) For $x_1 < x_2$ and $x_1, x_2 \in S_{XY}$

$$\frac{\tilde{g}(x_1)}{\tilde{f}(x_1)} = \frac{cg(x_1)}{f(x_1)} \leq \frac{cg(x_2)}{f(x_2)} = \frac{\tilde{g}(x_2)}{\tilde{f}(x_2)} \tag{6.70}$$

where the constant $c = \frac{\mathbb{E}[\tilde{X}]\mathbb{E}[Y]}{\mathbb{E}[\tilde{Y}]\mathbb{E}[X]} = \frac{\mathbb{P}(Y>0)}{\mathbb{P}(X>0)}$ corrects for the different probability mass of X and Y at zero. Thus, the likelihood ratio is still increasing, and the support on the positive half-line remains the same. This proves $X|_{X>0} \leq_{\mathrm{MLR}} Y|_{Y>0}$.

(*iii*) For $x_1 < x_2$ and $x_1, x_2 \in S_{XY}$

$$f(x_2)g(x_1) \leq f(x_1)g(x_2)$$

Integrating x_1 over $[0, x_2]$, and x_2 over $[x_1, \infty)$ gives, respectively

$$\begin{aligned}f(x_2)G(x_2) &\leq F(x_2)g(x_2) \\ (1 - F(x_1))g(x_1) &\leq f(x_1)(1 - G(x_1))\end{aligned}$$

This implies

$$\frac{1-F(x_1)}{1-G(x_1)} \le \frac{f(x_1)}{g(x_1)} \le \frac{f(x_2)}{g(x_2)} \le \frac{F(x_2)}{G(x_2)}$$

Choosing x_1 and x_2 arbitrarily close to any x shows after some rewriting that $G(x) \le F(x)$, which proves (iii).

(iv) From (i) and (iii), we have $X^{(n+j)} \le_{\text{st}} Y^{(n+j)}$ for all $n+j \in \mathbb{R}$. By Lemma 14(iv), and because $n+j$ is arbitrary, we have for all $j \in \mathbb{R}$, and $n = 0, 1, 2, \ldots$

$$-\frac{1}{n!} < s_n\left(-Y\right) \le s_n\left(-X\right) < 0 < s_n\left(X\right) \le s_n\left(Y\right)$$

Using (i), and repeating the steps with $\hat{X} := h(X) \le_{\text{MLR}} \hat{Y} := h(Y)$ results in (iv).

The equivalent condition for strictness of the inequalities follows from the definition of the MLR property and the proof.

□

Lemma 15 indicates that $X^{(i)} \le_{\text{MLR}} Y^{(i)}$ for some $i \in \mathbb{R}$ is sufficient to prove $X^{(j)}\big|_{X^{(j)}>0} \le_{\text{st}} Y^{(j)}\big|_{Y^{(j)}>0}$ for all $j \in \mathbb{R}$. The less restrictive $X^{(j)}\big|_{X^{(j)}>0} \le_{\text{st}} Y^{(j)}\big|_{Y^{(j)}>0}$ does not imply the inequality $X^{(j+1)}\big|_{X^{(j+1)}>0} \le_{\text{st}} Y^{(j+1)}\big|_{Y^{(j+1)}>0}$. For instance, consider the positive $X^{(j)} = X^{(j)}\big|_{X^{(j)}>0}$ and $Y^{(j)} = Y^{(j)}\big|_{Y^{(j)}>0}$ with

$$\mathbb{P}\Big(X^{(j)} = 1\Big) = \frac{1}{2} \qquad \mathbb{P}\Big(X^{(j)} = 4\Big) = \frac{1}{2}$$

$$\mathbb{P}\Big(Y^{(j)} = 2\Big) = \frac{1}{2} \qquad \mathbb{P}\Big(Y^{(j)} = 4\Big) = \frac{1}{2}$$

Using

$$\mathbb{P}\Big(X^{(j+1)} = i\Big) = \frac{i\,\mathbb{P}\big(X^{(j)} = i\big)}{\mathbb{E}\big[X^{(j)}\big]}$$

with means $\mathbb{E}\big[X^{(j)}\big] = \frac{5}{2}$ and $\mathbb{E}\big[Y^{(j)}\big] = 3$ gives

$$\mathbb{P}\Big(X^{(j+1)} = 1\Big) = \frac{1}{5} \qquad \mathbb{P}\Big(X^{(j+1)} = 4\Big) = \frac{4}{5}$$

$$\mathbb{P}\Big(Y^{(j+1)} = 2\Big) = \frac{1}{3} \qquad \mathbb{P}\Big(Y^{(j+1)} = 4\Big) = \frac{2}{3}$$

Thus, $X^{(j)}\big|_{X^{(j)}>0} \leq_{\text{st}} Y^{(j)}\big|_{Y^{(j)}>0}$, while it follows from the cdfs at $x = 2$ that $X^{(j+1)}\big|_{X^{(j+1)}>0} \not\leq_{\text{st}} Y^{(j+1)}\big|_{Y^{(j+1)}>0}$.

Vice versa, the inequality $X^{(j+1)}\big|_{X^{(j+1)}>0} \leq_{\text{st}} Y^{(j+1)}\big|_{Y^{(j+1)}>0}$ does not imply $X^{(j)}\big|_{X^{(j)}>0} \leq_{\text{st}} Y^{(j)}\big|_{Y^{(j)}>0}$. To see this, suppose

$$\begin{aligned} &\mathbb{P}\Big(X^{(j+1)} = 1\Big) = 1/2 \quad \mathbb{P}\Big(X^{(j+1)} = 2\Big) = 1/4 \quad \mathbb{P}\Big(X^{(j+1)} = 4\Big) = 1/4 \\ &\mathbb{P}\Big(Y^{(j+1)} = 1\Big) = 1/2 \qquad\qquad\qquad\qquad\qquad\qquad \mathbb{P}\Big(Y^{(j+1)} = 4\Big) = 1/2 \end{aligned}$$

We have $\mathbb{E}\big[1/X^{(j+1)}\big] = 11/16$, $\mathbb{E}\big[1/Y^{(j+1)}\big] = 5/8$, and $X^{(j+1)} \leq_{\text{st}} Y^{(j+1)}$. Using

$$\mathbb{P}\Big(X^{(j)} = i\Big) = \frac{\mathbb{P}\big(X^{(j+1)} = i\big)}{i} \frac{1}{\mathbb{E}\big[1/X^{(j+1)}\big]}$$

implies

$$\begin{aligned} &\mathbb{P}\Big(X^{(j)} = 1\Big) = 8/11 \quad \mathbb{P}\Big(X^{(j)} = 2\Big) = 2/11 \quad \mathbb{P}\Big(X^{(j)} = 4\Big) = 1/11 \\ &\mathbb{P}\Big(Y^{(j)} = 1\Big) = 4/5 \qquad\qquad\qquad\qquad\qquad\qquad \mathbb{P}\Big(Y^{(j)} = 4\Big) = 1/5 \end{aligned}$$

Here, $X^{(j+1)}\big|_{X^{(j+1)}>0} \leq_{\text{st}} Y^{(j+1)}\big|_{Y^{(j+1)}>0}$, but the cdfs at $x = 1$ imply $X^{(j)}\big|_{X^{(j)}>0} \not\leq_{\text{st}} Y^{(j)}\big|_{Y^{(j)}>0}$.

The next lemma applies some of the results of the preceding lemmas to a specific distribution.

Lemma 16. *Define the random variable $X_{\xi,l,u} = \frac{1}{2}R^2$ where R has density*

$$f_R(r; d, \xi, l, u) = \frac{|r|^{d-1}}{c(0; d, \xi, l, u)\sqrt{2\pi}} \exp\left(-\frac{(r-\xi)^2}{2}\right) 1_{l \leq r \leq u} \tag{6.71}$$

with

$$c(t; d, \xi, l, u) = \mathbb{E}\Big[|Z_{t,\xi}|^{d-1} 1_{l\sqrt{1-t} \leq Z_{t,\xi} \leq u\sqrt{1-t}}\Big] \quad Z_{t,\xi} \sim N\left(\frac{\xi}{\sqrt{1-t}}, 1\right) \quad t < 1$$

If $l \geq 0$, a recursive expression for $c(t; d, \xi, l, u)$ is in (6.17). *If $l < 0$, a recursive expression follows from* (6.17) *and the decomposition $c(t; d, \xi, l, u) = c(t; d, \xi, 0, l) + c(t; d, \xi, 0, u)$.*

(i) For $t < 1$, the moment generating function of $X_{\xi,l,u}$ is given by

$$m(t; d, \xi, l, u) = \frac{c(t; d, \xi, l, u)}{c(0; d, \xi, l, u)(1-t)^{d/2}} \exp\left(\frac{\xi^2 t}{2-2t}\right)$$

(ii) If $\lambda < 1$, $\lambda \neq 0$

$$s_n\left(\lambda X_{\xi,l,u}\right) = \frac{1}{\hat{\mu}_n}\left(m(\lambda; d, \xi, l, u) - \sum_{i=0}^{n} \frac{\hat{\mu}_i}{i!}\right)$$

where $m(\lambda; d, \xi, l, u)$ is from (i), and

$$\hat{\mu}_i = \left(\frac{\lambda}{2}\right)^i \frac{c(0; d+2i, \xi, l, u)}{c(0; d, \xi, l, u)},$$

with $c(0; d, \xi, l, u)$ recursively defined in (6.17).

(iii) If $l = 0$, $u \geq 0$, and $\lambda > 0$

For $x \in \left[0, \frac{u^2}{2}\right]$,

$$f_X(x; \xi, 0, u) = \frac{(2x)^{d/2-1}}{c(0; d, \xi, l, u)\sqrt{2\pi}} \exp\left(-\frac{(\sqrt{2x} - \xi)^2}{2}\right)$$

For $\xi_0 < \xi_1$,

$$s_n(-\lambda X_{\xi_1,0,u}) < s_n(-\lambda X_{\xi_0,0,u}) < 0 < s_n(\lambda X_{\xi_0,0,u}) < s_n(\lambda X_{\xi_1,0,u})$$

For $u_0 < u_1$,

$$s_n(-\lambda X_{\xi,0,u_1}) < s_n(-\lambda X_{\xi,0,u_0}) < 0 < s_n(\lambda X_{\xi,0,u_0}) < s_n(\lambda X_{\xi,0,u_1})$$

(iv) If $\xi = 0$, $l = 0$ and $u = \infty$, $\lambda < 1$, $\lambda \neq 0$

$$|\lambda| X_{0,0,\infty} \sim \text{Gamma}\left(\frac{d}{2}, |\lambda|\right)$$

$$\begin{aligned} s_n(\lambda X_{0,0,\infty}) &= \frac{1}{\lambda^n \Gamma(d/2+n)}\left(\frac{\Gamma(d/2)}{(1-\lambda)^{d/2}} - \sum\nolimits_{i=0}^{n} \frac{\lambda^i \Gamma\left(d/2+i\right)}{i!}\right) \\ &= \frac{1}{n!}\left({}_2F_1\left(1, n+\frac{d}{2}; n+1; \lambda\right) - 1\right) \end{aligned}$$

Proof of Lemma 16

Define $c_{t,d} = c(t; d, \xi, l, u)$ and $Z_t = Z_{t,\xi}$ for convenience. The recursive procedure to compute $c_{t,d}$ serves to handle the absolute operator of the moments of Z_t.

(*i*) The mgf $m(t)$ of $X_{\xi,l,u}$ is

$$\begin{aligned}
m(t) &= \mathbb{E}\left[\exp\left(\frac{t}{2}R^2\right)\right] \\
&= \frac{1}{c_{0,d}\sqrt{2\pi}}\int_l^u |r|^{d-1}\exp\left(\frac{t}{2}r^2 - \frac{(r-\xi)^2}{2}\right)\mathrm{d}r \\
&= \frac{1}{c_{0,d}\sqrt{2\pi}}\int_l^u |r|^{d-1}\exp\left(-\frac{1}{2}(1-t)r^2 + \xi r - \frac{\xi^2}{2}\right)\mathrm{d}r \\
&= \frac{1}{c_{0,d}\sqrt{2\pi}}\int_l^u |r|^{d-1}\exp\left(-\frac{1}{2}\left[r\sqrt{1-t} - \frac{\xi}{\sqrt{1-t}}\right]^2 + \frac{1}{2}\frac{\xi^2}{1-t} - \frac{\xi^2}{2}\right)\mathrm{d}r \\
&= \frac{\exp\left(\xi^2 t/(2-2t)\right)}{c_{0,d}\sqrt{1-t}\sqrt{2\pi}}\int_{l\sqrt{1-t}}^{u\sqrt{1-t}}\left(\frac{|s|}{\sqrt{1-t}}\right)^{d-1}\exp\left(-\frac{1}{2}\left[s - \frac{\xi}{\sqrt{1-t}}\right]^2\right)\mathrm{d}s \\
&= \frac{1}{c_{0,d}(1-t)^{d/2}}\exp\left(\frac{\xi^2 t}{2-2t}\right)\mathbb{E}\left[|Z_t|^{d-1}1_{l\sqrt{1-t}\le Z_t\le u\sqrt{1-t}}\right] \\
&= \frac{c_{t,d}}{c_{0,d}(1-t)^{d/2}}\exp\left(\frac{\xi^2 t}{2-2t}\right)
\end{aligned}$$

where $Z_t \sim N(\xi/\sqrt{1-t}, 1)$

(*ii*) Follows from the definition of s_n, and

$$\begin{aligned}
\hat{\mu}_i &= \mathbb{E}\left[(\lambda X)^i\right] = \frac{\lambda^i}{2^i}\mathbb{E}\left[R^{2i}\right] = \left(\frac{\lambda}{2}\right)^i\int_l^u \frac{r^{d-1+2i}}{c_{0,d}\sqrt{2\pi}}\exp\left(-\frac{(r-\xi)^2}{2}\right)\mathrm{d}r \\
&= \left(\frac{\lambda}{2}\right)^i\frac{1}{c_{0,d}}\mathbb{E}\left[|Z_0|^{d-1+2i}1_{l\le Z_0\le u}\right]
\end{aligned}$$

(*iii*) Using $l = 0$, we have $R \ge 0$, and the cdf of $X_{\xi,0,u}$ satisfies

$$\begin{aligned}
F_X(x;\xi,0,u) &= \mathbb{P}(X_{\xi,0,u} \le x) = \mathbb{P}\left(R^2 \le 2x\right) \\
&= \mathbb{P}\left(|R| \le \sqrt{2x}\right) = \mathbb{P}\left(R \le \sqrt{2x}\right)
\end{aligned}$$

Differentiating to x,

$$\begin{aligned}
f_X(x;\xi,0,u) &= \frac{1}{\sqrt{2x}}f_R\left(\sqrt{2x};\xi,0,u\right) \\
&= \frac{(2x)^{d/2-1}}{c_{0,d}\sqrt{2\pi}}\exp\left(-\frac{(\sqrt{2x}-\xi)^2}{2}\right)1_{0<\sqrt{2x}\le u}
\end{aligned}$$

While the parameter ξ affects the proportionality constant $c_{0,d}$, it does not change the dynamics of the likelihood ratio on the shared support of x. That is, for $0 < \sqrt{2x} \leq u$

$$\frac{f_X(x;\xi_1,0,u)}{f_X(x;\xi_0,0,u)} \propto \exp\left(-\frac{\left(\sqrt{2x}-\xi_1\right)^2}{2} + \frac{\left(\sqrt{2x}-\xi_0\right)^2}{2}\right)$$
$$= \exp\left(\sqrt{2x}\,(\xi_1-\xi_0) + \frac{\xi_0^2-\xi_1^2}{2}\right)$$

This likelihood ratio increases with x since $\xi_0 < \xi_1$. Furthermore, $f(x;\xi_0,0,u)$ has the same support as $f(x;\xi_1,0,u)$, which proves $X_{\xi_0,0,u} <_{\text{MLRP}} X_{\xi_1,0,u}$. The inequalities for $\xi_0 < \xi_1$ in (iii) follow from Lemma 15(iv) with $h(x) = \lambda x$.

If $0 < \sqrt{2x} \leq u_0$, the likelihood ratio function

$$\frac{f_X(x;\xi,0,u_1)}{f_X(x;\xi,0,u_0)} \propto \exp\left(-\frac{\left(\sqrt{2x}-\xi\right)^2}{2} + \frac{\left(\sqrt{2x}-\xi\right)^2}{2}\right) = 1$$

is constant, and, by $u_0 < u_1$, the support of $f(x;\xi,0,u_0)$ is strictly contained in the support of $f(x;\xi,0,u_1)$. This implies $X_{\xi,0,u_0} <_{\text{MLRP}} X_{\xi,0,u_1}$. Lemma 15$(iv)$ with $h(x) = \lambda x$ now gives the inequalities for $u_0 < u_1$.

(iv) Define $Y := \lambda X_{0,0,\infty}$. Because

$$F_{|Y|}(y;0,0,\infty) = \mathbb{P}(|\lambda| X_{0,0,\infty} \leq y) = F_X\left(\frac{y}{|\lambda|};0,0,\infty\right)$$

the density of $|Y|$ is given by

$$f_{|Y|}(y;0,0,\infty) = \frac{1}{|\lambda|} f_X\left(\frac{y}{|\lambda|};0,0,\infty\right) = \frac{(2y)^{d/2-1}}{|\lambda|^{d/2} c_{0,d}\sqrt{2\pi}} \exp\left(-\frac{y}{|\lambda|}\right)$$

where we used f_X from (iii). Notice that $Y \sim \text{Gamma}\left(\frac{d}{2},\lambda\right)$ if $\lambda > 0$, and $-Y \sim \text{Gamma}\left(\frac{d}{2},-\lambda\right)$ if $\lambda < 0$. The moment generating function of Y is $\mathbb{E}[tY] = (1-\lambda t)^{d/2}$. This gives for all $\lambda \neq 0$

$$s_n(\lambda X_{0,0,\infty}) = \frac{1}{\mathbb{E}[Y^n]} \sum_{i=1}^{\infty} \frac{\mathbb{E}\left[Y^{n+i}\right]}{(n+i)!}$$
$$= \frac{1}{\mathbb{E}[Y^n]} \left(\mathbb{E}[\exp(Y)] - \sum\nolimits_{i=0}^{n} \frac{\mathbb{E}\left[Y^i\right]}{i!}\right)$$

$$= \frac{\Gamma(d/2)}{\lambda^n \Gamma(d/2+n)} \left(\frac{1}{(1-\lambda)^{d/2}} - \frac{1}{\Gamma(d/2)} \sum\nolimits_{i=0}^{n} \frac{\lambda^i \Gamma(d/2+i)}{i!} \right)$$

The representation with the hypergeometric function follows from

$$\begin{aligned} \frac{n!}{\mathbb{E}[Y^n]} \sum_{i=1}^{\infty} \frac{\mathbb{E}\left[Y^{n+i}\right]}{(n+i)!} &= \frac{\Gamma(d/2)}{\lambda^n \Gamma(d/2+n)} \sum_{i=1}^{\infty} \frac{\lambda^{n+i} \Gamma(d/2+n+i)}{(n+1)\dots(n+i)\Gamma(d/2)} \\ &= \sum_{i=1}^{\infty} \lambda^i \prod\nolimits_{j=1}^{i} \left(\frac{d/2-1}{n+j} + 1 \right) \\ &= {}_2F_1\left(1, n+\frac{d}{2}; n+1; \lambda\right) - 1 \end{aligned}$$

□

Lemma 17 bounds the scaled remainder of a mixture of distributions in terms of the scaled remainder of the components of the mixture.

Lemma 17. *Consider a continuum of nonnegative random variables X_s on $s \in \mathcal{S}$ with $\mathcal{S}$ a collection with positive measure. The random variables are disjoint in the sense that $\mathbb{P}(X_s X_t = 0) = 1$ if $s \neq t$. For $X = \int_{s\in\mathcal{S}} X_s \,\mathrm{d}s$ and $n = 0, 1, 2, \dots$*

$$\min_{s\in\mathcal{S}} s_n(-X_s) \leq s_n(-X) < 0 < s_n(X) \leq \max_{s\in\mathcal{S}} s_n(X_s)$$

The inequalities are all strict if and only if there exists a set $S \subseteq \mathcal{S}$ with $\int_{s\in S} \mathbb{E}[X_s^n] \,\mathrm{d}s > 0$ that satisfies for all $\tilde{s} \in S$

$$\min_{s\in\mathcal{S}} s_n(X_s) < s_n(X_{\tilde{s}}) < \max_{s\in\mathcal{S}} s_n(X_s) \tag{6.72}$$

Proof of Lemma 17
Because $\mathbb{E}\left[X_s^i X_t^j\right] = 0$ for $s \neq t$ and $i, j \geq 1$, it follows that $\mathbb{E}\left[(X_s + X_t)^k\right] = \mathbb{E}\left[X_s^k\right] + \mathbb{E}\left[X_t^k\right]$. This means $\mathbb{E}[r_n(X_s + X_t)] = \mathbb{E}[r_n(X_s)] + \mathbb{E}[r_n(X_t)]$. Extending this result to the continuum $\mathcal{S}$, and using (6.31) and $s_n(X_s) > 0$ (see Lemma 14(*iii*))

$$\begin{aligned} 0 < s_n(X) = \frac{\mathbb{E}[r_n(X)]}{\mathbb{E}[X^n]} &= \frac{1}{\mathbb{E}[X^n]} \mathbb{E}\left[r_n\left(\int_{s\in\mathcal{S}} X_s \,\mathrm{d}s \right) \right] \\ &= \frac{1}{\mathbb{E}[X^n]} \int_{s\in\mathcal{S}} \mathbb{E}[r_n(X_s)] \,\mathrm{d}s = \int_{s\in\mathcal{S}} \frac{\mathbb{E}[X_s^n]}{\mathbb{E}[X^n]} \frac{\mathbb{E}[r_n(X_s)]}{\mathbb{E}[X_s^n]} \,\mathrm{d}s \end{aligned}$$

$$= \int_{s\in\mathcal{S}} \frac{\mathbb{E}[X_s^n]}{\mathbb{E}[X^n]} s_n(X_s)\,\mathrm{d}s \leq \max_{s\in\mathcal{S}} s_n(X_s)$$

The latter inequality holds strictly if and only if there exists a set $S \subseteq \mathcal{S}$ with positive measure $s_n(X_{\tilde{s}}) < \max_{s\in\mathcal{S}} s_n(X_s)$ for all $\tilde{s} \in S$. By $s_n(-X_s) < 0$ (see Lemma 14(*iii*))

$$0 > s_n(-X) = \frac{\mathbb{E}[r_n(-X)]}{\mathbb{E}[(-X)^n]} = \int_{s\in\mathcal{S}} \frac{\mathbb{E}[(-X_s)^n]}{\mathbb{E}[(-X)^n]} s_n\left(-X_s\right)\,\mathrm{d}s \geq \min_{s\in\mathcal{S}} s_n\left(-X_s\right)$$

The statement on strictness of the inequality is now straightforward. □

Lemma 18 enables us to find a simple bound of a standard optimization problem.

Lemma 18. *Consider the optimization problem*

$$\max_{\mathbf{x}} \|\mathbf{x}\| \tag{6.73}$$

subject to $\mathbf{x} \geq \mathbf{a}$ *and* $\mathbf{c}'\mathbf{x} \geq 0$ *where* $\mathbf{c} < \mathbf{0}$ *and* $\mathbf{a} \leq \mathbf{0}$ *are d-dimensional vectors.*

Define for $i = 1, \ldots, d$ *the d-dimensional vector* $\mathbf{x}^{(i)}$ *by*

$$x_k^{(i)} = \begin{cases} -\frac{1}{c_i}\sum_{j\neq i} a_j c_j & k = i \\ a_k & k \neq i \end{cases} \tag{6.74}$$

There exists an optimal $\mathbf{x}$ *that satisfies* $\mathbf{c}'\mathbf{x} = 0$*, and exactly one out of the d constraints in* $\mathbf{x} \geq \mathbf{a}$ *is not binding. The maximal value of* (6.73) *equals*

$$\max_{i=1\ldots d} \|\mathbf{x}^{(i)}\| = \max_i \sqrt{\left(-\frac{1}{c_i}\sum_{j\neq i} a_j c_j\right)^2 + \sum_{j\neq i} a_j^2}$$

Proof of Lemma 18

We first show that there exists an optimum where exactly one constraint in $\mathbf{x} \geq \mathbf{a}$ is not binding. The solution of the optimization problem (6.73) is identical with the convex objective function $\|\mathbf{x}\|^2 = \sum_k x_k^2$. By the convexity of this objective function, $\mathbf{x}$ cannot be an optimum if multiple constraints in $\mathbf{x} \geq \mathbf{a}$ are not binding.

Suppose all constraints $\mathbf{x} \geq \mathbf{a}$ are binding, i.e., $\mathbf{x} = \mathbf{a}$. This $\mathbf{x}$ is not the (unique) optimum. To see this, consider some i with $a_i c_i$ below the mean

$a_j c_j$: $a_i c_i \le \frac{1}{d}\mathbf{a}'\mathbf{c}$, and define $\mathbf{x}^{(i)}$ as in (6.74). Obviously, $\mathbf{c}'\mathbf{x}^{(i)} = 0$, and exactly one constraint in $\mathbf{x}^{(i)} \ge \mathbf{a}$ is not binding, namely $x_i^{(i)} > 0 \ge a_i$. It follows from

$$\sum_{j \ne i} a_j c_j = -a_i c_i + \mathbf{a}'\mathbf{c} \ge \left(-\frac{1}{d} + 1\right)\mathbf{a}'\mathbf{c} \ge (d-1)a_i c_i$$

and $-1/c_i > 0$, that

$$\|\mathbf{x}^{(i)}\|^2 = \sum_k \left(x_k^{(i)}\right)^2 \ge (d-1)^2 a_i^2 + \sum_{j \ne i} a_j^2 \ge \|\mathbf{a}\|^2$$

This excludes $\mathbf{x} = \mathbf{a}$ as the (unique) optimum. The bound is strict for $d > 2$, the case of our interest.

We stress that the optimum $\mathbf{x}^{(i)}$ does not necessarily satisfy $a_i c_i \le \frac{1}{d}\mathbf{a}'\mathbf{c}$. For instance, suppose $\mathbf{c} = (-4, -1, -1)'$ and $\mathbf{a} = (-0.1, -1, -1)'$. The maximum is at $\mathbf{x}^{(2)}$ and $\mathbf{x}^{(3)}$, while $a_2 c_2 = a_3 c_3 > \frac{1}{3}\mathbf{a}'\mathbf{c}$. Still, $a_1 c_1 \le \frac{1}{3}\mathbf{a}'\mathbf{c}$ and $\|\mathbf{x}^{(1)}\| > \|\mathbf{a}\|$ hold, which serves our purpose to show that $\mathbf{x} = \mathbf{a}$ is not optimal.

The constraint $\mathbf{c}'\mathbf{x} \ge 0$ binds at the optimal $\mathbf{x}$. If $\mathbf{c}'\mathbf{x} > 0$, we can further increase the unique x_i with $x_i > 0 > a_i$, since $c_i < 0$. This implies that the optimal $\mathbf{x}$ is one of the d vectors $\mathbf{x}^{(i)}$. Collecting previous results gives the lemma.

□

The preceding lemmas enable us to prove the theorem.

Proof of Theorem 8

The random variable X has the following density in polar coordinates

$$f\left(\frac{1}{2}\mathbf{s}'\mathbf{B}\mathbf{s}r^2\right) = \frac{1}{\mathbb{P}(\mathbf{a} \le Z \le \mathbf{b})} \frac{1}{|S_d|} \frac{1_{\mathbf{a} \le r\mathbf{s} \le \mathbf{b}}}{\sqrt{2\pi}} r^{d-1} \exp\left(-\frac{1}{2}\|r\mathbf{s} - \mu\|^2\right) \tag{6.75}$$

where $|S_d|$ is the surface of the d-dimensional sphere S_d, $\mathbf{s} \in S_d$, $r \ge 0$ is the distance to the origin $\tilde{\nu} = \mathbf{0}$ where $Z = \mathbf{0}$, and $1/\mathbb{P}(\mathbf{a} \le Z \le \mathbf{b})$ is a proportionality constant. For example, in the two dimensional example in the right panel in Figure 6.1 on p.130, we have $r = \|\tilde{\mathbf{x}}\|$ as the distance to the origin, $\mathbf{s} = \tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|$ as the direction from the origin to $\tilde{\mathbf{x}}$, and $\|r\mathbf{s} - \mu\| = \|\tilde{\mathbf{x}} - \tilde{\mu}\|$ as the distance from $\tilde{\mathbf{x}}$ to $\tilde{\mu}$ which determines the probability mass at $\tilde{\mathbf{x}}$ under the new measure. Using $\|\mathbf{s}\| = 1$,

$$\|r\mathbf{s} - \mu\|^2 = \|r\mathbf{s}\|^2 + \|\mu\|^2 - 2r\mu'\mathbf{s}$$

$$= r^2 + \|\mu\|^2 - 2r\mu'\mathbf{s}$$
$$= \left(r - \mu'\mathbf{s}\right)^2 - (\mu'\mathbf{s})^2 + \|\mu\|^2 \qquad (6.76)$$

Let $\lambda(\mathbf{s}) = \mathbf{s}'\mathbf{B}\mathbf{s}$, and $u(\mathbf{s}) \geq 0$ is the maximal distance from the origin to the boundaries of the region enclosed by $\mathbf{a}$ and $\mathbf{b}$. Since $\mathbf{a} \leq 0\mathbf{s} \leq \mathbf{b}$, the density of X along each $\mathbf{s}$ is proportional to

$$f\left(\frac{1}{2}\lambda(\mathbf{s})r^2; \mathbf{s}\right) \propto 1_{0\leq r\leq u(\mathbf{s})} \exp\left(-\frac{1}{2}\left[\left(r - \mu'\mathbf{s}\right)^2 - (\mu'\mathbf{s})^2 + \|\mu\|^2\right]\right)$$

Therefore, $X \sim \frac{1}{2}\lambda(S)R_S^2$ with S uniform on S_d, and by (6.75) and (6.76) the probability density function of $R_\mathbf{s}$ is proportional to

$$r^{d-1} \exp\left(-\frac{1}{2}\left(r - \mu'\mathbf{s}\right)^2\right) 1_{0\leq r\leq u(\mathbf{s})} \qquad (6.77)$$

Using the notation introduced in Lemma 16, let the random variable $X_\mathbf{s} \sim \frac{1}{2}\lambda(\mathbf{s})R_\mathbf{s}^2 = \frac{1}{2}\lambda(\mathbf{s})X_{\xi(\mathbf{s}),0,u(\mathbf{s})}$ represent the conditional distribution $X|_{S=\mathbf{s}}$. Then,

$$X^- = \max(-X, 0) = \frac{1}{|S_d|}\int_{s\in\mathcal{S}^-} -X_s \mathrm{d}\mathbf{s} \qquad \mathcal{S}^- = \{\mathbf{s} \in S_d : \lambda(\mathbf{s}) < 0\}$$
$$X^+ = \max(X, 0) = \frac{1}{|S_d|}\int_{s\in\mathcal{S}^+} X_s \mathrm{d}\mathbf{s} \qquad \mathcal{S}^+ = \{\mathbf{s} \in S_d : \lambda(\mathbf{s}) \geq 0\}$$

Applying Lemma 17 to the nonnegative mixtures X^- and X^+,

$$\min_{\mathbf{s}\in\mathcal{S}^-} s_n(X_\mathbf{s}) \leq s_n(-X^-) < 0 < s_n(X^+) \leq \max_{\mathbf{s}\in\mathcal{S}^+} s_n(X_\mathbf{s}). \qquad (6.78)$$

We neglect unit vectors $\mathbf{s} \in S_d$ with $u(\mathbf{s}) = 0$ since the probability mass along such vectors degenerates at zero in (6.77), or equivalently, $X_\mathbf{s} \equiv 0$. To identify such vectors, note that $u(\mathbf{s}) = 0$ corresponds to an origin at the boundary of the region in $[\mathbf{a}, \mathbf{b}]$ and $\mathbf{s}$ pointing outward this region. Therefore, such $\mathbf{s}$ satisfy (i) $s_i < 0$ for some i with $a_i = 0$, or (ii) $s_i > 0$ for some i with $b_i = 0$. Neglecting such vectors gives the following bounds on $\lambda(\mathbf{s})$:

$$\underline{\lambda} = \min_{\|\mathbf{s}\|\in\mathcal{S}-} \{\lambda(\mathbf{s}) : s_i \geq 0 \text{ if } a_i = 0\}$$
$$\bar{\lambda} = \max_{\|\mathbf{s}\|\in\mathcal{S}-} \{\lambda(\mathbf{s}) : s_i \leq 0 \text{ if } b_i = 0\}$$

We certainly obtain for $\underline{\lambda}$ and $\bar{\lambda}$ the minimal and maximal eigenvalue of $\mathbf{B}$ if the origin is in the interior of the region. That is, when $\mathbf{a}$ and $\mathbf{b}$ are strictly negative and strictly positive, respectively.

By Lemma 14(*ii*), $\underline{\lambda} \leq 0$ provides a (negative) lower bound on $s_n(X_\mathbf{s})$ for each $\mathbf{s} \in \mathcal{S}^-$. Similarly, $\bar{\lambda} \geq 0$ corresponds to an upper bound on $\max_{\mathbf{s}\in\mathcal{S}^+} s_n(X_\mathbf{s})$ for each $\mathbf{s} \in \mathcal{S}^+$.

Note that (6.77) is similar to (6.71) on p.174 with mean $\xi = \mu'\mathbf{s}$. Lemma 16(*iii*) indicates that upper bounds on $\xi(\mathbf{s}) := \mu'\mathbf{s}$ and $u(\mathbf{s})$ bound each $s_n(X_\mathbf{s})$. Depending on $\xi(\mathbf{s})$, the bounds on $s_n(X_\mathbf{s})$ are of two different types:

(*i*) $\xi(\mathbf{s}) < 0$: Using $\xi(\mathbf{s}) < 0$ and $u(\mathbf{s}) \leq \infty$:

$$s_n\left(\underline{\lambda} X_{0,0,\infty}\right) \leq s_n(X_\mathbf{s}) \qquad \mathbf{s} \in \mathcal{S}^- \tag{6.79}$$

$$s_n(X_\mathbf{s}) \leq s_n\left(\bar{\lambda} X_{0,0,\infty}\right) \qquad \mathbf{s} \in \mathcal{S}^+ \tag{6.80}$$

(*ii*) $\xi(\mathbf{s}) > 0$: The mean $\xi(\mathbf{s}) = \mu'\mathbf{s}$ of $R_\mathbf{s}$ is maximal at S_d for $\mathbf{s} = \mu/\|\mu\|$. Thus, $\xi(\mathbf{s}) \leq \|\mu\|$. The function $u(\mathbf{s})$ measures the distance along $\mathbf{s}$ from the origin to the boundaries of the integration region T determined by $\mathbf{a}$ and $\mathbf{b}$. There exists at least one $s_i < 0$, because $\mu \leq 0$, and by assumption $\xi(\mathbf{s}) = \mu'\mathbf{s} > 0$. This implies $u(\mathbf{s}) < \infty$ since all elements in $\mathbf{a}$ are finite. The parameter $\bar{u}$ is an upper bound on $u(\mathbf{s})$ for all $\mathbf{s}$ that satisfy $\mu'\mathbf{s} \geq 0$. Ideally, it equals

$$\sup_{\mathbf{a}\leq\mathbf{x}\leq\mathbf{b},\mu'\mathbf{x}\geq 0} \|\mathbf{x}\| < \infty$$

A simple upper bound on this optimization problem follows by dropping the upper bound $\mathbf{b}$ on $\mathbf{x}$. The closed-form solution of Lemma 18 implies the upper bound (6.37). We have for $\xi(\mathbf{s}) > 0$

$$s_n\left(\underline{\lambda} X_{\|\mu\|,0,\bar{u}}\right) \leq s_n(X_\mathbf{s}) \qquad \mathbf{s} \in \mathcal{S}^- \tag{6.81}$$

$$s_n(X_\mathbf{s}) \leq s_n\left(\bar{\lambda} X_{\|\mu\|,0,\bar{u}}\right) \qquad \mathbf{s} \in \mathcal{S}^+ \tag{6.82}$$

Combining (6.79)–(6.82) yields

$$\min\left(s_n\left(\underline{\lambda} X_{0,0,\infty}\right), s_n\left(\underline{\lambda} X_{\|\mu\|,0,\bar{u}}\right)\right) \leq \min_{\mathbf{s}\in\mathcal{S}^-} s_n(X_\mathbf{s})$$

$$\max_{\mathbf{s}\in\mathcal{S}^+} s_n(X_\mathbf{s}) \leq \max\left(s_n\left(\bar{\lambda} X_{0,0,\infty}\right), s_n\left(\bar{\lambda} X_{\|\mu\|,0,\bar{u}}\right)\right)$$

Applying these inequalities with the lower bound $t_n(\underline{m}_{n+1}^+)$ on $s_n(X^+)$ from Lemma 14(*iii*) to (6.78) gives the bounds in Theorem 8. Lemma 14(*ii*) gives the properties of s_n in Theorem 8(*iv*). In the statement of the Theorem, we removed the subscript $l = 0$ from $X_{\xi,l,u}$ for convenience.

□

6.A.4 Proofs Section 6.5

Proof of Lemma 9

We find $\tilde{\mu}_i$ such that $Z_i = Z_0 + \tilde{\mu}_i \sim N(\tilde{\mu}_i, 1)$,

$$\begin{aligned}\mathbb{E}[U_i] &= \mathbb{E}\left[\tilde{Z}_i 1_{\tilde{a}_i \le \tilde{Z}_i \le \tilde{b}_i}\right] \\ &= \mathbb{E}\left[(Z_0 + \tilde{\mu}_i) 1_{\tilde{a}_i - \tilde{\mu}_i \le Z_0 \le \tilde{b}_i - \tilde{\mu}_i}\right] \\ &= \phi(\tilde{a}_i - \tilde{\mu}_i) - \phi(\tilde{b}_i - \tilde{\mu}_i) + \left[\Phi(\tilde{b}_i - \tilde{\mu}_i) - \Phi(\tilde{a}_i - \tilde{\mu}_i)\right] \tilde{\mu}_i \\ &= \phi\left(\frac{a_i - \mu_i}{D_{ii}}\right) - \phi\left(\frac{b_i - \mu_i}{D_{ii}}\right) + \left[\Phi\left(\frac{b_i - \mu_i}{D_{ii}}\right) - \Phi\left(\frac{a_i - \mu_i}{D_{ii}}\right)\right] \tilde{\mu}_i \qquad (6.83) \\ &= \phi\left(\hat{a}_i - \hat{\mu}_i\right) - \phi\left(\hat{b}_i - \hat{\mu}_i\right) + \left[\Phi\left(\hat{b}_i - \hat{\mu}_i\right) - \Phi\left(\hat{a}_i - \hat{\mu}_i\right)\right] \tilde{\mu}_i \qquad (6.84) \\ &= 0\end{aligned}$$

As before, $\hat{\mathbf{x}} = \mathbf{D}^{-1}\mathbf{x}$ and $\tilde{\mathbf{x}} = \mathbf{D}^{-1}(\mathbf{x} - \nu)$ for any vector $\mathbf{x}$. Using (6.8), we may find

$$\hat{\mu} = \mathbf{D}^{-1}\mu = \left(\mathbf{D}^{-1} - \mathbf{D}\boldsymbol{\Sigma}^{-1}\right)\nu = \mathbf{B}\mathbf{D}^{-1}\nu = \mathbf{B}\hat{\nu} \qquad (6.85)$$

and

$$\tilde{\mu} = \mathbf{D}^{-1}(\mu - \nu) = (\mathbf{B} - \mathbf{I})\hat{\nu} \qquad (6.86)$$

Substituting (6.85) and (6.86) into (6.84) leads to

$$\frac{\phi_i\left(\hat{\mathbf{a}} - \mathbf{B}\hat{\nu}\right) - \phi_i\left(\hat{\mathbf{b}} - \mathbf{B}\hat{\nu}\right)}{\Phi_i\left(\hat{\mathbf{b}} - \mathbf{B}\hat{\nu}\right) - \Phi_i\left(\hat{\mathbf{a}} - \mathbf{B}\hat{\nu}\right)} = \mathbf{e}_i(\mathbf{I} - \mathbf{B})\hat{\nu}$$

where $\mathbf{e}_i$ is the vector with one nonzero entry at element i, and that entry is equal to 1. Thus, we want to find the fixed-point $\mathbf{f}(\hat{\nu}) = \hat{\nu}$ with

$$f_i(\hat{\nu}) = \frac{\phi_i\left(\hat{\mathbf{a}} - \mathbf{B}\hat{\nu}\right) - \phi_i\left(\hat{\mathbf{b}} - \mathbf{B}\hat{\nu}\right)}{\Phi_i\left(\hat{\mathbf{b}} - \mathbf{B}\hat{\nu}\right) - \Phi_i\left(\hat{\mathbf{a}} - \mathbf{B}\hat{\nu}\right)} + \mathbf{e}_i\mathbf{B}\hat{\nu} \qquad i = 1, \ldots, d$$

For existence and uniqueness, it suffices by Banach fixed-point theorem to show that $\mathbf{f}$ is a contraction mapping. To prove this, notice that any $\hat{\nu} = \mathbf{B}^{-1}\hat{\mu}$, the component f_i equals the conditional expectation,

$$f_i(\mathbf{B}^{-1}\hat{\nu}) = \mathbb{E}\left[Z_0 \,\middle|\, \hat{a}_i - \hat{\mu}_i < Z_0 < \hat{b}_i - \hat{\mu}_i\right] + \hat{\mu}_i$$

$$= \mathbb{E}\left[Z_0 + \hat{\mu}_i \,\middle|\, \hat{a}_i < Z_0 + \hat{\mu}_i < \hat{b}_i\right]$$

This function has an absolute derivative with respect to $\hat{\mu}_i$ that is smaller than one. Since $\hat{\mu} = \mathbf{B}\hat{\nu}$, the absolute derivative of the elements in $\mathbf{f}$ is along any direction $\hat{\nu}$ at most $\max_i |\lambda_i(\mathbf{B})|$:

$$\left|\frac{\mathrm{d}f_i}{\mathrm{d}\hat{\nu}_j}\right| = \left|\frac{\mathrm{d}f_i}{\mathrm{d}\hat{\mu}}' \frac{\mathrm{d}\hat{\mu}}{\mathrm{d}\hat{\nu}_j}\right| = \left|\frac{\mathrm{d}f_i}{\mathrm{d}\hat{\mu}}' \mathbf{B}\mathbf{e}_j\right| \leq \max_i |\lambda_i(\mathbf{B})|$$

This shows that $\mathbf{f}$ is a contraction mapping:

$$\|\mathbf{f}(\hat{\nu}^{(k)}) - \mathbf{f}(\hat{\nu}^{(m)})\| \leq \left\|\hat{\nu}^{(k)} - \hat{\nu}^{(m)}\right\| \max_i |\lambda_i(\mathbf{B})| < \|\hat{\nu}^{(k)} - \hat{\nu}^{(m)}\|$$

For the sequence $\hat{\nu}^{(k)} = \mathbf{f}(\hat{\nu}^{(k-1)})$, the distance $\|\hat{\nu}^{(k)} - \hat{\nu}^*\|$ to the fixed point $\hat{\nu}^*$ decreases each step with at least the factor $\max_i |\lambda_i(\mathbf{B})|$.

□

Proof of Lemma 10

Using symmetry, $a_i = a$, $b_i = \infty$, $D_{ii} = \sqrt{\omega}$, and (6.49), we obtain

$$\tilde{\mu}_1 \mathbf{1} = \frac{\mu_1}{\sqrt{\omega}}\left(\mathbf{I} - \mathbf{B}^{-1}\right)\mathbf{1} = \frac{\mu_1}{\sqrt{\omega}}\left(1 - \frac{1}{\lambda_{\mathbf{1}}(\mathbf{B})}\right)\mathbf{1} = \frac{2\mu_1}{\sqrt{\omega}}\frac{\rho - 1}{\rho d}\mathbf{1}.$$

This implies for (6.83)

$$\phi\left(\frac{a-\mu_1}{\sqrt{\omega}}\right) + \left(1 - \Phi\left(\frac{a-\mu_1}{\sqrt{\omega}}\right)\right)\frac{2\mu_1}{\sqrt{\omega}}\frac{\rho-1}{\rho d} = 0$$

Letting

$$y := \frac{a - \mu_1}{\sqrt{\omega}} \tag{6.87}$$

gives

$$\phi(y) + (1 - \Phi(y))\frac{2}{\sqrt{\omega}}\frac{\rho - 1}{\rho d}\left(a - y\sqrt{\omega}\right) = 0$$

We obtain (6.52) after some rewriting.

Next, we show that for any ρ, a unique y solves (6.52). The left-hand side (LHS) of (6.52) is linearly increasing in y with a constant slope of one. On the right-hand side (RHS) is the inverse Mills' ratio $\phi/(1-\Phi)$ scaled by the scalar $\rho d/[2(\rho - 1)]$. The inverse Mills' ratio monotonically increases, which means that the RHS decreases from 0 to $-\infty$ if the scalar is negative, i.e., if $\rho > 0$. This confirms the uniqueness of the solution y for $\rho > 0$.

If $\rho < 0$, we obtain from $\rho > -1/(d-1)$ that $1 - 1/\rho > d$ such that

$$\frac{\rho d}{2(\rho - 1)} = \frac{d}{2(1 - 1/\rho)} < \frac{1}{2}$$

The derivative on the RHS in (6.52) is then between 0 and 1/2, because the inverse Mills' ratio has derivative between zero and one. This proves that y is unique for $\rho < 0$. Uniqueness of y is trivial for the remaining case $\rho = 0$.

Using (6.87), the inequality $\mu_1 > a$ is equivalent to $y < 0$ which means that the LHS of (6.52) exceeds the RHS at $y = 0$. Equivalently,

$$-\frac{a}{\sqrt{\omega}} > \frac{\rho d}{2(\rho - 1)} \frac{\phi(0)}{1/2}$$

$$a < -\frac{\rho d}{\rho - 1}\sqrt{\frac{\omega}{2\pi}} = \rho d \sqrt{\frac{1 + (d-1)\rho}{\pi\,(1-\rho)\,(2 + (d-2)\rho)}}.$$

□

Samenvatting (Summary in Dutch)

Sinds het faillissement van Lehman Brothers in september 2008 is er veel aandacht voor het gevaar van systeemrisico in de financiële sector. Dit heeft geleid tot een aantal hervormingen in de financiële sector. Een zichtbaar resultaat is de invoering van het zogenaamde Single Supervisory Mechanism in november 2014. Hierdoor is de Europese Centrale Bank nu de belangrijkste toezichthouder op het bankwezen binnen de eurozone. Dit grensoverschrijdende toezicht ondervangt het probleem van verminderd nationaal toezicht op buitenlandse takken binnen de eurozone en is daarmee een belangrijke stap in de vorming van een Europese bankenunie.

Het derde financiële steunpakket voor Griekenland is een andere recente gebeurtenis waarbij systeemrisico een prominente rol speelde. De hervormingen die aan dit pakket zijn verbonden lijken sterk op de voorstellen die slechts een week voor de overeenkomst van dit steunpakket waren afgewezen in een Grieks referendum. Met de instemming voor het steunpakket negeerden de Griekse beleidsmakers niet alleen de uitslag van het referendum, maar schonden ze ook verkiezingsbeloftes bij de verkiezingen van nog geen zes maanden eerder. Wanbetaling, of erger, Griekenland uit de eurozone (een Grexit) kan echter leiden tot een onvoorspelbare kettingreactie, met name in de Griekse economie. Zo kan een plotselinge herinvoering van een eigen munt resulteren in een nieuwe uitstroom van kapitaal, een sterke depreciatie van de nieuwe munt, bankruns, rechtszaken over contractuele verplichtingen en andere systeemrelevante problemen. In een uiterst geval kan het zelfs leiden tot het uiteenvallen van de eurozone.

De afgelopen jaren daalde echter het grensoverschrijdende risico van een eventueel faillissement van Griekenland. Toen de eerste problemen in Griekenland aan het licht kwamen in 2009, streefde de Trojka (EC, ECB en het IMF) er nog naar om een Grexit hoe dan ook te voorkomen. Belangrijk daarbij was de eerdere systeemcrisis na het faillissement van Lehman Brothers en de volatiliteit op financiële markten die duidde op een sterke nervositeit onder beleggers. Sindsdien zijn buitenlandse investeringen in Griekenland sterk afgenomen en zijn het Griekse bankwezen en de Trojka de belangrijkste investeerders in Griekse staatsobligaties. Dit maakt het grensoverschrijdende effect van een Griekse wanbetaling een stuk beter voorspelbaar. Buiten Griekenland reageerden de financiële markten dan ook gelaten op het politieke tumult rondom het derde steunpakket. Zo steeg de rente op Portugese tienjaars staatsobligaties naar 3,2%, wat zeer beperkt is in vergelijking met de 17,3% in januari 2012. Daarom heerst nu een consensus dat het systeemrisico na een Grieks faillissement vooral lokale gevolgen heeft. De Grieken stonden hierdoor met de rug tegen de muur tijdens de onderhandelingen.

Hoe belangrijk systeemrisico ook is, toch bestaat er geen eenduidige definitie van systeemrisico. Wel is duidelijk dat het gaat om het risico op de ineenstorting van een geheel systeem. Om zulke risico's te beoordelen is een diepgaande analyse nodig van vele kenmerken van het systeem. Uiteraard duiden meerdere zwakke componenten met onderling sterke verbindingen op een zeer kwetsbaar systeem. Het is echter vaak minder eenvoudig om van te voren te bepalen hoe sterk een verbinding tussen componenten *tijdens* een sterke schok is. Verder is moeilijk om vooraf in te schatten wat een extra component zal doen met de kwetsbaarheid van het systeem. Leidt bijvoorbeeld de toetreding van een land tot de eurozone tot een stabielere economie in de gehele eurozone? Zou het opheffen van de handelsboycot tussen Rusland en een groot deel van Europa het gebied als geheel economisch stabieler maken? Bij het beoordelen van zulke kwesties is het van belang om het besmettingsgevaar binnen het systeem zo goed mogelijk in kaart te brengen. Zo'n analyse beperkt zich niet tot het financieel-economische systeem. Het is ook toepasbaar op het inschatten van de sociaal-economische gevolgen van een cyberaanval, een stroomuitval en een uitbraak van een virus als ebola.

Systeemrisico is niet hetzelfde als *systematisch* risico. In financiële portefeuilletheorie verwijst systematisch risico naar het risico dat niet kan worden geëlimineerd door diversificatie met effecten zoals aandelen en obligaties. Aandelen hebben bijvoorbeeld een gemeenschappelijke blootstelling aan terreuraanslagen, natuurrampen en andere rampen. Tenzij een belegger een tegengestelde positie inneemt met derivaten is systematisch risico niet volledig te elimineren. Systematisch risico gaat dus over de gemeenschappelijke blootstelling van afzonderlijke componenten aan externe schokken. Dus al kan systematisch risico tot een systeemcrisis leiden, het is slechts één mogelijke vorm van systeemrisico.

Dit proefschrift gaat over systeemrisico in de financiële sector. Specifieker gezegd gaat het in op meerdere aspecten van het ontstaan van systeemrisico in de financiële sector en de besmettingseffecten daarvan. Het is daarbij een bouwsteen voor een verdere analyse van het effect van systeemrisico op de reële economie. Om systeemrisico vast te stellen in de financiële sector *meten* we systeemrisico in Europese aandelen (hoofdstuk 2) en aandelen in de Verenigde Staten (hoofdstuk 3) met twee verschillende methodes. Na meting van een aanzienlijke mate van systeemrisico in de financiële sector, beschouwen we in hoofdstuk 2 de *gevolgen* van systeemrisico op de prijsvorming van aandelen. Daarnaast identificeren we *determinanten* van systeemrisico, zowel empirisch (hoofdstuk 3) als theoretisch (hoofdstuk 4). Hoofdstuk 5 beschouwt een eerlijke *verdeling* van systeemrisico over de onderliggende componenten, dus banken in een bancaire context. Tot slot presenteren we in hoofdstuk 6 een nauwkeuriger methode voor het *berekenen* en *simuleren* van systeemrisico met de multivariate normale verdeling.

De hoofdstukken zijn geordend op basis van de link naar het dagelijks leven. De eerste twee hoofdstukken zijn vooral empirisch georiënteerd, de andere drie hoofdstukken in toenemende mate theoretisch. Hoewel ieder hoofdstuk over systeemrisico gaat, zijn de hoofdstukken ook afzonderlijk te lezen. Er volgt nu een korte samenvatting van ieder hoofdstuk.

Hoofdstuk 2 toont aan dat binnen de Europese financiële sector grote instellingen een significant lager rendement behalen na correctie voor standaard risicofactoren. Dit patroon is in geen enkele andere bedrijfstak in Europa te vinden. Het patroon is te interpreteren als een indicatie voor een latente staatsgarantie. Deze garantie beschermt alleen grote financiële instellingen tegen grote schokken en blijkt vrij sterk te zijn. Zelfs tijdens de schuldencrisis overstemde het grotere individuele faillissementsrisico de twijfels over staatsgaranties. Een andere bevinding is dat de financiële factor een meer marktconforme beprijzing van aandelen geeft dan de momentum factor. De financiële factor verklaart namelijk een aantal anomalieën op grote en waardeaandelen in het Europese Fama en French (2012)-model.

In hoofdstuk 3 definiëren we staartafhankelijkheid op bedrijfstakniveau als maat voor systeemrisico op financiële markten. We meten de staartafhankelijkheid in de 48 Amerikaanse Fama en French (1997) bedrijfstakken door het tellen van simultane extreme negatieve rendementen binnen een bedrijfstak. De vijf industrieën die het hoogst scoren op staartafhankelijkheid zijn het bankwezen, petroleum & aardgas, nutsbedrijven, financiële handel en verzekeringen. Alle andere bedrijfstakken scoren beduidend lager. Een grote marktbeta, een grote marktkapitalisatie en een kleine volatiliteit zijn belangrijke determinanten van staartafhankelijkheid. Daarnaast is staartafhankelijkheid een belangrijke indicator voor staartafhankelijkheid in het volgende jaar.

Hoofdstuk 4 analyseert in een theoretisch model hoe de maatschappelijke welvaart afhangt van het gezamenlijke effect van de balansomvang, het bankkapitaal en de onderlinge afhankelijkheid van vermogenstitels. Het model suggereert dat banken voor minder kapitaal kiezen dan sociaal wenselijk is, terwijl banken voor een te grote balansomvang kiezen als afhankelijkheden klein zijn. Strenger kapitaalbeleid resulteert in grotere banken bij lage faillissementskosten, hoge financieringskosten en hoge belastingtarieven. Voor de maatschappelijke welvaart is niettemin kapitaalbeleid effectiever dan beleid dat zich richt op een kleinere balansomvang. Dit geldt met name als afhankelijkheden laag zijn, dus in tijden van economische voorspoed. Onze resultaten ondersteunen het anticyclische kapitaalbeleid in de voorstellen van Basel III en impliceren een tegengestelde striktheid van monetair en prudentieel beleid.

Hoofdstuk 5 identificeert een eenvoudige voldoende voorwaarde om systeemrisico eerlijk te verdelen op basis van de Shapley waarde. Onze voorwaarde houdt in dat twee risicomaten op bankniveau in Segoviano and Goodhart (2009) en Zhou (2010) een Shapley waarde zijn. De veel gebruikte *MES* in Acharya et al. (2012) en $\Delta CoVaR$ in Adrian and Brunnermeier (2011) zijn echter geen Shapley waarde.

De staarten van de multivariate normale verdeling spelen een belangrijke rol om systeemrisico te simuleren en te berekenen, hetzij direct of indirect. Zo is de t-verdeling een mix van oneindig veel normale verdelingen. Zelfs voor de multivariate normale verdeling gaat bij bestaande methodes veel nauwkeurigheid verloren bij de modellering van de multivariate staarten. Hoofdstuk 6 stelt daarom een nieuwe analytische methode voor om de kansen van de multivariate normale verdeling te begrenzen. Deze methode combineert een Taylor expansie met een verandering van kansmaat. De verkregen grenzen zijn (*i*) strikter dan die van bestaande methodes, (*ii*) bij een aanzienlijk deel van de gesimuleerde cases met een lage dimensie nauwkeuriger dan de simulatiemethode van Genz (1992), en (*iii*) het nauwkeurigst als het conditiegetal van de covariantiematrix bijna één is. De relatieve fout van de grenzen daalt in de staart waardoor onze methode in het bijzonder bruikbaar is in het multivariate staartgebied. Dit is dus belangrijk voor systeemrisico aangezien systeemgebeurtenissen realisaties zijn in dit staartgebied.

Bibliography

Acharya, V. V. (2009). A theory of systemic risk and design of prudential bank regulation. *Journal of Financial Stability*, 5(3):224–255.

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. P. (2012). Measuring systemic risk. *CEPR Discussion Papers*, 8824.

Adrian, T. and Brunnermeier, M. K. (2011). CoVar. *NBER Working Paper*, 17454.

Allen, F., Carletti, E., and Marquez, R. (2014). Deposits and bank capital structure. *Working paper.*

Allen, L. and Rai, A. (1996). Operational efficiency in banking: An international comparison. *Journal of Banking & Finance*, 20(4):655–672.

Ang, A. and Kristensen, D. (2012). Testing conditional factor models. *Journal of Financial Economics*, 106(1):132–156.

Angeloni, I. and Faia, E. (2013). Capital regulation and monetary policy with fragile banks. *Journal of Monetary Economics*, 60(3):311–324.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2):277–97.

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and momentum everywhere. *Journal of Finance*, 68(3):929–985.

Baker, M. and Wurgler, J. (2015). Do strict capital requirements raise the cost of capital? Bank regulation, capital structure, and the low-risk anomaly. *American Economic Review*, 105(5):315–320.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1):3–18.

Barry, T. A., Lepetit, L., and Tarazi, A. (2011). Ownership structure and risk in publicly held and privately owned banks. *Journal of Banking & Finance*, 35(5):1327–1340.

Bekaert, G., Hodrick, R. J., and Zhan, X. (2009). International stock return comovements. *Journal of Finance*, 64(6):2591–2626.

Benoit, S., Colletaz, G., Hurlin, C., and Pérignon, C. (2013). A theoretical and empirical comparison of systemic risk measures. *HEC Paris Research Paper*, FIN-2014-1030.

Bisias, D., Flood, M., Lo, A. W., and Valavanis, S. (2012). A survey of systemic risk analytics. *Annual Review of Financial Economics*, 4(1):255–296.

Bradley, M., Jarrell, G. A., and Kim, E. (1984). On the existence of an optimal capital structure: Theory and evidence. *Journal of Finance*, 39(3):857–878.

Brennan, M. J. and Schwartz, E. S. (1978). Corporate income taxes, valuation, and the problem of optimal capital structure. *Journal of Business*, 51(1):103–114.

Bühler, W. and Prokopczuk, M. (2010). Systemic risk: Is the banking sector special? *SSRN Working Paper*.

Cao, Z. (2014). Multi-CoVar and Shapley value: A systemic risk measure. *Working Paper*.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.

Chan, K. and Chen, N.-F. (1991). Structural and return characteristics of small and large firms. *Journal of Finance*, 46(4):1467–1484.

Chordia, T. and Shivakumar, L. (2002). Momentum, business cycle, and time-varying expected returns. *Journal of Finance*, 57(2):985–1019.

Cifuentes, R., Ferrucci, G., and Shin, H. S. (2005). Liquidity risk and contagion. *Journal of the European Economic Association*, 3(2-3):556–566.

Davies, R. and Tracey, B. (2014). Too big to be efficient? The impact of implicit subsidies on estimates of scale economies for banks. *Journal of Money, Credit and Banking*, 46(s1):219–253.

De Bandt, O. and Hartmann, P. (2000). Systemic risk: A survey. *ECB Working Paper*, 35.

De Nicolò, G., Gamba, A., and Lucchetta, M. (2014). Microprudential regulation in a dynamic model of banking. *Review of Financial Studies*, 27(7):2097–2138.

De Vries, C. G. (2005). The simple economics of bank fragility. *Journal of Banking & Finance*, 29(4):803–825.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51(3):393–414.

Diamond, D. W. and Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91(3):401–419.

Drehmann, M. and Tarashev, N. (2013). Measuring the systemic importance of interconnected banks. *Journal of Financial Intermediation*, 22(4):586–607.

Fama, E. and French, K. (1997). Industry costs of equity. *Journal of Financial Economics*, 43(2):153–193.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, 105(3):457–472.

Fayed, H. and Atiya, A. (2014). An evaluation of the integral of the product of the error function and the normal probability density with application to the bivariate normal integral. *Mathematics of Computation*, 83(285):235–250.

Fratzscher, M. (2002). Financial market integration in Europe: On the effects of EMU on stock markets. *International Journal of Finance & Economics*, 7(3):165–193.

Furfine, C. H. (2003). Interbank exposures: Quantifying the risk of contagion. *Journal of Money, Credit and Banking*, pages 111–128.

Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *Quarterly Journal of Economics*, 127(2):645–700.

Galati, G. and Moessner, R. (2013). Macroprudential policy – a literature review. *Journal of Economic Surveys*, 27(5):846–878.

Gandhi, P. and Lustig, H. (2015). Size anomalies in U.S. bank stock returns. *Journal of Finance*, 70(2):733–768.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1(2):141–149.

Genz, A. (1993). Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics*, 25:400–405.

Genz, A. and Bretz, F. (2009). *Computation of multivariate normal and t probabilities*, volume 195. Springer.

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57(5):1121–1152.

Glasserman, P. and Li, J. (2005). Importance sampling for portfolio credit risk. *Management Science*, 51(11):1643–1656.

Graham, J. R. (1999). Do personal taxes affect corporate financing decisions? *Journal of Public Economics*, 73(2):147–185.

Haldane, A. G. and May, R. M. (2011). Systemic risk in banking ecosystems. *Nature*, 469:351–355.

Han, C.-H. and Wu, C.-T. (2010). Efficient importance sampling for estimating lower tail probabilities under Gaussian and Student's t distributions. *Working paper.*

Hardouvelis, G. A., Malliaropulos, D., and Priestley, R. (2006). EMU and European stock market integration. *Journal of Business*, 79(1):365–392.

Hartmann, P., Straetmans, S., and de Vries, C. G. (2004). Asset market linkages in crisis periods. *Review of Economics and Statistics*, 86(1):313–326.

Hashorva, E. and Hüsler, J. (2003). On multivariate gaussian tails. *Annals of the Institute of Statistical Mathematics*, 55(3):507–522.

Hou, K., Karolyi, G. A., and Kho, B.-C. (2011). What factors drive global stock returns? *Review of Financial Studies*, 24(8):2527–2574.

Huang, X. (1992). Statistics of bivariate extreme values. PhD thesis, Tinbergen Institute.

Huang, X., Zhou, H., and Zhu, H. (2012). Assessing the systemic risk of a heterogeneous portfolio of banks during the recent financial crisis. *Journal of Financial Stability*, 8(3):193–205.

Hughes, J. P. and Mester, L. J. (1993). A quality and risk-adjusted cost function for banks: Evidence on the "too-big-to-fail" doctrine. *Journal of Productivity Analysis*, 4(3):293–315.

Ibragimov, R., Jaffee, D., and Walden, J. (2011). Diversification disasters. *Journal of Financial Economics*, 99(2):333–348.

Ince, O. S. and Porter, R. B. (2006). Individual equity return data from Thomson Datastream: Handle with care! *Journal of Financial Research*, 29(4):463–479.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance*, 45(3):881–898.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Kalai, E. and Samet, D. (1987). On weighted Shapley values. *International Journal of Game Theory*, 16:205–222.

Kaufman, G. G. and Scott, K. E. (2003). What is systemic risk, and do bank regulators retard or contribute to it? *Independent Review*, 7(3):371–391.

Kim, S. J., Moshirian, F., and Wu, E. (2005). Dynamic stock market integration driven by the European Monetary Union: An empirical analysis. *Journal of Banking & Finance*, 29(10):2475–2502.

Kinlaw, W., Kritzman, M., and Turkington, D. (2012). Toward determining systemic importance. *Journal of Portfolio Management*, 38(4):100–111.

Kraus, A. and Litzenberger, R. H. (1973). A state-preference model of optimal financial leverage. *Journal of Finance*, 28(4):911–922.

Kritzman, M., Li, Y., Page, S., and Rigobon, R. (2011). Principal components as a measure of systemic risk. *Journal of Portfolio Management*, 37(4):112–126.

Laeven, L., Ratnovski, L., and Tong, H. (2014). Bank size and systemic risk. *IMF Staff Discussion Note*, 14/04.

Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, 96(2):175–194.

McNeil, A. J., Frey, R., and Embrechts, P. (2010). *Quantitative Risk Management: Concepts, techniques, and tools.* Princeton University Press.

Miao, J. (2005). Optimal capital structure and industry dynamics. *Journal of Finance*, 60(6):2621–2659.

Modigliani, F. and Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *American Economic review*, 48(3):261–297.

Moran, P. A. P. (1983). A new expansion for the multivariate normal distribution. *Australian Journal of Statistics*, 25(2):339–344.

Myerson, R. B. (1980). Conference structures and fair allocation rules. *International Journal of Game Theory*, 9:169–182.

Nomura, N. (2014). Computation of multivariate normal probabilities with polar coordinate systems. *Journal of Statistical Computation and Simulation*, 84(3):491–512.

Pais, A. and Stork, P. A. (2013). Bank size and systemic risk. *European Financial Management*, 19(3):429–451.

Phinikettos, I. and Gandy, A. (2011). Fast computation of high-dimensional multivariate normal probabilities. *Computational Statistics & Data Analysis*, 55(4):1521–1529.

Poon, S.-H., Rockinger, M., and Tawn, J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *Review of Financial Studies*, 17(2):581–610.

Repullo, R. and Suarez, J. (2013). The procyclical effects of bank capital regulation. *Review of Financial Studies*, 26(2):452–490.

Santos, J. A. (2001). Bank capital regulation in contemporary banking theory: A review of the literature. *Financial Markets, Institutions & Instruments*, 10(2):41–84.

Segoviano, M. A. and Goodhart, C. (2009). Banking stability measures. *IMF Discussion Paper*, 627.

Shapley, L. S. (1953). A value for n-person games. In Kuhn, H. and Tucker, A., editors, *Contributions to the Theory of Games*, volume 2, pages 307–317. Princeton University Press.

Strebulaev, I. A. (2007). Do tests of capital structure theory mean what they say? *Journal of Finance*, 62(4):1747–1787.

Tarashev, N. A., Borio, C. E., and Tsatsaronis, K. (2013). Risk attribution using the Shapley value: Methodology and policy applications. *BIS Working Paper*, 308.

Taylor, J. B. (2010). Defining systemic risk operationally. *Board of Trustees of the Leland Stanford Junior University*, pages 33–57.

Titman, S. and Tsyplakov, S. (2007). A dynamic model of optimal capital structure. *Review of Finance*, 11(3):401–451.

Van Dijk, M. A. (2011). Is size dead? A review of the size effect in equity returns. *Journal of Banking & Finance*, 35(12):3263–3274.

Viale, A. M., Kolari, J. W., and Fraser, D. R. (2009). Common risk factors in bank stocks. *Journal of Banking & Finance*, 33(3):464–472.

Wachter, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market volatility? *Journal of Finance*, 68(3):987–1035.

Wagner, W. (2010). Diversification at financial institutions and systemic crises. *Journal of Financial Intermediation*, 19(3):373–386.

Wagner, W. (2011). Systemic liquidation risk and the diversity–diversification trade-off. *Journal of Finance*, 66(4):1141–1175.

Winter, E. (2002). The Shapley value. In Aumann, R. J. and Hart, S., editors, *Handbook of Game Theory with Economic Applications*, volume 3, chapter 53, pages 2025–2054. Elsevier.

Zeng, L., Yong, H. H. A., Treepongkaruna, S., and Faff, R. (2014). Is there a banking risk premium in the US stock market? *Journal of Financial Management, Markets and Institutions*, 2(1):27–42.

Zhou, C. (2010). Are banks too big to fail? Measuring systemic importance of financial institutions. *International Journal of Central Banking*, 6:205–250.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

594 M.P. WOJTOWICZ, *Pricing Credits Derivatives and Credit Securitization*
595 R.S. SAYAG, *Communication and Learning in Decision Making*
596 S.L. BLAUW, *Well-to-do or doing well? Empirical studies of wellbeing and development*
597 T.A. MAKAREWICZ, *Learning to Forecast: Genetic Algorithms and Experiments*
598 P. ROBALO, *Understanding Political Behavior: Essays in Experimental Political Economy*
599 R. ZOUTENBIER, *Work Motivation and Incentives in the Public Sector*
600 M.B.W. KOBUS, *Economic Studies on Public Facility use*
601 R.J.D. POTTER VAN LOON, *Modeling non-standard financial decision making*
602 G. MESTERS, *Essays on Nonlinear Panel Time Series Models*
603 S. GUBINS, *Information Technologies and Travel*
604 D. KOPÁNYI, *Bounded Rationality and Learning in Market Competition*
605 N. MARTYNOVA, *Incentives and Regulation in Banking*
606 D. KARSTANJE, *Unraveling Dimensions: Commodity Futures Curves and Equity Liquidity*
607 T.C.A.P. GOSENS, *The Value of Recreational Areas in Urban Regions*
608 Ł.M. MARĆ, *The Impact of Aid on Total Government Expenditures*
609 C. LI, *Hitchhiking on the Road of Decision Making under Uncertainty*
610 L. ROSENDAHL HUBER, *Entrepreneurship, Teams and Sustainability: a Series of Field Experiments*
611 X. YANG, *Essays on High Frequency Financial Econometrics*
612 A.H. VAN DER WEIJDE, *The Industrial Organization of Transport Markets: Modeling pricing, Investment and Regulation in Rail and Road Networks*
613 H.E. SILVA MONTALVA, *Airport Pricing Policies: Airline Conduct, Price Discrimination, Dynamic Congestion and Network Effects.*
614 C. DIETZ, *Hierarchies, Communication and Restricted Cooperation in Cooperative Games*
615 M.A. ZOICAN, *Financial System Architecture and Intermediation Quality*
616 G. ZHU, *Three Essays in Empirical Corporate Finance*

617 M. PLEUS, *Implementations of Tests on the Exogeneity of Selected Variables and their Performance in Practice*

618 B. VAN LEEUWEN, *Cooperation, Networks and Emotions: Three Essays in Behavioral Economics*

619 A.G. KOPÁNYI-PEUKER, *Endogeneity Matters: Essays on Cooperation and Coordination*

620 X. WANG, *Time Varying Risk Premium and Limited Participation in Financial Markets*

621 L.A. GORNICKA, *Regulating Financial Markets: Costs and Trade-offs*

622 A. KAMM, *Political Actors playing games: Theory and Experiments*

623 S. VAN DEN HAUWE, *Topics in Applied Macroeconometrics*

624 F.U. BRÄUNING, *Interbank Lending Relationships, Financial Crises and Monetary Policy*

625 J.J. DE VRIES, *Estimation of Alonso's Theory of Movements for Commuting*

626 M. POPŁAWSKA, *Essays on Insurance and Health Economics*

627 X. CAI, *Essays in Labor and Product Market Search*

628 L. ZHAO, *Making Real Options Credible: Incomplete Markets, Dynamics, and Model Ambiguity*

629 K. BEL, *Multivariate Extensions to Discrete Choice Modeling*

630 Y. ZENG, *Topics in Trans-boundary River sharing Problems and Economic Theory*

631 M.G. WEBER, *Behavioral Economics and the Public Sector*

632 E. CZIBOR, *Heterogeneity in Response to Incentives: Evidence from Field Data*

633 A. JUODIS, *Essays in Panel Data Modelling*

634 F. ZHOU, *Essays on Mismeasurement and Misallocation on Transition Economies*

635 P. MULLER, *Labor Market Policies and Job Search*

636 N. KETEL, *Empirical Studies in Labor and Education Economics*

637 T.E. YENILMEZ, *Three Essays in International Trade and Development*

638 L.P. DE BRUIJN, *Essays on Forecasting and Latent Values*

639 S. VRIEND, *Profiling, Auditing and Public Policy: Applications in Labor and Health Economics*

640 M.L. ERGUN, *Fat Tails in Financial Markets*

641 T. HOMAR, *Intervention in Systemic Banking Crises*

642 R. LIT, *Time Varying Parameter Models for Discrete Valued Time Series*

643 R.H. KLEIJN, *Essays on Bayesian Model Averaging using Economic Time Series*

Major policy decisions dealing with financial markets can create substantial uncertainty about systemic risk, sometimes with disastrous consequences. The decision to allow Lehman Brothers to fail, for example, led to the financial meltdown in 2008, causing the Great Recession. In a similar vein, recent fears of a breakup of the eurozone have resulted in panic on the financial markets due to concerns about the potential systemic effects of a breakup.
Several essential aspects of systemic risk are still unresolved issues for academics, policy makers, and the public at large. Is systemic risk really a significant factor in the financial industry? What are the determinants of this risk? How should a policy maker deal with systemic risk? Can we trust the outcome of a standard simulation on the likelihood of multiple simultaneous extreme events? This thesis addresses these important issues.

Sander Muns received an MSc in Applied Mathematics (specialization in Probability, Risk and Statistics) from Delft University of Technology (2006), an MSc in Economics & Business (spec. Financial Economics) from the Erasmus University Rotterdam (2008), and an MSc (MPhil) in Economics (spec. Econometrics and Finance) from Tinbergen Institute (2011). Since 2008, he is affiliated to CPB Netherlands Bureau for Economic Policy Analysis. He started his PhD research in 2011 at the Erasmus University Rotterdam and Tinbergen Institute.