Title
Essays in Financial Econometrics, Asset Pricing and Corporate Finance

Author
Pelger, Markus

Publication Date
2015

Peer reviewed|Thesis/dissertation

Essays in Financial Econometrics, Asset Pricing and Corporate Finance

by

Markus Pelger

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Robert M. Anderson, Chair
Professor Martin Lettau
Professor Michael Jansson

Summer 2015

Essays in Financial Econometrics, Asset Pricing and Corporate Finance

Abstract

Essays in Financial Econometrics, Asset Pricing and Corporate Finance

by

Markus Pelger

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Robert M. Anderson, Chair

My dissertation explores how tail risk and systematic risk affects various aspects of risk management and asset pricing. My research contributions are in econometric and statistical theory, in finance theory and empirical data analysis. In Chapter 1 I develop the statistical inferential theory for high-frequency factor modeling. In Chapter 2 I apply these methods in an extensive empirical study. In Chapter 3 I analyze the effect of jumps on asset pricing in arbitrage-free markets. Chapter 4 develops a general structural credit risk model with endogenous default and tail risk and analyzes the incentive effects of contingent capital. Chapter 5 derives various evaluation models for contingent capital with tail risk.

Chapter 1 develops a statistical theory to estimate an unknown factor structure based on financial high-frequency data. I derive a new estimator for the number of factors and derive consistent and asymptotically mixed-normal estimators of the loadings and factors under the assumption of a large number of cross-sectional and high-frequency observations. The estimation approach can separate factors for normal "continuous" and rare jump risk. The estimators for the loadings and factors are based on the principal component analysis of the quadratic covariation matrix. The estimator for the number of factors uses a perturbed eigenvalue ratio statistic. The results are obtained under general conditions, that allow for a very rich class of stochastic processes and for serial and cross-sectional correlation in the idiosyncratic components.

Chapter 2 is an empirical application of my high-frequency factor estimation techniques. Under a large dimensional approximate factor model for asset returns, I use high-frequency data for the S&P 500 firms to estimate the latent continuous and jump factors. I estimate four very persistent continuous systematic factors for 2007 to 2012 and three from 2003 to 2006. These four continuous factors can be approximated very well by a market, an oil, a finance and an electricity portfolio. The value, size and momentum factors play no significant role in explaining these factors. For the time period 2003 to 2006 the finance factor seems to disappear. There exists only one persistent jump factor, namely a market jump factor. Using implied volatilities from option price data, I analyze the systematic factor structure of the volatilities. There is only one persistent market volatility factor, while during the financial

crisis an additional temporary banking volatility factor appears. Based on the estimated factors, I can decompose the leverage effect, i.e. the correlation of the asset return with its volatility, into a systematic and an idiosyncratic component. The negative leverage effect is mainly driven by the systematic component, while it can be non-existent for idiosyncratic risk.

In Chapter 3 I analyze the effect of jumps on asset pricing in arbitrage-free markets and I show that jumps have to come as a surprise in an arbitrage-free market. I model asset prices in the most general sensible form as special semimartingales. This approach allows me to also include jumps in the asset price process. I show that the existence of an equivalent martingale measure, which is essentially equivalent to no-arbitrage, implies that the asset prices cannot exhibit predictable jumps. Hence, in arbitrage-free markets the occurrence and the size of any jump of the asset price cannot be known before it happens. In practical applications it is basically not possible to distinguish between predictable and unpredictable discontinuities in the price process. The empirical literature has typically assumed as an identification condition that there are no predictable jumps. My result shows that this identification condition follows from the existence of an equivalent martingale measure, and hence essentially comes for free in arbitrage-free markets.

Chapter 4 is joint work with Behzad Nouri, Nan Chen and Paul Glasserman. Contingent capital in the form of debt that converts to equity as a bank approaches financial distress offers a potential solution to the problem of banks that are too big to fail. This chapter studies the design of contingent convertible bonds and their incentive effects in a structural model with endogenous default, debt rollover, and tail risk in the form of downward jumps in asset value. We show that once a firm issues contingent convertibles, the shareholders' optimal bankruptcy boundary can be at one of two levels: a lower level with a lower default risk or a higher level at which default precedes conversion. An increase in the firm's total debt load can move the firm from the first regime to the second, a phenomenon we call *debt-induced collapse* because it is accompanied by a sharp drop in equity value. We show that setting the contractual trigger for conversion sufficiently high avoids this hazard. With this condition in place, we investigate the effect of contingent capital and debt maturity on capital structure, debt overhang, and asset substitution. We also calibrate the model to past data on the largest U.S. bank holding companies to see what impact contingent convertible debt might have had under the conditions of the financial crisis.

Chapter 5 develops and compares different modeling approaches for contingent capital with tail risk, debt rollover and endogenous default. In order to apply contingent convertible capital in practice it is desirable to base the conversion on observable market prices that can constantly adjust to new information in contrast to accounting triggers. I show how to use credit spreads and the risk premium of credit default swaps to construct the conversion trigger and to evaluate the contracts under this specification.

To My Parents Ilse and Wilhelm Pelger

Contents

Acknowledgments

There are numerous people who deserve special thanks. Without their help, support, guidance and friendship this dissertation would not exist. They made the five years of my Ph.D. in Berkeley a unique and enjoyable experience. I would not have been able to write this thesis without the tremendous support of my advisors, colleagues, my family and friends.

I would like to express my immense appreciation and thanks to my advisor, Robert M. Anderson. It is difficult to convey in words my gratitude towards him. His invaluable guidance, constant encouragement and endless patience have not only made this dissertation possible, but have also allowed me to grow as a researcher. Bob's advice drawing from his years of experience helped me to develop the ideas which I have turned into my dissertation. He supported every step in my academic career unconditionally. In the uncountable number of meetings, he never lacked advice on any aspect or concern during the Ph.D.

I am deeply indebted to my dissertation committee members Martin Lettau and Michael Jansson for their guidance, support, patience and help. They are all exceptional advisors and their support was above and beyond the call of duty. Martin has shaped my empirical research and provided invaluable support and guidance in the final part of my Ph.D. studies. I thank Michael for his very insightful discussions and sparking my interest in econometrics and for his support for my Ph.D. application to Berkeley which made it possible for me to study at this exceptional university.

I also wish to express a special thanks to Lisa Goldberg for the many helpful discussions and for providing a practitioner's perspective on my research questions. I would like to specifically thank Ulrike Malmendier for always guiding me towards improving the economic motivation behind my research questions. I also want to express my gratitude towards Steve Evans who supported me before and during my whole Ph.D. and directed my research towards credit risk modeling.

In addition to those mentioned above, many Berkeley faculty members have enriched my experience with discussions, guidance, and support. I would like to specially thank Nicolae Gârleanu, Richard Stanton, Johan Walden, Demian Pouzo, Jim Powell, Bryan Graham, Alexei Tchistyi, Dwight M. Jaffee and Martha Olney. I am very indebted to my coauthors Paul Glasserman, Nan Chen, Behzad Nouri and An Chen for their extensive and fruitful cooperation. I also would like to thank Jason Yue Zhu for excellent research assistance with my projects on factor modeling.

Finally, I am grateful to the incredible administrative staff at Berkeley who makes our life as researchers so much easier. Thank you Patrick Allen, Vicky Lee, Joe Sibol, Phil Walz, Emil Schissel, Camille Fernandez and Hong Nguyen.

I deeply appreciate the financial support of the Center for Risk Management Research at UC Berkeley.

I have also had the chance to interact with and learn from fantastic collegues, many of whom are close friends by now. You all made the last five years special in many respects and greatly contributed to the completion of this thesis. I especially want to thank Raymond Leung, Santiago Pereda Fernández, Paolo Zacchia, Kaushik Krishnan, Eric Auerbach, David Schönholzer, Albert Hu, Michael Weber, Matthew Botsch, Farshad Haghpanah, Victoria Vanasco, Pierre Bachas, Moises Yi, Francesco D'Acunto, Samim Ghamami, Johannes Wieland, Slavik Sheremirov, Aniko Öery, Estefania Santacreu-Vasut, Valentina Paredes and in particular Edson Severini. Special thanks also to Tom Zimmermann, Frank Schilbach and Simon Hilpert, who started their research careers together with me as exchange students at Berkeley.

A special thanks goes to my friends from the International House and the track and field activities, who have been a pillar to me all of these years, in ways I cannot express with words. Their friendship, energy, enthusiasm, and view on life have filled these last 5 years with happy moments and have helped me sail through the ups and downs of being a graduate student. Special thanks goes to our trainer Mark Jellison and my training partners Samuel Raymundus Butarbutar, Rea Kolb, Bunsak Pra, Luca Moreschini, Youness Bennani and Roy Bez. I will miss our TAC sessions. I also want to give special thanks to Eugen Solowjow, Gerd Grau, Ryan Olver, Charles Shi, Max Wallot, Kamila Demkova, Han Jin, Josh Renner, Yoon Jung Jeong, Matthias Müller, Darko Cotoras, Jorge Andres Barrios, Gerd Brandstetter, Sunny Mistry, Jessy Jingxue Sheng and Yuka Matsutani with whom I have shared so many nice memories at the International House.

Throughout all these years in Berkeley, I have successfully maintained a parallel virtual life with my amazing group of friends from Germany, who have also been important in this process. Thank you Stefan Lehner, Alexander Rütters, Philipp Hagen, Mareike Mink, Blanka Horvath, Michael Horwath, Lionel Kapff, Sven Hüschemenger, Timm Holtermann, Jürgen and Anke Nietsch and Sebastian Ebert.

I am deeply grateful to Adelina Yanyue Wang who has been a very special person in my life.

I would like to conclude this acknowledgment by thanking my family who stood by my side in good and bad times. First of all, I want to thank my grandparents, who have sacrificed so much for our family and have always been there for me. I only wish that I could have seen them more often. Finally, I want to thank my parents - Ilse and Wilhelm Pelger - for their unconditional love and faithful support, for raising me to believe in myself and pursue my dreams, for everything they taught me about life, and all the time and effort they have invested in me. This thesis is dedicated to you.

Chapter 1

Large-Dimensional Factor Modeling Based on High-Frequency Observations

1.1 Introduction

1.1.1 Motivation and Modeling Framework

Financial economists are now in the fortunate situation of having a huge amount of high-frequency financial data for a large number of assets. Over the past fifteen years the econometric methods to analyze the high-frequency data for a small number of assets has grown exponentially. At the same time the field of large dimensional data analysis has exploded providing us with a variety of tools to analyze a large cross-section of financial assets over a long time horizon. This paper merges these two literatures by developing statistical methods for estimating the systematic pattern in high frequency data for a large cross-section. One of the most popular methods for analyzing large cross-sectional data sets is factor analysis. Some of the most influential economic theories, e.g. the arbitrage pricing theory of Ross (1976) are based on factor models. While there is a well-developed inferential theory for factor models of large dimension with long time horizon and for factor models of small dimension based on high-frequency observations, the inferential theory for large dimensional high-frequency factor models is absent.

This chapter develops the statistical inferential theory for factor models of large dimensions based on high-frequency observations as illustrated in Figure 1.1. Conventional factor analysis requires a long time horizon, while this methodology works with short time horizons, e.g. a week. My approach also allows for non-stationary price processes. If a large cross-section of firms and sufficiently many high-frequency asset prices are available, we can estimate the number of systematic factors and derive consistent and asymptotically mixed-normal estimators of the latent loadings and factors. These results are obtained for very general stochastic processes, namely Itô semimartingales, and allow for weak serial and cross-sectional correlation in the idiosyncratic errors. The estimation approach can separate factors for systematic large sudden movements, so-called jumps factors, from continuous factors.

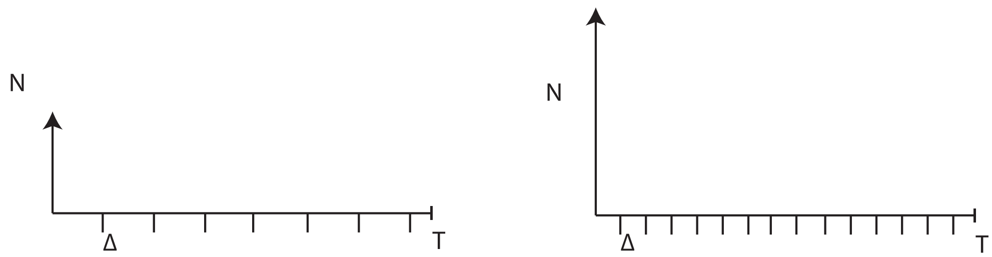


Figure 1.1: The number of cross-sectional observations N goes to infinity and the observed time increments Δ go to zero, while the time horizon T stays constant.

This methodology has many important applications. First, we obtain guidance on how many factors might explain the systematic movements and see how this number changes over short time horizons. Second, we can analyze how loadings and factors change over short time horizons and study their persistence. Third, we can analyze how continuous systematic risk factors, which capture the variation during "normal" times, are different from jump factors, which can explain systematic tail events. Fourth, after identifying the systematic and idiosyncratic components we can apply these two components separately to previous empirical high-frequency studies to see if there is a different effect for systematic versus nonsystematic movements. For example we can examine which components drive the leverage effect. Although most of my motivations are for financial high-frequency data, the methodology can be applied to any problem where we can observe a large number of cross-sectional and high-frequency observations. In Chapter 2 I apply my estimation method to a large high-frequency data set of the S&P500 firms to test these questions empirically.

My estimation approach combines the different fields of high-frequency econometrics and large-dimensional factor analysis. It allows to estimate an unknown factor structure for general continuous-time processes based on high-frequency data and provides an inferential theory. In contrast to conventional factor analysis, which applies principal component analysis to the covariance matrix of the data, I use a spectral decomposition of the quadratic covariation process. Using a truncation approach, the continuous and jump components of the price processes can be separated, and the quadratic covariation matrix for the continuous movements and for the jump movements can be estimated. Then applying principal component analysis to this "jump covariance" and a "continuous risk covariance" matrix separately we can estimate the systematic jump and continuous factors. The number of factors is estimated by analyzing the ratio of perturbed eigenvalues, which is a novel idea to the literature.

For general continuous-time processes neither conventional long-horizon factor analysis nor small dimensional high-frequency factor analysis can be used to analyze large dimensional high-frequency factor models. We cannot apply the results of conventional long-horizon factor analysis to our high-frequency setup for the following reasons: (1) Long-horizon factor analysis is based on the covariance matrix, which cannot be estimated for a fixed short time horizon. (2) After rescaling the increments, we can interpret the quadratic covariation estimator as a sample covariance estimator. However, in contrast to the covariance estimator, the limiting object will be a random variable and the asymptotic distribution results have to be formulated in terms of stable convergence in law, which is stronger than convergence in distribution. (3) Models with jumps have "heavy-tailed rescaled increments" which cannot be accommodated in the relevant long-horizon factor models. (4) In stochastic volatility or stochastic intensity jump models the data is non-stationary. Some of the results in large dimensional factor analysis do not apply to non-stationary data. (5) In contrast to long-horizon factor analysis the asymptotic distribution of my estimators have a mixed Gaussian limit and so will generally have heavier tails than a normal distribution. On the other hand the high-frequency factor analysis of small cross-sections does not allow us to estimate an unknown factor structure. Essentially, the existing results in high-frequency factor analysis extend the framework of classical regression theory to the high-frequency setup, which requires the number of potential factors to be small and to be known. However, our problem is to estimate the unknown factors for which a large cross-section is necessary.

My approach requires only relatively weak assumptions. First, the individual asset price dynamics are modeled as Itô-semimartingales, the most general class of stochastic processes for which the general results of high-frequency econometrics are available. It includes many processes, for example stochastic volatility processes or jump-diffusion processes with stochastic intensity rate. Second, the dependence between the assets is modeled by an approximate factor structure. The idiosyncratic risk can be serially correlated and weakly cross-sectionally correlated and hence allows for a very general specification. The main identification criterion for the systematic risk is that the quadratic covariation of the idiosyncratic risk has bounded eigenvalues, while the quadratic covariation matrix of the systematic factor part has unbounded eigenvalues. For this reason the principal component analysis can relate the eigenvectors of the exploding eigenvalues to the loadings of the factors. Third, in order to separate continuous systematic risk from jump risk, I allow only finite activity jumps, i.e. there are only finitely many jumps in the asset price processes. Many of my results work without this restriction and it is only needed for the separation of these two components. This still allows for a very rich class of models and for example general compound poisson processes with stochastic intensity rates can be accommodated. Fourth, for the asymptotic mixed-normality of my estimators we need some restrictions on the tail-behavior of the idiosyncratic risk. Last but not least, we work under the simultaneous limit of a growing number of high-frequency and cross-sectional observations. I do not restrict the path of how these two parameters go to infinity. However, my results break down if one of the two parameters stays finite. In this sense the "curse of dimensionality" turns into a "blessing".

I extend my model into two directions. First, I include microstructure noise and develop an estimator for the variance of microstructure noise and for the impact of microstructure noise on the spectrum of the factor estimator, allowing us to test if a frequency is sufficiently coarse to neglect the noise. Second, I derive a statistic to determine if the estimated statistical factors can be explained by a set of observed factors. The challenge is that factor models are only identified up to invertible transformations. I provide a new measure for the distance between two sets of factors and develop its asymptotic distribution under the same weak assumptions as for the estimation of the factors. Based on this I develop a new test to determine if a set of estimated statistical factors can be written as a linear combination of observed economic variables.

1.1.2 Related Literature and Contribution

My work builds on the fast growing literatures in the two separate fields of large-dimensional factor analysis and high-frequency econometrics. Bai and Ng (2008) provide a good overview of large dimensional factor analysis. The notion of an "approximate factor model" was introduced by Chamberlain and Rothschild (1983), which allowed for a non-diagonal covariance matrix of the idiosyncratic component. They applied principal component analysis to the population covariance. Connor and Korajczyk (1986, 1988, 1993) study the use of principal component analysis in the case of an unknown covariance matrix, which has to be estimated. The general case of a static large dimensional factor model is treated in Bai (2003). He develops an inferential theory for factor models for a large cross-section and long time horizons based on a principal component analysis of the sample covariance matrix. His paper is the closest to mine from this literature. As pointed out before we cannot map the high-frequency problem into the long horizon model. However, many of my arguments are close to his derivation. Forni, Hallin, Lippi and Reichlin (2000) introduced the dynamic principal component method. Fan, Liao and Mincheva (2013) study an approximate factor structure with sparsity. High-frequency econometrics is a also a relatively young and very fast growing field. An excellent and very up-to-date textbook treatment of high-frequency econometrics is Aït-Sahalia and Jacod (2014). Many of my asymptotic results for the estimation of the quadratic covariation are based on Jacod (2007), where he develops the asymptotic properties of realized power variations and related functionals of semimartingales. In an influential series of papers, Barndorff-Nielsen and Shephard (2004, 2006) and Barndorff-Nielsen, Shephard, and Winkel (2006b) introduce the concept of (bi-) power variation - a simple but effective technique to identify and measure the variation of jumps from intraday data. Aït-Sahalia and Jacod (2009) and Mancini (2004, 2009) introduce a threshold estimator for separating the continuous from the jump variation, which I use in this paper. Todorov and Bollerslev (2010) develop the theoretical framework for high-frequency factor models for a low dimension. Their results are applied empirically in Bollerslev, Li and Todorov (2015).

So far there are relatively few papers combing high-frequency analysis with high-dimensional regimes, but this is an active and growing literature. Important recent papers include Wang and Zou (2010), Tao, Wang and Chen (2013), and Tao, Wang and Zhou (2013) who establish results for large sparse matrices estimated with high-frequency observations. Fan, Furger and Xiu (2014) estimate a large-dimensional covariance matrix with high-frequency data for a given factor structure. My results were derived simultaneously and independently to results in the two papers by Aït-Sahalia and Xiu (2015a+b). Their papers and my work both address the problem of finding structure in high-frequency financial data, but proceed in somewhat different directions and achieve complementary results. In their first paper Aït-Sahalia and Xiu (2015a) develop the inferential theory of principal component analysis applied to a low-dimensional cross-section of high-frequency data. Their work is different from mine as they consider a low-dimensional regime without any factor structure imposed on the data, while I work in a large-dimensional setup which requires the additional structure of a factor model. In addition their analysis focuses on the continuous structure whereas I analyze both the continuous and jump structures. Their second paper Aït-Sahalia and Xiu (2015b) considers a large-dimensional high-frequency factor model and they derive consistent estimators for the factors based on continuous processes. Their second paper essentially extends Fan, Liao and Mincheva's (2013) framework to high-frequency data. Their work is different from mine as their main identification condition is a sparsity assumption on the idiosyncratic covariance matrix. In my paper I also allow for jumps and derive the asymptotic distribution theory of the estimators.

This chapter develops a new estimator for the number of factors that can distinguish between the number of continuous and jump factors and requires only weak assumptions.[1] The most relevant estimators for the number of factors in large-dimensional factor models based on long-horizons are the Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013) estimators.[2] In simulations the performance of the last two estimators seems to dominate the first one[3], but none of the arguments of these two can be transferred to our high-frequency problem without imposing unrealistically strong assumptions on the processes. The Bai and Ng (2002) paper uses an information criterion, while Onatski applies an eigenvalue difference estimator and Ahn and Horenstein an eigenvalue ratio approach. The estimation approach of Onatski and Ahn and Horenstein crucially depends on results from random matrix theory. Hence, they need to make strong assumptions on the underlying stochastic processes for the residuals. The basic idea in all estimation approaches is that the systematic eigenvalues of the estimated covariance matrix or estimated quadratic covariation matrix will explode, while the other eigenvalues of the idiosyncratic part will be bounded. Under additional strong assumptions random matrix theory implies that a certain fraction of the small eigenvalues will be bounded from below and above and the largest residual eigenvalues will cluster, i.e. will be almost the same. Onatksi analyses the difference in eigenvalues. As long as the eigenvalue difference is small, it is likely to be part of the residual spectrum because of the clustering effect. The first time the eigenvalue difference is above a threshold, it indicates the beginning of the systematic spectrum. The Ahn and Horenstein method looks for the maximum in the

[1] Aït-Sahalia and Xiu (2015b) develop simultaneously and independently from me an estimator for the number of factors which is essentially an extension of the Bai and Ng (2002) estimator to high-frequency data. Aït-Sahalia and Xiu's techniques assume continuous processes. I also allow for jumps and my approach can deal with strong and weak factors.

[2] There are many alternative methods, e.g. Hallin and Lisak (2007), Aumengual and Watson (2007), Alessi et al. (2010) or Kapetanious (2010), but in simulations they do not outperform the above methods.

[3] See for example the numerical simulations in Onatski (2010) and Ahn and Horenstein (2013).

eigenvalue ratios. As the smallest systematic eigenvalue is unbounded, while up to a certain index the nonsystematic eigenvalues are bounded from above and below, consistency follows. However, if the first systematic factor is stronger than the other weak systematic factors the Ahn and Horenstein method often fails in simulations with realistic values.[4] In this sense the Onatksi estimator is more appealing as it focusses on the residual spectrum and tries to identify when the spectrum is unlikely to be due to residual terms. I propose a perturbation method. All the eigenvalues are perturbed by adding a particular value. Then, as long as the eigenvalue ratio of the perturbed eigenvalues is close to one, the spectrum is due to the residuals. The perturbation value is chosen such that it dominates the residual eigenvalues but is of a smaller order than the systematic eigenvalues. The perturbed eigenvalues will always be bounded from below and hence we do not need the strong assumptions of random matrix theory. The eigenvalue ratio of perturbed eigenvalues will cluster at 1 due to a rate argument instead of a random matrix theory argument. As we are focussing on the residual spectrum we do not run into the problem of strong versus weak factors. The approach is robust to the choice of the perturbation value. Simulations illustrate the excellent performance of my new estimator.

While my estimation theory is derived under the assumption of synchronous data with negligible microstructure noise, I extend the model to estimate the effect of microstructure noise on the spectrum of the factor estimator. Inference on the volatility of a continuous semimartingale under noise contamination can be pursued using smoothing techniques. Several approaches have been developed, prominent ones by Zhang (2006), Barndorff-Nielsen et al. (2008) and Jacod et al. (2009) in the one-dimensional setting and generalizations for a noisy non-synchronous multi-dimensional setting by Aït-Sahalia et al. (2010), Podolskij and Vetter (2009), Barndorff-Nielsen et al. (2011) and Bibinger and Winkelmann (2014) among others. However, neither the microstructure robust estimators nor the non-synchronicity robust estimators can be easily extended to our large dimensional problem. The main results of my paper assume synchronous data with negligible microstructure noise. Using for example 5-minute sampling frequency as commonly advocated in the literature on realized volatility estimation, e.g. Andersen et al. (2001) and the survey by Hansen and Lunde (2006), seems to justify this assumption and still provides enough high-frequency observations to apply my estimator to a weekly or monthly horizon. In this paper I extend my model to include microstructure noise and develop an estimator for the variance of microstructure noise and for the impact of microstructure noise on the spectrum of the factor estimator, allowing us to test if a frequency is sufficiently coarse to neglect the noise. This novel estimator for the variance of the microstructure noise is also the first to use the information contained in a large cross-section of high-frequency data.

The rest of the chapter is organized as follows. Section 2.2 introduces the factor model. In Section 2.3 I explain my estimators. Section 1.4 summarizes the assumptions and the

[4]Their proposal to demean the data which is essentially the same as projecting out an equally weighted market portfolio does not perform well in simulations with a strong factor. The obvious extension to project out the strong factors does also not really solve the problem as it is unclear how many factors we have to project out.

asymptotic consistency results for the estimators of the factors, loadings and common components. In Subsection 1.4.3 I also deal with the separation into continuous and jump factors. In Section 1.5 I show the asymptotic mixed-normal distribution of the estimators and derive the consistent estimators for the covariance matrices occurring in the limiting distributions. In Section 1.6 I develop the estimator for the number of factors. The extension to microstructure noise is treated in Section 1.7. The test for comparing two sets of factors is presented in Section 1.8. Section 1.9 presents some simulation results. Concluding remarks are provided in Section 2.6. All the proofs are deferred to the appendices.

1.2 Model Setup

1.2.1 Factor Model

Assume the N-dimensional stochastic process $X(t)$ can be explained by a factor model, i.e.

$$X_i(t) = \Lambda_i^\top F(t) + e_i(t) \qquad i = 1, ..., N \text{ and } t \in [0, T]$$

where Λ_i is a $K \times 1$ dimensional vector and $F(t)$ is a K-dimensional stochastic process. The loadings Λ_i describe the exposure to the systematic factors F, while the residuals e_i are stochastic processes that describe the idiosyncratic component. $X(t)$ will typically be the log-price process. However, we only observe the stochastic process X at M discrete time observations in the interval $[0, T]$. If we use an equidistant grid[5], we can define the time increments as $\Delta_M = t_{j+1} - t_j = \frac{T}{M}$ and observe

$$X_i(t_j) = \Lambda_i^\top F(t_j) + e_i(t_j) \qquad i = 1, ..., N \text{ and } j = 1, ..., M$$

or in vector notation

$$X(t_j) = \Lambda F(t_j) + e(t_j) \qquad j = 1, ..., M.$$

with $\Lambda = (\Lambda_1, ..., \Lambda_N)^\top$. In my setup the number of cross-sectional observations N and the number of high-frequency observations M is large, while the time horizon T and the number of systematic factors K is fixed. The loadings Λ, factors F, residuals e and number of factors K are unknown and have to be estimated.

1.2.2 Differences to Long-Horizon Factor Models

For general continuous-time processes the high-frequency factor model cannot be estimated using long-horizon estimation techniques. Conventional factor analysis applies principal component analysis to the covariance matrix of the data. As the sample covariance matrix

[5]Most of my results would go through under a time grid that is not equidistant as long as the largest time increment goes to zero with speed $O\left(\frac{1}{M}\right)$.

$\widehat{Cov}(X) = \frac{1}{T}\sum_{t=1}^{T}(X(t) - \bar{X})^2$ with $\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X(t)$ can only estimate the population covariance matrix consistently for $T \to \infty$, it does not work in our case of a fixed time horizon T. The are two more reasons, why we cannot transform our problem into one that can be solved by a conventional large-dimensional factor model. First, realistic models for financial price data have stochastic volatility and also potentially stochastic jump intensity rates. In both cases the data is non-stationary and a covariance estimator with rescaled increments converges to a random variable. In addition the asymptotic distribution of a the rescaled covariance estimator has a mixed Gaussian limit which typically has heavier tails than a normal distribution. Second, even under the restrictive assumption of independent and stationary increments, the rescaled increments would not satisfy the basic assumptions of the Bai (2003) framework if we allow for jumps. Models with jumps have "heavy-tailed rescaled increments" which violate necessary moment conditions.

A key element of this paper is to replace the covariance matrix with the quadratic covariation matrix. The quadratic covariation matrix of a stochastic process can be estimated consistently for a very general class of stochastic processes with the number of high-frequency observations M going to infinity for a fixed time horizon T. Similar to the covariance, the quadratic covariation is a bilinear form, which will allow us to identify the factors under suitable assumptions.

I denote by $\Delta_j X$ the jth observed increment of the process X, i.e. $\Delta_j X = X(t_{j+1}) - X(t_j)$ and write $\Delta X(t) = X(t) - X(t-)$ for the jumps of the process X. Of course, $\Delta X(t) = 0$ for all $t \in [0, T]$ if the process is continuous. For a very general class of stochastic processes, so-called semimartingales, the sum of squared increments converges to the quadratic covariation for $M \to \infty$:

$$\sum_{j=1}^{M} (\Delta_j X_i)^2 \xrightarrow{p} [X_i, X_i] \qquad \sum_{j=1}^{M} \Delta_j X_i \Delta_j X_k \xrightarrow{p} [X_i, X_k].$$

The predictable quadratic covariation $\langle X_i, X_k \rangle$ is the predictable conditional expectation of $[X_i, X_k]$, i.e. it is the so-called compensator process. It is the same as the realized quadratic covariation $[X_i, X_k]$ for a continuous process, but differs if the processes have jumps. The realized quadratic covariation $[X_i, X_k]_t$ and the conditional quadratic covariation $\langle X_i, X_k \rangle_t$ are themselves stochastic processes. If we leave out the time index t, it means that we are considering the quadratic covariation evaluated at the terminal time T, which is a random variable. For more details see Rogers (2004) or Jacod and Shiryaev (2002).

The simplest case for my factor model assumes that all stochastic processes are Brownian motions and is given by

$$X_T = \begin{pmatrix} \Lambda_{11} & \cdots & \Lambda_{1K} \\ \vdots & \ddots & \vdots \\ \Lambda_{1K} & \cdots & \Lambda_{NK} \end{pmatrix} \begin{pmatrix} W_{F_1}(t) \\ \vdots \\ W_{F_K}(t) \end{pmatrix} + \begin{pmatrix} \sigma_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{NN} \end{pmatrix} \begin{pmatrix} W_{e_1}(t) \\ \vdots \\ W_{e_N}(t) \end{pmatrix}$$

where all Brownian motions W_{F_k} and W_{e_i} are independent of each other. In this case the quadratic covariation equals

$$[X, X] = \Lambda[F, F]\Lambda^\top + [e, e] = \Lambda\Lambda^\top T + \begin{pmatrix} \sigma_{11}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{NN}^2 \end{pmatrix} T$$

Under standard assumptions $\Lambda\Lambda^\top$ is a $N \times N$ matrix of rank K and its eigenvalues will go to infinity for $N \to \infty$. On the other hand $[e, e]$ has bounded eigenvalues. Using a notion that is similar to the approximate factor model of Chamberlain and Rothschild (1983) the systematic factors and loadings can be linked to the exploding eigenvalues of $[X, X]$. The approximate factor model does not require a diagonal structure for the residual matrix $[e, e]$ but only relies on the boundedness of its eigenvalues. Hence, it could accommodate weak serial and cross-sectional dependence in the error terms and factors.

The problem is the estimation of the unobserved quadratic covariation matrix $[X, X]$ for large N. Although, we can estimate each entry of the matrix with a high precision, the estimation errors will sum up to a non negligible quantity if N is large. In the case of a large-dimensional sample covariance matrix Bai (2003) has solved the problem. I apply similar ideas to our setup.

This work is different from Bai's (2003) paper as we combine high-frequency econometrics with large dimensional matrix theory. In the simple case, where all stochastic processes are driven by Brownian motions, we could actually map the problem into the framework of the Bai (2003) paper. If we divide the increments by the square root of the length of the time increments $\Delta_M = T/M$, we end up with a conventional covariance estimator:

$$\sum_{j=1}^{M} (\Delta_j X_i)^2 = \frac{T}{M} \sum_{j=1}^{M} \left(\frac{\Delta_j X_i}{\sqrt{\Delta_M}} \right)^2 \qquad \text{with } \frac{\Delta_j X_i}{\sqrt{\Delta_M}} \sim i.i.d.\ N(0, \Lambda_i \Lambda_i^\top + \sigma_{ii}^2).$$

Hence, for this simple, but from a practical perspective irrelevant example, there is no need for a new statistical theory as it can be mapped into an already existing framework.

However, for general stochastic process we need to develop a new inferential theory for the large dimensional high-frequency factor model. Assume that the underlying stochastic processes have stochastic volatility and jumps. Both are features that are necessary to model asset prices realistically.

$$F(t) = \int_0^t \sigma_F(s) dW_F(s) + \sum_{s \le t} \Delta F(s) \qquad e(t) = \int_0^t \sigma_e(s) dW_e(s) + \sum_{s \le t} \Delta e(s).$$

The quadratic covariation matrices evaluated at time T will now be random variables given by

$$[F, F] = \int_0^T \sigma_F^\top(s) \sigma_F(s) ds + \sum_{s \le T} \Delta F^2(s) \qquad [e, e] = \int_0^T \sigma_e^\top(s) \sigma_e(s) ds + \sum_{s \le T} \Delta e^2(s).$$

The quadratic covariation of X is still given by $[X,X] = \Lambda[F,F]\Lambda^\top + [e,e]$ and under the assumption of bounded eigenvalues for $[e,e]$ and some weak assumptions on the loadings and factors, we could still identify the systematic part through the exploding eigenvalues. Under some weak assumptions the estimator for the quadratic covariation is $\sqrt{M}$ consistent with an asymptotic mixed-Gaussian law:

$$\sqrt{M}\left(\sum_{j=1}^{M}(\Delta_j X_i)^2 - [X_i, X_i]\right) \stackrel{L-s}{\to} N\left(0, 2\int_0^T \sigma_{X_i}^4(s)ds + 4\sum_{s\leq T}\Delta X_i(s)^2\sigma_{X_i}^2(s-)\right).$$

Here the mode of convergence is stable convergence in law, which is stronger than simple convergence in distribution. For more details see Aït-Sahalia and Jacod (2014). Note, that the variance in the normal distribution is a random variable. Stable convergence in law allows us to replace the random variance by a consistent estimator $\hat{\Gamma}_i \stackrel{p}{\to} 2\int_0^T \sigma_{X_i}^4(s)ds + 4\sum_{s\leq T}\Delta X_s^2 \sigma_{X_i}^2(s-)$ and to obtain convergence in distribution for the normalized estimator:

$$\sqrt{M}\frac{\left(\sum_{j=1}^{M}(\Delta_j X_i)^2 - [X_i, X_i]\right)}{\sqrt{\hat{\Gamma}_i}} \stackrel{D}{\to} N(0,1).$$

It should be not surprising that if the asymptotic distribution for the quadratic covariation estimator is different from the asymptotic distribution of the sample covariance matrix, then the estimators for the loadings and factors based on high-frequency observations will be different from the results in the Bai (2003) paper. In particular, the limiting objects will be random variables and the asymptotic distribution results use stable convergence in law. The difference between the long-time horizon factor analysis and high-frequency factor analysis is not only the asymptotic distribution theory, but also the conditions that have to be satisfied by the stochastic processes. If X_i is allowed to have jumps, then it is easy to show that the rescaled increments $\frac{\Delta_j X_i}{\sqrt{\Delta_M}}$ do not have fourth moments. However, Bai (2003) requires the random variables to have at least 8 moments, which is another reason why we cannot simply map the high-frequency problem into the long time horizon framework. Last but not least, from a conceptional point of view my high-frequency estimator is based on path-wise arguments for the stochastic processes, while Bai's estimator is based on population assumptions.

1.3 Estimation Approach

We have M observations of the N-dimensional stochastic process X in the time interval $[0,T]$. For the time increments $\Delta_M = \frac{T}{M} = t_{j+1} - t_j$ we denote the increments of the stochastic processes by

$$X_{j,i} = X_{t_{j+1},i} - X_{t_j,i} \qquad F_j = F_{t_{j+1}} - F_{t_j} \qquad e_{j,i} = e_{t_{j+1},i} - e_{t_j,i}.$$

In matrix notation we have

$$\underset{(M\times N)}{X} = \underset{(M\times K)}{F}\underset{(K\times N)}{\Lambda^\top} + \underset{(M\times N)}{e}.$$

For a given K our goal is to estimate Λ and F. As in any factor model where only X is observed Λ and F are only identified up to K^2 parameters as $F\Lambda^\top = FAA^{-1}\Lambda^\top$ for any arbitrary invertible $K \times K$ matrix A. Hence, for my estimator I impose the K^2 restrictions that $\frac{\hat{\Lambda}^\top\hat{\Lambda}}{N} = I_K$ which gives us $\frac{K(K+1)}{2}$ restrictions and that $\hat{F}^\top\hat{F}$ is a diagonal matrix, which yields another $\frac{K(K-1)}{2}$ restrictions.

Denote the K largest eigenvalues of $\frac{1}{N}X^\top X$ by V_{MN}. The estimator for the loadings $\hat{\Lambda}$ is defined as the K eigenvectors of V_{MN} multiplied by $\sqrt{N}$. The estimator for the factor increments is $\hat{F} = \frac{1}{N}X\hat{\Lambda}$. Note that $\frac{1}{N}X^\top X$ is an estimator for $\frac{1}{N}[X,X]$ for a finite N. We study the asymptotic theory for $M,N \to \infty$. As in Bai (2003) we consider a simultaneous limit which allows (N,M) to increase along all possible paths.

The systematic component of $X(t)$ is the part that is explained by the factors and defined as $C(t) = \Lambda F(t)$. The increments of the systematic component $C_{j,i} = F_j\Lambda_i^\top$ are estimated by $\hat{C}_{j,i} = \hat{F}_j\hat{\Lambda}_i^\top$.

We are also interested in estimating the continuous and jump component of the factors and the volatility of the factors. Denoting by F^C the factors that have a continuous component and by F^D the factor processes that have a jump component, we can write

$$X(t) = \Lambda^C F^C(t) + \Lambda^D F^D(t) + e(t).$$

Note, that for factors that have both, a continuous and a jump component, the corresponding loadings have to coincide. In the following we assume a non-redundant representation of the K^C continuous and K^D jump factors. For example if we have K factors which have all exactly the same jump component but different continuous components, this results in K different total factors and $K^C = K$ different continuous factors, but in only $K^D = 1$ jump factor.

Intuitively under some assumptions we can identify the jumps of the process $X_i(t)$ as the big movements that are larger than a specific threshold. Set the threshold identifier for jumps as $\alpha\Delta_M^{\bar{\omega}}$ for some $\alpha > 0$ and $\bar{\omega} \in \left(0, \frac{1}{2}\right)$ and define $\hat{X}_{j,i}^C = X_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ and $\hat{X}_{j,i}^D = X_{j,i}\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}$. The estimators $\hat{\Lambda}^C$, $\hat{\Lambda}^D$, $\hat{F}^C$ and $\hat{F}^D$ are defined analogously to $\hat{\Lambda}$ and $\hat{F}$, but using $\hat{X}^C$ and $\hat{X}^D$ instead of X. [6]

The quadratic covariation of the factors can be estimated by $\hat{F}^\top\hat{F}$ and the volatility component of the factors by $\hat{F}^{C\top}\hat{F}^C$. I show that the estimated increments of the factors $\hat{F}, \hat{F}^C$ and $\hat{F}^D$ can be used to estimate the quadratic covariation with any other process.

The number of factors can be consistently estimated through the perturbed eigenvalue ratio statistic and hence, we can replace the unknown number K by its estimator $\hat{K}$. Denote the ordered eigenvalues of $X^\top X$ by $\lambda_1 \geq ... \geq \lambda_N$. We choose a slowly increasing sequence

[6]For the jump threshold I recommend the *TOD* specification of Bollerslev, Li and Todorov (2013).

$g(N,M)$ such that $\frac{g(N,M)}{N} \to 0$ and $g(N,M) \to \infty$. Based on simulations a good choice for the perturbation term g is the median eigenvalue rescaled by $\sqrt{N}$. Then, we define perturbed eigenvalues $\hat{\lambda}_k = \lambda_k + g(N,M)$ and the perturbed eigenvalue ratio statistic

$$ER_k = \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}} \qquad \text{for } k = 1, ..., N-1.$$

The estimator for the number of factors is defined as the first time that the perturbed eigenvalue ratio statistic does not cluster around 1 any more:

$$\hat{K}(\gamma) = \max\{k \le N-1 : ER_k > 1+\gamma\} \qquad \text{for } \gamma > 0.$$

If $ER_k < 1+\gamma$ for all k, then set $\hat{K}(c) = 0$. The definition of $\hat{K}^C(\gamma)$ and $\hat{K}^D(\gamma)$ is analogous but using λ_i^C respectively λ_i^D of the matrices $\hat{X}^{C\top}\hat{X}^C$ and $\hat{X}^{D\top}\hat{X}^D$. Based on extensive simulations a constant γ between 1.1 and 1.2 seems to be good choice.

1.4 Consistency Results

1.4.1 Assumptions on Stochastic Processes

All the stochastic processes considered in this paper are locally bounded special Itô semimartingales as defined in Definition A.1 and explained in more detail in Appendix A.2. These particular semimartingales are the most general stochastic processes for which we can develop an asymptotic theory for the estimator of the quadratic covariation. A d-dimensional locally bounded special Itô semimartingale Y can be represented as

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_E \delta(s,x)(\mu - \nu)(ds, dx)$$

where b_s is a locally bounded predictable drift term, σ_s is an adapted cádlág volatility process, W is a d-dimensional Brownian motion and $\int_0^t \int_E \delta(s,x)(\mu-\nu)(ds,dx)$ describes a jump martingale. μ is a Poisson random measure on $\mathbb{R}_+ \times E$ with $(E, \mathbb{E})$ an auxiliary measurable space on the space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\ge 0}, \mathbb{P})$. The predictable compensator (or intensity measure) of μ is $\nu(ds, dx) = ds \times v(dx)$ for some given finite or sigma-finite measure on $(E, \mathbb{E})$. These dynamics are very general and completely non-parametric. They allow for correlation between the volatility and asset price processes. I only impose some week regularity conditions in Definition A.1. The model includes many well-known continuous-time models as special cases: for example stochastic volatility models like the CIR or Heston model, the affine class of models in Duffie, Pan and Singleton (2000), Barndorff-Nielsen and Shephard's (2002) Ornstein-Uhlenbeck stochastic volatility model with jumps or Andersen, Benzoni, and Lund's (2002) stochastic volatility model with log-normal jumps generated by a non-homogenous Poisson process.

1.4.2 Consistency

The key assumption for obtaining a consistent estimator for the loadings and factors is an approximate factor structure. It requires that the factors are systematic in the sense that they cannot be diversified away, while the idiosyncratic residuals are nonsystematic and can be diversified away. The approximate factor structure assumption uses the idea of appropriately bounded eigenvalues of the residual quadratic covariation matrix, which is analogous to Chamberlain and Rothschild (1983) and Chamberlain (1988). Let $\|A\| = (tr(A \top A))^{1/2}$ denote the norm of a matrix A and $\lambda_i(A)$ the i's largest singular value of the matrix A, i.e. the square-root of the i's largest eigenvalue of $A^\top A$. If A is a symmetric matrix then λ_i is simply the i's largest eigenvalue of A.

Assumption 1.1. ***Factor structure assumptions***

1. ***Underlying stochastic processes***
   F and e_i are Itô-semimartingales as defined in Definition A.1
$$F(t) = F(0) + \int_0^t b_F(s)ds + \int_0^t \sigma_F(s)dW_s + \sum_{s \leq t} \Delta F(s)$$
$$e_i(t) = e(0) + \int_0^t b_{e_i}(s)ds + \int_0^t \sigma_{e_i}(s)dW_s + \sum_{s \leq t} \Delta e_i(s)$$
   In addition each e_i is a square integrable martingale.

2. ***Factors and factor loadings***
   The quadratic covariation matrix of the factors Σ_F is positive definite a.s.
$$\sum_{j=1}^{M} F_j F_j^\top \xrightarrow{p} [F, F]_T =: \Sigma_F$$
   and
$$\|\frac{\Lambda^\top \Lambda}{N} - \Sigma_\Lambda\| \to 0.$$
   where the matrix Σ_Λ is also positive definite. The loadings are bounded, i.e. $\|\Lambda_i\| < \infty$ for all $i = 1, ..., N$.

3. ***Independence of F and e***
   The factor process F and the residual processes e are independent.

4. ***Approximate factor structure***
   The largest eigenvalue of the residual quadratic covariation matrix is bounded in probability, i.e.
$$\lambda_1([e, e]) = O_p(1)$$

As the predictable quadratic covariation is absolutely continuous, we can define the instantaneous predictable quadratic covariation as

$$\frac{d\langle e_i, e_k\rangle_t}{dt} = \sigma_{e_{i,k}}(t) + \int \delta_{i,k}(z)v_t(z) =: G_{i,k}(t)$$

We assume the largest eigenvalue of the matrix $G(t)$ is almost surely bounded for all t:

$$\lambda_1(G(t)) < C \qquad \text{a.s. for all } t \text{ for some constant } C.$$

5. ***Identification condition*** *All Eigenvalues of $\Sigma_\Lambda \Sigma_F$ are distinct a.s..*

The most important part of Assumption 1.1 is the approximate factor structure in point 4. It implies that the residual risk can be diversified away. Point 1 states that we can use the very general class of stochastic processes defined in Definition A.1. The assumption that the residuals are martingales and hence do not have a drift term is only necessary for the asymptotic distribution results. The consistency results do not require this assumption. Point 2 implies that the factors affect an infinite number of assets and hence cannot be diversified away. Point 3 can be relaxed to allow for a weak correlation between the factors and residuals. This assumption is only used to derive the asymptotic distribution of the estimators. The approximate factor structure assumption in point 4 puts a restriction on the correlation of the residual terms. It allows for cross-sectional (and also serial) correlation in the residual terms as long as it is not too strong. We can relax the approximate factor structure assumption. Instead of almost sure boundedness of the predictable instantaneous quadratic covariation matrix of the residuals it is sufficient to assume that

$$\frac{1}{N}\sum_{i=1}^N \sum_{k\neq i}^N \Lambda_i G_{i,k}(t)\Lambda_k^\top < C \qquad \text{a.s. for all } t$$

Then, all main results except for Theorem 5 and 8 continue to hold. Under this weaker assumption we do not assume that the diagonal elements of G are almost surely bounded. By Definition A.1 the diagonal elements of G are already locally bounded which is sufficient for most of our results.

Note that point 4 puts restrictions on both the realized and the conditional quadratic covariation matrix. In the case of continuous residual processes, the conditions on the conditional quadratic covariation matrix are obviously sufficient. However, in our more general setup it is not sufficient to restrict only the conditional quadratic covariation matrix.

Assumption 1.2. ***Weak dependence of error terms***
The row sum of the quadratic covariation of the residuals is bounded in probability:

$$\sum_{i=1}^N \|[e_k, e_i]\| = O_p(1) \qquad \forall k = 1, ..., N$$

Assumption 1.2 is stronger than $\lambda_1([e,e]) = O_p(1)$ in Assumption 1.1. As the largest eigenvector of a matrix can be bounded by the largest absolute row sum, Assumption 1.2 implies $\lambda_1([e,e]) = O_p(1)$. If the residuals are cross-sectionally independent it is trivially satisfied. However it allows for a weak correlation between the residual processes. For example, if the residual part of each asset is only correlated with a finite number of residuals of other assets, it will be satisfied.

As pointed out before, the factors F and loadings Λ are not separately identifiable. However, we can estimate them up to an invertible $K \times K$ matrix H. Hence, my estimator $\hat{\Lambda}$ will estimate ΛH and $\hat{F}$ will estimate $FH^{\top^{-1}}$. Note, that the common component is well-identified and $\hat{F}\hat{\Lambda}^\top = \hat{F}H^{\top^{-1}}H^\top\Lambda^\top$. For almost all purposes knowing ΛH or $FH^{\top^{-1}}$ is as good as knowing Λ or F as what is usually of interest is the vector space spanned by the factors. For example testing the significance of F or $FH^{\top^{-1}}$ in a linear regression yields the same results.[7]

In my general approximate factor models we require N and M to go to infinity. The rates of convergence will usually depend on the smaller of these two values denoted by $\delta = \min(N, M)$. As noted before we consider a simultaneous limit for N and M and not a path-wise or sequential limit. Without further assumptions the asymptotic results do not hold for a fixed N or M. In this sense the large dimension of our problem, which makes the analysis more complicated, also helps us to obtain more general results.

Note that F_j is the increment $\Delta_j F$ and goes to zero for $M \to \infty$ for almost all increments. It can be shown that in a specific sense we can also consistently estimate the factor increments, but the asymptotic statements will be formulated in terms of the stochastic process F evaluated at a discrete time point t_j. For example $F_T = \sum_{j=1}^M F_j$ denotes the factor process evaluated at time T. Similarly we can evaluate the process at any other discrete time point $T_m = m \cdot \Delta_M$ as long as $m \cdot \Delta_M$ does not go to zero. Essentially m has to be proportional to M. For example, we could chose T_m equal to $\frac{1}{2}T$ or $\frac{1}{4}T$. The terminal time T can always be replaced by the time T_m in all the theorems. The same holds for the common component.

Theorem 1.1. *Consistency of estimators*
Define the rate $\delta = \min(N, M)$ *and the invertible matrix* $H = \frac{1}{N}\left(F^\top F\right)\left(\Lambda^\top \hat{\Lambda}\right) V_{MN}^{-1}$. *Then the following consistency results hold:*

1. *Consistency of loadings estimator: Under Assumption 1.1 it follows that*

$$\hat{\Lambda}_i - H^\top \Lambda_i = O_p\left(\frac{1}{\sqrt{\delta}}\right).$$

2. *Consistency of factor estimator and common component: Under Assumptions 1.1 and 1.2 it follows that*

$$\hat{F}_T - H^{-1}F_T = O_p\left(\frac{1}{\sqrt{\delta}}\right), \qquad \hat{C}_{T,i} - C_{T,i} = O_p\left(\frac{1}{\sqrt{\delta}}\right).$$

[7]For a more detailed discussion see Bai (2003) and Bai and Ng (2008).

3. *Consistency of quadratic variation: Under Assumptions 1.1 and 1.2 and for any stochastic process $Y(t)$ satisfying Definition A.1 we have for $\frac{\sqrt{M}}{N} \to 0$ and $\delta \to \infty$:*

$$\sum_{j=1}^{M} \hat{F}_j \hat{F}_j^\top = H^{-1}[F,F]_T H^{-1^\top} + o_p(1) , \qquad \sum_{j=1}^{M} \hat{F}_j Y_j = H^{-1}[F,Y]_T + o_p(1)$$

$$\sum_{j=1}^{M} \hat{e}_{j,i}\hat{e}_{j,k} = [e_i, e_k] + o_p(1) , \qquad \sum_{j=1}^{M} \hat{e}_{j,i} Y_j = [e_i, Y] + o_p(1)$$

$$\sum_{j=1}^{M} \hat{C}_{j,i}\hat{C}_{j,k} = [C_i, C_k] + o_p(1) , \qquad \sum_{j=1}^{M} \hat{C}_{j,i} Y_j = [C_i, Y] + o_p(1).$$

for $i, k = 1, ..., N$.

This statement only provides a pointwise convergence of processes evaluated at specific times. A stronger statement would be to show weak convergence for the stochastic processes. However, weak convergence of stochastic processes requires significantly stronger assumptions[8] and will in general not be satisfied under my assumptions.

1.4.3 Separating Continuous and Jump Factors

Using a thresholding approach we can separate the continuous and jump movements in the observable process X and estimate the systematic continuous and jump factors. The idea is that with sufficiently many high-frequency observations, we can identify the jumps in X as the movements that are above a certain threshold. This allows us to separate the quadratic covariation matrix of X into its continuous and jump component. Then applying principal component analysis to each of these two matrices we obtain our separate factors. A crucial assumption is that the thresholding approach can actually identify the jumps:

Assumption 1.3. ***Truncation identification***
F and e_i have only finite activity jumps and factor jumps are not "hidden" by idiosyncratic jumps:

$$\mathbb{P}\left(\Delta X_i(t) = 0 \text{ if } \Delta(\Lambda_i^\top F(t)) \neq 0 \text{ and } \Delta e_i(t) \neq 0\right) = 0.$$

The quadratic covariation matrix of the continuous factors $[F^C, F^C]$ and of the jump factors $[F^D, F^D]$ are each positive definite a.s. and the matrices $\frac{\Lambda^{C^\top}\Lambda^C}{N}$ and $\frac{\Lambda^{D^\top}\Lambda^D}{N}$ each converge in probability to positive definite matrices.

Assumption 1.3 has three important parts. First, we require the processes to have only finite jump activity. This mean that on every finite time interval there are almost surely only

[8] See for example Prigent (2003)

finitely many jumps. With infinite activity jump processes, i.e. each interval can contain infinitely many small jumps, we cannot separate the continuous and discontinuous part of a process. Second, we assume that a jump in the factors or the idiosyncratic part implies a jump in the process X_i. The reverse is trivially satisfied. This second assumption is important to identify all the times of discontinuities of the unobserved factors and residuals. This second part is always satisfied as soon as the Lévy measure of F_i and e_i have a density, which holds in most models used in the literature. The third statement is a non-redundancy condition and requires each systematic jump factor to jump at least once in the data. This is a straightforward and necessary condition to identify any jump factor. Hence, the main restriction in Assumption 1.3 is the finite jump activity. For example compound poisson processes with stochastic intensity rate fall into this category.

Theorem 1.2. ***Separating continuous and jump factors:***
Assume Assumptions 1.1 and 1.3 hold. Set the threshold identifier for jumps as $\alpha\Delta_M^{\bar{\omega}}$ for some $\alpha > 0$ and $\bar{\omega} \in \left(0, \frac{1}{2}\right)$ and define $\hat{X}^C_{j,i} = X_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}}$ and $\hat{X}^D_{j,i} = X_{j,i}\mathbb{1}_{\{|X_{j,i}|> \alpha\Delta_M^{\bar{\omega}}\}}$. The estimators $\hat{\Lambda}^C$, $\hat{\Lambda}^D$, $\hat{F}^C$ and $\hat{F}^D$ are defined analogously to $\hat{\Lambda}$ and $\hat{F}$, but using $\hat{X}^C$ and $\hat{X}^D$ instead of X.
Define $H^C = \frac{1}{N}\left({F^C}^\top F^C\right)\left({\Lambda^C}^\top \hat{\Lambda}^C\right){V^C_{MN}}^{-1}$ and $H^D = \frac{1}{N}\left({F^D}^\top F^D\right)\left({\Lambda^D}^\top \hat{\Lambda}^D\right){V^D_{MN}}^{-1}$.

1. *The continuous and jump loadings can be estimated consistently:*

$$\hat{\Lambda}^C_i = {H^C}^\top \Lambda^C_i + o_p(1) \qquad , \qquad \hat{\Lambda}^D_i = {H^D}^\top \Lambda^D_i + o_p(1).$$

2. *Assume that additionally Assumption 1.2 holds. The continuous and jump factors can only be estimated up to a finite variation bias term*

$$\hat{F}^C_T = {H^C}^{-1} F^C_T + o_p(1) + \text{finite variation term}$$
$$\hat{F}^D_T = {H^D}^{-1} F^D_T + o_p(1) + \text{finite variation term.}$$

3. *Under the additional Assumption 1.2 we can estimate consistently the covariation of the continuous and jump factors with other processes. Let $Y(t)$ be an Itô-semimartingale satisfying Definition A.1. Then we have for $\frac{\sqrt{M}}{N} \to 0$ and $\delta \to \infty$:*

$$\sum_{j=1}^{M} \hat{F}^C_j Y_j = {H^C}^{-1}[F^C, Y]_T + o_p(1) \qquad , \qquad \sum_{j=1}^{M} \hat{F}^D_j Y_j = {H^D}^{-1}[F^D, Y]_T + o_p(1).$$

The theorem states that we can estimate the factors only up to a finite variation term, i.e. we can only estimate the martingale part of the process correctly. The intuition behind this problem is very simple. The truncation estimator can correctly separate the jumps from the continuous martingale part. However, all the drift terms will be assigned to the continuous component. If a jump factor also has a drift term, this will now appear in the continuous part and as this drift term affects infinitely many cross-sectional X_i, it cannot be diversified away.

1.5 Asymptotic Distribution

1.5.1 Distribution Results

The assumptions for asymptotic mixed-normality of the estimators are stronger than those needed for consistency. Asymptotic mixed-normality of the loadings does not require additional assumptions, while the asymptotic normality of the factors needs substantially stronger assumptions. This should not be surprising as essentially all central limit theorems impose restrictions on the tail behavior of the sampled random variables.

In order to obtain a mixed Gaussian limit distribution for the loadings we need to assume that there are no common jumps in σ_F and e_i and in σ_{e_i} and F. Without this assumption the estimator for the loadings still converges at the same rate, but it is not mixed-normally distributed any more. Note that Assumption 1.1 requires the independence of F and e, which implies the no common jump assumption.

Theorem 1.3. ***Asymptotic distribution of loadings***
Assume Assumptions 1.1 and 1.2 hold and define $\delta = min(N, M)$. Then

$$\sqrt{M}\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) = V_{MN}^{-1}\left(\frac{\hat{\Lambda}^\top \Lambda}{N}\right)\sqrt{M} F^\top e_i + O_p\left(\frac{\sqrt{M}}{\delta}\right)$$

1. *If $\frac{\sqrt{M}}{N} \to 0$, then*

$$\sqrt{M}(\hat{\Lambda}_i - H^\top \Lambda_i) \xrightarrow{L-s} N\left(0, V^{-1} Q \Gamma_i Q^\top V^{-1}\right)$$

where V is the diagonal matrix of eigenvalues of $\Sigma_\Lambda^{\frac{1}{2}} \Sigma_F \Sigma_\Lambda^{\frac{1}{2}}$ and $\underset{N,M\to\infty}{plim} \frac{\hat{\Lambda}^\top \Lambda}{N} = Q = V^{\frac{1}{2}} \Upsilon^\top \sigma_F^{\frac{1}{2}}$ with Υ being the eigenvectors of V. The entry $\{l, g\}$ of the $K \times K$ matrix Γ_i is given by

$$\Gamma_{i,l,g} = \int_0^T \sigma_{F^l,F^g} \sigma_{e_i}^2 ds + \sum_{s\le T} \Delta F^l(s) \Delta F^g(s) \sigma_{e_i}^2(s) + \sum_{s'\le T} \Delta e_i^2(s') \sigma_{F^g,F^l}(s').$$

F^l denotes the l-th component of the the K dimensional process F and σ_{F^l,F^g} are the entries of its $K \times K$ dimensional volatility matrix.

2. *If $\liminf \frac{\sqrt{M}}{N} \ge \tau > 0$, then $N(\hat{\Lambda}_i - \Lambda_i H) = O_p(1)$.*

The limiting distributions of the loadings is obviously different from the distribution in conventional factor analysis. Here we can see very clearly how the results from high-frequency econometrics impact the estimators in our factor model.

Assumption 1.4. ***Asymptotically negligible jumps of error terms***
Assume Z is some continuous square integrable martingale with quadratic variation $\langle Z, Z \rangle_t$. Assume that the jumps of the martingale $\frac{1}{\sqrt{N}} \sum_{i=1}^N e_i(t)$ are asymptotically negligible in the sense that

$$\frac{\Lambda^\top [e,e]_t \Lambda}{N} \xrightarrow{p} \langle Z, Z \rangle_t \ , \qquad \frac{\Lambda^\top \langle e^D, e^D \rangle_t \Lambda}{N} \xrightarrow{p} 0 \qquad \forall t > 0.$$

Assumption 1.4 is needed to obtain an asymptotic mixed-normal distribution for the factor estimator. It means that only finitely many residual terms can have a jump component. Hence, the weighted average of residual terms has a quadratic covariation that depends only on the continuous quadratic covariation. This assumption is essentially a Lindeberg condition. If it is not satisfied and under additional assumptions the factor estimator converges with the same rate to a distribution with the same variance, but with heavier tails than a mixed-normal distribution.

Assumption 1.5. ***Weaker dependence of error terms***

- ***Assumption 5.1: Weak serial dependence***
  The error terms exhibit weak serial dependence if and only if

$$\|\mathbb{E}\left[e_{ji} e_{jr} \sum_{l \neq j} e_{li} \sum_{s \neq j} e_{sr} \right] \| \leq C \|\mathbb{E}[e_{ji} e_{jr}]\| \|\mathbb{E}\left[\sum_{l \neq j} e_{li} \sum_{s \neq j} e_{lr} \right] \|$$

  for some finite constant C and for all $i, r = 1, ..., N$ and for all partitions $[t_1, ..., t_M]$ of $[0, T]$.

- ***Assumption 5.2: Weak cross-sectional dependence***
  The error terms exhibit weak cross-sectional dependence if and only if

$$\sum_{j=1}^M \frac{1}{N^2} \sum_{i=1}^N \sum_{r=1}^N \mathbb{E}\left[e_{ji}^2 e_{jr}^2 \right] = O\left(\frac{1}{\delta} \right)$$

  for all $i, r = 1, ..., N$ and for all partitions $[t_1, ..., t_M]$ of $[0, T]$ for $M, N \to \infty$ and

$$\sum_{i=1}^N |G_{k,i}(t)| \leq C \qquad \text{a.s. for all } k = 1, ..., N \text{ and } t \in (0, T] \text{ and some constant } C.$$

Assumption 1.5 is only needed to obtain the general rate results for the asymptotic distribution of the factors. If $\frac{N}{M} \to 0$, we don't need it anymore. Lemma 1.1 gives sufficient conditions for this assumption. Essentially, if the residual terms are independent and "almost" continuous then it holds. Assumption 1.5 is not required for any consistency results.

Lemma 1.1. ***Sufficient conditions for weaker dependence***
Assume Assumptions 1.1 and 1.2 hold and that

1. e_i *has independent increments.*
2. e_i *has 4th moments.*
3. $\mathbb{E}\left[\sum_{i=1}^N \langle e_i^D, e_i^D \rangle\right] \leq C$ *for some constant* C *and for all* N.
4. $\sum_{i=1}^N |G_{k,i}(t)| \leq C$ *a.s. for all* $k = 1, ..., N$ *and* $t \in (0,T]$ *and some constant* C.

Then Assumption 1.5 is satisfied.

Theorem 1.4. ***Asymptotic distribution of the factors:***
Assume Assumptions 1.1 and 1.2 hold. Then

$$\sqrt{N}\left(\hat{F}_T - H^{-1}F_T\right) = \frac{1}{\sqrt{N}} e_T \Lambda H + O_P\left(\frac{\sqrt{N}}{\sqrt{M}}\right) + O_p\left(\frac{\sqrt{N}}{\delta}\right)$$

If Assumptions 1.4 and 1.5 hold and $\frac{\sqrt{N}}{M} \to 0$ *or only Assumption 1.4 holds and* $\frac{N}{M} \to 0$*:*

$$\sqrt{N}\left(\hat{F}_T - H^{-1}F_T\right) \xrightarrow{L-s} N\left(0, {Q^{-1}}^\top \Phi_T Q^{-1}\right)$$

with $\Phi_T = \underset{N\to\infty}{plim} \frac{\Lambda^\top [e] \Lambda}{N}$.

The assumptions needed for Theorem 1.4 are stronger than for all the other theorems. Although they might not always be satisfied in practice, simulations indicate that that the asymptotic distribution results still seem to provide a very good approximation even if the conditions are violated. As noted before it is possible to show that under weaker assumptions the factor estimators have the same rate and variance, but an asymptotic distribution that is different from a mixed-normal distribution.

The next theorem about the common components essentially combines the previous two theorems.

Theorem 1.5. ***Asymptotic distribution of the common components***
Define $C_{T,i} = \Lambda_i^\top F_T$ *and* $\hat{C}_{T,i} = \hat{\Lambda}_i^\top \hat{F}_T$ *Assume that Assumptions 1.1 - 1.4 hold.*

1. *If Assumption 1.5 holds, i.e. weak serial dependence and cross-sectional dependence, then for any sequence* N, M

$$\frac{\sqrt{\delta}\left(\hat{C}_{T,i} - C_{T,i}\right)}{\sqrt{\frac{\delta}{N} W_{T,i} + \frac{\delta}{M} V_{T,i}}} \xrightarrow{D} N(0,1)$$

2. Assume $\frac{N}{M} \to 0$ (but we do not require Assumption 1.5)

$$\frac{\sqrt{N}\left(C_{T,i} - \hat{C}_{T,i}\right)}{\sqrt{W_{T,i}}} \xrightarrow{D} N(0,1)$$

with

$$\begin{aligned} W_{T,i} &= \Lambda_i^\top \Sigma_\Lambda^{-1} \Phi_T \Sigma_\Lambda^{-1} \Lambda_i \\ V_{T,i} &= F_T^\top \Sigma_F^{-1} \Gamma_i \Sigma_F^{-1} F_T. \end{aligned}$$

1.5.2 Estimating Covariance Matrices

The asymptotic covariance matrix for the estimator of the loadings can be estimated consistently under relatively weak assumptions, while the asymptotic covariance of the factor estimator requires stricter conditions. In order to estimate the asymptotic covariance for the loadings, we cannot simply apply the truncation approach to the estimated processes. The asymptotic covariance matrix of the factors runs into a dimensionality problem, which can only be solved under additional assumptions.

Theorem 1.6. ***Feasible estimator of covariance matrix of loadings***
Assume Assumptions 1.1 and 1.2 hold and $\frac{\sqrt{M}}{N} \to 0$. Define the asymptotic covariance matrix of the loadings as $\Theta_{\Lambda,i} = V^{-1} Q \Gamma_i Q^\top V^{-1}$. Take any sequence of integers $k \to \infty$, $\frac{k}{M} \to 0$. Denote by $I(j)$ a local window of length $\frac{2k}{M}$ around j. Define the $K \times K$ matrix $\hat{\Gamma}_i$ by

$$\begin{aligned} \hat{\Gamma}_i =& M \sum_{j=1}^{M} \left(\frac{\hat{X}_j^C \hat{\Lambda}}{N}\right)\left(\frac{\hat{X}_j^C \hat{\Lambda}}{N}\right)^\top \left(\hat{X}_{j,i}^C - \frac{\hat{X}_j^C \hat{\Lambda}}{N}\hat{\Lambda}_i\right)^2 \\ &+ \frac{M}{2k} \sum_{j=k+1}^{M-k} \left(\frac{\hat{X}_j^D \hat{\Lambda}}{N}\right)\left(\frac{\hat{X}_j^D \hat{\Lambda}}{N}\right)^\top \left(\sum_{h \in I(j)} \left(\hat{X}_{h,i}^C - \frac{\hat{X}_h^C \hat{\Lambda}}{N}\hat{\Lambda}_i\right)^2\right) \\ &+ \frac{M}{2k} \sum_{j=k+1}^{M-k} \left(\hat{X}_{j,i}^D - \frac{\hat{X}_j^D \hat{\Lambda}}{N}\hat{\Lambda}_i\right)^2 \left(\sum_{h \in I(j)} \left(\frac{\hat{X}_h^C \hat{\Lambda}}{N}\right)\left(\frac{\hat{X}_h^C \hat{\Lambda}}{N}\right)^\top\right) \end{aligned}$$

Then a feasible estimator for $\Theta_{\Lambda,i}$ is $\hat{\Theta}_{\Lambda,i} = V_{MN}^{-1} \hat{\Gamma}_i V_{MN}^{-1} \xrightarrow{p} \Theta_{\Lambda,i}$ and

$$\sqrt{M} \hat{\Theta}_{\Lambda,i}^{-1/2} (\hat{\Lambda}_i - H^\top \Lambda_i) \xrightarrow{D} N(0, I_K)$$

Theorem 1.7. ***Consistent estimator of covariance matrix of factors***
Assume the Assumptions of Theorem 1.4 hold and $\sqrt{N}\left(\hat{F}_T - H^{-1} F_T\right) \xrightarrow{L-s} N(0, \Theta_F)$

with $\Theta_F = \underset{N,M\to\infty}{plim} H^\top \frac{\Lambda^\top [e] \Lambda}{N} H$. *Assume that the error terms are cross-sectionally independent. Denote the estimator of the residuals by* $\hat{e}_{j,i} = X_{j,i} - \hat{C}_{j,i}$. *Then a consistent estimator is* $\hat{\Theta}_F = \frac{\sum_{i=1}^N \hat{\Lambda}_i \hat{e}_i^\top \hat{e}_i \hat{\Lambda}_i^\top}{N} \xrightarrow{p} \Theta_F$ *and*

$$\sqrt{N}\hat{\Theta}_F^{-1/2}(\hat{F}_T - H^{-1}F_T) \xrightarrow{D} N(0, I_K).$$

The assumption of cross-sectional independence here is somewhat at odds with our general approximate factor model. The idea behind the approximate factor model is exactly to allow for weak dependence in the residuals. However, without further assumptions the quadratic covariation matrix of the residuals cannot be estimated consistently as its dimension is growing with N. Even if we knew the true residual process $e(t)$ we would still run into the same problem. Assuming cross-sectional independence is the simplest way to reduce the number of parameters that have to be estimated. We could extend this theorem to allow for a parametric model capturing the weak dependence between the residuals or we could impose a sparsity assumption similar to Fan, Liao and Mincheva (2013). In both cases the theorem would continue to hold.

Theorem 1.8. ***Consistent estimator of covariance matrix of common components***
Assume Assumptions 1.1-1.5 hold and that the residual terms e *are cross-sectionally independent. Then for any sequence* N, M

$$\left(\frac{1}{N}\hat{W}_{T,i} + \frac{1}{M}\hat{V}_{T,i}\right)^{-1/2} \left(\hat{C}_{T,i} - C_{T,i}\right) \xrightarrow{D} N(0,1)$$

with $\hat{W}_{T,i} = \hat{\Lambda}_i^\top \hat{\Theta}_F \hat{\Lambda}_i$ *and* $\hat{V}_{T,i} = \hat{F}_T^\top \left(\hat{F}^\top \hat{F}\right)^{-1} \hat{\Gamma}_i \left(\hat{F}^\top \hat{F}\right)^{-1} \hat{F}_T$.

1.6 Estimating the Number of Factors

I have developed a consistent estimator for the number of total factors, continuous factors and jump factors, that does not require stronger assumptions than those needed for consistency. Intuitively the large eigenvalues are associated with the systematic factors and hence the problem of estimating the number of factors is roughly equivalent to deciding which eigenvalues are considered to be large with respect to the rest of the spectrum. Under the assumptions that we need for consistency I can show that the first K "systematic" eigenvalues of $X^\top X$ are $O_p(N)$, while the nonsystematic eigenvalues are $O_p(1)$. A straightforward estimator for the number of factors considers the eigenvalue ratio of two successive eigenvalues and associates the number of factors with a large eigenvalue ratio. However, without very strong assumptions we cannot bound the small eigenvalues from below, which could lead to exploding eigenvalue ratios in the nonsystematic spectrum. I propose a perturbation method to avoid this problem. As long as the eigenvalue ratios of the perturbed eigenvalues cluster, we are in the nonsystematic spectrum. As soon as we do not observe this clustering any more, but a large eigenvalue ratio of the perturbed eigenvalues, we are in the systematic spectrum.

Theorem 1.9. ***Estimator for number of factors***
Assume Assumption 1.1 holds and $O\left(\frac{N}{M}\right) \leq O(1)$. *Denote the ordered eigenvalues of* $X^\top X$ *by* $\lambda_1 \geq ... \geq \lambda_N$. *Choose a slowly increasing sequence* $g(N,M)$ *such that* $\frac{g(N,M)}{N} \to 0$ *and* $g(N,M) \to \infty$. *Define perturbed eigenvalues*

$$\hat{\lambda}_k = \lambda_k + g(N,M)$$

and the perturbed eigenvalue ratio statistics:

$$ER_k = \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}} \qquad \text{for } k = 1, ..., N-1$$

Define

$$\hat{K}(\gamma) = \max\{k \leq N-1 : ER_k > 1+\gamma\}$$

for $\gamma > 0$. *If* $ER_k < 1+\gamma$ *for all* k, *then set* $\hat{K}(\gamma) = 0$. *Then for any* $\gamma > 0$

$$\hat{K}(\gamma) \xrightarrow{p} K.$$

Assume in addition that Assumption 1.3 holds. Set the threshold identifier for jumps as $\alpha\Delta_M^{\bar{\omega}}$ *for some* $\alpha > 0$ *and* $\bar{\omega} \in \left(0, \frac{1}{2}\right)$ *and define* $\hat{X}_{j,i}^C = X_{j,i}\mathbb{1}_{\{|X_{j,i}| \leq \alpha\Delta_M^{\bar{\omega}}\}}$ *and* $\hat{X}_{j,i}^D = X_{j,i}\mathbb{1}_{\{|X_{j,i}| > \alpha\Delta_M^{\bar{\omega}}\}}$. *Denote the ordered eigenvalues of* $\hat{X}^{C\top}\hat{X}^C$ *by* $\lambda_1^C \geq ... \geq \lambda_N^C$ *and analogously for* $\hat{X}^{D\top}\hat{X}^D$ *by* $\lambda_1^D \geq ...\lambda_N^D$. *Define* $\hat{K}^C(\gamma)$ *and* $\hat{K}^D(\gamma)$ *as above but using* λ_i^C *respectively* λ_i^D. *Then for any* $\gamma > 0$

$$\hat{K}^C(\gamma) \xrightarrow{p} K^C \qquad \hat{K}^D(\gamma) \xrightarrow{p} K^D$$

where K^C *is the number of continuous factors and* K^D *is the number of jump factors.*

My estimator depends on two choice variables: the perturbation g and the cutoff γ. In contrast to Bai and Ng, Onatski or Ahn and Horenstein we do not need to choose some upper bound on the number of factors. Although consistency follows for any g or γ satisfying the necessary conditions, the finite sample properties will obviously depend on them. As a first step for understanding the factor structure I recommend plotting the perturbed eigenvalue ratio statistic. In all my simulations the transition from the idiosyncratic spectrum to the systematic spectrum is very apparent. Based on simulations a good choice for the perturbation is $g = \sqrt{N} \cdot median(\{\lambda_1, ..., \lambda_N\})$. Obviously this choice assumes that the median eigenvalue is bounded from below, which is not guaranteed by our assumptions but almost always satisfied in practice. In the simulations I also test different specifications for g, e.g.

$\log(N) \cdot median(\{\lambda_1, ..., \lambda_N\})$. My estimator is very robust to the choice of the perturbation value. A more delicate issue is the cutoff γ. Simulations suggest that γ between 1.1 and 1.2 performs very well. As we are actually only interested in detecting a deviation from clustering around 1, we can also define γ to be proportional to a moving average of perturbed eigenvalue ratios.

What happens if we employ my eigenvalue ratio estimator with a constant perturbation or no perturbation at all? Under stronger assumptions on the idiosyncratic processes, the eigenvalue ratio estimator is still consistent as Proposition 1.1 shows:

Proposition 1.1. ***Onatski-type estimator for number of factors***
Assume Assumptions 1.1 and 1.3 hold and $\frac{N}{M} \to c > 0$. In addition assume that

1. *The idiosyncratic terms follow correlated Brownian motions:*

$$e(t) = A\epsilon(t)$$

   where $\epsilon(t)$ is a vector of N independent Brownian motions.

2. *The correlation matrix A satisfies:*
   a) *The eigenvalue distribution function $\mathcal{F}^{AA^\top}$ converges to a probability distribution function $\mathcal{F}_A$.*
   b) *The distribution $\mathcal{F}_A$ has bounded support, $u(\mathcal{F}) = min(z : \mathcal{F}(z) = 1)$ and $u(\mathcal{F}^{AA^\top}) \to u(\mathcal{F}_A) > 0$.*
   c) $\liminf_{z \to 0} z^{-1} \int_{u(\mathcal{F}_A)-z}^{u(\mathcal{F}_A)} d\mathcal{F}_A(\lambda) = k_A > 0$.

Denote the ordered eigenvalues of $X^\top X$ by $\lambda_1 \geq ... \geq \lambda_N$. Define

$$\hat{K}^{ON}(\gamma) = \max\left\{k \leq K_{\max}^{ON} : \frac{\lambda_k}{\lambda_{k+1}} \geq \gamma\right\}$$

for any $\gamma > 0$ and slowly increasing sequence $K_{\max}^{ON}$ s.t. $\frac{K_{\max}^{ON}}{N} \to 0$. Then

$$\hat{K}^{ON}(\gamma) \xrightarrow{p} K$$

The estimator in Theorem 1.9 follows a similar logic as the Onatski-type estimator in Proposition 1.1, but uses different statistical arguments and much weaker assumptions. Under the assumptions of the Onatski estimator the largest eigenvalues of the idiosyncratic part cluster and are asymptotically very close to each other. Hence, the eigenvalue ratio for adjacent nonsystematic eigenvalues converges to 1. Under the Onatski assumptions in Proposition 1.1, we could also set $g = C$ to some constant, which is independent of N and M. We would get

$$ER_K = O_p(N)$$
$$ER_k = \frac{\lambda_k + C}{\lambda_{k+1} + C} \xrightarrow{p} 1 \qquad k \in [K+1, K_{\max}^{ON}]$$

However, the Onatski-type estimator in Proposition 1.1 fails if we use the truncated data $\hat{X}^C$ or $\hat{X}^D$.

1.7 Microstructure Noise

Asset prices observed at very high frequencies are contaminated by microstructure noise and in this section I provide an estimator for the impact of noise on the spectral distribution. A distinct characteristic of high-frequency financial data is that they are observed with noise, and that this noise interacts with the sampling frequency in complex ways. So far, we have assumed that $X(t)$ is the true log asset price, which might be justified in a perfect market with no trading imperfections, frictions, or informational effects. By contrast, market microstructure noise summarizes the discrepancy between the efficient log-price and the observed log-price, as generated by the mechanics of the trading process. Source of noise can be a collection of market microstructure effects, either information or non-information related, such as the presence of a bid-ask spread and the corresponding bounces, the differences in trade sizes and the corresponding differences in representativeness of the prices, the different informational content of price changes due to informational asymmetries of traders, the gradual response of prices to a block trade, the strategic component of the order flow, inventory control effects, the discreteness of price changes in markets that are subject to a tick size, etc., all summarized into the noise term. That these phenomena are real and important is an accepted fact in the market microstructure literature, both theoretical and empirical.

Here I show how the microstructure noise affects the largest eigenvalue of the residual matrix. The estimation of the number of factors crucially depends on the size of this largest eigenvalue. This theorem can be used to show that the estimator for the number of factors does not change in the presence of micro structure noise. It can also be used to derive an estimator for the variance of the microstructure noise. If we do not use microstructure noise robust estimators for the quadratic covariation matrix, the usual strategy is to use a lower sampling frequency that trades off the noise bias with the estimation variance. This theorem can provide some guidance if the frequency is sufficiently low to neglect the noise.

In a series of papers Jacod et al (2009) and Barndorff-Nielsen et al. (2011) among others have developed microstructure noise robust estimators for the quadratic covariation matrix. However, it is beyond the scope of this paper to develop the asymptotic theory for these more general estimators in the context of a large dimensional factor model and I leave this to future research.

Theorem 1.10. ***Upper bound on impact of noise***
Assume we observe the true asset price with noise:

$$Y_i(t_j) = X_i(t_j) + \tilde{\epsilon}_{j,i}$$

where the noise $\tilde{\epsilon}_{j,i}$ is i.i.d. $(0, \sigma_\epsilon^2)$ and independent of X and has finite fourth moments. Furthermore assume that Assumption 1.1 holds and that $\frac{N}{M} \to c < 1$. Denote increments *of the noise by $\epsilon_{j,i} = \tilde{\epsilon}_{j+1,i} - \tilde{\epsilon}_{j,i}$. Then we can bound the impact of noise on the largest eigenvalue of the idiosyncratic spectrum:*

$$\lambda_1\left(\frac{(e+\epsilon)^\top(e+\epsilon)}{N}\right) - \lambda_1\left(\frac{e^\top e}{N}\right) \leq \min_{s\in[K+1,N-K]}\left(\lambda_s\left(\frac{Y^\top Y}{N}\right)\frac{1}{1+cos\left(\frac{s+r+1}{N}\pi\right)}\right) \cdot 2\left(\frac{1+\sqrt{c}}{1-\sqrt{c}}\right)^2 + o_p(1).$$

The variance of the microstructure noise is bounded by

$$\sigma_\epsilon^2 \leq \frac{c}{2(1-\sqrt{c})^2}\min_{s\in[K+1,N-K]}\left(\lambda_s\left(\frac{Y^\top Y}{N}\right)\frac{1}{1+cos\left(\frac{s+r+1}{N}\pi\right)}\right) + o_p(1)$$

where $\lambda_s\left(\frac{Y^\top Y}{N}\right)$ denotes the sth largest eigenvalue of a symmetric matrix $\frac{Y^\top Y}{N}$.

Remark 1.1. *For $s = \frac{1}{2}N - K - 1$ the inequality simplifies to*

$$\lambda_1\left(\frac{(e+\epsilon)^\top(e+\epsilon)}{N}\right) - \lambda_1\left(\frac{e^\top e}{N}\right) \leq \lambda_{1/2N-K-1}\left(\frac{Y^\top Y}{N}\right) \cdot 2\left(\frac{1+\sqrt{c}}{1-\sqrt{c}}\right)^2 + o_p(1)$$

respectively

$$\sigma_\epsilon^2 \leq \frac{c}{2(1-\sqrt{c})^2} \cdot \lambda_{1/2N-K-1}\left(\frac{Y^\top Y}{N}\right) + o_p(1).$$

Hence, the contribution of the noise on the largest eigenvalue of the idiosyncratic part and the microstructure noise variance can be bounded by approximately the median eigenvalue of the observed quadratic covariation matrix multiplied by a constant that depends only on the ratio of M and N.

1.8 Identifying the Factors

This section develops a new estimator for testing if a set of estimated statistical factors is the same as a set of observable economic variables. As we have already noted before, factor models are only identified up to invertible transformations. We need a measure to describe how close two vector spaces are to each other. I provide a new measure for the distance between two sets of factors and develop its asymptotic distribution. Based on this I develop a new test to determine if a set of estimated statistical factors can be written as a linear combination of observed economic variables.

A natural measure for the closeness of two factors is the correlation. Denote the time increments of a candidate economic factor $G_1(t)$ by the $M \times 1$ vector G_1. Then $\frac{F_1^\top G_1}{\sqrt{F_1^\top F_1 G_1^\top G_1}}$

is a measure for $\frac{[F_1(t),G_1(t)]}{\sqrt{[F_1(t),F_1(t)][G_1(t),G_1(t)]}}$. If the correlation is equal to 1, then the two vectors are parallel to each other and describe the same factor. If we want to compare two sets of factors, we need a measure that is invariant to linear transformations. As proposed by Bai and Ng (2006) the generalized correlation is a natural candidate measure. For a K_F-dimensional sets of factors and a K_G-dimensional set of observable variables the generalized correlations are defined as the square roots of the $min(K_F, K_G)$ largest eigenvalues of the matrix $[F, F]^{-1}[F, G][G, G]^{-1}[G, F]$. The estimators for the generalized correlations are the square roots of the largest eigenvalues of $(F^\top F)^{-1}(F^\top G)(G^\top G)^{-1}(G^\top F)$. If the two sets of factors span the same vector spaces, the generalized correlations are all equal to 1. Otherwise they denote the highest possible correlations that can be achieved through linear combinations of the subspaces. If for example for $K_F = K_G = 3$ the generalized correlations are $\{1, 1, 0\}$ it implies that there exists a linear combination of G that can replicate two of the three factors in F.

Although labeling the measure as a correlation, we do not demean the data. This is because the drift term essentially describes the mean of a semimartingale and when calculating or estimating the quadratic covariation it is asymptotically negligible. Hence, my generalized correlation measure is based only on inner products and the generalized correlations correspond to the singular values of the matrix $[F, G]$ if F and G are orthonormalized with respect to the inner product $[.,.]$.

The generalized correlation can also be used to measure the distance between two sets of loadings. In statistical factor analysis loadings can be interpreted as portfolio weights. If two sets of loadings span the same vector space, it implies that they represent the same factors. The distance between two loading matrices Λ and $\tilde{\Lambda}$ with dimension $N \times K$ respectively $N \times \tilde{K}$ is estimated as the square root of the $\min(K, \tilde{K})$ largest eigenvalues of $(\Lambda^\top\Lambda)^{-1}\Lambda^\top\tilde{\Lambda}(\tilde{\Lambda}^\top\tilde{\Lambda})^{-1}\tilde{\Lambda}^\top\Lambda$. It describes to which degree Λ is a linear transformation of $\tilde{\Lambda}$.

It is well-known that if F and G are observed and *i.i.d.* normally distributed then $\frac{\sqrt{M}(\hat{\rho}_k^2-\rho_k^2)}{2\rho_k(1-\rho_k^2)} \xrightarrow{D} N(0,1)$ for $k = 1, ..., \min(K_F, K_G)$ where ρ_k is the kth generalized correlation.[9]. The result can also be extended to elliptical distributions. However, the normalized increments of stochastic processes that can realistically model financial time series are neither normally nor elliptically distributed. Hence, we cannot directly make use of these results as for example in Bai and Ng (2006).

I propose a new estimator that can be applied to essentially any stochastic process satisfying Definition 1. The total generalized correlation denoted by $\bar{\rho}$ is defined as the sum of the squared generalized correlations $\bar{\rho} = \sum_{k=1}^{\min(K_F,K_G)} \rho_k^2$. It is equal to

$$\bar{\rho} = trace\left([F, F]^{-1}[F, G][G, G]^{-1}[G, F]\right).$$

The estimator for the total generalized correlation is defined as

$$\hat{\bar{\rho}} = trace\left((\hat{F}^\top\hat{F})^{-1}(\hat{F}^\top G)(G^\top G)^{-1}(G^\top\hat{F})\right).$$

[9]See for example Anderson (1984)

As the trace operator is a differentiable function and the quadratic covariation estimator is asymptotically mixed-normally distributed we can apply a delta method argument to show that $\sqrt{M}(\hat{\bar{\rho}} - \bar{\rho})$ is asymptotically mixed-normally distributed as well.

A test for equality of two sets tests if $\bar{\rho} = \min(K_F, K_G)$. As an example consider $K_F = K_G = 3$ and the total generalized correlation is equal to 3. In this case $F(t)$ is a linear transformation of $G(t)$ and both describe the same factor model. Based on the asymptotic normal distribution of $\hat{\bar{\rho}}$ we can construct a test statistic and confidence intervals. The null hypothesis is $\bar{\rho} < \min(K_F, K_G)$.

In the simple case of $K_F = K_G = 1$ the squared generalized correlation and hence also the total generalized correlation correspond to a measure of R^2, i.e. it measures the amount of variation that is explained by G_1 in a regression of F_1 on G_1. My measure of total generalized correlations can be interpreted as a generalization of R^2 for a regression of a vector space on another vector space.

Theorem 1.11. ***Asymptotic distribution for total generalized correlation***
Assume $F(t)$ is a factor process as in Assumption 1.1. Denote by $G(t)$ a K_G-dimensional process satisfying Definition A.1. The process G is either (i) a well-diversified portfolio of X, i.e. it can be written as $G(t) = \frac{1}{N}\sum_{i=1}^{N} w_i X_i(t)$ with $\|w_i\|$ bounded for all i or (ii) G is independent of the residuals $e(t)$. Furthermore assume that $\frac{\sqrt{M}}{N} \to 0$. The $M \times K_G$ matrix of increments is denoted by G. Assume that[10]

$$\sqrt{M}\left(\begin{pmatrix} F^\top F & F^\top G \\ G^\top F & G^\top G \end{pmatrix} - \begin{pmatrix} [F,F] & [F,G] \\ [G,F] & [G,G] \end{pmatrix}\right) \overset{L-s}{\to} N(0,\Pi)$$

Denote the total generalized correlation by $\bar{\rho} = trace\left([F,F]^{-1}[F,G][G,G]^{-1}[G,F]\right)$ and its estimator by $\hat{\bar{\rho}} = trace\left((\hat{F}^\top \hat{F})^{-1}(\hat{F}^\top G)(G^\top G)^{-1}(G^\top \hat{F})\right)$. Then

$$\sqrt{M}\left(\hat{\bar{\rho}} - \bar{\rho}\right) \overset{L-s}{\to} N(0,\Xi) \qquad \textit{and} \qquad \frac{\sqrt{M}}{\sqrt{\Xi}}\left(\hat{\bar{\rho}} - \bar{\rho}\right) \overset{D}{\to} N(0,1)$$

with $\Xi = \xi^\top \Pi \xi$ and

$$\xi = vec\left(\begin{pmatrix} -\left([F,F]^{-1}[F,G][G,G]^{-1}[G,F][F,F]^{-1}\right)^\top & [F,F]^{-1}[F,G][G,G]^{-1} \\ [G,G]^{-1}[G,F][F,F]^{-1} & -\left([G,G]^{-1}[G,F][F,F]^{-1}[F,G][G,G]^{-1}\right)^\top \end{pmatrix}\right).$$

Theorem 1.12. ***A feasible central limit theorem for the generalized continuous correlation***

[10] As explained in for example Barndorff-Nielsen and Shephard (2004) the statement should be read as $\sqrt{M}\left(vec\left(\begin{pmatrix} F^\top F & F^\top G \\ G^\top F & G^\top G \end{pmatrix}\right) - vec\left(\begin{pmatrix} [F,F] & [F,G] \\ [G,F] & [G,G] \end{pmatrix}\right)\right) \overset{L-s}{\to} N(0,\Pi)$, where *vec* is the vectorization operator. Inevitably the matrix Π is singular due to the symmetric nature of the quadratic covariation. A proper formulation avoiding the singularity uses *vech* operators and elimination matrices (See Magnus (1988)).

Assume Assumptions 1.1 to 1.3 hold. The process G is either (i) a well-diversified portfolio of X, i.e. it can be written as $G(t) = \frac{1}{N}\sum_{i=1}^{N} w_i X_i(t)$ with $\|w_i\|$ bounded for all i or (ii) G is independent of the residuals $e(t)$. Furthermore assume that $\frac{\sqrt{M}}{N} \to 0$. Denote the threshold estimators for the continuous factors as $\hat{F}^C$ and for the continuous component of G as $\hat{G}^C$. The total generalized continuous correlation is

$$\bar{\rho}^C = trace\left([F^C, F^C]^{-1}[F^C, G^C][G^C, G^C]^{-1}[G^C, F^C]\right)$$

and its estimator is

$$\hat{\bar{\rho}}^C = trace\left((\hat{F^C}^\top \hat{F}^C)^{-1}(\hat{F^C}^\top \hat{G}^C)(\hat{G^C}^\top \hat{G}^C)^{-1}(\hat{G^C}^\top \hat{F}^C)\right).$$

Then

$$\frac{\sqrt{M}}{\sqrt{\hat{\Xi}^C}}\left(\hat{\bar{\rho}}^C - \bar{\rho}^C\right) \xrightarrow{D} N(0,1)$$

Define the $M \times (K_F + K_G)$ matrix $Y = \begin{pmatrix}\hat{F}^C & \hat{G}^C\end{pmatrix}$. Choose a sequence satisfying $k \to \infty$ and $\frac{k}{M} \to 0$ and estimate spot volatilities as

$$\hat{v}_j^{i,r} = \frac{M}{k}\sum_{l=1}^{k-1} Y_{j+l,i}Y_{j+l,r}.$$

The estimator of the $(K_F + K_G) \times (K_F + K_G)$ quarticity matrix $\hat{\Pi}^C$ has the elements

$$\hat{\Pi}^C_{r+(i-1)(K_F+K_G),n+(m-1)(K_F+K_G)} = \frac{1}{M}\left(1 - \frac{2}{k}\right)\sum_{j=1}^{M-k+1}\left(v_j^{i,r}v_j^{m,n} + v_j^{i,n}v_j^{r,m}\right)$$

for $i, r, m, n = 1, ..., K_F + K_G$. Estimate $\hat{\xi}^C = vec(S)$ for the matrix S with block elements

$$\begin{aligned}
S_{1,1} &= -\left(\left(\hat{F}^{C\top}\hat{F}^C\right)^{-1}\hat{F}^{C\top}\hat{G}^C\left(\hat{G}^{C\top}\hat{G}^C\right)^{-1}\hat{G}^{C\top}\hat{F}^C\left(\hat{F}^{C\top}\hat{F}^C\right)^{-1}\right)^\top \\
S_{1,2} &= \left(\hat{F}^{C\top}\hat{F}^C\right)^{-1}\hat{F}^{C\top}\hat{G}^C\left(\hat{G}^{C\top}\hat{G}^C\right)^{-1} \\
S_{2,1} &= \left(\hat{G}^{C\top}\hat{G}^C\right)^{-1}\hat{G}^{C\top}\hat{F}^C\left(\hat{F}^{C\top}\hat{F}^C\right)^{-1} \\
S_{2,2} &= -\left(\left(\hat{G}^{C\top}\hat{G}^C\right)^{-1}\hat{G}^{C\top}\hat{F}^C\left(\hat{F}^{C\top}\hat{F}^C\right)^{-1}\hat{F}^{C\top}\hat{G}^C\left(\hat{G}^{C\top}\hat{G}^C\right)^{-1}\right)^\top.
\end{aligned}$$

The estimator for the covariance of the total generalized correlation estimator is $\hat{\Xi}^C = \hat{\xi}^{C\top}\hat{\Pi}^C\hat{\xi}^C$.

The assumption that G has to be a well-diversified portfolio of the underlying asset space is satisfied by essentially all economic factors considered in practice, e.g. the market factor or the value, size and momentum factors. Hence, practically it does not impose a restriction on the testing procedure. This assumption is only needed to obtain the same distribution theory for the quadratic covariation of G with the the estimated factors as with the true factors.

1.9 Simulations

This section considers the finite sample properties of my estimators through Monte-Carlo simulations. In the first subsection I use Monte-Carlo simulations to analyze the distribution of my estimators for the loadings, factors and common components. In the second subsection I provide a simulation study of the estimator for the number of factors and compare it to the most popular estimators in the literature.

My benchmark model is a Heston-type stochastic volatility model with jumps. In the general case I assume that the K factors are modeled as

$$\begin{aligned} dF_k(t) &= (\mu - \sigma_{F_k}^2(t))dt + \rho_F \sigma_{F_k}(t) dW_{F_k}(t) + \sqrt{1-\rho_F^2}\sigma_{F_k}(t) d\tilde{W}_{F_k}(t) + J_{F_k} dN_{F_k}(t) \\ d\sigma_{F_k}^2(t) &= \kappa_F \left(\alpha_F - \sigma_{F_k}^2(t)\right) dt + \gamma_F \sigma_{F_k}(t) d\tilde{W}_{F_k}(t) \end{aligned}$$

and the N residual processes as

$$\begin{aligned} de_i(t) &= \rho_e \sigma_{e_i}(t) dW_{e_i}(t) + \sqrt{1-\rho_e^2}\sigma_{e_i}(t) d\tilde{W}_{e_i}(t) + J_{e_i} dN_{e_i}(t) - \mathbb{E}[J_{e_i}]\nu_e dt \\ d\sigma_{e_i}^2(t) &= \kappa_e \left(\alpha_e - \sigma_{e_i}^2(t)\right) dt + \gamma_e \sigma_{e_i}(t) d\tilde{W}_{e_i}(t) \end{aligned}$$

The Brownian motions $W_F, \tilde{W}_F, W_e, \tilde{W}_e$ are assumed to be independent. I set the parameters to values typically used in the literature: $\kappa_F = \kappa_e = 5$, $\gamma_F = \gamma_e = 0.5$, $\rho_F = -0.8$, $\rho_e = -0.3$, $\mu = 0.05$, $\alpha_F = \alpha_e = 0.1$. The jumps are modeled as a compound Poisson process with intensity $\nu_F = \nu_e = 6$ and normally distributed jumps with $J_{F_k} \sim N(-0.1, 0.5)$ and $J_{e_i} \sim N(0, 0.5)$. The time horizon is normalized to $T = 1$.

In order to separate continuous from discontinuous movements I use the threshold $3\hat{\sigma}_X(j) \cdot \Delta_M^{0.48}$. The spot volatility is estimated using Barndorff-Nielsen and Shephard's (2006) bi-power volatility estimator on a window of $\sqrt{M}$ observations. Under certain assumptions the bi-power estimator is robust to jumps and estimates the volatility consistently.

In order to capture cross-sectional correlations I formulate the dynamics of X as

$$X(t) = \Lambda F(t) + Ae(t)$$

where the matrix A models the cross-sectional correlation. If A is an identity matrix, then the residuals are cross-sectionally independent. The empirical results suggest that it is very important to distinguish between strong and weak factors. Hence the first factor is multiplied by the scaling parameter $\sigma_{dominant}$. If $\sigma_{dominant} = 1$ then all factors are equally strong. In practice, the first factor has the interpretation of a market factor and has a significantly larger variance than the other weaker factors. Hence, a realistic model with several factors should set $\sigma_{dominant} > 1$.

The loadings Λ are drawn from independent standard normal distributions. All Monte-Carlo simulations have 1000 repetitions. I first simulate a discretized model of the continuous time processes with 2000 time steps representing the true model and then use the data which is observed on a coarser grid with $M = 50, 100, 250$ or 500 observations. My results are robust to changing the number of Monte-Carlo simulations or using a finer time grid for the "true" process.

1.9.1 Asymptotic Distribution Theory

In this subsection I consider only one factor in order to assess the properties of the limiting distribution, i.e. $K = 1$ and $\sigma_{dominant} = 1$. I consider three different cases:

1. **Case 1: Benchmark model with jumps.** The correlation matrix A is a Toplitz matrix with parameters $(1, 0.2, 0.1)$, i.e. it is a symmetric matrix with diagonal elements 1 and the first two off-diagonals have elements 0.2 respectively 0.1.

2. **Case 2: Benchmark model without jumps.** This model is identical to case 1 but without the jump component in the factors and residuals.

3. **Case 3: Toy model.** Here all the stochastic processes are standard Brownian motions

$$X(t) = \Lambda W_F(t) + W_e(t)$$

   After rescaling the model is identical to the simulation study considered in Bai (2003).

Obviously, we can only estimate the continuous and jump factors in case 1.

In order to assess the accuracy of the estimators I calculate the correlations of the estimator for the loadings and factors with the true values. If jumps are included, we have additionally correlations for the continuous and jump estimators. In addition for $t = T$ and $i = N/2$ I calculate the asymptotic distribution of the rescaled and normalized estimators:

$$\begin{aligned} CLT_C &= \left(\frac{1}{N}\hat{V}_{T,i} + \frac{1}{M}\hat{W}_{T,i}\right)^{-1/2} \left(\hat{C}_{T,i} - C_{T,i}\right) \\ CLT_F &= \sqrt{N}\hat{\Theta}_F^{-1/2}(\hat{F}_T - H^{-1}F_T) \\ CLT_\Lambda &= \sqrt{M}\hat{\Theta}_{\Lambda,i}^{-1/2}(\hat{\Lambda}_i - H^\top \Lambda_i) \end{aligned}$$

Table 1.1 reports the mean and standard deviation of the correlation coefficients between $\hat{F}_T$ and F_T and $\hat{\Lambda}_i$ and Λ_i based on 1000 simulations. In case 1 I also estimate the continuous and jump part. The correlation coefficient can be considered as a measure of consistency.

	N=200, M=250					N=100, M=100				
	Case 1			Case 2	Case 3	Case 1			Case 2	Case 3
	Total	Cont.	Jump			Total	Cont.	Jump		
Corr. F_T	0.994	0.944	0.972	0.997	0.997	0.986	0.789	0.943	0.994	0.997
SD F_T	0.012	0.065	0.130	0.001	0.000	0.037	0.144	0.165	0.002	0.000
Corr. Λ	0.995	0.994	0.975	0.998	0.998	0.986	0.966	0.949	0.994	0.998
SD Λ	0.010	0.008	0.127	0.001	0.000	0.038	0.028	0.157	0.002	0.000
	N=500, M=50					N=50, M=500				
	Case 1			Case 2	Case 3	Case 1			Case 2	Case 3
	Total	Cont.	Jump			Total	Cont.	Jump		
Corr. F_T	0.997	0.597	0.926	0.999	0.999	0.973	0.961	0.954	0.988	0.990
SD F_T	0.006	0.196	0.151	0.001	0.000	0.067	0.028	0.141	0.005	0.002
Corr. Λ	0.979	0.921	0.906	0.987	0.990	0.991	0.997	0.974	0.999	0.999
SD Λ	0.027	0.051	0.175	0.005	0.002	0.053	0.002	0.128	0.001	0.000

Table 1.1: Mean and standard deviations of estimated correlation coefficients between $\hat{F}_T$ and F_T and $\hat{\Lambda}_i$ and Λ_i based on 1000 simulations.

For the factor processes the correlation is based on the quadratic covariation between the true and the estimated processes. I run the simulations for four combinations of N and M: $N = 200, M = 250$, $N = 100, M = 100$, $N = 500, M = 50$ and $N = 50, M = 500$. The correlation coefficients in all cases are very close to one, indicating that my estimators are very precise. Note, that we can only estimate the continuous and jump factor up to a finite variation part. However, when calculating the correlations, the drift term is negligible. For a small number of high-frequency observations M the continuous and the jump factors are estimated with a lower precision as the total factor. This is mainly due to an imprecision in the estimation of the jumps. In all cases the loadings can be estimated very precisely. The simpler the processes, the better the estimators work. For sufficiently large N and M, increasing N improves the estimator for the loadings, while increasing M leads to a better estimation of the factors. Overall, the finite sample properties for consistency are excellent.

Table 1.2 and Figures 1.2 to 1.4 summarize the simulation results for the normalized estimators CLT_C, CLT_F and CLT_Λ. The asymptotic distribution theory suggests that they should be $N(0,1)$ distributed. The tables list the means and standard deviations based on 1000 simulations. For the toy model in case 3 the mean is close to 0 and the standard deviation almost 1, indicating that the distribution theory works. Figure 1.4 depicts the histograms overlaid with a normal distribution. The asymptotic theory provides a very good approximation to the finite sample distributions. Adding stochastic volatility and weak cross-sectional correlation still provides a good approximation to a normal distribution. The common component estimator is closer to the asymptotic distribution than the factor or

N=200,	M=250	CLT_C	CLT_F	CLT_Λ	N=100,	M=100	CLT_C	CLT_F	CLT_Λ
Case 1	Mean	0.023	0.015	0.051	Case 1	Mean	-0.047	0.025	-0.006
	SD	1.029	1.060	1.084		SD	0.992	1.139	1.045
Case 2	Mean	0.004	-0.007	-0.068	Case 2	Mean	-0.005	0.030	0.041
	SD	1.040	1.006	1.082		SD	1.099	1.046	1.171
Case 3	Mean	0.000	0.002	0.003	Case 3	Mean	0.024	-0.016	-0.068
	SD	1.053	1.012	1.049		SD	1.039	1.060	1.091
N=500,	M=50	CLT_C	CLT_F	CLT_Λ	N=50,	M=500	CLT_C	CLT_F	CLT_Λ
Case 1	Mean	-0.026	-0.012	-0.029	Case 1	Mean	-0.005	-0.044	0.125
	SD	0.964	1.308	1.002		SD	1.055	4.400	1.434
Case 2	Mean	-0.028	-0.009	0.043	Case 2	Mean	0.012	-0.018	-0.020
	SD	1.120	1.172	1.178		SD	0.989	1.038	1.178
Case 3	Mean	-0.064	0.003	0.018	Case 3	Mean	0.053	0.030	-0.013
	SD	1.079	1.159	1.085		SD	1.015	1.042	1.141

Table 1.2: Mean and standard deviation of normalized estimators for the common component, factors and loadings based on 1000 simulations

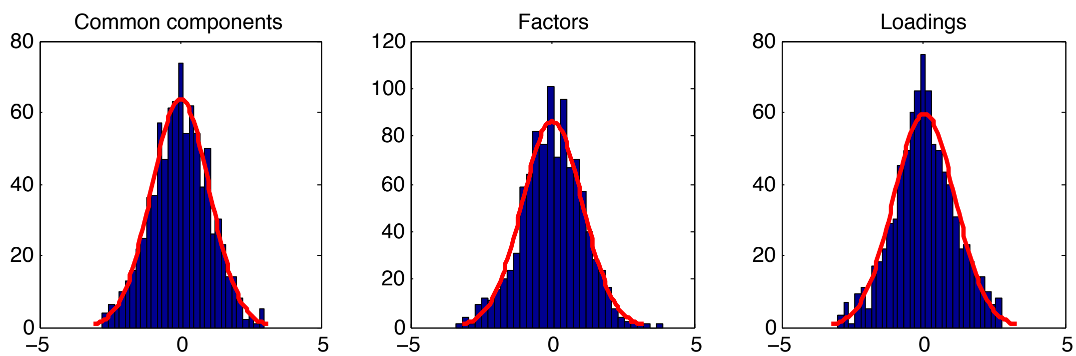


Figure 1.2: Case 1 with $N = 200$ and $M = 250$. Histogram of standardized common components CLT_C, factors CLT_F and loadings CLT_Λ. The normal density function is superimposed on the histograms.

loading estimator. Even in case 1 with the additional jumps the approximation works well. The common component estimator still performs the best. Without an additional finite sample correction the loading estimator in case 1 would have some large outliers. In more detail, the derivations for case 1 assume that the time increments are sufficiently small such that the two independent processes $F(t)$ and $e_i(t)$ do not jump during the same time increment. Whenever this happens the rescaled loadings statistic explodes. For very few of the 1000 simulations in case 1 we observe this problem and exclude these simulations. I have set the length of the local window in the covariance estimation of the loadings estimator to $k = \sqrt{M}$.

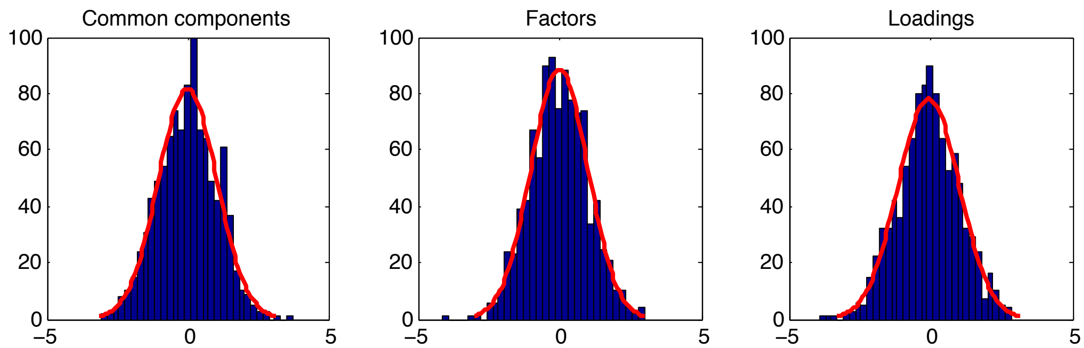


Figure 1.3: Case 2 with $N = 200$ and $M = 250$. Histogram of standardized common components CLT_C, factors CLT_F and loadings CLT_Λ. The normal density function is superimposed on the histograms.

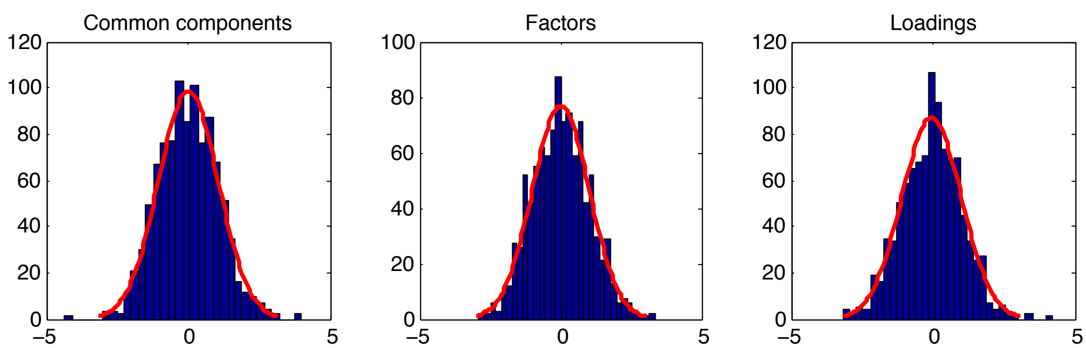


Figure 1.4: Case 3 with $N = 200$ and $M = 250$. Histogram of standardized common components CLT_C, factors CLT_F and loadings CLT_Λ. The normal density function is superimposed on the histograms.

The estimator for the covariance of the factors assumes cross-sectional independence, which is violated in the simulation example as well as Assumption 1.5. Nevertheless in the simulations the normalized statistics approximate a normal distribution very well. Overall, the finite sample properties for the asymptotic distribution work well.

1.9.2 Number of Factors

In this subsection I analyze the finite sample performance of my estimator for the number of factors and show that it outperforms the most popular estimators in the literature. One of the main motivations for developing my estimator is that the assumptions needed for the Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013) estimator cannot be extended to the general processes that we need to consider. In particular all three estimators assume essentially that the residuals can be written in the form BEA, where B is a $T \times T$ matrix capturing serial correlation, A is a $N \times N$ matrix modeling the cross-sectional correlation and E is a $T \times N$ matrix of i.i.d. random variables with finite fourth moments. Such a formulation rules out jumps and a complex stochastic volatility structure.

In the first part of this section we work with a variation of the toy model such that we can apply all four estimators and compare them:

$$X(t) = \Lambda W_F(t) + \theta A W_e(t)$$

where all the Brownian motions are independent and the $N \times N$ matrix A models the cross-sectional dependence, while θ captures the signal-to-noise ratio. The matrix A is a Toplitz matrix with parameters $(1, a, a, a, a^2)$, i.e. it is a symmetric matrix with diagonal element 1 and the first four off-diagonals having the elements a, a, a and a^2. A dominant factor is modeled with $\sigma_{dominant} > 1$. Note, that after rescaling this is the same model that is also considered in Bai and Ng, Onatski and Ahn and Horenstein. Hence, these results obviously extend to the long horizon framework. In the following simulations we always consider three factors, i.e. $K = 3$.

I simulate four scenarios:

1. Scenario 1: Dominant factor, large noise-to signal ratio, cross-sectional correlation $\sigma_{dominant} = \sqrt{10}$, $\theta = 6$ and $a = 0.5$.
2. Scenario 2: No dominant factor, large noise-to signal ratio, cross-sectional correlation $\sigma_{dominant} = 1$, $\theta = 6$ and $a = 0.5$.
3. Scenario 3: No dominant factor, small noise-to signal ratio, cross-sectional correlation $\sigma_{dominant} = 1$, $\theta = 1$ and $a = 0.5$.
4. Scenario 4: Toy model $\sigma_{dominant} = 1$, $\theta = 1$ and $a = 0$.

My empirical studies in Chapter 2 suggest that in the data the first systematic factor is very dominant with a variance that is 10 times larger then the those of the other weaker factors. Furthermore the idiosyncratic part seems to have a variance that is at least as large the variance of the common components. Both findings indicate that scenario 1 is the most realistic case and any estimator of practical relevance must also work in this scenario.

My perturbed eigenvalue ratio statistic has two choice parameters: the perturbation $g(N, M)$ and the cutoff γ. In the simulations I set the cutoff equal to $\gamma = 1.2$. For the perturbation I consider the two choices $g(N, M) = \sqrt{N} \cdot median\{\lambda_1, ..., \lambda_N\}$ and $g(N, M) = \log(N) \cdot median\{\lambda_1, ..., \lambda_N\}$. The first estimator is denoted by $ERP1$, while the second is $ERP2$. All our results are robust to these choice variables. The Onatski (2010) estimator is denoted by *Onatski* and I use the same parameters as in his paper. The Ahn and Horenstein (2013) estimator is labeled as *Ahn*. As suggested in their paper, for their estimator I first demean the data in the cross-sectional and time dimension before applying principal component analysis. *Bai* denotes the BIC3 estimator of Bai and Ng (2002). The BIC3 estimator outperforms the other versions of the Bai and Ng estimators in simulations. For the last three estimators, we need to define an upper bound on the number of factors, which I set equal to $k_{max} = 20$. The main results are not affected by changing k_{max}. For $ERP1$ and $ERP2$ we consider the whole spectrum. The figures and plots are based on 1000 simulations.

Obviously there are more estimators in the literature, e.g. Harding (2013), Alessi, Barigozzi and Capasso (2010) and Hallin and Liska (2007). However, the simulation studies in their papers indicate that the Onatski and Ahn and Horenstein estimators dominate most other estimators.

Figures 1.5 to 1.8 plot the root-mean squared error for the different estimators for a growing number $N = M$ and show that my estimators strongly outperform or are at least as good as the other estimators. In the most relevant Scenario 1 depicted in Figure 1.5 only the $ERP1$, $ERP2$ and *Onatski* estimator are reliable. This is because these three estimators focus on the residual spectrum and are not affected by strong factors. Although we apply the demeaning as proposed in Ahn and Horenstein, their estimator clearly fails. Table 1.3 shows the summary statistics for this scenario. *Ahn* and *Bai* severely underestimate the number of factors, while the $ERP1$ and $ERP2$ estimators are the best. Note, that the maximal error for both ERP estimators is smaller than for *Onatski*. In Figure 1.6 we remove the strong factor and the performance of *Ahn* drastically improves. However $ERP1$ and $ERP1$ still show a comparable performance. In the less realistic Scenarios 3 and 4, all estimators are reliable and perform equally well.

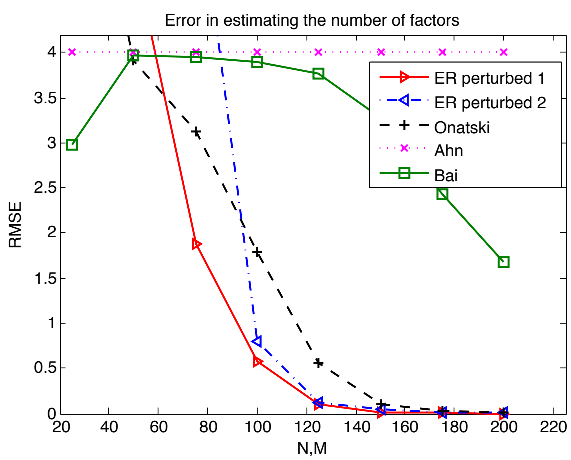


Figure 1.5: RMSE (root-mean squared error) for the number of factors in scenario 1 for different estimators with $N = M$.

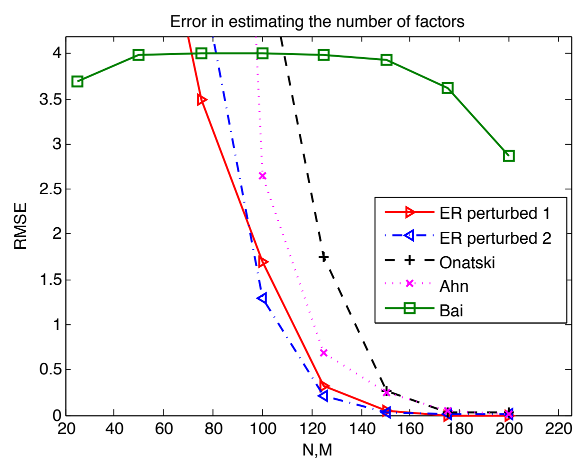


Figure 1.6: RMSE (root-mean squared error) for the number of factors in scenario 2 for different estimators with $N = M$.

	ERP1	ERP2	Onatski	Ahn	Bai
RMSE	0.32	0.18	0.49	4.00	3.74
Mean	2.79	2.88	2.76	1.00	1.09
Median	3	3	3	1	1
SD	0.52	0.41	0.66	0.00	0.28
Min	1	1	1	1	1
Max	3	4	5	1	2

Table 1.3: Scenario 1: $N = M = 125$, $K = 3$.

	ERP1	ERP2	Onatski	Ahn	Bai
RMSE	1.48	0.87	1.99	0.73	3.99
Mean	2.39	2.62	2.31	2.56	1.00
Median	3	3	3	3	1
SD	1.05	0.85	1.23	0.73	0.06
Min	0	0	0	1	1
Max	4	4	6	4	2

Table 1.4: Scenario 2: $N = M = 125$, $K = 3$.

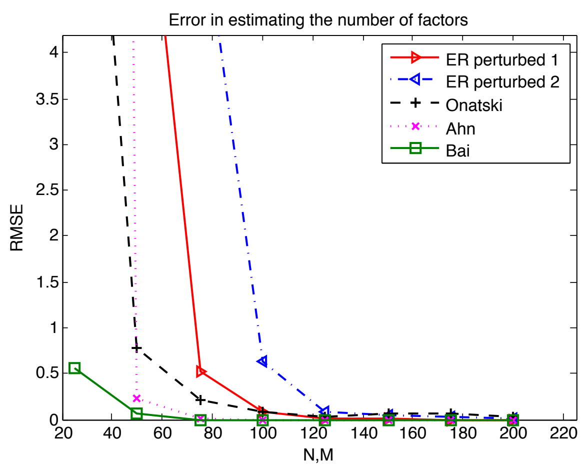


Figure 1.7: RMSE (root-mean squared error) for the number of factors in scenario 3 for different estimators with $N = M$.

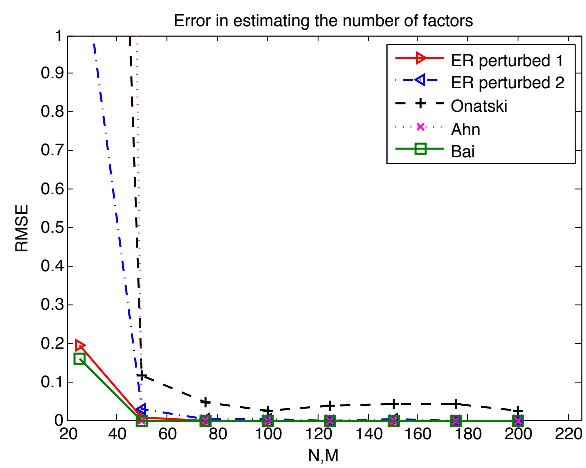


Figure 1.8: RMSE (root-mean squared error) for the number of factors in scenario 4 for different estimators with $N = M$.

	ERP1	ERP2	Onatski	Ahn	Bai
RMSE	0.00	0.01	0.06	0.00	0.00
Mean	3.00	3.01	3.03	3.00	3.00
Median	3	3	3	3	3
SD	0.03	0.08	0.24	0.00	0.00
Min	3	3	3	3	3
Max	4	4	7	3	3

Table 1.5: Scenario 3: $N = M = 125$, $K = 3$.

	ERP1	ERP2	Onatski	Ahn	Bai
RMSE	0.00	0.00	0.05	0.00	0.00
Mean	3.00	3.00	3.03	3.00	3.00
Median	3	3	3	3	3
SD	0.00	0.03	0.22	0.00	0.00
Min	3	3	3	3	3
Max	3	4	7	3	3

Table 1.6: Scenario 4: $N = M = 125$, $K = 3$.

Figures 1.9 and 1.10 show $ERP1$ applied to the benchmark model Case 1 from the last subsection. The first dominant factor has a continuous and a jump component, while the

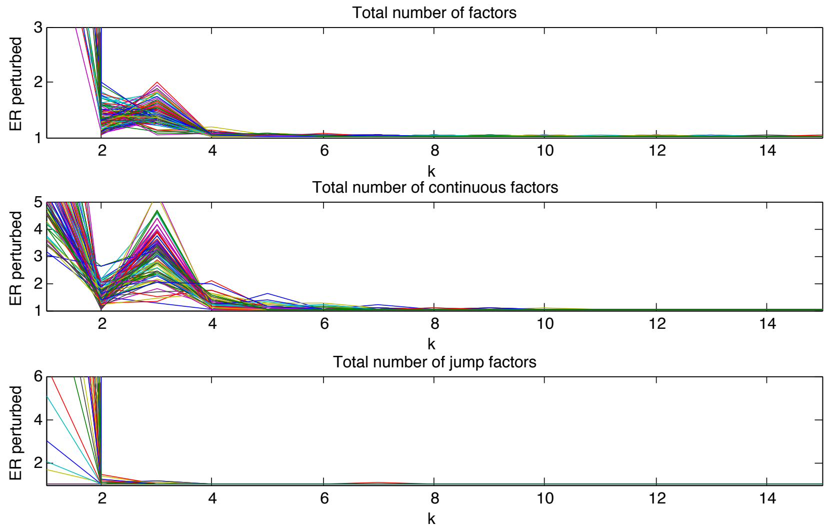


Figure 1.9: Perturbed eigenvalue ratios (ERP1) in the benchmark case 1 with $K = 3$, $K^C = 3$, $K^D = 1$, $\sigma_{dominant} = 3$, $N = 200$ and $M = 250$ for 100 simulated paths.

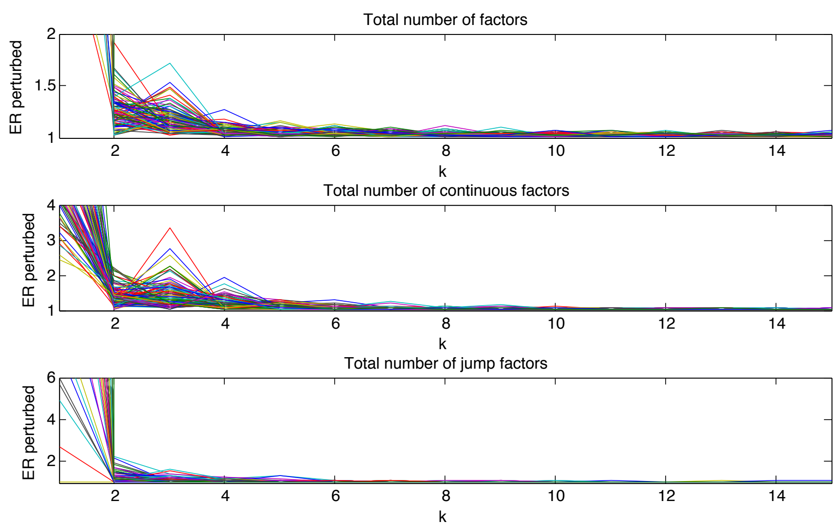


Figure 1.10: Perturbed eigenvalue ratios (ERP1) in the benchmark case 1 with $K = 3$, $K^C = 3$, $K^D = 1$, $\sigma_{dominant} = 3$, $N = 100$ and $M = 100$ for 100 simulated paths.

other two weak factors are purely continuous. Hence, we have $K = 3$, $K^C = 3$, $K^D = 1$ and $\sigma_{dominant} = 3$. I simulate 100 paths for the perturbed eigenvalue ratio and try to estimate K, K^C and K^D. We can clearly see that $ERP1$ clusters for $k > 3$ in the total and continuous case respectively $k > 1$ in the jump case and increases drastically at the true number of factors. How the cutoff threshold γ has to be set, depends very much on the data set. The choice of $\gamma = 1.2$, that worked very well in my previous simulations, would potentially not have been the right choice for Figures 1.9 and 1.10. Nevertheless, just by looking at the plots it is very apparent what the right number of factors should be. Therefore, I think plotting the perturbed eigenvalue ratios is a very good first step for understanding the potential factor structure in the data.

1.10 Conclusion

This chapter studies factor models in the new setting of a large cross section and many high-frequency observations under a fixed time horizon. I propose a principal component estimator based on the increments of the observed time series, which is a simple and feasible estimator. For this estimator I develop the asymptotic distribution theory. Using a simple truncation approach the same methodology allows to estimate continuous and jump factors. My results are obtained under very general conditions for the stochastic processes and allow for cross-sectional and serial correlation in the residuals. I also propose a novel estimator for the number of factors, that can also consistently estimate the number of continuous and jump factors under the same general conditions.

In an extensive empirical study in Chapter 2 I apply the estimation approaches developed in this chapter to 5 minutes high-frequency price data of S&P 500 firms from 2003 to 2012. I can show that the continuous factor structure is highly persistent in some years, but there is also time variation in the number and structure of factors over longer horizons. For the time period 2007 to 2012 I estimate four continuous factors which can be approximated very well by a market, oil, finance and electricity factor. The value, size and momentum factors play no significant role in explaining these factors. From 2003 to 2006 one continuous systematic factor disappears. Systematic jump risk also seems to be different from systematic continuous risk. There seems to exist only one persistent jump factor, namely a market jump factor.

Arbitrage pricing theory links risk premiums to systematic risk. In future projects I want to analyze the ability of the high-frequency factors to price the cross-section of returns. Furthermore I would like to explore the possibility to use even higher sampling frequencies by developing a microstructure noise robust estimation method.

Chapter 2

Understanding Systematic Risk: A High-Frequency Approach

2.1 Introduction

One of the most popular methods for modeling and estimating systematic risk are factor models. This paper employs the new statistical methods developed in Chapter 1 to estimate and analyze an unknown factor structure in a large cross-section of high-frequency equity data. Conventional factor analysis requires long time horizons, while this new methodology works with short time horizons, e.g. a month. The question of how to capture systematic risk is one of the most fundamental questions in asset pricing. This paper enhances our understanding about systematic risk by answering the following questions: (1) What is a good number of factors to explain the systematic movements and how does this number change over time? (2) What are the factors and how persistent is the factor structure over time? (3) Are continuous systematic risk factors, which capture the variation during "normal" times, different from jump factors, which can explain systematic tail events? (4) How does the leverage effect, i.e. the correlation of asset returns with its volatility, depend on systematic and nonsystematic risk.

The important contribution of this paper is that it does not use a pre-specified (and potentially miss-specified) set of factors. Instead I estimate the statistical factors, which can explain most of the common comovement in a large cross-section of high-frequency data. As the high-frequency data allows me to analyze different short time horizons independently, I do not impose restrictions on the potential time-variation in the factors. For a pre-specified set of factors studies have already shown that time-varying systematic risk factors capture the data better.[1] Empirical evidence also suggests that for a given factor structure systemic

[1]The idea of time-varying systematic risk factors contains the conditional version of the CAPM as a special case, which seems to explain systematic risk significantly better than its constant unconditional version. Contributions to this literature include for example Jagannathan and Wang (1996) and Lettau and Ludvigson (2001). Bali, Engle, and Tang (2014) have also shown that GARCH-based time-varying conditional betas help explain the cross-sectional variation in expected stock returns.

risk associated with discontinuous price movements is different from continuous systematic risk.[2] I confirm and extend these results to a latent factor structure.

The statistical theory underlying my estimations is very general and developed in Chapter 1. It combines the two fields of high-frequency econometrics and large-dimensional factor analysis. Under the assumption of an approximate factor model it estimates an unknown factor structure for general continuous-time processes based on high-frequency data. Using a truncation approach, I can separate the continuous and jump components of the price processes, which I use to construct a "jump covariance" and a "continuous risk covariance" matrix. The latent continuous and jump factors can be separately estimated by principal component analysis. The number of total, continuous and jump factors is estimated by analyzing the ratio of perturbed eigenvalues, which is a novel idea to the literature and shows an excellent performance in simulations. A new generalized correlation test allows me to compare the statistical factors with observed economic factors.

My empirical investigations are based on a novel high-frequency data set of 5-minutes prices for the S&P 500 firms from 2003 to 2012. My estimation approach indicates that the number and the factors do change over time. I estimate four very persistent continuous systematic factors for 2007 to 2012 and three from 2003 to 2006. These continuous factors can be approximated very well by an equally-weighted market portfolio and three industry factors, namely an oil, finance and electricity factor.[3] The value, size and momentum factors play no significant role in explaining these factors. For the time period 2003 to 2006 the finance factor seems to disappear, while the remaining factor structure stays persistent. For the whole time period there seems to exist only one persistent jump factor, namely a market jump factor. My results are robust to the sampling frequency and microstructure noise.[4]

Table 2.1 illustrates these findings, where I try to replicate the statistical factors with industry portfolios and the Fama-French Carhart factors. The number of generalized correlations close to 1 are a measure of how many factors the two sets have in common. The industry factors can approximate the persistent continuous factors very well, while size, value and momentum factors achieve only low correlations. The jump structure is different from the continuous structure.

[2]Empirical studies supporting this hypothesis include Bollerslev, Li and Todorov (2015), Pan (2002), Eraker, Johannes and Polson (2003), Bollerslev and Todorov (2011) and Gabaix (2012).

[3]The industry factors are constructed as portfolios with equally-weighted returns for firms in the oil and gas industry, the banking and insurance industry and the electricity and electric utility industry

[4]I can show that the estimated monthly and yearly factor structures are essentially identical based on 5 minutes data. Changes in the factor structure seem to occur only for different years. Within a year the estimated factor structure is basically the same if we use 5 minutes, 15 minutes or daily data. Microstructure noise becomes only relevant for high-frequencies. The fact that my results are robust to different time horizons indicates that they are robust to microstructure noise.

Generalized correlations of 4 continuous factors with industry continuous factors			
1.00	**0.98**	**0.95**	**0.80**
Generalized correlations of 4 jump factors with industry jump factors			
0.99	0.75	0.29	0.05
Generalized correlations of 4 continuous factors with Fama-French Carhart Factors			
0.95	0.74	0.60	0.00

Table 2.1: Interpretation of statistical factors. Generalized correlations of first four largest statistical factors for 2007-2012 with industry factors (market, oil, finance and electricity factors) and Fama-French Carhart factors. Results are taken from Tables 2.6, 2.8 and 2.11. Values larger than 0.8 are in bold.

Using short-maturity, at-the-money implied volatilities from option price data, I create a data set of daily volatilities for the same S&P 500 firms from 2003 to 2012 and analyze the systematic factor structure of the volatilities. There seems to be only one persistent market volatility factor, while during the financial crisis an additional temporary banking volatility factor appears.

This paper contributes to the understanding of the leverage effect by separating this effect into its systematic and idiosyncratic component based on the estimated latent factors. The leverage effect, which describes the generally negative correlation between an asset return and its volatility changes, is one of the most important empirical stylized facts about the volatility. High-frequency data is particularly suited for analyzing the leverage effect as it allows to estimate changes in the unobserved volatility. There is no consensus on the economic explanation for this statistical effect. The magnitude of the effect seems to be too large to be explained by financial leverage. Alternative economic interpretations use a risk-premium argument. An anticipated rise in volatility increases the risk premium and hence requires a higher rate of return from the asset. This leads to a fall in the asset price. The causality for these two interpretations is different. These different explanations have been tested by Bekaert and Wu (2000) who use a parametric conditional CAPM model under a GARCH specification to obtain results consistent with the risk-premium story. I estimate the leverage effect completely non-parametrically and decompose it into its systematic and nonsystematic part based on my general statistical factors. I show that the leverage effect appears predominantly for systematic risk, while it is smaller and can even be non-existent for idiosyncratic risk. These findings rule out the financial leverage story, as that explanation does not distinguish between different sources of risk.

As an illustration I plot the cross-sectional distribution for different components of the leverage effect in Figure 2.1. The continuous returns and the volatilties are first decomposed into a systematic and idiosyncratic component based on the 4 continuous return factors and the largest volatility factor. Then I calculate the correlations between the different components, where for example $(syst, idio)$ denotes the correlation between the systematic returns and the idiosyncratic volatility. The largest negative leverage effect holds for the

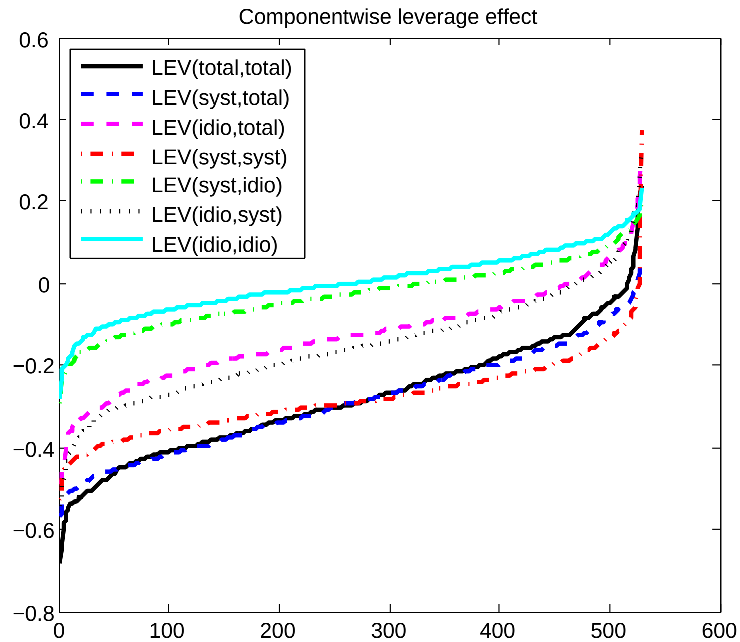


Figure 2.1: Componentwise leverage effect in 2012: Sorted correlations between total, systematic and idiosyncratic log-prices with total, systematic and idiosyncratic implied volatility. 4 asset factors and 1 volatility factor.

systematic components, while the correlation with the idiosyncratic volatility is on average zero.

My paper contributes to the central question in empirical and theoretical asset pricing what constitutes systematic risk. There are essentially three common ways of selecting which factors and how many describe the systematic risk. The first approach is based on theory and economic intuition. The capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) with the market as the only common factor falls into this category. The second approach bases factors on firm characteristics with the three-factor model of Fama and French (1993) as its most famous example. My approach falls into the third category where factor selection is statistical. This approach is motivated by the arbitrage pricing theory (APT) of Ross (1976). Factor analysis can be used to analyze the covariance structure of returns. This approach yields estimates of factor exposures as well as returns to underlying factors, which are linear combinations of returns on underlying assets. The notion of an "approximate factor model" was introduced by Chamberlain and Rothschild (1983), which allowed for a non-diagonal covariance matrix of the idiosyncratic component. Connor and Korajczyk (1986, 1988, 1993) study the use of principal component analysis in the case of an unknown covariance matrix, which has to be estimated.[5] One distinctive feature of the factor literature described above is that it uses long-horizon data. The advantages of using high-frequency data are apparent as it provides more information for more precise estimation and there are sufficiently many data points for estimating a factor structure that varies over longer time-horizons. For example the factor analysis can be pursued on a monthly basis to test how the factor structure changes over time.[6]

So far most of the empirical literature that utilizes the tools of high-frequency econometrics to analyze a factor structure is limited to a pre-specified set of factors. For example, Bollerslev, Li and Todorov (2015) estimate the betas for a continuous and jump market factor. Fan, Furger and Xiu (2014) estimate a large-dimensional covariance matrix with high-frequency data for a given factor structure. My work goes further as I estimate the unknown continuous and jump factor structure in a large cross-section. An exception is Aït-Sahalia and Xiu (2015a) who apply nonparametric principal component analysis to a low-dimensional cross-section of high-frequency data.[7]

My results were derived simultaneously and independently to results by Aït-Sahalia and Xiu (2015b). Both of our papers consider the estimation of a large-dimensional factor model based on high-frequency observations. From the theoretical side their work is different from my work as I also include jumps and provide a distribution theory. My main identification condition is a bounded eigenvalue condition on the idiosyncratic covariance matrix, while their identification is based on a sparsity assumption on the idiosyncratic covariance matrix. In the empirical part they consider a similar data set with 15 minutes data and show that 4 statistical factors are sufficient to obtain a block-diagonal pattern in the idiosyncratic covariance matrix. Their study focuses on estimating the continuous covariance matrix, while my work tries to explain the factor structure itself and also considers the factor structures in jumps and volatilities.

The rest of the paper is organized as follows. Section 2.2 introduces the factor model. In Section 2.3, I explain the estimation method. Section 2.4 analyzes the systematic pattern in equity data based on high-frequency data. Section 2.5 is an empirical application to volatility data and includes the analysis of the leverage effect. Concluding remarks are provided in

[5]The general case of a static large dimensional factor model is treated in Bai (2003) and Bai and Ng (2002). Forni, Hallin, Lippi and Reichlin (2000) introduced the dynamic principal component method. Fan, Liao and Mincheva (2013) study an approximate factor structure with sparsity.

[6]A disadvantage of working with high-frequency data is the relatively short time horizon for which appropriate data for a large cross-section is available. Arbitrage pricing theory links risk premiums to systematic risk. Factors that explain most of the comovements should also explain most of the risk premia. Unfortunately, the short-time horizon of 10 years puts restrictions on a reliable estimation of the risk premium and hence for testing this statement. Hence, this paper focusses on interpreting and understanding the properties of factors that explain most of the common comovements in the data without testing the asset pricing implications.

[7]My results were derived simultaneously and independently to results in Aït-Sahalia and Xiu (2015a). They find that the first three continuous principal components explain a large fraction of the variation in the S&P100 index. Their work is different from mine as they consider a low-dimensional regime for continuous processes, whereas I work in a large-dimensional regime and analyze both the continuous and jump structures.

Section 2.6. All the mathematical statements and additional empirical results are deferred to the appendices.

2.2 Factor Model

The theoretical foundation for my empirical results assumes an asymptotic framework in which the number of cross-sectional and high-frequency observations both go to infinity. The high number of cross-sectional observations makes the large dimensional covariance analysis challenging, but under the assumption of a general approximate factor structure the "curse of dimensionality" turns into a "blessing" as it becomes necessary for estimating the systematic factors. I argue that my data set with around 20,000 yearly observations for each of the 500 cross-sectional assets is sufficiently large for invoking asymptotic theory.[8]

This paper assumes that log asset prices can be modeled by an approximate factor model. Hence most co-movements in asset prices are due to a systematic factor component. In more detail assume that we have N assets with log prices denoted by $X_i(t)$.[9] Assume the N-dimensional stochastic process $X(t)$ can be explained by a factor model, i.e.

$$X_i(t) = \Lambda_i^\top F(t) + e_i(t) \qquad i = 1, ..., N \text{ and } t \in [0, T]$$

where Λ_i is a $K \times 1$ dimensional vector and $F(t)$ is a K-dimensional stochastic process. The loadings Λ_i describe the exposure to the systematic factors F, while the residuals e_i are stochastic processes that describe the idiosyncratic component. However, we only observe the stochastic process X at M discrete time observations in the interval $[0, T]$. If we use an equidistant grid[10], we can define the time increments as $\Delta_M = t_{j+1} - t_j = \frac{T}{M}$ and observe

$$X(t_j) = \Lambda F(t_j) + e(t_j) \qquad j = 1, ..., M.$$

with $\Lambda = (\Lambda_1, ..., \Lambda_N)^\top$ and $X(t) = (X_1(t), ..., X_N(t))^\top$. In our setup the number of cross-sectional observations N and the number of high-frequency observations M is large, while the time horizon T and the number of systematic factors K is fixed. The loadings Λ, factors F, residuals e and number of factors K are unknown and have to be estimated.

[8] I have run many robustness tests where I vary the number of cross-sectional and high-frequency observations and my general findings are not affected. The availability of reliable intra-day data for a large cross-section limits my study to the data that I am using in this thesis. I cannot rule out the possibility that with more data I could find additional factors that are persistent. My estimations indicate that there are four respectively three strong factors in the equity data and they seem to follow a very strong pattern, which makes me believe that this is not a pure data-mining but real economic phenomena. It is possible that other factors, e.g. a value factor, do not explain much of the correlation for my cross-section and hence are not identified as a systematic factor.

[9] Later in this paper I will also use volatilities for the process $X_i(t)$.

[10] My results would go through under a time grid that is not equidistant as long as the largest time increment goes to zero with speed $O\left(\frac{1}{M}\right)$.

We are also interested in estimating the continuous and jump component of the factors and the volatility of the factors. Denoting by F^C the factors that have a continuous component and by F^D the factor processes that have a jump component, we can write

$$X(t) = \Lambda^C F^C(t) + \Lambda^D F^D(t) + e(t).$$

Note, that for factors that have both, a continuous and a jump component, the corresponding loadings have to coincide. In the following we assume a non-redundant representation of the K^C continuous and K^D jump factors. For example if we have K factors which have all exactly the same jump component but different continuous components, this results in K different total factors and $K^C = K$ different continuous factors, but in only $K^D = 1$ jump factor.

My approach requires only very weak assumptions which are summarized in Appendix B.2 First, the individual asset price dynamics are modeled as Itô-semimartingales, which is the most general class of stochastic processes, for which the general results of high-frequency econometrics are available. It includes many processes, for example stochastic volatility processes or jump-diffusion processes with stochastic intensity rate. Second, the dependence between the assets is modeled by an approximate factor structure similar to Chamberlain and Rothschild (1983). The idiosyncratic risk can be serially correlated and weakly cross-sectionally correlated and hence allows for a very general specification. The main identification criterion for the systematic risk is that the quadratic covariation matrix of the idiosyncratic risk has bounded eigenvalues, while the quadratic covariation matrix of the systematic factor part has unbounded eigenvalues. For this reason the principal component analysis can relate the eigenvectors of the exploding eigenvalues to the loadings of the factors. Third, in order to separate continuous systematic risk from jump risk, we allow only finite activity jumps, i.e. there are only finitely many jumps in the asset price processes. Many of my results work without this restriction and it is only needed for the separation of these two components. This still allows for a very rich class of models and for example general compound poisson processes with stochastic intensity rates can be accommodated. Last but not least, we work under the simultaneous limit of a growing number of high-frequency and cross-sectional observations. We do not restrict the path of how these two parameters go to infinity. However, my results break down if one of the two parameters stays finite. In this sense the "curse of dimensionality" becomes a "blessing".

2.3 Estimation

2.3.1 Estimating the Factors

I employ the estimation techniques developed in Chapter 1. We have M observations of the N-dimensional stochastic process X in the time interval $[0, T]$. For the time increments $\Delta_M = \frac{T}{M} = t_{j+1} - t_j$ we denote the increments of the stochastic processes by

$$X_{j,i} = X_i(t_{j+1}) - X_i(t_j) \qquad F_j = F(t_{j+1}) - F(t_j) \qquad e_{j,i} = e_i(t_{j+1}) - e_i(t_j).$$

In matrix notation we have

$$\underset{(M\times N)}{X} = \underset{(M\times K)}{F}\underset{(K\times N)}{\Lambda^\top} + \underset{(M\times N)}{e}.$$

For a given K our goal is to estimate Λ and F. As in any factor model where only X is observed Λ and F are only identified up to invertible transformations. I impose the standard normalization that $\frac{\hat{\Lambda}^\top\hat{\Lambda}}{N} = I_K$ and that $\hat{F}^\top\hat{F}$ is a diagonal matrix.

The estimator for the loadings $\hat{\Lambda}$ is defined as the eigenvectors associated with the K largest eigenvalues of $\frac{1}{N}X^\top X$ multiplied by $\sqrt{N}$. The estimator for the factor increments is $\hat{F} = \frac{1}{N}X\hat{\Lambda}$. Note that $\frac{1}{N}X^\top X$ is an estimator for the quadratic covariation $\frac{1}{N}[X,X]$ for a finite N. The asymptotic theory is applied for $M, N \to \infty$. The systematic component of $X(t)$ is the part that is explained by the factors and defined as $C(t) = \Lambda F(t)$. The increments of the systematic component $C_{j,i} = F_j\Lambda_i^\top$ are estimated by $\hat{C}_{j,i} = \hat{F}_j\hat{\Lambda}_i^\top$.

Intuitively under some assumptions we can identify the jumps of the process $X_i(t)$ as the big movements that are larger than a specific threshold. I set the threshold identifier for jumps as $\alpha\Delta_M^{\bar{\omega}}$ for some $\alpha > 0$ and $\bar{\omega} \in \left(0, \frac{1}{2}\right)$ and define $\hat{X}^C_{j,i} = X_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ and $\hat{X}^D_{j,i} = X_{j,i}\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}$. The estimators $\hat{\Lambda}^C$, $\hat{\Lambda}^D$, $\hat{F}^C$ and $\hat{F}^D$ are defined analogously to $\hat{\Lambda}$ and $\hat{F}$, but using $\hat{X}^C$ and $\hat{X}^D$ instead of X. [11]

The quadratic covariation of the factors can be estimated by $\hat{F}^\top\hat{F}$ and the volatility component of the factors by $\hat{F}^{C\top}\hat{F}^C$. I show that the estimated increments of the factors $\hat{F}, \hat{F}^C$ and $\hat{F}^D$ can be used to estimate the quadratic covariation with any other process.

As I have already noted before, factor models are only identified up to invertible transformations. Two sets of factors represent the same factor model if the factors span the same vector space. When trying to interpret estimated factors by comparing them with economic factors, we need a measure to describe how close two vector spaces are to each other. As proposed by Bai and Ng (2006) the generalized correlation is a natural candidate measure. Let F be our K-dimensional set of factor processes and G be a K_G-dimensional set of economic candidate factor processes. We want to test if a linear combination of the candidate factors G can replicate some or all of the true factors F. The first generalized correlation is the highest correlation that can be achieved through a linear combination of the factors F and the candidate factors G. For the second generalized correlation we first project out the subspace that spans the linear combination for the first generalized correlation and then determine the highest possible correlation that can be achieved through linear combinations of the remaining $K-1$ respectively $K_G - 1$ dimensional subspaces. This procedure continues until we have calculated the $min(K, K_G)$ generalized correlation. Mathematically the generalized correlations are the square root of the $min(K, K_G)$[12] largest eigenvalues of the matrix $[F,G]^{-1}[F,F][G,G]^{-1}[G,F]$. If $K = K_G = 1$ it is simply the correlation as measured by the quadratic covariation. Similarly the distance between two loading matrices Λ and $\tilde{\Lambda}$

[11] For the jump threshold I use the TOD specification of Bollerslev, Li and Todorov (2013).

[12] Using $min(K, K_G)$ instead of $max(K, K_G)$ is just a labeling convention. All the generalized correlations after $min(K, K_G)$ are zero and hence usually neglected.

with dimension $N \times K$ respectively $N \times \tilde{K}$ is measured as the square root of the $\min(K, \tilde{K})$ largest eigenvalues of $(\Lambda^\top \Lambda)^{-1}\Lambda^\top \tilde{\Lambda}(\tilde{\Lambda}^\top \tilde{\Lambda})^{-1}\tilde{\Lambda}^\top \Lambda$. If the two matrices span the same vector spaces, the generalized correlations are all equal to 1. Otherwise they denote the highest possible correlations that can be achieved through linear combinations of the subspaces. If for example for $K = K_G = 3$ the generalized correlations are $\{1, 1, 0\}$ it implies that there exists a linear combination of the three factors in G that can replicate two of the three factors in F.[13] I have shown that under general conditions the estimated factors $\hat{F}$, $\hat{F}^C$ and $\hat{F}^D$ can be used instead of the true unobserved factors for calculating the generalized correlations. Unfortunately, in this high-frequency setting there does not exist a theory for confidence intervals for the individual generalized correlations. However, I have developed an asymptotic distribution theory for the sum of squared generalized correlations, which I label as total generalized correlation. I use the total generalized correlation test described in Section 1.8 to test if a set of economic factors represents the same factor model as the statistical factors.

The theorems and assumptions are collected in Chapter 1.

2.3.2 Estimating the Number of Factors

In Chapter 1 I also develop a new estimator for the number of factors, that can also distinguish between the number of continuous and jump factors. This estimator uses only the same weak assumptions that are needed for the consistency of my factor estimator. It can also easily be extended to long time horizon factor models and in simulations it outperforms the existing estimators while maintaining significantly weaker assumptions. Intuitively the large eigenvalues are associated with the systematic factors and hence the problem of estimating the number of factors is roughly equivalent to deciding which eigenvalues are considered to be large with respect to the rest of the spectrum. Under the assumptions that we need for consistency I can show that the first K "systematic" eigenvalues of $X^\top X$ are $O_p(N)$, while the nonsystematic eigenvalues are $O_p(1)$. A straightforward estimator for the number of factors considers the eigenvalue ratio of two successive eigenvalues and associates the number of factors with a large eigenvalue ratio. However, without very strong assumptions we cannot bound the small eigenvalues from below, which could lead to exploding eigenvalue ratios in the nonsystematic spectrum. I propose a perturbation method to avoid this problem. As long as the eigenvalue ratios of the perturbed eigenvalues cluster, we are in the nonsystematic spectrum. As soon as we do not observe this clustering any more, but a large eigenvalue ratio of the perturbed eigenvalues, we are in the systematic spectrum.

[13] Although labeling the measure as a correlation, we do not demean the data. This is because the drift term essentially describes the mean of a semimartingale and when calculating or estimating the quadratic covariation it is asymptotically negligible. Hence, the generalized correlation measure is based only on inner products and the generalized correlations correspond to the singular values of the matrix $[F, G]$ if F and G are orthonormalized with respect to the inner product $[.,.]$. The generalized correlation between two sets of loadings is a measure of how well we can describe one set as a linear combination of the other set.

The number of factors can be consistently estimated through the perturbed eigenvalue ratio statistic and hence, we can replace the unknown number K by its estimator $\hat{K}$. Denote the ordered eigenvalues of $X^\top X$ by $\lambda_1 \geq ... \geq \lambda_N$. We choose a slowly increasing sequence $g(N, M)$ such that $\frac{g(N,M)}{N} \to 0$ and $g(N, M) \to \infty$. Based on simulations a good choice for the perturbation term g is the median eigenvalue rescaled by $\sqrt{N}$, but the results are very robust to different choices of the perturbation.[14] Then, we define perturbed eigenvalues $\hat{\lambda}_k = \lambda_k + g(N, M)$ and the perturbed eigenvalue ratio statistic

$$ER_k = \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}} \qquad \text{for } k = 1, ..., N-1.$$

The estimator for the number of factors is defined as the first time that the perturbed eigenvalue ratio statistic does not cluster around 1 any more:

$$\hat{K}(\gamma) = \max\{k \leq N-1 : ER_k > 1+\gamma\} \qquad \text{for } \gamma > 0.$$

The definition of $\hat{K}^C(\gamma)$ and $\hat{K}^D(\gamma)$ is analogous but using λ_i^C respectively λ_i^D of the matrices $\hat{X}^{C\top}\hat{X}^C$ and $\hat{X}^{D\top}\hat{X}^D$. The results in my empirical analysis are robust to a wide range of values for the threshold γ.

2.4 High-Frequency Factors in Equity Data

2.4.1 Data

I use intraday log-prices from the Trade and Quote (TAQ) database for the time period from January 2003 to December 2012 for all the assets included in the S&P 500 index at any time between January 1993 and December 2012. In order to strike a balance between the competing interests of utilizing as much data as possible and minimizing the effect of microstructure noise and asynchronous returns, I choose to use 5-minute prices.[15] More details about the data selection and cleaning procedures are in Appendix B.1. For each of the 10 years we have on average 250 trading days with 77 log-price increments per day. Within each year we have a cross-section N between 500 and 600 firms.[16] The exact number for each year is in Table 2.2. After applying the cleaning procedure the intersection of the firms for the time period 2007 to 2012 is 498, while the intersection of all firms for the 10 years is only 304. The yearly results use all the available firms in that year, while the analysis over longer horizons uses the cross-sectional intersection.

[14] I estimate the number of factors using the perturbed eigenvalue ratio estimator with $g(N, M) = \sqrt{N} \cdot median\{\lambda_1, ..., \lambda_N\}$. For robustness I also use an unperturbed eigenvalue ratio test and $g(N, M) = log(N) \cdot median\{\lambda_1, ..., \lambda_N\}$. The results are the same.

[15] I have run robustness tests with 15min and daily data and the main results do not change.

[16] I do not extend my analysis to the time before 2003 as there are too many missing high-frequency observations for the large cross-section.

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Original	614	620	622	612	609	606	610	603	587	600
Cleaned	446	540	564	577	585	598	608	597	581	593
Dropped	27.36%	12.90%	9.32%	5.72%	3.94%	1.32%	0.33%	1.00%	1.02%	1.17%

Table 2.2: Observations after data cleaning

When identifying jumps, we face the tradeoff of finding all discontinuous movements against misclassifying high-volatility regimes as jumps. Therefore, the threshold should take into account changes in volatilities and intra-day volatility patterns. I use the TOD estimator of Bollerslev, Li and Todorov (2013) for separating the continuous from the jump movements. Hence the threshold is set as $a \cdot 77^{-0.49}\hat{\sigma}_{j,i}$, where $\hat{\sigma}_{j,i}$ estimates the daily volatility of asset i at time j by combining an estimated Time-of-Day volatility pattern with a jump robust bipower variation estimator for that day. Intuitively I classify all increments as jumps that are beyond a standard deviations of a local estimator of the stochastic volatility. For my analysis I use $a = 3$, $a = 4$ and $a = 4.5$.

Table 2.3 lists the fraction of increments identified as jumps for different thresholds. Depending on the year for $a = 3$ more than 99% of the observations are classified as continuous, while less than 1% are jumps. In 2012, 99.2% of the movements are continuous and explain around 85% of the total quadratic variation, while the 0.8% jumps explain the remaining 15% of the total quadratic covariation. Changing the threshold, we can either classify more or less movements as jumps.[17] All the results for the continuous factors are extremely robust to this choice. However, the results for the jump factors are sensitive to the threshold. Therefore, I am very confident about the results for the continuous factors, while the jump factor results have to be interpreted with caution. If not noted otherwise, the threshold is set to $a = 3$ in the following.

As a first step Table 2.3 lists for each year the fraction of the total continuous variation explained by the first four continuous factors and the fraction of the jump variation explained by the first jump factor. As expected systematic risk varies over time and is larger during the financial crisis. The systematic continuous risk with 4 factors accounts for around 40-47% of the total correlation from 2008 to 2011, but explains only around 20-31% in the other years.[18] A similar pattern holds for the jumps where the first jump factor explains up to 10 times more of the correlation in 2010 than in the years before the financial crisis.

I have applied the factor estimation to the quadratic covariation and the quadratic correlation matrix, which corresponds to using the covariance or the correlation matrix in

[17]There is no consensus on the number of jumps in the literature. Christensen, Oomen and Podolskij (2014) use ultra high-frequency data and estimate that the jump variation accounts for about 1% of total variability. Most studies based on 5 minutes data find that the jump variation should be around 10 - 20% of the total variation. My analysis considers both cases.

[18]The percentage of correlation explained by the first four factors is calculated as the sum of the first four eigenvalues divided by the sum of all eigenvalues of the continuous quadratic correlation matrix.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Percentage of increments identified as jumps										
a=3	0.011	0.011	0.011	0.010	0.010	0.009	0.008	0.008	0.007	0.008
a=4	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001
a=4.5	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.001
Variation explained by jumps										
a=3	0.19	0.19	0.19	0.16	0.21	0.16	0.16	0.15	0.12	0.15
a=4	0.07	0.07	0.07	0.05	0.10	0.06	0.06	0.06	0.03	0.05
a=4.5	0.05	0.04	0.05	0.04	0.08	0.04	0.05	0.05	0.02	0.04
Percentage of jump correlation explained by first 1 jump factor										
a=3	0.05	0.03	0.03	0.03	0.06	0.07	0.08	0.19	0.12	0.06
a=4	0.03	0.02	0.02	0.04	0.08	0.06	0.08	0.25	0.09	0.08
a=4.5	0.03	0.03	0.02	0.05	0.09	0.06	0.08	0.22	0.12	0.09
Percentage of continuous correlation explained by first 4 continuous factors										
	0.26	0.20	0.21	0.22	0.29	0.45	0.40	0.40	0.47	0.31

Table 2.3: (1) Fraction of increments identified as jumps for different thresholds. (2) Fraction of total quadratic variation explained by jumps for different thresholds. (3) Systematic jump correlation as measured by the fraction of the jump correlation explained by the first jump factor for different thresholds. (4) Systematic continuous correlation as measured by the fraction of the continuous correlation explained by the first four continuous factors.

long-horizon factor modeling. For the second estimator I rescale each asset for the time period under consideration by the square-root of its quadratic covariation. Of course, the resulting eigenvectors need to be rescaled accordingly in order to obtain estimators for the loadings and factors. All my results are virtually identical for the covariation and the correlation approach, but the second approach seems to provide slightly more robust estimators for shorter time horizons. Hence, all results reported in this paper are based on the second approach.

2.4.2 Continuous Factors

Number of Factors

I estimate four continuous factors for each of the years from 2007 to 2012 and three continuous factors for the years 2003 to 2006. Figure 2.2 shows the estimation results for the numbers of continuous factors. Starting from the right we are looking for a visible strong increase in the perturbed eigenvalue ratio. Asymptotically any critical value larger than 1 should indicate the beginning of the systematic spectrum. However, for our finite sample we need to choose a critical value. In the plots I set the critical value γ equal to 1.08. Fortunately there are very visible humps at 4 for the years 2007 to 2012 and strong increases at 3 for the years 2003 to 2006, which can be detected for a wide range of critical values. Therefore, my estimator strongly indicates that there are 4 continuous factors from 2007 to 2012 and three continuous factors from 2003 to 2006. As a robustness test in Figure B.1 I also use an unperturbed eigenvalue ratio statistic. The results are the same.[19]

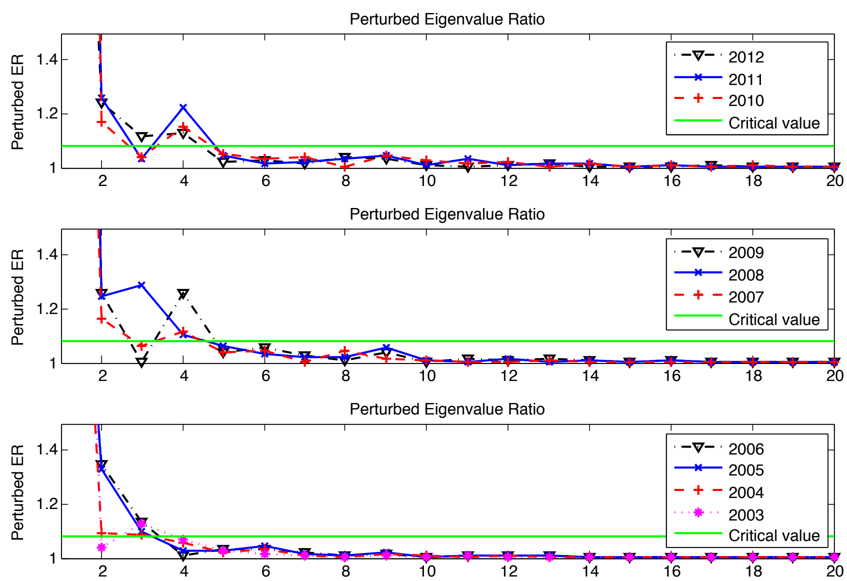


Figure 2.2: Number of continuous factors

In Figure 2.3 I apply the same analysis without separating the data into a continuous and jump component and obtain the same number of factors as in the continuous case. The perturbed eigenvalue ratios stop to cluster at the value 4 for 2007 to 2012 and at the value 3 for 2003 to 2006. This implies either that the continuous and jump factors are the same or that the continuous factors dominate the jump factors.

Persistence of Factors

The first four continuous factors are highly persistent for the time period 2007 to 2012, while there are three highly persistent factors for the time period 2003 to 2006. When comparing the systematic factor structures over time, I am interested if two sets of factors span the same vector space. I call a factor structure persistent if the vector spaces spanned by the factors stay constant. Persistence does not mean that the betas from a regression stay constant, which they usually do not do. If the factors estimated over a longer time horizon (e.g. 10 years) span the same vector space as factors estimated over all shorter

[19] I have conducted the same analysis for more perturbation functions with the same findings. The results are available upon request.

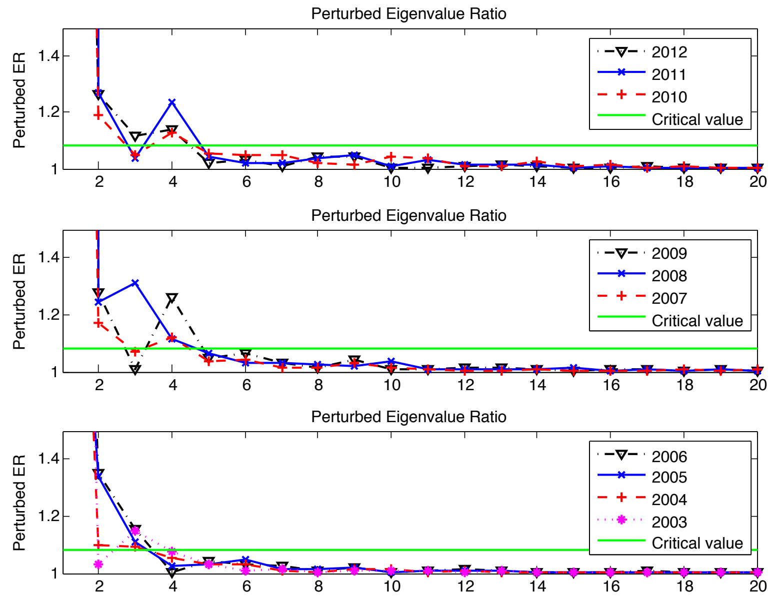


Figure 2.3: Number of total factors

horizons (e.g. 1 year) included in the longer period, persistence follows. The difficulty in comparing the factor structure over time is that the same set of factors can lead to different principal components. An economic factor that explains a large fraction of the variation in one year and hence is associated with a large eigenvalue, might explain less variation and be linked to a smaller eigenvalue in another year. The generalized correlations allow us to compare the vector spaces that are spanned by different sets of factors. In Table 2.4 I calculate the generalized correlations for the first four largest statistical factors based on yearly quadratic correlations matrices and on a six-year quadratic correlation matrix. The number of generalized correlations that are close to one essentially suggests how many of the factors in the two sets are the same. The results indicate that it does not matter if we use a one year or six years horizon for the time period 2007 to 2012 for estimating the factors. In the same table I also compare the yearly loadings with the six-year loadings, which are essentially the same and hence represent the same portfolio space. As the loadings could be interpreted as portfolio weights, the same set of loadings implies the same factors for the same time period. Hence we have two ways to show the persistence in the factor structure.

In Table 2.5, I show that the first four yearly statistical factors and loadings are essentially identical to the first four monthly statistical factors and loadings in the year 2011. Identical results hold for the other years. This is another strong indication for the persistence of the first four continuous factors.[20]

However, when doing the same analysis for the longer horizon 2003 to 2012 in Table 2.4,

[20]I have done the same analysis for all the years and I will provide the results upon request.

we observe that one factor disappears in 2003 to 2006. The first three generalized correlations are close to one, indicating that the two sets of factors share at least a three dimensional subspace, i.e. three of the factors coincide. The fourth generalized correlation for 2003 to 2006 however is significantly smaller implying that one of the four yearly factors cannot be written as a linear combination of the four factors estimated based on the 10 year horizon. This result is in line with my estimation results for the number of factors, where one factor seems to disappear before 2007. We observe exactly the same pattern for the loadings.

Factors, N=498					2007	2008	2009	2010	2011	2012
1. Generalized Correlation					1.00	1.00	1.00	1.00	1.00	1.00
2. Generalized Correlation					1.00	1.00	1.00	1.00	1.00	1.00
3. Generalized Correlation					0.99	0.99	1.00	1.00	1.00	0.98
4. Generalized Correlation					0.97	0.96	0.98	0.99	0.99	0.98
Loadings, N=498					2007	2008	2009	2010	2011	2012
1. Generalized Correlation					0.99	1.00	1.00	1.00	1.00	0.99
2. Generalized Correlation					0.97	0.99	0.99	0.99	0.99	0.97
3. Generalized Correlation					0.94	0.98	0.98	0.98	0.98	0.95
4. Generalized Correlation					0.93	0.97	0.96	0.97	0.95	0.93
Factors, N=302	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1. Generalized Correlation	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. Generalized Correlation	0.97	0.99	0.99	0.99	1.00	1.00	1.00	1.00	0.99	0.99
3. Generalized Correlation	0.95	0.97	0.98	0.99	0.99	0.99	0.97	0.98	0.99	0.98
4. Generalized Correlation	**0.47**	**0.63**	**0.17**	**0.67**	0.99	0.99	0.94	0.92	0.97	0.96
Loadings, N=302	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1. Generalized Correlation	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	0.99
2. Generalized Correlation	0.91	0.96	0.97	0.97	0.98	0.99	0.99	0.99	0.98	0.96
3. Generalized Correlation	0.86	0.92	0.93	0.95	0.97	0.97	0.95	0.96	0.96	0.94
4. Generalized Correlation	**0.34**	**0.52**	**0.16**	**0.57**	0.95	0.96	0.90	0.88	0.93	0.91

Table 2.4: Persistence of continuous factors: Generalized correlations of the first four largest yearly continuous factors and their loadings with the first four statistical continuous factors and loadings for 2007-2012 respectively 2003-2012.

Interpretation of Factors

The four persistent continuous factors for 2007 to 2012 can be approximated very well by industry factors. The loading estimators can essentially be interpreted as portfolio weights for the factor construction. Simple eyeballing indicates that the first statistical factor seems

1	2	3	4	5	6	7	8	9	10	11	12
Generalized correlations of monthly with yearly continuous factors											
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.99	0.99	0.99	0.99	0.99	1.00	1.00	0.99	0.99	1.00	0.99	0.99
0.98	0.93	0.99	0.97	0.98	0.98	0.98	0.99	0.99	0.96	0.90	0.96
Generalized correlations of monthly with yearly continuous loadings											
0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.96	0.96	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.99	0.98	0.97
0.95	0.95	0.96	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.96
0.94	0.86	0.94	0.90	0.93	0.94	0.94	0.95	0.96	0.90	0.84	0.91

Table 2.5: Persistence of continuous factors in 2011. Generalized correlation of monthly continuous factors and loadings with yearly continuous factors and loadings. The yearly number of factors is $K = 4$.

to be an equally weighted market portfolio, a result which has already been confirmed in many studies. The loadings for the second to fourth statistical factors have a very particular pattern: Banks and insurance companies have very large loadings with the same sign, while firms related to oil and gas have large loadings with the opposite sign. Firms related to electricity seem to have their own pattern unrelated to the previous two. Motivated by these observations I construct four economic factors as

- Market (equally weighted)
- Oil and gas (40 equally weighted assets)
- Banking and Insurance (60 equally weighted assets)
- Electricity (24 equally weighted assets)

The details are in Appendix B.1.1.

The generalized correlations of the market, oil and finance factors with the first four largest statistical factors for 2007 to 2012 are very high as shown in the first analysis of Table 2.6. This indicates that three of the four statistical factors can almost perfectly be replicated by the three economic factors. This relationship is highly persistent over time. In Table 2.6 the top of the first column uses the factors and generalized correlations based on a 6 year horizon, while in the last six columns I estimate the yearly statistical factors and calculate their generalized correlations with the yearly market, oil and finance factors. The generalized correlations close to one indicates that at least three of the statistical factors do not change over time and are persistent.

Identifying the fourth continuous factor is challenging and the closest approximation seems to be an electricity factor. The second analysis in Table 2.6 shows the generalized correlations of the four continuous statistical factors for 2007 to 2012 with the four economic factors. The fourth generalized correlation essentially measures how well the additional electricity factor can explain the remaining statistical factor. The fourth yearly generalized correlation takes values between 0.75 and 0.87, which means that the electricity factor can help substantially to explain the statistical factors, but it is not sufficient to perfectly replicate them. The first column shows the result for the total six year time horizon while the last six columns list the yearly results. In conclusion it seems that the relationship between the four economic and statistical factors is persistent over time.

The results in Subsection 2.4.2 indicate that the factor structure in 2003 to 2006 might be different compared to the later period. Based on the intersection of all the firms for 2003 to 2012 I analyze the generalized correlations of the first four yearly continuous statistical factors with the four yearly continuous industry factors. The third analysis in Table 2.6 shows that as expected one factor disappears in the early four years. A fourth generalized correlation between 0.16 and 0.35 for 2003 to 2006 suggests strongly that the statistical factors and industry factors have at most three factors in common. The fourth, fifth and sixth analyses in Table 2.6 try to identify the disappearing factor. Looking at the fifth analysis it seems that dropping the finance factor for the time period 2003 to 2006 leads to the smallest reduction in generalized correlations, i.e. the three statistical factors for 2003 to 2006 are not well-explained by a finance factor. On the other hand this finance factor is crucial for explaining the statistical factors for 2007 to 2012.

As a robustness test I extend the analysis to daily data and also include daily times-series of the Fama-French-Carhart factors. First I calculate daily continuous returns by adding up the continuous log-price increments for each day and creating this way the four continuous daily statistical factors. Then I calculate generalized correlations of the daily continuous factors with economic daily factors. I always include the daily continuous market, oil and finance factors in the analysis and in addition include either

1. Case 1: no additional factor
2. Case 2: a daily continuous electricity factor
3. Case 3: size, value and momentum factors
4. Case 4: daily continuous electricity, size, value and momentum factors

The results are summarized in Table 2.7. Obviously, the size, value and momentum factor do not explain much variation beyond the industry factors. The fourth generalized correlation in case 2 is with 0.81 almost the same as in case 4. In particular, the fourth factor seems to be much better explained by an electricity factor than by a size, value or momentum factor, which only account for a fourth generalized correlation of 0.43 in case 3.

Gen. corr. of 4 continuous factors with market, oil and finance factors										
N=498	2007-2012				2007	2008	2009	2010	2011	2012
1. Gen. Corr.	**1.00**				1.00	1.00	1.00	1.00	1.00	1.00
2. Gen. Corr.	**0.98**				0.98	0.97	0.98	0.97	0.98	0.93
3. Gen. Corr.	**0.95**				0.91	0.95	0.94	0.93	0.97	0.87
Gen. corr. of 4 continuous factors with market, oil, finance and electricity factors										
N=498	2007-2012				2007	2008	2009	2010	2011	2012
1. Gen. Corr.	**1.00**				1.00	1.00	1.00	1.00	1.00	1.00
2. Gen. Corr.	**0.98**				0.98	0.97	0.99	0.97	0.98	0.93
3. Gen. Corr.	**0.95**				0.91	0.95	0.95	0.93	0.94	0.90
4. Gen. Corr.	**0.80**				0.87	0.78	0.75	0.75	0.80	0.76
Gen. corr. of 4 continuous factors with market, oil, finance and electricity factors										
N=302	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1. Gen. Corr.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. Gen. Corr.	0.97	0.99	1.00	1.00	0.99	0.97	0.98	0.96	0.98	0.95
3. Gen. Corr.	0.57	0.75	0.77	0.89	0.85	0.92	0.95	0.92	0.93	0.83
4. Gen. Corr.	**0.10**	**0.23**	**0.16**	**0.35**	0.82	0.74	0.72	0.68	0.78	0.78
Gen. corr. of 4 continuous factors with market, oil and finance factors										
N=302	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1. Gen. Corr.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. Gen. Corr.	0.97	0.99	1.00	1.00	0.99	0.97	0.98	0.96	0.97	0.94
3. Gen. Corr.	**0.46**	**0.49**	**0.47**	**0.49**	0.84	0.92	0.94	0.89	0.93	0.83
Gen. corr. of 4 continuous factors with market, oil and electricity factors										
N=302	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1. Gen. Corr.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. Gen. Corr.	0.97	0.99	1.00	1.00	0.98	0.97	0.95	0.94	0.96	0.93
3. Gen. Corr.	**0.36**	**0.64**	**0.97**	**0.84**	0.83	0.76	0.73	0.69	0.78	0.78
Gen. corr. of 4 continuous factors with market, finance and electricity factors										
N=302	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1. Gen. Corr.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. Gen. Corr.	0.57	0.75	0.98	0.89	0.88	0.92	0.98	0.94	0.95	0.85
3. Gen. Corr.	**0.19**	**0.27**	**0.57**	**0.45**	0.83	0.74	0.73	0.72	0.78	0.78

Table 2.6: Interpretation of statistical continuous factors. Generalized correlation of economic factors (market, oil, finance and electricity factors) with first four largest statistical factors for different time periods.

	Case 1	Case 2	Case 3	Case 4
1. Generalized Correlation	1.00	1.00	1.00	1.00
2. Generalized Correlation	0.99	0.99	0.98	0.99
3. Generalized Correlation	0.95	0.95	0.95	0.95
4. Generalized Correlation		**0.81**	**0.43**	**0.83**

Table 2.7: Generalized correlations of the 4 daily continuous factors with daily economic factors for 2007-2012. N=498.

R^2 for daily excess returns of portfolios based on continuous loadings				
N=498	1. Factor	2. Factor	3. Factor	4. Factor
Short	1.00	0.96	0.95	0.97
Long	1.00	0.97	0.95	0.97
Generalized correlations with 4 economic industry factors				
	1.00	**0.97**	**0.92**	**0.79**
Generalized correlations with 4 Fama-French Carhart Factors				
	0.95	0.74	0.60	0.00

Table 2.8: (1) R^2 for portfolios of daily CRSP excess returns based on continuous loadings $\hat{\Lambda}^C$. (2) Generalized correlations of daily CRSP excess returns based on continuous loadings with 4 economic factors. (3) Generalized correlations of daily CRSP excess returns based on continuous loadings with 4 Fama-French Carhart Factors. The time period is 2007-2012 and N=498.

However, the Fama-French-Carhart factors are based on daily excess returns, which also include jumps and overnight returns. Additionally, daily returns are also mathematically different from daily increments in log asset prices. Hence the comparison with daily continuous returns might be misleading. Hence, I construct the 4 statistical and four economic industry factors using daily excess returns from CRSP. The estimated continuous loadings serve again as the portfolio weights for the statistical factors. Based on the daily excess returns from 2007 to 2012, I run simple OLS regressions in order to explain the four statistical factors. The short regression uses a market, oil, finance and electricity factor, while the long regression applies the same regressors and additionally the size, value and momentum factors. Table 2.8 shows that almost all the variation can be explained by the industry factors, while adding the Fama-French-Carhart factors does not change the explanatory power. In the second part of Table 2.8 I repeat a similar analysis as in Table 2.7 but using the daily excess returns. As before the generalized correlations with the 4 economic factors are very large indicating that a linear combination of daily excess returns of the industry portfolios can approximate the excess returns of the statistical factors very well. On the other hand the best linear combination of daily excess returns of the Fama-French Carhart Factors provides only a poor approximation to the statistical factor returns.

	4 statistical and 3 economic factors			4 statistical and 4 economic factors		
	$\hat{\bar{\rho}}$	SD	95% CI	$\hat{\bar{\rho}}$	SD	95% CI
2007-2012	2.72	0.001	(2.71, 2.72)	3.31	0.003	(3.30, 3.31)
2007	2.55	0.06	(2.42, 2.67)	3.21	0.01	(3.19, 3.22)
2008	2.66	0.08	(2.51, 2.81)	3.18	0.29	(2.62, 3.75)
2009	2.86	0.10	(2.67, 3.05)	3.42	0.15	(3.14, 3.71)
2010	2.80	0.04	(2.72, 2.88)	3.38	0.01	(3.37, 3.39)
2011	2.82	0.00	(2.82, 2.82)	3.47	0.06	(3.35, 3.58)
2012	2.62	0.03	(2.56, 2.68)	3.25	0.01	(3.24, 3.26)

Table 2.9: Total generalized correlations (=sum of squared generalized correlations) with standard deviations and confidence intervals for the four statistical factors with three economic factors (market, oil and finance) and four economic factors (additional electricity factor). Number of assets $N = 498$.

As a last step I apply the statistical test of Section 1.8 to test if the three respectively four continuous economic factors can perfectly replicate the statistical factors. So far I have not provided confidence intervals for the generalized correlations. Unfortunately, in this high-frequency setting there does not exist a theory for confidence intervals for the individual generalized correlations. However, I have developed an asymptotic distribution theory for the sum of squared generalized correlations, which I label as total generalized correlation. The left part of Table 2.9 lists the total generalized correlation for different time periods for three economic factors. A total generalized correlation of three indicates that three of the four statistical factors can be perfectly replicated by the three economic factors. Only in 2009 we cannot reject the null hypothesis of a perfect factor replication. In the right half of Table 2.9 I apply the same test to four economic factors. Now a total generalized correlation of four implies that the four statistical factors are identical to the four economic factors. We reject the null hypothesis of a perfect linear combination. Hence, although my set of economic factors approximates the statistical factors very well, there seems to be a missing component.

2.4.3 Jump Factors

There seems to be a lower number of jump factors, which do not coincide with the continuous factors. Only the jump market factor seems to be persistent, while neither the number nor the structure of the other jump factors have the same persistence as for the continuous counterpart. Figures B.2, B.3 and B.4 estimate the number of jump factors for different thresholds. In most years the estimator indicates only one jump factor. Under almost all specifications there seems to be at most four jump factors and hence I will restrict the following analysis to the first four largest jump factors.

	Yearly vs. 6-year jump factors						Yearly vs. 6-year jump loadings					
	2007	2008	2009	2010	2011	2012	2007	2008	2009	2010	2011	2012
a=3	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.95**	**0.99**	**0.98**	**0.98**	**0.97**	**0.95**
	0.96	**1.00**	**0.95**	**0.98**	0.76	**0.85**	**0.84**	**0.97**	**0.84**	**0.85**	0.49	0.68
	0.81	**0.88**	**0.84**	0.69	0.59	0.70	0.58	0.74	0.60	0.28	0.40	0.46
	0.12	**0.81**	0.14	0.25	0.05	0.17	0.05	0.63	0.07	0.15	0.03	0.07
a=4	**1.00**	**1.00**	**0.99**	**1.00**	**0.99**	**0.99**	**0.88**	**0.97**	**0.81**	**0.95**	**0.85**	**0.85**
	0.66	**0.99**	0.63	**1.00**	0.51	0.43	0.30	**0.87**	0.20	**0.93**	0.16	0.13
	0.34	0.52	0.09	**0.97**	0.43	0.20	0.10	0.23	0.02	0.66	0.11	0.05
	0.14	0.03	0.05	0.17	0.13	0.03	0.05	0.01	0.01	0.04	0.03	0.01
a=4.5	**0.99**	**0.99**	**0.98**	**1.00**	**0.99**	**0.99**	**0.85**	**0.95**	0.72	**0.95**	0.75	**0.82**
	0.79	**0.97**	0.77	**1.00**	0.40	0.49	0.30	0.78	0.24	**0.93**	0.10	0.14
	0.28	0.44	0.28	**0.96**	0.26	0.24	0.08	0.13	0.06	0.67	0.06	0.05
	0.05	0.16	0.01	0.53	0.11	0.07	0.01	0.05	0.00	0.12	0.03	0.02

Table 2.10: Generalized correlations of 4 largest yearly jump factors with 4 jump factors for 2007-2012 and generalized correlations of 4 yearly jump loadings with 4 jump loadings for 2007-2012 for different thresholds. Here $K = 4$ and $N = 498$. Values larger than 0.8 are in bold.

In Table 2.10 I analyze the persistence of the jump factors by comparing the estimation based on 6 years with the estimation based on yearly data. For the smallest threshold there seems to be two persistent factors as the first two generalized correlations are close to 1, but the structure is much less persistent than for the continuous data. For the largest threshold only the first generalized correlation is close to one suggesting only one persistent factor. In Table B.1 we see that for the shorter time horizon of a month only one factor is persistent independently of the threshold. This could be explained by the fact that the systematic jumps do not necessarily happen during every month and hence the systematic structure measured on a monthly basis can be very different from longer horizons.

My estimator for identifying the jumps might erroneously classify high volatility time periods as jumps. Increasing the threshold in the estimator reduces this error, while we might misclassify small jumps as continuous movements. Increasing the threshold, reduces the persistence in the jump factors up to the point where only a market jump factors remains. It is unclear if the persistence for small jump thresholds is solely due to misclassified high volatility movements.

Table 2.11 confirms that the jump factors are different from the continuous factors. Here I estimate the generalized correlations of the first four statistical jump factors with the market, oil, finance and electricity jump factors for 2007 to 2012. I can show that the first statistical jump factor is essentially the equally weighted market jump factor which is responsible for the first generalized correlation to be equal to 1. However, the correlations between the other statistical factors and the industry factors are significantly lower.

Generalized correlations of 4 economic jump factors with 4 statistical jump factors							
	2007-2012	2007	2008	2009	2010	2011	2012
a=3	**1.00**	**1.00**	**1.00**	**0.99**	**1.00**	**1.00**	**1.00**
	0.85	**0.95**	0.62	**0.86**	**0.81**	**0.86**	**0.83**
	0.61	0.77	0.40	0.76	0.31	0.61	0.59
	0.21	0.10	0.22	0.50	0.10	0.20	0.28
a=4	**0.99**	**0.99**	**0.95**	**0.94**	**1.00**	**0.99**	**0.99**
	0.74	0.53	0.41	0.59	**0.90**	0.53	0.57
	0.31	0.35	0.29	0.44	0.39	0.35	0.42
	0.03	0.19	0.20	0.09	0.05	0.14	0.16
a=4.5	**0.99**	**0.99**	**0.91**	**0.91**	**1.00**	**0.98**	**0.99**
	0.75	0.54	0.41	0.56	**0.93**	0.55	0.75
	0.29	0.35	0.30	0.40	0.68	0.38	0.29
	0.05	0.18	0.22	0.04	0.08	0.03	0.05

Table 2.11: Generalized correlations of market, oil, finance and electricity jump factors with first 4 jump factors from 2007-2012 for N=498 and for different thresholds. Values larger than 0.8 are in bold.

2.4.4 Comparison with Daily Data and Total Factors

The continuous factors dominate the jump factors and daily returns give similar but noisier estimators than the continuous high-frequency analysis. In this section I compare the estimators based on continuous, jump and total high-frequency data and daily CRSP returns. As I make the comparisons for each year separately, I can use my largest cross-sectional sample as listed in Table 2.2. I compare the loadings based on daily, total and jump high-frequency loadings with the continuous loadings. As the loadings can be interpreted as portfolio weights, the same set of loadings also implies the same factors. However, some assets might be close substitutes in which case different portfolios might still describe the same factors. Thus, I use the loadings estimated from the different data sets to construct continuous factors and estimate the distance between the different sets of continuous factors.

Figure B.5 shows the yearly estimators for the number of factors based on approximately 250 daily observations. The pattern is similar to the continuous estimators, but much noisier. This can be either due to the fact that the number of observations is much smaller then the approximate 20,000 in the high-frequency case, but also because the daily returns include the jumps and overnight movements. In Table B.2 we observe that the continuous factors and loadings are close to but different from those based on daily CRSP returns. This is a positive finding as it indicates that my results are in some sense robust to the frequency. On the other hand the high-frequency estimator seems to estimate the pattern in the data more precisely and there is a gain from moving from daily to intra-day data.

In Table B.2 I also show that the total factors and continuous high-frequency factors are essentially identical. This result has two consequences. First, it confirms that my findings about the continuous factors are robust to the jump threshold. Even if all the movements are classified as continuous, we obtain essentially the same estimators for the loadings. Second, the systematic continuous pattern dominates the systematic jump pattern. The first jump factor is a market jump factor and hence is described by the same loadings as the first continuous factor. Even if there are systematic jump factors that are different from the second to fourth continuous factors their impact on the spectrum is so small, that it is not detected when considering only the first four total factors. This is also partly due to the fact that the jump quadratic covariation is only a small fraction of the total quadratic covariation.

Finally in Table B.3 I confirm that the systematic jump factors are different from the systematic continuous factors. The higher the jump threshold the less likely it is that the large increments are due to continuous movements with high volatility. Thus for a larger jump threshold the correlation between the jump factors and continuous factors decreases up to the point where only the market factor remains as having a common continuous and jump component.

2.4.5 Microstructure Noise

Non-synchronicity and microstructure noise are two distinguishing characteristics of high-frequency financial data. First, the time interval separating successive observations can be random, or at least time varying. Second, the observations are subject to market microstructure noise, especially as the sampling frequency increases. The fact that this form of noise interacts with the sampling frequency distinguishes this from the classical measurement error problem in statistics. Inference on the volatility of a continuous semimartingale under noise contamination can be pursued using smoothing techniques.[21] However, neither the microstructure robust estimators nor the non-synchronicity robust estimators can be easily extended to our large dimensional problem. The main results of my paper assume synchronous data with negligible microstructure noise. Using for example 5-minute sampling frequency as commonly advocated in the literature on realized volatility estimation, e.g. Andersen et al. (2001) and the survey by Hansen and Lunde (2006), seems to justify this assumption.

[21] Several approaches have been developed, prominent ones by Zhang (2006), Barndorff-Nielsen et al. (2008) and Jacod et al. (2009) in the one-dimensional setting and generalizations for a noisy non-synchronous multi-dimensional setting by Aït-Sahalia et al. (2010), Podolskij and Vetter (2009), Barndorff-Nielsen et al. (2011) and Bibinger and Winkelmann (2014) among others.

Volatility signature plots as used in Hansen and Lunde (2006) are very useful tools for identifying frequencies that are contaminated by noise. One common approach in the literature is to sample at lower frequencies to minimize the contamination by microstructure noise at the cost of using less data. Clearly when estimating the quadratic covariation without applying microstructure noise corrections there is a tradeoff between the higher noise variance and higher precision of using a finer frequency. An important question in this respect is the variance of the unobservable noise. For example Hansen and Lunde (2006) have estimated the microstructure noise variance for different assets. In Chapter 1 I propose an estimator for the microstructure noise that utilizes the information in the cross-section. Under the assumptions outlined in Theorem 1.10, the increments of microstructure noise create a very specific spectral pattern. This allows us to derive upper bounds on the variance of the microstructure noise. These bounds are solely functions of the estimated eigenvalues and the ratio $\frac{M}{N}$. From a practical perspective it is ambiguous how to choose M for a given time horizon. For example Lee and Mykland (2009) use a year as the reference horizon for high to low frequencies, i.e. in our case M would be around $250 \cdot 77 = 19,250$. One could also argue that a month is a better cutoff between high and low frequencies which would set M to around $21 \cdot 77 = 1,617$. Obviously the results of my estimator for the microstructure noise variance depend on this choice.

Table 2.12 shows the estimation results for the different years. For a monthly reference level for high to low frequencies the upper bounds are very similar to the estimates in Hansen and Lunde (2006). For a yearly reference level they are significantly lower. In either case we would have microstructure noise contamination that can be neglected when using 5 minute data. The fact that my results are robust to different time horizons, e.g. 5 minutes, 15 minutes and daily horizons, further confirms that the results are robust to microstructure noise.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Median eigenvalue	0.048	0.036	0.033	0.034	0.040	0.119	0.070	0.030	0.032	0.025
Var of MN (T=12)	0.029	0.034	0.034	0.037	0.046	0.141	0.088	0.036	0.036	0.030
Var of MN (T=1)	0.001	0.001	0.001	0.001	0.001	0.003	0.002	0.001	0.001	0.001

Table 2.12: Estimation of the upper bound of the variance of microstructure noise for different years and different reference levels. (1) Median eigenvalue of yearly quadratic covariation matrix. (2) Upper bound on microstructure noise variance if the reference level for high and low frequency is a month (i.e. M is around 1,617). (3) Upper bound on microstructure noise variance if the reference level for high and low frequency is a year (i.e. M is around 19,250).

2.5 Empirical Application to Volatility Data

Using implied volatilities from option price data, I analyze the systematic factor structure of the volatilities. There is only one persistent market volatility factor, while during the financial crisis an additional temporary banking volatility factor appears. Based on the estimated factors, we can decompose the leverage effect, i.e. the correlation of the asset return with its volatility, into a systematic and an idiosyncratic component. The negative leverage effect is mainly driven by the systematic component, while the idiosyncratic component can be positively correlated with the volatility. These findings are important as they can rule out popular explanations of the leverage effect, which do not distinguish between systematic and non-systematic risk.

2.5.1 Volatility Factors

As the volatility of asset price processes is not observed, we cannot directly apply our factor analysis approach to the data. There are essentially two ways to estimate the volatility. The first is based on high-frequency asset prices and estimates the spot volatility using the realized quadratic covariations for a short time window. The second approach infers the volatility under the risk-neutral measure using option price data. The VIX is the most prominent example for the second approach. I have pursued both approaches, but most of the results reported in this chapter are based on the second one.

Using the realized quadratic variation for a short horizon, e.g. a day, we can obtain estimators for the spot volatilities, which we can use for our large-dimensional factor analysis. The details for the estimation of the spot volatilities and the construction of volatility of volatility estimators can be found in chapter 8 of Aït-Sahalia and Jacod (2014). The volatility of volatility estimator requires a bias correction. These estimators appear to be very noisy in practice and that is why I prefer an alternative approach.

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Original	408	479	507	525	543	557	565	561	549	565
Cleaned	399	465	479	495	508	528	536	530	523	529
Dropped	2.21%	2.92%	5.52%	5.71%	6.45%	5.21%	5.13%	5.53%	4.74%	6.37%

Table 2.13: Observations after data cleaning.

Medvedev and Scaillet (2007) show that under a jump-diffusion stochastic volatility model the Black-Scholes implied volatility of an at-the-money option with a small time-to-maturity is close to the unobserved volatility. Using this insight I use implied volatilities for my factor estimation approach.[22] Under relatively general conditions estimating the factor structure

[22]A rigorous study would require us to take into account the estimation error for the implied volatility and

(and also later the leverage effect) under the risk-neutral measure yields the same results as under the physical measure.

In some sense I am trying to create a VIX-type times-series for all the assets in the cross-section. The older version of VIX, the VXO, was actually a measure of implied volatility calculated using 30-day S&P100 index at-the-money options. The VIX uses the concept of generalized implied volatility. I have also created a panel of generalized implied volatilities for my cross-section. However, the theoretical justification of this approach assumes an infinite number of strike prices for each asset and the quality of the estimator deteriorates for a small number of available strikes as it is the case for many assets in my sample. Therefore, the Black-Scholes implied volatility appears to be a much more robust estimator. For the assets in our sample, where we have a large number of strikes available, the generalized volatility and simple implied-volatility are very close, while for those with only few strike prices, the generalized volatility estimators seem to be unreliable.

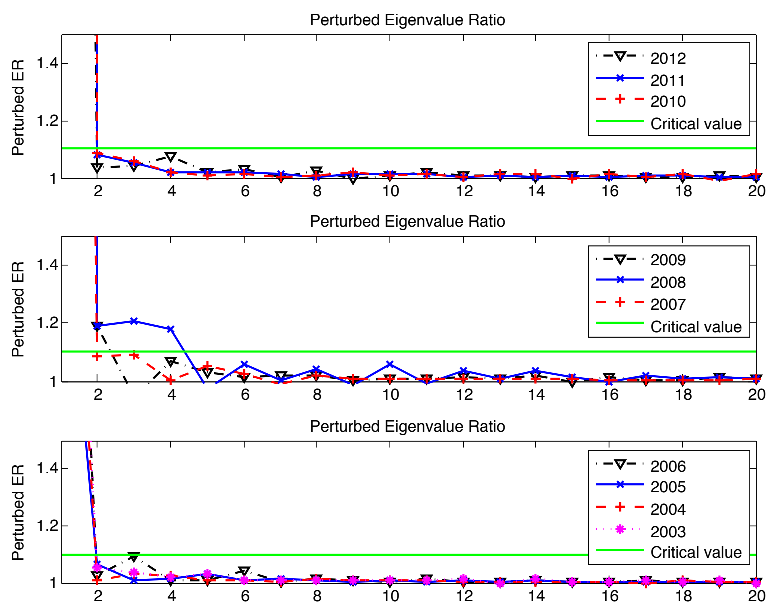


Figure 2.4: Number of volatility factors

Using daily implied volatilities from OptionMetrics for the same assets and time period as in the previous section, I apply my factor analysis approach. OptionMetrics provides implied volatilities for 30 days at-the-money standard call and put options using a linearly interpolated volatility surface. I average the implied call and put volatilities for each asset and each day. More details about the data and some results are in Appendix B.1.5. Unfor-

to derive the simultaneous limit of M, N, the at-the-moneyness and the maturities of the options. As such an extension is beyond the scope of the paper, I treat the implied volatilities of short-maturity at-the-money options as the true observed volatilities under the risk-neutral measure.

tunately, the data is not available to construct intra-day implied volatilities for the whole cross-section. However, as the previous section has illustrated, the results with daily data seem to capture similar results as with higher frequency data. Table 2.13 reports the data after the data cleaning.

Generalized correlations of 4 economic volatility factors with 4 statistical volatility factors										
					2007	2008	2009	2010	2011	2012
1. Gen. Corr.					1.00	1.00	1.00	1.00	1.00	1.00
2. Gen. Corr.					0.19	**0.90**	**0.92**	0.33	0.67	0.28
3. Gen. Corr.					0.07	0.34	0.13	0.06	0.11	0.05
4. Gen. Corr.					0.01	0.05	0.00	0.00	0.01	0.01
Generalized correlations between volatility and continuous loadings										
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1. Gen. Corr.	0.95	0.91	0.84	0.92	0.96	0.98	0.97	0.98	0.98	0.97
2. Gen. Corr.	0.59	0.68	0.84	0.34	0.69	0.79	0.57	0.37	0.31	0.21
3. Gen. Corr.	0.46	0.56	0.31	0.34	0.08	0.57	0.52	0.09	0.31	0.10
4. Gen. Corr.	0.14	0.11	0.14	0.03	0.08	0.12	0.02	0.08	0.02	0.05
Generalized correlations between volatility and jump loadings										
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1. Gen. Corr.	0.92	0.90	0.88	0.82	0.93	0.94	0.96	0.93	0.93	0.93
2. Gen. Corr.	0.47	0.61	0.72	0.33	0.21	0.20	0.46	0.15	0.38	0.09
3. Gen. Corr.	0.29	0.52	0.27	0.31	0.16	0.19	0.46	0.15	0.11	0.09
4. Gen. Corr.	0.29	0.10	0.04	0.07	0.16	0.19	0.03	0.10	0.03	0.01
Generalized correlations between volatility and daily loadings										
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1. Gen. Corr.	0.95	0.93	0.84	0.92	0.96	0.98	0.96	0.98	0.97	0.96
2. Gen. Corr.	0.44	0.64	0.84	0.34	0.61	0.64	0.46	0.29	0.49	0.23
3. Gen. Corr.	0.44	0.64	0.39	0.23	0.29	0.64	0.46	0.19	0.26	0.22
4. Gen. Corr.	0.22	0.06	0.21	0.18	0.02	0.38	0.02	0.07	0.03	0.09

Table 2.14: Generalized correlations between four economic (market, oil, finance and electricity) and four statistical volatility factors and between the loadings of the volatility factors and the loadings for continuous, jump and daily data.

Figure 2.4 estimates the number of volatility factors. There seems to be only one strong persistent factor, which is essentially a market volatility factor and very highly correlated with the VIX. In 2008 there seems to be a second temporary volatility factor. The volatility factors seem to be different from the continuous, jump and daily factors. Table 2.14 shows the generalized correlations of the volatility loadings with the continuous, jump and daily loadings for each year. The volatility loadings cannot be interpreted as a portfolio of assets, but rather as a portfolio of volatilities. There does not seem to be a strong correlation with the other factors. Table 2.14 also calculates the generalized correlation of the first four volatility factors with market, oil, finance and electricity volatility factors constructed as equally weighted portfolios of volatilities for these industries. In 2008 and 2009 there seems to appear a temporary finance factor as the second generalized correlation jumps up.[23] This finding is in line with the results for the number of factors and not surprising given the financial crisis.

2.5.2 Leverage Effect

One of the most important empirical stylized facts about the volatility is the leverage effect, which describes the generally negative correlation between an asset return and its volatility changes. The term "leverage" originates in one possible economic interpretation of this phenomenon, developed in Black (1976) and Christie (1982). When asset prices decline, companies become mechanically more leveraged as equity is a residual claim and the relative value of their debt rises relative to that of their equity. Therefore, their stock should become riskier, and as a consequence more volatile. Although this is only a hypothesis, this explanation has coined the term "leverage effect" to describe the statistical regularity in the correlation between asset return and volatility.[24]

There is no consensus on the economic explanation for this statistical effect. The magnitude of the effect seems to be too large to be explained by financial leverage.[25] Alternative economic interpretations as suggested for example by French et al. (1987) and Campbell and Hentschel (1992) use a risk-premium argument. An anticipated rise in volatility increases the risk premium and hence requires a higher rate of return from the asset. This leads to a fall in the asset price. The causality for these two interpretations is different: The leverage hypothesis claims that return shocks lead to changes in volatility, while the risk premium story implies that return shocks are caused by changes in conditional volatility. Showing that the leverage effect appears only for systematic priced risk but not for unpriced non-systematic risk could rule out the leverage story. These different explanations have been tested by Bekaert and Wu (2000) who use a parametric conditional CAPM model under a GARCH specification to obtain results consistent with the risk-premium story. I estimate the leverage effect completely non-parametrically and decompose it into its systematic and

[23] I have calculated the generalized correlations between the first two statistical volatility factors and different combinations of the four economic volatility factors. It seems that the finance factor can explain most of the second statistical volatility factor.

[24] There are studies (e.g., Nelson (1991) and Engle and Ng (1993)) showing that the effect is generally asymmetric. Declines in stock prices are usually accompanied by larger increases in volatility than the declines in volatility with rising stock markets. Yu (2005) has estimated various discrete-time models with a leverage effect.

[25] Figlewski and Wang (2000) raise the question whether the effect is linked to financial leverage at all. They show that there is only an effect on volatility when leverage changes due to changes in stock prices but not when leverage changes because of a change in debt or number of shares.

nonsystematic part based on my general statistical factors. I show that the leverage effect appears predominantly for systematic risk, while it can be non-existent for idiosyncratic risk.

The estimation of the leverage effect is difficult because volatility is unobservable. As in the previous subsection there are essentially two non-parametric approaches to estimate the correlation between asset returns and the changes in their volatility. First, the common approach is to conduct preliminary estimation of the volatility over small windows, then to compute the correlation between returns and the increments of the estimated volatility. Wang and Mykland (2012), Aït-Sahalia, Fan and Li (2013) and Aït-Sahalia and Jacod (2014) are examples of this approach. Such estimators appear to be very noisy in practice. Second, Kalnina and Xiu (2014) use volatility instruments based on option data, such as the VIX or Black-Scholes implied volatilities. Their approach seems to provide much better estimates than the first one.

Based on my factor estimation approach I decompose the leverage effect into a systematic and idiosyncratic component. The estimated high-frequency factors allow us to separate the returns and volatilities into a systematic and idiosyncratic component, which we can then use to calculate the different components of the leverage effect. I use two different methodologies. First, I employ only the high-frequency equity data and apply Aït-Sahalia and Jacod's (2014) approach as described in Theorem B.1. As already noted in Kalnina and Xiu (2014), this estimator for correlation leads to a downward bias unless a long time horizon with a huge amount of high-frequency data is used. However, my main findings, namely that systematic risk drives the leverage effect is still apparent. Second, I estimate the correlation between daily continuous returns and daily implied volatilities. This approach follows the same reasoning as Kalnina and Xiu, except that they use intra-day data and and develop a bias reduction technique for volatility instruments that are unknown functions of the unobserved volatility. For a limited number of assets we have high-frequency prices of volatility instruments, for example the VIX. For these assets using simple daily increments in implied volatilities or the more sophisticated high-frequency bias-reduced estimators with volatility instruments yield very close results. Thus, I am confident that my main findings are robust to the estimation approach employed.

In this paper I use an average leverage effect, where we measure the leverage effect with the continuous quadratic covariation for the time horizon T[26]:

$$LEV = \frac{[\sigma_i^2, X_i]_T^C}{\sqrt{[X_i, X_i]_T^C}\sqrt{[\sigma_i^2, \sigma_i^2]_T^C}}.$$

Based on systematic factors, we can decompose this average leverage effect into a systematic and idiosyncratic part:

$$LEV_i^{syst} = \frac{[\sigma_i^2, X_i^{syst}]_T^C}{\sqrt{[X_i, X_i]_T^C}\sqrt{[\sigma_i^2, \sigma_i^2]_T^C}} \qquad LEV_i^{idio} = \frac{[\sigma_i^2, X_i^{idio}]_T^C}{\sqrt{[X_i, X_i]_T^C}\sqrt{[\sigma_i^2, \sigma_i^2]_T^C}}$$

[26] Aít-Sahalia and Jacod (2014) and Kalnina and Xiu (2014) also work with an aggregate leverage effect.

where $X_i^{syst}(t) = \Lambda_i^\top F(t)$ and $X_i^{idio}(t) = e_i(t) = X_i(t) - \Lambda_i^\top F(t)$ for asset i. My estimator for this simple decomposition based on implied volatilities is therefore

$$\widehat{LEV}_i = \frac{\hat{\sigma}_i^{2\top}\hat{X}_i^C}{\sqrt{\hat{X}_i^{C\top}\hat{X}_i^C}\sqrt{\hat{\sigma}_i^{2\top}\hat{\sigma}_i^2}}, \quad \widehat{LEV}_i^{syst} = \frac{\hat{\sigma}_i^{2\top}\hat{F}^C\hat{\Lambda}_i^C}{\sqrt{\hat{X}_i^{C\top}\hat{X}_i^C}\sqrt{\hat{\sigma}_i^{2\top}\hat{\sigma}_i^2}}, \quad \widehat{LEV}_i^{idio} = \frac{\hat{\sigma}_i^{2\top}\hat{e}_i^C}{\sqrt{\hat{X}_i^{C\top}\hat{X}_i^C}\sqrt{\hat{\sigma}_i^{2\top}\hat{\sigma}_i^2}}$$

where $\hat{\Lambda}^C$ is obtained from the high-frequency data, $\hat{X}^C$, $\hat{F}^C$ and $\hat{e}_i^C$ are based on the accumulated daily continuous increments and $\hat{\sigma}^2$ are the daily increments of an estimator of the implied volatility. In the following I use our four continuous statistical factors for estimating the systematic continuous part of $\hat{X}^C$. The decomposition of the leverage effect based on spot volatilities applies the systematic and non-systematic returns to Theorem B.1.

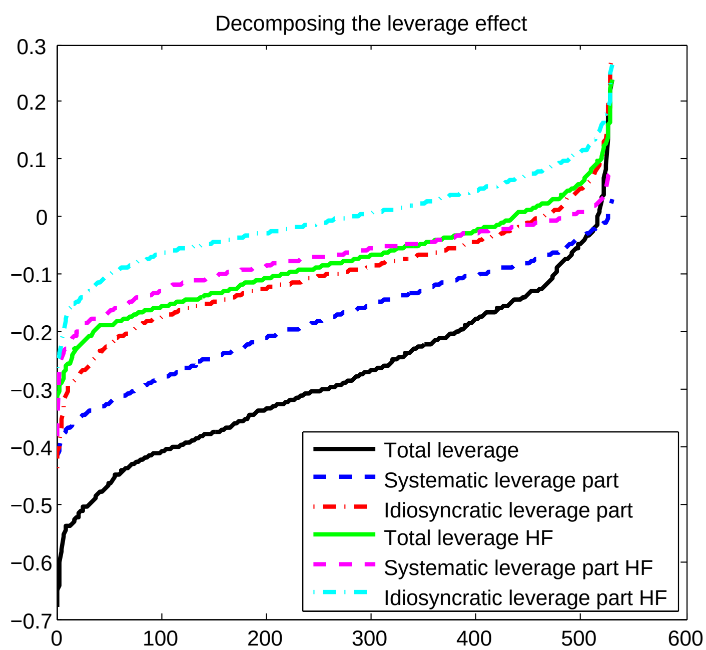


Figure 2.5: Decomposition of the leverage effect in 2012 using implied volatilities and high-frequency volatilities. I use 4 continuous asset factors.

Figure 2.5 plots the sorted decomposition of the leverage effect based on implied and realized volatility estimators. For each type of leverage I have sorted the values separately. Hence the different curves should be interpreted as describing the cross-sectional distribution of the leverage effect for the different components.[27] There is clearly a difference between the systematic leverage and idiosyncratic leverage. The total leverage is close to the systematic

[27]For the same value on the x-axis different curves usually represent different firms.

leverage. Note however, that the absolute magnitude of high-frequency estimations of the leverage effects are significantly below the estimates based on the implied-volatility, which are more in line with the values usually assumed in the literature. In Appendix B.1.5 the Figures B.26 to B.35 show the corresponding plots for all the years. We observe the same pattern: The systematic part of the leverage effect is larger than the idiosyncratic part and high-frequency based volatilities underestimate the leverage effect while implied volatility based estimates have the correct size.

The previous results could be driven by the fraction of total risk explained by the systematic part. Even if the idiosyncratic part of the return has the same correlation with the volatility, it can lead to a low covariance if it is only a small part of the total variation. For example from 2003 to 2006 the systematic factors explain only a small fraction of the total variation as can be seen in Table 2.3. This can explain the downward shift in the systematic leverage curves in Figures B.32 to B.35. A more meaningful measure is therefore the componentwise leverage effect. I decompose X_i and σ_i^2 into a systematic and idiosyncratic part: $X_i^{syst}, X_i^{idio}, {\sigma^2}_i^{syst}$ and ${\sigma^2}_i^{idio}$ based on the results in Section 2.4 and the previous subsection. X_i^{total} and ${\sigma}_i^{2total}$ denote the total asset price respectively total volatility. Then for $y, z = syst$, $idio$ and $total$ I calculate the *componentwise leverage effect*

$$LEV_i^{y,z} = \frac{[X_i^y, {\sigma^2}_i^z]_T^C}{\sqrt{[X_i^y, X_i^y]_T^C}\sqrt{[{\sigma^2}_i^z, {\sigma^2}_i^z]_T^C}}$$

and obtain $LEV^{total,total}$, $LEV^{syst,total}$, $LEV^{idio,total}$, $LEV^{syst,syst}$, $LEV^{syst,idio}$, $LEV^{idio,syst}$ and $LEV^{idio,idio}$. For the componentwise leverage effect I use the implied volatility data, the four continuous statistical asset factors and the largest volatility factor.

Figure 2.1 depicts the sorted results for 2012. The other years are in Appendix B.1.5 in Figures B.6 to B.15. There are three main findings:

1. $LEV^{total,total}$, $LEV^{syst,total}$ and $LEV^{syst,syst}$ yield the highest values and are similar to each other.

2. $LEV^{idio,total}$ and $LEV^{idio,syst}$ take intermediate values and are also very similar to each other.

3. $LEV^{syst,idio}$ and $LEV^{idio,idio}$ are on average zero and very close to each other.

In conclusion it seems that the largest leverage effect is due to the systematic asset price and systematic volatility part. Without the systematic market volatility factor the leverage effect basically disappears. One interpretation of this finding is that the main contributor to the leverage effect is non-diversifiable risk. This finding lends support to the risk-premium explanation of the leverage effect and is a counterargument for the financial leverage story.

The strongest result is the very small correlation of idiosyncratic volatility with any part of the return. In the years from 2007 to 2012 the correlation of the systematic return with the total volatility is much larger than the correlation of the idiosyncratic return with the total volatility. This particular pattern becomes weaker for 2003 to 2012. The results are robust to different variations of the leverage effect estimation. In Figure 2.6 and Figures B.16 to B.25 I calculate the componentwise leverage effect based on implied and high-frequency volatilities. As expected the high-frequency estimator underestimates the leverage effect, but the pattern is exactly the same and hence robust to the estimation methodology.

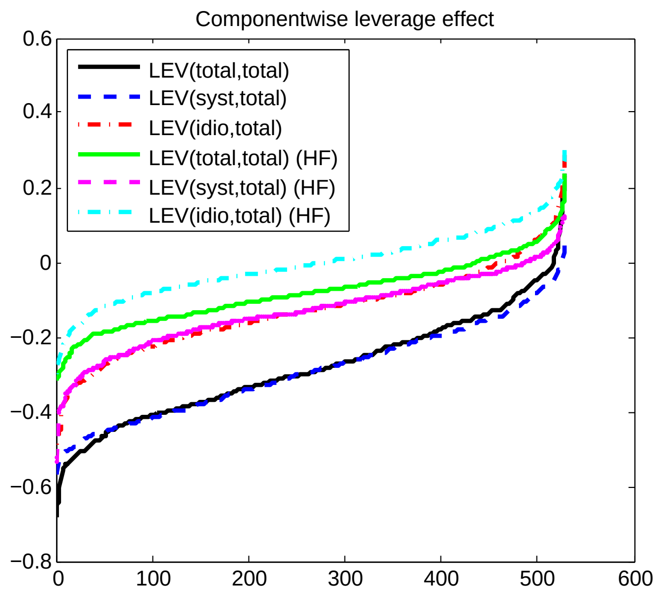


Figure 2.6: Componentwise leverage effect in 2012 based on implied and high-frequency volatilities. 4 continuous asset factors.

For the implied-volatility based leverage estimator I calculate the correlation between the daily accumulated continuous log price increments and the increments of daily implied volatilities. Figure 2.7 and Figures B.36 to B.45 depict the componentwise leverage effect if we replace the accumulated continuous increments by daily CRSP returns. The results are essentially the same. In Figure 2.8 and Figures B.46 to B.55 I replace the four statistical factors based on continuous loadings applied to daily CRSP excess returns by the 4 Fama-French-Carhart factors. The pattern in the systematic and idiosyncratic leverage effect are not affected. This seems surprising at first as factors based on the continuous loadings are different from the Fama-French-Carhart factors except for the market factor. I can show that the leverage effect results are mainly driven by the market factor. If we replaced the four factors in our analysis by merely the market factor, the observed pattern would be very similar.

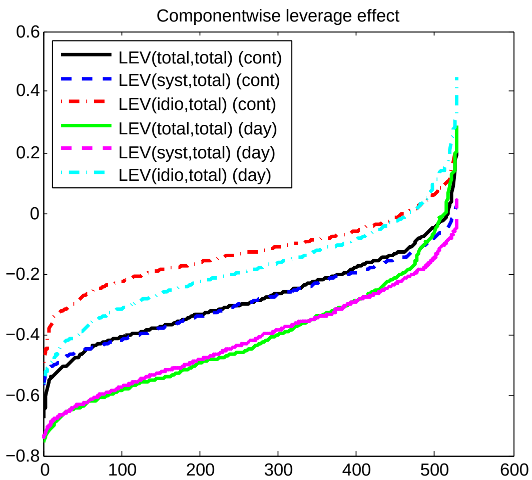


Figure 2.7: Componentwise leverage effect in 2012 with daily continuous log price increments $LEV(cont)$ and daily returns $LEV(day)$ and 4 asset factors.

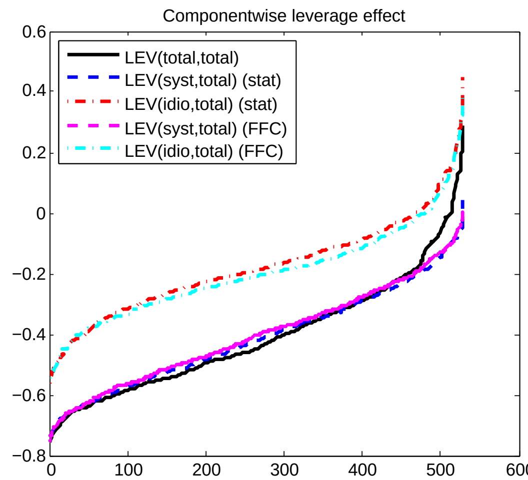


Figure 2.8: Componentwise leverage effect in 2012 with 4 continuous factors $LEV(stat)$ or 4 Fama-French-Carhart factors $LEV(FFC)$.

2.6 Conclusion

This paper studies factor models in the new setting of a large cross section and many high-frequency observations under a fixed time horizon. In an extensive empirical study I can show that the continuous factor structure is highly persistent. For the time period 2007 to 2012 I estimate four continuous factors which can be approximated very well by a market, oil, finance and electricity factor. The value, size and momentum factors play no significant role in explaining these factors. From 2003 to 2006 one continuous systematic factor disappears. There seems to exist only one persistent jump factor, namely a market jump factor. Using implied volatilities from option price data, I analyze the systematic factor structure of the volatilities. There seems to be only one persistent market volatility factor, while during the financial crisis an additional temporary banking volatility factor appears. Based on the estimated factors, I can decompose the leverage effect, i.e. the correlation of the asset return with its volatility, into a systematic and an idiosyncratic component. The negative leverage effect is mainly driven by the systematic component, while the idiosyncratic component can be positively correlated with the volatility.

Arbitrage pricing theory links risk premiums to systematic risk. In future projects I want to analyze the ability of the high-frequency factors to price the cross-section of returns. Furthermore I would like to explore the possibility to use even higher sampling frequencies by developing a microstructure noise robust estimation method.

Chapter 3

No Predictable Jumps in Arbitrage-Free Markets

3.1 Introduction

Semimartingales are the most general processes for which a stochastic integral can be defined. Thus, semimartingales are the most general stochastic processes used in asset-pricing models.

If an arbitrage existed in an asset-pricing model, then a trader would exploit it, and as a result the asset price process would change. Hence, all sensible asset-pricing models assume the absence of arbitrage. Ansel and Stricker (1991) show that a suitable formulation of absence of arbitrage implies that security gains must be special semimartingales, i.e. semimartingales that have finite conditional means.

Since the absence of arbitrage effectively implies that asset price processes are special semimartingales, there is no substantive loss of generality in restricting asset-pricing models to be special semimartingales. Indeed, Back (1990) and Schweizer (1992) model asset prices as special semimartingales. They derive a formula for the local risk premium of an asset which is proportional to its covariance with the state price density process.

A stochastic process is predictable if it is measurable with respect to the σ-field generated by the left-continuous, adapted processes. Intuitively, the realization of a predictable jump is known just before it happens. Both Back (1990) and Schweizer (1992) implicitly allow for asset prices to exhibit predictable jumps. In particular, the predictable finite variation part of the asset price process, which is usually called the "drift" term, can have discontinuities.

Empirically, it is not possible to distinguish a predictable jump from a non-predictable jump. However, these two jumps have different properties which can have a huge effect on econometric estimators. Hence, the econometrics literature for models with discontinuities generally excludes predictable jumps from the asset price processes. For example Barndorff-Nielsen and Shephard (2004b) have shown, under the assumption that there are no predictable jumps, that the realized power variation and its extension the realized bipower variation can be used to separately estimate the integrated volatility of the continuous and the jump component of a certain class of stochastic processes. Similarly, Barndorff-Nielsen and Shephard (2004a), Barndorff-Nielsen and Shephard (2006) and Aït-Sahalia and Jacod (2009), all assume, directly or indirectly (i.e. as a consequence of other assumptions) the absence of predictable jumps. In particular, all these financial econometrics papers assume that the "drift" term has to be continuous, which in the case of a special semimartingale is equivalent to the absence of predictable jumps.

In this chapter, we show that a suitable formulation of the absence of arbitrage implies that asset prices, in addition to being special semimartingales, do not have predictable jumps. Just as Ansel and Stricker (1991) show that there is no substantive economic loss of generality in restricting asset prices to special semimartingales, our finding shows that there is no substantive economic loss of generality in restricting asset prices to special semimartingales without predictable jumps. In particular, the absence of arbitrage implies that the drift term of an asset price process has to be continuous, so that assumption in the empirical finance literature involves no substantive loss of generality.

The idea of the proof is based on two facts. First, a local martingale does not have any predictable jumps. Hence, in particular we need only to show that the predictable finite variation part of the asset price process cannot have predictable jumps. Second, the existence of an equivalent martingale measure puts restrictions on the predictable finite variation part. We obtain a CAPM like representation, where the predictable finite variation part is proportional to the "covariance" between the state price density and the local martingale part of the asset price. As these both processes are also local martingales, they cannot have predictable jumps, which in turn implies that the predictable finite variation part cannot have any jumps at all.

3.2 The Model

We have essentially the same model as in Back (1990) and Schweizer (1992). We refer to those papers for the underlying motivation and to Kallenberg (1997) for the probabilistic concepts. Assume a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, satisfying the usual conditions, is given. The discounted gains process is denoted by $X(t)$. We could start by modeling an asset price process, dividend process and discount-rate process individually, but we do not lose generality by directly starting with $X(t)$. Under the real-world measure $\mathbb{P}$ the discounted gains process $X(t)$ is assumed to be a special semimartingale

$$X(t) = X_0 + A(t) + M^{\mathbb{P}}(t)$$

where $A(t)$ is the predictable finite variation part and $M^{\mathbb{P}}(t)$ is a local martingale. The predictable σ-field is the σ-field generated by the left-continuous, adapted processes. We say that the stochastic process $A(t)$ is predictable, if it is measurable with respect to the predictable σ-field. In particular, if $A(t)$ has a predictable discontinuity, it is known right before it happens. The decomposition of a special semimartingale into a predictable finite variation part and a local martingale is unique. In a general semimartingale, $A(t)$ is not assumed to be unique. It is the predictability of A(t) that implies the uniqueness of the decomposition of a semimartingale.

For example consider

$$X(t) = a(t) + B(t) + \sum_{i=1}^{N(t)} Y_i$$

where $B(t)$ is a Brownian motion, $\sum_{i=1}^{N(t)} Y_i$ a compound Poisson process independent of B and $a(t)$ is a predictable finite variation process. The Poisson process $N(t)$ has intensity λ and the jump sizes Y_i are i.i.d. with $\mathbb{E}[Y] = \kappa < \infty$. As $\sum_{i=1}^{N(t)} Y_i$ is of finite variation, $X(t)$ is a semimartingale with $B(t)$ being the local martingale part. The unique special semimartingale representation takes the form:

$$X(t) = \underbrace{B(t) + \left(\sum_{i=1}^{N} (t) Y_i - t\lambda\kappa \right)}_{M^{\mathbb{P}}(t)} + \underbrace{t\lambda\kappa + a(t)}_{A(t)}$$

The compensated jump process is now part of the local martingale, while the compensator of the jump process plus $a(t)$ form the predictable finite variation process. The only assumption that we have made about $a(t)$ is that it is a predictable finite variation process. In particular, $a(t)$ could be a discontinuous process, i.e. it could have predictable jumps. The main contribution of this paper is to show that the predictable process $a(t)$ has to be continuous in arbitrage-free markets.

Following Back's (1990) heuristic, the predictable finite variation part corresponds to the the conditional mean:

$$\mathbb{E}_t \left[dX(t) \right] = \mathbb{E}_t \left[dA(t) \right] + \mathbb{E}_t \left[dM^{\mathbb{P}} \right] = dA(t)$$

as the differential of the predictable finite variation part is known just before t. Of course, this is just an heuristic as any rigorous statement would involve stochastic integrals.

As it is well-known the existence of an equivalent martingale measure is essentially equivalent to the absence of arbitrage opportunities. "Essentially" means that this statement depends on the precise definition of arbitrage opportunities; see Kreps (1981) and Stricker (1990) for a discussion. An equivalent martingale measure $\mathbb{Q}$ for X is a probability measure that is equivalent to $\mathbb{P}$ (i.e. $\mathbb{P}$ and $\mathbb{Q}$ have the same null sets) and has the property that X is a martingale with respect to $\mathbb{Q}$. The equivalence implies the existence of the Radon-Nikodym derivative $Z_T = \frac{d\mathbb{Q}}{d\mathbb{P}}$, which defines a strictly positive martingale Z with $Z_0 = 1$:

$$Z_t = \mathbb{E}^{\mathbb{P}} \left[Z_T | \mathfrak{F}_t \right] = \frac{d\mathbb{Q}}{d\mathbb{P}} | \mathfrak{F}_t$$

The martingale property of X under $\mathbb{Q}$ is equivalent to the statement that XZ is a $\mathbb{P}-$martingale. A more general concept is a martingale density (introduced by Schweizer (1992)):

Definition 3.1. *A local $\mathbb{P}$-martingale Z with $Z_0 = 1$ is called a martingale density for X if the process XZ is a local $\mathbb{P}$-martingale. Z is called a strict martingale density, if, in addition, Z is strictly positive.*

All the results that we can derive for a martingale density will of course hold for an equivalent martingale measure.[1]

A key concept in working with semimartingales are the predictable covariation process $\langle .,. \rangle$ and the covariation process $[.,.]$. The conditional covariation process can be interpreted as a conditional covariance. If M_1 and N_2 are two local martingales such that the product MN is a special semimartingale, then the conditional covariation process $\langle M, N \rangle$ is the predictable finite variation part in the canonical decomposition of MN. If $X_1 = A_1 + M_1$ and $X_2 = A_2 + M_2$ are special semimartingales, and if X_1X_2 is a special semimartingale, then $\langle X_1, X_2 \rangle$ is defined as

$$\langle X_1, X_2 \rangle(t) = \langle M_1, M_2 \rangle(t) + \sum_{0 \leq s \leq t} \Delta X_1(s) \Delta X_2(s)$$

The jumps are denoted by $\Delta X(s) = X(s) - X(s-) \neq 0$. The jumps in the above representation are predictable and we will show in the following that the existence of an equivalent martingale measure implies that such jumps cannot occur. The conditional quadratic covariation should not be confused with the quadratic covariation process $[.,.]$. $\langle .,. \rangle$ is the $\mathbb{P}$-compensator of $[.,.]$. If M_1 and M_2 are semimartingales the quadratic covariation process is defined by

$$[M_1, M_2] = M_1M_2 - \int M_1(t-)dM_2(t) - \int M_2(t-)dM_1(t)$$

We use the following result from Kallenberg (1997):

Proposition 3.1. *A local martingale is predictable iff it is a.s. continuous.*

We conclude, that a predictable local martingale cannot have any jumps:

Corollary 3.1. *A local martingale does not have predictable jumps.*

As we will refer several times to Yoerup's lemma (Dellacherie and Meyer (1982), VII.36), we state it here for convenience:

Lemma 3.1. *Let M be a local martingale and A a predictable process of finite variation. Then the quadratic variation process $[M, A]$ is a local martingale.*

We can now state our main theorem:

[1]The martingale density assumption is weaker than the assumption about an equivalent martingale measure, as XZ does not need to be a real martingale. Hence for a martingale density, $\mathbb{Q}$ is in general only a sub-probability, i.e. $\mathbb{Q}(\Omega) \leq 1$.

Theorem 3.1. *Let Z be a strict martingale density for X. If XZ is a special semimartingale, then X cannot have predictable jumps.*

Proof. By Yoerup's lemma, $[Z, A]$ is a local $\mathbb{P}$-martingale. Hence, ZA is a special semimartingale:

$$d(ZA) = Z_{-}dA + A_{-}dZ + d[Z, A]$$

As ZX is a special semimartingale, $ZM^{\mathbb{P}}$ is one as well and thus $\langle Z, M^{\mathbb{P}}\rangle$ exists. Next, we apply the product rule to XZ:

$$\begin{aligned} d(XZ) &= X_{-}dZ + Z_{-}dX + [Z, X] \\ &= X_{-}dZ + Z_{-}dM^{\mathbb{P}} + Z_{-}dA + d[Z, A] + d[Z, M^{\mathbb{P}}] - d\langle Z, M^{\mathbb{P}}\rangle + d\langle Z, M^{\mathbb{P}}\rangle \\ &= \text{local } \mathbb{P}\text{-martingale } + Z_{-}dA + d\langle Z, M^{\mathbb{P}}\rangle \end{aligned}$$

In the last line we have used Yoerup's lemma again. But as XZ is a local martingale by assumption, the two last terms, which are predictable and of finite variation, must vanish. Hence, we conclude

$$dA = -\frac{1}{Z_{-}} d\langle Z, M^{\mathbb{P}}\rangle.$$

Hence, for all predictable jumps Δ in $A(t)$ one must have

$$\Delta A(t) = \Delta\left(-\frac{1}{Z_(t-)} d\langle Z(t), M^{\mathbb{P}(t)}\rangle\right) = \frac{\Delta Z(t)}{Z(t-)}\Delta M^{\mathbb{P}}(t)$$

As Z and $M^{\mathbb{P}}$ are local $\mathbb{P}$-martingales, they cannot have any predictable jumps, and thus neither can A. In conclusion, X cannot have any predictable jumps. □

Corollary 3.2. *Assume that $X = X_0 + A + M^{\mathbb{P}}$ is a special semimartingale and that there exists an equivalent martingale measure $\mathbb{Q}$ for X with respect to $\mathbb{P}$, which is defined by the Radon-Nikodym derivative Z. Assume that both X and Z are locally square-integrable. Then X cannot have any predictable jumps.*

Proof. The local square-integrability ensures that $\langle M^{\mathbb{P}}, Z\rangle$ is well-defined. By definition $dM^{\mathbb{P}}Z - d\langle M^{\mathbb{P}}, Z\rangle$ is a local martingale. Hence, $M^{\mathbb{P}}Z$ is a special semimartingale with decomposition $(dM^{\mathbb{P}}Z - d\langle M^{\mathbb{P}}, Z\rangle) + d\langle M^{\mathbb{P}}, Z\rangle$. As AZ is a special semimartingale, we conclude that XZ is a special semimartingale. If Z defines an equivalent martingale measure, it is also a strict martingale density and hence we can apply Theorem 1. □

3.3 Conclusion

We model asset prices in the most general sensible form as special semimartingales. This approach allows us to also include jumps in the asset price process. We show that the existence of an equivalent martingale measure, which is essentially equivalent to no-arbitrage, implies that the asset prices cannot exhibit predictable jumps. Hence, in arbitrage-free markets the occurrence and the size of any jump of the asset price cannot be known before it happens. In practical applications it is basically not possible to distinguish between predictable and unpredictable discontinuities in the price process. The empirical literature has typically assumed as an identification condition that there are no predictable jumps. Our result shows that this identification condition follows from the existence of an equivalent martingale measure, and hence essentially comes for free in arbitrage-free markets.

Chapter 4

Contingent Capital, Tail Risk, and Debt-Induced Collapse

4.1 Introduction

The problem of banks that are too big to fail plays out as an unwillingness on the part of governments to impose losses on bank creditors for fear of the disruptive consequences to the financial system and the broader economy. Higher capital requirements and restrictions on business practices may reduce the likelihood of a bank becoming insolvent, but they do not commit the regulators, managers or investors to a different course of action conditional on a bank approaching insolvency.

Contingent capital addresses this problem through a contractual commitment to have bond holders share some of a bank's downside risk without triggering failure. Contingent convertibles (CoCos) and bail-in debt are the two main examples of contingent capital. Both are debt that converts to equity under adverse conditions. CoCos provide "going concern" contingent capital, meaning that they are designed to convert well before a bank would otherwise default. Bail-in debt is "gone-concern" contingent capital and converts when the bank is no longer viable, wiping out the original shareholders and transferring ownership to the bailed-in creditors.

These instruments are increasingly important elements of reforms to enhance financial stability. Prominent examples are major issuances by Lloyds Banking Group, Credit Suisse, and BBVA. Rabobank, UBS, and Barclays have issued alternative structures in which debt is automatically written down rather than converted. The Swiss banking regulator has increased capital requirements for Swiss banks to 19% of risk-weighted assets, of which 9% can take the form of CoCos. The European Commission's proposed resolution framework relies on bail-in debt as one of its primary tools. In the U.S., bail-in is central to the implementation of the FDIC's authority to resolve large complex financial institutions granted by the Dodd-Frank act.

The logic of contingent capital is compelling. Raising new equity from private investors is particularly difficult for a bank nearing financial distress, which strengthens arguments for government support once a crisis hits; contingent capital solves this problem by committing creditors to provide equity through conversion of their claims. Nevertheless, the relative complexity of these instruments has raised some questions about whether they can be designed to function as expected and whether they might have unintended consequences.

The goal of this chapter is to analyze the design of contingent capital and to investigate the incentives these instruments create for shareholders. This work makes several contributions. First, our analysis reveals a new phenomenon we call *debt-induced collapse*. With CoCos on its balance sheet, a firm operates in one of two regimes: one in which the CoCos function as intended or another in which the equity holders optimally declare bankruptcy before conversion, effectively reducing the CoCos to straight debt. A transition from the first regime to the second is precipitated by an increase in the firm's debt load, and its consequences include a sharp increase in the firm's default probability and a drop in the value of its equity. This is the sense in which debt induces a collapse. We show that this hazard is avoided by setting the trigger for conversion at a sufficiently high level.

Once debt-induced collapse is precluded, we can investigate the incentive effects of CoCos — effects that would be lost in the alternative regime in which CoCos degenerate to straight debt. We investigate how the value of equity responds to various changes in capital structure and find, perhaps surprisingly, that equity holders often have a positive incentive to issue CoCos. We also find that CoCos can be effective in mitigating the problem of debt overhang — the reluctance of equity holders to inject additional capital into an ailing firm when most of the resulting increase in firm value is captured by debt holders. CoCos can create a strong positive incentive for shareholders to invest additional equity to stave off conversion. We also examine how CoCos affect the sensitivity of equity value to the riskiness of the firm's assets. This sensitivity is always positive in simple models, creating an incentive for asset substitution by shareholders once they have issued debt. We will see that this is not necessarily the case in a richer setting in which new debt is issued as old debt matures.

We develop our analysis in a structural model of the type introduced in Leland (1994) and Leland and Toft (1996), as extended by Chen and Kou (2009) to include jumps. The key state variable is the value of the firm's underlying assets, and equity and debt values are derived as functions of this state variable. CoCo conversion is triggered by a function of this state variable, such as a capital ratio. The model has three particularly important features. First, default is endogeneous and results from the optimal behavior of equity holders. This feature is essential to the analysis of incentive effects and to the emergence of the two default regimes described above. Second, the firm's debt has finite maturity and must be rolled over as it matures. This, too, is crucial in capturing incentive effects. In a classical single-period model of the type in Merton (1974), all the benefits of reducing default risk accrue to bond holders — equity holders always prefer riskier assets and are always deterred from further investment by the problem of debt overhang. But in a model with debt rollover, reducing default risk allows the firm to issue debt at a higher market price, and part of this increase in firm value is captured by equity holders, changing their incentives. This feature also allows us to investigate how debt maturity interacts with the efficacy of CoCos. Finally, jumps are also essential to understanding incentive effects. Downward jumps generate a higher asset yield (in the form of an increase in the risk-neutral drift) but expose the firm to tail risk. CoCos can increase equity holders' incentive to take on tail risk because equity holders would prefer a dilutive conversion at a low asset value over one at a high asset value. For the same reason, CoCos are more effective in mitigating debt overhang when asset value is subject to downward jumps.

After demonstrating these implications through a mix of theoretical and numerical results, we calibrate the model to data on the largest U.S. bank holding companies for the period 2004–2011. Some of the comparative statics in our numerical examples depend on parameter values, so the purpose of the calibration is to investigate the model's implications at parameter values representative of the large financial institutions that would be the main candidates for CoCo issuance. We calculate the model-implied increase in loss absorption that would have resulted from replacing 10% of each firm's debt with CoCos, estimate which firm's would have triggered conversion and when, and compare the impact on debt overhang costs at three dates during the financial crisis. Overall, this counterfactual exploration suggests that CoCos would have had a beneficial effect, had they been issued in advance of the crisis.

Albul, Jaffee, and Tchistyi (2010) also develop a structural model for the analysis of contingent capital; their model has neither jumps nor debt rollover (they consider only infinite maturity debt), and their analysis and conclusions are quite different from ours. Pennacchi's (2010) model includes jumps and instantaneous maturity debt; he studies the model through simulation, taking default as exogenous, and thus does not investigate the structure of shareholders' optimal default. Hilscher and Raviv (2011), Himmelberg and Tsyplakov (2012), and Koziol and Lawrenz (2012) investigate other aspects of contingent capital in rather different models. Glasserman and Nouri (2012) jointly model capital ratios based on accounting and asset values; they value debt that converts progressively, rather than all at once, as a capital ratio deteriorates. None of the previous literature combines the key features of our analysis — endogenous default, debt rollover, jumps, and analytical tractability — nor does previous work identify the phenomenon of debt-induced collapse.

Much of the current interest in contingent capital stems from Flannery (2005). Flannery(2005) proposed reverse convertible debentures (called contingent capital certificates in Flannery (2009)) that would convert from debt to equity based on a bank's stock price rather than an accounting measure. Sundaresan and Wang (2014b) raise conceptual concerns about market-based triggers. Several authors have proposed various alternative security designs; these include Bolton and Samama (2012), Calomiris and Herring (2011), Duffie (2010), Madan and Schoutens (2010), McDonald (2013), Pennacchi, Vermaelen, and Wolf (2010), and Squam Lake Working Group (2009); see Pazarbasioglu et al. (2011) for an overview.

The rest of this chapter is organized as follows. Section 4.2 details the structural model and derives values for the firm's liabilities. Section 4.3 characterizes the endogenous default barrier and includes our main theoretical results describing debt-induced collapse. Sections 4.4–4.6 investigate the impact of debt rollover and incentive effects on debt overhang and asset substitution. Section 4.7 presents the calibration to bank data. Technical details are deferred to an appendix.

4.2 The Model

4.2.1 Firm Asset Value

Much as in Leland (1994), Leland and Toft (1996), and Goldstein, Ju, and Leland (2001), consider a firm generating cash through its investments and operations continuously at rate $\{\delta_t, t \geq 0\}$. This income flow is exposed to both diffusive and jump risk, with dynamics given by

$$\frac{d\delta_t}{\delta_{t-}} = \tilde{\mu}dt + \tilde{\sigma}d\tilde{W}_t + d\left(\sum_{i=1}^{\tilde{N}_t}(\tilde{Y}_i - 1)\right). \tag{4.1}$$

Here, $\tilde{\mu}$ and $\tilde{\sigma}$ are constants, $\{\tilde{W}_t, t \geq 0\}$ is a standard Brownian motion, and we write δ_{t-} to indicate the value of δ just prior to a possible jump at time t. Jumps are driven by a Poisson process $\{\tilde{N}_t, t \geq 0\}$ with intensity $\tilde{\lambda}$. The jump sizes $\{\tilde{Y}_i, i = 1, 2, \dots\}$, $\tilde{N}$, and $\tilde{W}$ are all independent of each other. Since we are mainly concerned with the impact of downside shocks to the firm's business, we assume that the $\tilde{Y}_i$ are all less than 1. The common distribution of the $\tilde{Y}_i$ is set by positing $\tilde{Z} := -\log(\tilde{Y})$, for tractability, to have an exponential distribution, $f_{\tilde{Z}}(z) = \tilde{\eta}\exp(-\tilde{\eta}z)$, $z \geq 0$, for some $\tilde{\eta} > 0$. We assume a constant risk-free interest rate r.

In a rational expectations framework with a representative agent having HARA utility, the equilibrium price of any claim on the future income of the firm can be shown to be the expectation of the discounted payoff of the claim under a "risk-neutral" probability measure $\mathbb{Q}$; see Naik and Lee (1990) and Kou (2002) for justification of this assertion in the jump-diffusion setting. The value of the firm's assets is the present value of the future cash flows they generate,

$$V_t = \mathbb{E}^{\mathbb{Q}}\left[\int_t^\infty e^{-r(u-t)}\delta_u du \middle| \delta_t\right],$$

for all $t \geq 0$. Following Naik and Lee (1990) and Kou (2002) we can easily establish that $\delta := V_t/\delta_t$ is a constant and V_t evolves as a jump-diffusion process

$$\frac{dV_t}{V_{t-}} = \left(r - \delta + \frac{\lambda}{1+\eta}\right)dt + \sigma\, dW_t + d\left(\sum_{i=1}^{N_t}(Y_i - 1)\right), \tag{4.2}$$

Under $\mathbb{Q}$, $\{W_t\}$ in (4.2) is a standard Brownian motion and $\{N_t\}$ is a Poisson processes with intensity λ. The distribution of the jump size Y_i has the same form as before, but now with parameter η. Kou (2002) gives explicit expressions for the parameters in (4.2) in terms of the parameters in (4.1). We will value pieces of the firm's capital structure as contingent claims on the asset value process V, taking expectations under $\mathbb{Q}$ and using the dynamics in (4.2).

4.2.2 The Capital Structure and Endogenous Default

The firm finances its assets through straight debt, contingent convertible debt (CoCos), and equity. We detail these in order of seniority.

Straight Debt

We use the approach of Leland and Toft (1996) to model the firm's senior debt. The firm continuously issues straight debt with par value $p_1 dt$ in $(t, t+dt)$. The maturity of newly issued debt is exponentially distributed with mean $1/m$; that is, a portion $m\exp(-ms)ds$ of the total amount $p_1 dt$ matures during the time interval $(t+s, t+s+ds)$, for each $s \geq 0$. The debt pays a continuous coupon at rate c_1 per unit of par value. In the case of bank deposits with no stated term, the maturity profile reflects the distribution of time until depositors withdraw their funds.

The exponential maturity profile and the constant issuance rate keep the total par value of debt outstanding constant at

$$P_1 = \int_t^{\infty} \left(\int_{-\infty}^{t} p_1 m e^{-m(s-u)}\, du \right) ds = \frac{p_1}{m}.$$

Thus, the firm continuously settles and reissues debt at a fixed rate.[1] This debt rollover will be important to our analysis through its effect on incentives for equity holders.

Tax deductibility of coupon payments lowers the cost of debt service to the firm. Bank deposits have special features that can generate additional funding benefits.[2] Deposit insurance creates a funding benefit if the premium charged includes an implicit government subsidy. Customers value the safety and ready availability of bank deposits and are willing to pay (or accept a lower interest rate) for this convenience. DeAngelo and Stulz (2013) and Sundaresan and Wang (2014a) model this effect as a liquidity spread that lowers the net cost of deposits to the bank. In Allen, Carletti, and Marquez (2013), the funding benefit of deposits results from market segmentation. We model these funding benefits by introducing a factor κ_1, $0 \leq \kappa_1 < 1$, such that the net cost of coupon payments is $(1-\kappa_1)c_1 P_1$. In the special case of tax benefits, κ_1 would be the firm's marginal tax rate.

Contingent Convertibles

We use the same basic framework to model the issuance and maturity of CoCos as we use for straight debt. In both cases, we would retain tractability if we replaced the assumption of an exponential maturity profile with consols, but we would then lose the effect of debt

[1]Diamond and He (2014, p.750) find this mechanism well suited to modeling banks. One alternative would be to have a bank shrink its balance sheet after a negative shock to assets. As stressed by Hanson, Kashyap, and Stein (2011), this would run counter to the macroprudential objective of maintaining the supply of credit in an economic downturn, which, as they further note, is one of the objectives of contingent capital.

[2]See Appendix C.1 for a discussion of the application of Leland (1994) and its extensions to banks.

rollover. We denote by P_2 the par value of CoCos outstanding, which remains constant prior to conversion or default and pays a continuous coupon at rate c_2. The mean maturity is assumed to be the same as for the straight debt, $1/m$, and new debt is issued at rate $p_2 = mP_2$. It is straightforward to generalize the model to different maturities for CoCos and straight debt, and we use different maturity levels in our numerical examples later in the paper. As in the case of straight debt, we introduce a factor κ_2 that captures any funding benefit of CoCos. The tax treatment of CoCo coupons varies internationally.

Conversion of CoCos from debt to equity is triggered when a function of the state variable V_t reaches a threshold. As long as the function is invertible, we can model this as conversion the first time V_t itself falls below an exogenously specified threshold V_c. Thus, conversion occurs at

$$\tau_c = \inf\{t \geq 0 : V_t \leq V_c\}. \tag{4.3}$$

In particular, we can implement a capital ratio trigger by having CoCos convert the first time

$$(V_t - P_1 - P_2)/V_t \leq \rho,$$

with $\rho \in (0, 1)$ equal to, say, 5%. The numerator on the left is an accounting measure of equity, and dividing by asset value yields a capital ratio.[3] To put this in the form of (4.3), we set

$$V_c = (P_1 + P_2)/(1 - \rho). \tag{4.4}$$

Another choice that fits within our framework would be to base conversion on the level of earnings δV_t, as in Koziol and Lawrenz (2012). For the derivations in this section we will keep the value of V_c general, except to assume that $V_0 > V_c$ so that conversion does not occur at time zero.

At the instant of conversion, the CoCo liability is erased and CoCo investors receive Δ shares of the firm's equity for every dollar of principal, for a total of ΔP_2 shares. We normalize the number of shares to 1 prior to conversion. Thus, following conversion, the CoCo investors own a fraction $\Delta P_2/(1 + \Delta P_2)$ of the firm. In the bail-in case, $\Delta = \infty$, so the original shareholders are wiped out and the converted investors take control of the firm.[4]

Endogenous Default

The firm has two types of cash inflows and two types of cash outflows. The inflows are the income stream $\delta_t dt = \delta V_t dt$ and the proceeds from new bond issuance $b_t dt$, where b_t is the total market value of bonds issued at time t. The cash outflows are the net coupon payments

[3]This approximates a tangible common equity ratio. If CoCos are treated as Tier 1 capital, we could define a trigger based on a Tier 1 capital ratio through the condition $(V_t - P_1)/V_t \leq \rho$ and thus $V_c = P_1/(1-\rho)$.

[4]We do not distinguish between contractual and statutory conversion. Under the former, conversion is an explicit contractual feature of the debt. The statutory case refers to conversion imposed on otherwise standard debt at the discretion of a regulator granted explicit legal authority to force such a conversion.

and the principal due $(p_1+p_2)dt$ on maturing debt.[5] The net coupon payments, factoring in tax deductibility and any other funding benefits, are given by $A_t = (1-\kappa_1)c_1P_1+(1-\kappa_2)c_2P_2$.

Let $\bar{p}$ denote the total rate of issuance (and retirement) of par value of debt, just as b_t denotes the total rate of issuance measured at market value. We have $\bar{p} = p_1 + p_2$ prior to conversion of any CoCos and $\bar{p} = p_1$ after conversion. Whenever

$$b_t + \delta V_t > A_t + \bar{p}, \tag{4.5}$$

the firm has a net inflow of cash, which is distributed to equity holders as a dividend flow. When the inequality is reversed, the firm faces a cash shortfall. The equity holders then face a choice between making further investments in the firm — in which case they invest just enough to make up the shortfall — or abandoning the firm and declaring bankruptcy. Bankruptcy then occurs at the first time the asset level is at or below V_b^*, with V_b^* chosen optimally by the equity holders. In fact, it would be more accurate to say that V_b^* is determined simultaneously with b_t, because the market value of debt depends on the timing of default, just as the firm's ability to raise cash through new debt influences the timing of default.

The equity holders choose a bankruptcy policy to maximize the value of equity. To be feasible, a policy must be consistent with limited liability, meaning that it ensures that equity value remains positive prior to default. This formulation is standard and follows Leland (1994) and Leland and Toft (1996) and, in the jump-diffusion case, Chen and Kou (2009).

However, the presence of CoCos creates a distinctive new feature, driven by whether default occurs before or after conversion. Depending on the parameters of the model, the equity holders may find either choice to be optimal. If they choose to default before conversion, then the CoCos effectively degenerate to junior straight debt. Importantly, we will see that positive incentive effects from CoCo issuance are lost in this case. Indeed, the behavior of the model and, in particular, the value of equity, are *discontinuous* as we move from a regime in which conversion precedes default to a regime in which the order is reversed. We will see that this change can result from an increase in debt — either straight debt or CoCos — so we refer to this phenomenon as *debt-induced collapse*.

Upon default, we assume that a fraction $(1-\alpha)$, $0 \leq \alpha \leq 1$, of the firm's asset value is lost to bankruptcy and liquidation costs. Letting τ_b denote the time at which bankruptcy is declared and V_{τ_b} the value of the firm's assets at that moment leaves the firm with αV_{τ_b} after bankruptcy costs. These remaining assets are used first to repay creditors. If default occurs after conversion, only the straight debt remains at bankruptcy. If default occurs before conversion, the CoCos degenerate to junior debt and are repaid from any assets that remain after the senior debt is repaid.

---

[5]Our discussion of cash flows is informal and used to provide additional insight into the model. For a rigorous formulation of the Leland-Toft model through cash flows, see Décamps and Villeneuve (2014).

4.2.3 Liability Valuation

Our model yields closed-form expressions for the values of the firm's liabilities. We proceed by taking the level of the default boundary V_b as given and valuing each layer of the capital structure. We then derive the optimal level V_b^*, leading to the concept of debt-induced collapse.

We begin by limiting attention to the case $V_b \le V_c$, which ensures that the firm does not default before conversion.[6] With V_b fixed, the default time τ_b is the first time the asset value V_t is at or below V_b. To value a unit of straight debt at time t that matures at time $t+T$, we discount the coupon stream earned over the interval $[t,(t+T)\wedge\tau_b]$ and the (partial) principal received at $(t+T)\wedge\tau_b$ to get a market value of

$$\begin{aligned} b(V_t;T;V_b) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}\mathbf{1}_{\{\tau_b>T+t\}}|V_t\right] \quad \text{(principal payment if no default)} \\ &\quad +\mathbb{E}^{\mathbb{Q}}\left[e^{-r\tau_b}\mathbf{1}_{\{\tau_b\le T+t\}}\cdot\frac{\alpha V_{\tau_b}}{P_1}\Big|V_t\right] \quad \text{(payment at default)} \\ &\quad +\mathbb{E}^{\mathbb{Q}}\left[\int_0^{\tau_b\wedge(T+t)} c_1e^{-r(u-t)}du\Big|V_t\right]. \quad \text{(coupon payments)} \end{aligned} \tag{4.6}$$

To simplify notation, we will henceforth take $t=0$ and omit the conditional expectation given V_t, though it should be understood that the value of each liability is a function of the current value V of the firm's assets.

Recall that the debt maturity T is exponentially distributed with density $m\exp(-mT)$, and the total par value is P_1. The total market value of straight debt outstanding is then

$$\begin{aligned} B(V;V_b) &= P_1\int_0^\infty b(V_t;T;V_b)me^{-mT}dT \\ &= P_1\left(\frac{m+c_1}{m+r}\right)\mathbb{E}^{\mathbb{Q}}\left[1-e^{-(m+r)\tau_b}\right]+\mathbb{E}^{\mathbb{Q}}\left[e^{-(m+r)\tau_b}\alpha V_{\tau_b}\right]. \end{aligned} \tag{4.7}$$

The market value of a CoCo combines the value of its coupons, its principal, and its potential conversion to equity. To distinguish the equity value the CoCo investors receive after conversion from equity value before conversion or without the possibility of conversion, we adopt the following notation:

- E^{BC} denotes equity value before conversion for the original firm, one with P_1 in straight debt and P_2 in CoCos;
- E^{PC} denotes post-conversion equity value and thus refers to a firm with P_1 in straight debt and no CoCos;

[6] In a model with jumps, the default time τ_b and conversion time τ_c may coincide, even if $V_b < V_c$. We adopt the convention that events occur in the order implied by the barrier levels, so in this case the CoCos would be treated as having converted when the firm's assets are liquidated in bankruptcy. For the case $V_b = V_c$, context determines the assumed order of events as follows: when we discuss $V_b \le V_c$, we mean that conversion precedes bankruptcy, and when we discuss $V_b \ge V_c$ we mean the opposite.

- E^{NC} denotes no-conversion equity value, which refers to a firm with P_1 in straight debt and P_2 in non-convertible junior debt.

Each of these is a function of the current asset value V and a default barrier V_b. We will use the same superscripts to differentiate total firm value and other quantities as needed.

With this convention, a CoCo with maturity T and unit face value has market value

$$\begin{aligned} d(V;T;V_b) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}\mathbf{1}_{\{\tau_c>T\}}\right] + \mathbb{E}^{\mathbb{Q}}\left[\int_0^{T\wedge\tau_c} c_2 e^{-rs}ds\right] \\ &\quad + \frac{\Delta}{1+\Delta P_2}\mathbb{E}^{\mathbb{Q}}\left[e^{-r\tau_c}E^{\mathsf{PC}}(V_{\tau_c};V_b)\mathbf{1}_{\{\tau_c<T\}}\right]. \quad \text{(conversion value)} \end{aligned}$$

In writing $E^{\mathsf{PC}}(V_{\tau_c};V_b)$, we are taking the value of post-conversion equity when the underlying asset value is at V_{τ_c} and the default barrier remains at V_b. At conversion, the CoCo investors collectively receive ΔP_2 shares of equity, giving them a fraction $\Delta P_2/(1+\Delta P_2)$ of the firm; dividing this by P_2 yields the amount that goes to a CoCo with a face value of 1. The total market value of CoCos outstanding is then

$$\begin{aligned} D(V;V_b) &= P_2\int_0^{\infty} d(V;T;V_b)\, me^{-mT}dT \\ &= P_2\left(\frac{c_2+m}{m+r}\right)\left(1-\mathbb{E}^{\mathbb{Q}}\left[e^{-(r+m)\tau_c}\right]\right) + \frac{\Delta P_2}{1+\Delta P_2}\mathbb{E}^{\mathbb{Q}}\left[e^{-(r+m)\tau_c}E^{\mathsf{PC}}(V_{\tau_c};V_b)\right]. \end{aligned} \tag{4.8}$$

To complete the calculation in (4.8), it remains to determine the post-conversion equity value $E^{\mathsf{PC}}(V_{\tau_c};V_b)$. We derive this value by calculating total firm value and subtracting the value of debt. After conversion, the firm has only one class of debt, so

$$E^{\mathsf{PC}}(V_{\tau_c};V_b) = F^{\mathsf{PC}}(V_{\tau_c};V_b) - B(V_{\tau_c};V_b), \tag{4.9}$$

where $F^{\mathsf{PC}}(V_{\tau_c};V_b)$ is the total firm value after conversion:

$$\begin{aligned} F^{\mathsf{PC}}(V_{\tau_c};V_b) &= \underbrace{V_{\tau_c}}_{\text{unleveraged firm value}} + \underbrace{\mathbb{E}^{\mathbb{Q}}\left[\int_{\tau_c}^{\tau_b}\kappa_1 c_1 P_1 e^{-rs}ds|V_{\tau_c}\right]}_{\text{funding benefits}} - \underbrace{\mathbb{E}^{\mathbb{Q}}\left[e^{-r(\tau_b-\tau_c)}(1-\alpha)V_{\tau_b}|V_{\tau_c}\right]}_{\text{bankruptcy costs}} \\ &= V_{\tau_c} + \frac{\kappa_1 c_1 P_1}{r}\left(1-\mathbb{E}^{\mathbb{Q}}\left[e^{-r(\tau_b-\tau_c)}|V_{\tau_c}\right]\right) - \mathbb{E}^{\mathbb{Q}}\left[e^{-r(\tau_b-\tau_c)}(1-\alpha)V_{\tau_b}|V_{\tau_c}\right] \\ &=: V_{\tau_c} + FB_1 - BCOST. \end{aligned}$$

The conversion of the CoCos does not affect the value of the senior debt, so the valuation expression in (4.7) applies to $B(V_{\tau_c};V_b)$ in (4.9).

To find the value of equity before conversion, we again derive the total firm value and subtract the debt value. We continue to limit attention to the case $V_b \leq V_c$. Any funding benefit from CoCos terminates at the conversion time τ_c. So, the firm value before conversion is

$$\begin{aligned} F^{\mathsf{BC}}(V;V_b) &= V + \underbrace{\frac{\kappa_1 c_1 P_1}{r}\left(1-\mathbb{E}^{\mathbb{Q}}\left[e^{-r\tau_b}\right]\right)}_{\text{funding benefits from straight debt}} + \underbrace{\frac{\kappa_2 c_2 P_2}{r}\left(1-\mathbb{E}^{\mathbb{Q}}\left[e^{-r\tau_c}\right]\right)}_{\text{funding benefits from CoCos}} \\ &\quad -\mathbb{E}^{\mathbb{Q}}\left[e^{-r\tau_b}(1-\alpha)V_{\tau_b}\right] \\ &=: V + FB_1 + FB_2 - BCOST. \end{aligned} \tag{4.10}$$

The market value of the firm's equity is given by

$$E^{\mathsf{BC}}(V;V_b) = F^{\mathsf{BC}}(V;V_b) - B(V;V_b) - D(V;V_b). \tag{4.11}$$

A similar calculation leads to closed-form liability evaluation if conversion does not occur prior to bankruptcy, i.e., $V_b > V_c$. In this case, CoCos degenerate to junior debt in bankruptcy. Upon default, CoCo holders are repaid from whatever assets remain after liquidation and payment of senior debt.[7] Before default, the total market value of straight debt is

$$B(V;V_b) = P_1\left(\frac{m+c_1}{m+r}\right)\mathbb{E}^{\mathbb{Q}}\left[1-e^{-(m+r)\tau_b}\right] + \mathbb{E}^{\mathbb{Q}}\left[e^{-(m+r)\tau_b}(\alpha V_{\tau_b}\wedge P_1)\right] \tag{4.12}$$

leaving a CoCo value of

$$D(V;V_b) = P_2\left(\frac{m+c_2}{m+r}\right)\mathbb{E}^{\mathbb{Q}}\left[1-e^{-(m+r)\tau_b}\right] + \mathbb{E}^{\mathbb{Q}}\left[e^{-(m+r)\tau_b}(\alpha V_{\tau_b}-P_1)^+\right]. \tag{4.13}$$

Total firm value in this case is given by

$$F^{\mathsf{BC}}(V;V_b) = V + \left(\frac{\kappa_1 c_1 P_1}{r} + \frac{\kappa_2 c_2 P_2}{r}\right)\left(1-\mathbb{E}^{\mathbb{Q}}\left[e^{-r\tau_b}\right]\right) - \mathbb{E}^{\mathbb{Q}}\left[e^{-r\tau_b}(1-\alpha)V_{\tau_b}\right].$$

The only difference between this expression and (4.10) lies in the funding benefit provided by the CoCo coupon payments, which now terminates at default rather than conversion. Equity value in the case $V_b > V_c$ now follows from (4.11) using these expressions.

All pieces (4.7)–(4.13) of the capital structure of the firm can be explicitly evaluated through expressions for the joint transforms of hitting times τ_b or τ_c and asset value V given explicitly by Kou (2002) and Kou and Wang (2003). The appendix contains additional details.

In (4.7) and (4.13) we have implicitly made a standard assumption that the asset value recovered in bankruptcy does not exceed the total amount due to bond holders. Indeed,

[7] One might alternatively suppose that the bankruptcy court or resolution authority would treat the CoCo investors as shareholders, in which case the last term in (4.13) would be dropped. Our analysis goes through under either formulation. Because CoCos are ordinarily considered debt instruments, we adopt the partial recovery assumption in (4.13) to be concrete.

Chen and Kou (2009) show that this property holds at the endogenous default time chosen by shareholders. With the addition of CoCos, we make the further assumption that the conversion ratio satisfies

$$1/\Delta \geq \alpha V_c - P_1 - P_2. \tag{4.14}$$

The expression on the left is the price per share applied to the CoCos at conversion; the expression on the right is the price per share the original equity holders could get by liquidating the firm and paying off all debt at $V = V_c$. If (4.14) were violated, the original equity holders might be motivated to liquidate the firm even when it has strictly positive equity value. Ordinarily, we expect the right side of (4.14) to be negative and the condition therefore satisfied by any $\Delta > 0$.

4.2.4 The Bail-In Case

In the bail-in case, conversion of debt to equity occurs when the firm would not otherwise be viable, rather than at an exogenously specified trigger. We model this by taking $V_c = V_b$, with the understanding that conversion occurs just before what would otherwise be bankruptcy. We set $\Delta = \infty$ so the original shareholders are wiped out, and the firm is taken over by the bail-in investors. As bankruptcy is avoided, we assume that no bankruptcy costs are incurred, so $\alpha = 1$. Just after conversion, the firm continues to operate, now with just P_1 in debt outstanding.

4.3 Optimal Default and Debt-Induced Collapse

Having valued the firm's equity at an arbitrary default barrier V_b, we now proceed to derive the equity holder's optimal default barrier V_b^* and to investigate its implications.

4.3.1 Endogenous Default Boundary

As in Section 4.2, we denote by $E^{\mathsf{PC}}(V; V_b)$ the post-conversion equity value for a firm with asset value V and default barrier V_b. After conversion, we are dealing with a conventional firm, meaning one without CoCos. In such a firm, the equity holders choose the default barrier V_b to maximize the value of equity subject to the constraint that equity value can never be negative; that is, they solve

$$\max_{V_b} E^{\mathsf{PC}}(V; V_b) \tag{4.15}$$

subject to the limited liability constraint

$$E^{\mathsf{PC}}(V'; V_b) \geq 0, \quad \text{for all } V' \geq V_b.$$

The limited liability constraint ensures that the chosen V_b is feasible. Without this condition, a choice of V_b that maximizes $E^{\mathsf{PC}}(V; V_b)$ at the current asset level V might entail sustaining a negative value of equity at some asset level between V_b and V, which is infeasible. Denote the solution to this problem by V_b^{PC}.

Before conversion, when the firm's liabilities include CoCos, equity value is given by $E^{\mathsf{BC}}(V;V_b)$, and the shareholders would like to choose V_b to maximize this value. If they choose $V_b < V_c$, conversion will precede bankruptcy, and following conversion they — and the new shareholders who were formerly CoCo holders — will face an equity maximization problem of the type in (4.15). Hence, before conversion the equity holders face a commitment problem, in the sense that they cannot necessarily commit to holding V_b at the same level after conversion that they would have chosen before conversion. Anticipating this effect, they will choose $V_b = V_b^{\mathsf{PC}}$ if they choose $V_b < V_c$. Thus, before conversion, equity holders will choose V_b to solve

$$\max_{V_b} E^{\mathsf{BC}}(V;V_b)$$

subject to the limited liability constraint

$$E^{\mathsf{BC}}(V';V_b) \geq 0, \quad \text{for all } V' \geq V_b$$

and the commitment condition that $V_b = V_b^{\mathsf{PC}}$ if $V_b < V_c$. Let V_b^* denote the solution to this problem.

Chen and Kou (2009) have solved the optimal default barrier problem with only straight debt, and this provides the solution for the post-conversion firm: $V_b^{\mathsf{PC}} = P_1\epsilon_1$, where ϵ_1 depends on c_1, m, κ_1 and α but is independent of the capital structure and V. See equation (C.1) in the appendix for an explicit expression. Recall that $E^{\mathsf{NC}}(V;V_b)$ denotes the value of equity if the P_2 in CoCos is replaced with non-convertible junior debt in the original firm. Extending Chen and Kou (2009), we can express the optimal default barrier for this firm as $V_b^{\mathsf{NC}} = P_1\epsilon_1 + P_2\epsilon_2$, where ϵ_2 is defined analogously to ϵ_1 using c_2 instead of c_1; see (C.2). We always have $V_b^{\mathsf{PC}} \leq V_b^{\mathsf{NC}}$ because increasing the amount of non-convertible debt while holding everything else fixed raises the default barrier. We can now characterize the optimal default barrier with CoCos.

Theorem 4.1. *For a firm with straight debt and with CoCos that convert at V_c, the optimal default barrier V_b^* has the following property: Either*

$$V_b^* = V_b^{PC} \leq V_c \quad \text{or} \quad V_b^* = V_b^{NC} \geq V_c. \tag{4.16}$$

Moreover, V_b^{PC} is optimal whenever it is feasible, meaning that it preserves the limited liability of equity.

This result reduces the possible default barriers for a firm with CoCos to two candidates, each of which corresponds to the default barrier for a firm without CoCos. The second case is a candidate only if, without the conversion feature, it would be optimal to default at an asset level higher than the trigger V_c. This can occur only if the first case does not yield a feasible solution.

We will see that a firm can sometimes move from the first case in (4.16) to the second case by increasing its debt load. The transition is discontinuous, creating a jump up in the default barrier and a drop in equity value. We refer to this phenomenon as *debt-induced collapse*. This phenomenon is not present without CoCos (or with bail-in debt). Moreover, we will see that the positive incentive effects that result from CoCos under the first case in (4.16) disappear following the collapse.

To illustrate, we consider an example. The heavy solid line in Figure 4.1 shows equity value as a function of asset value for the NC firm, in which the CoCos are replaced by junior debt. The optimal default barrier V_b^{NC} is at 93, and the NC equity value and its derivative are equal to zero at this point. If the conversion trigger V_c is below 93 (two cases are considered in the figure), then $V_b = V_b^{\mathsf{NC}} = 93$ is a feasible default level for the original firm because the resulting equity values are consistent with limited liability. The optimal post-conversion default barrier is $V_b^{\mathsf{PC}} = 58$. Suppose the conversion trigger is at $V_c = 65$, and suppose the original shareholders of the original firm with CoCos attempt to set the default barrier at $V_b = 58$. The dashed line shows the resulting equity value. At higher asset values, the dashed lines is above the solid line, suggesting that equity holders would prefer to set the default barrier at 58 than at 93. However, the dashed line is not a feasible choice because it creates negative equity values at lower asset levels; the best the shareholders can do in this case is to set $V_b = 93$. If the conversion trigger were at $V_c = 75$, a default barrier of $V_b = V_b^{\mathsf{PC}} = 58$ would be feasible because the resulting equity values (the dash-dot line) remain positive; in fact, this choice would then be optimal. If we imagine starting with the conversion trigger at 75 and gradually decreasing it toward 65, at some level of V_c in between the default barrier jumps up from 58 to 93, and the equity curve collapses down to the heavy solid line showing the equity curve for the NC firm.

In the bail-in case, the original equity holders are effectively choosing V_c because their default is a conversion that transfers ownership to the new shareholders. After conversion, the new shareholders will choose default barrier V_b^{PC}. Before conversion, the original equity value is given by E^{BC}, evaluated with $\Delta = \infty$. In maximizing the value of their claim, the original equity holders will choose a level of V_c consistent with limited liability, $E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) \geq 0$, for all $V \geq V_c$. The value of equity changes continuously with V_c and with the debt levels P_1 and P_2 (this can be seen from the expression (C.3) given in the appendix) so there is no phenomenon of debt-induced collapse.

4.3.2 Constraints on Debt Levels

We now analyze the effect of changing the debt levels P_1 and P_2 and the limits imposed by Theorem 5.3. For purposes of illustration, we start with a simple case in which V_c is held fixed as we vary P_1 and P_2. This allows us to isolate individual effects of changes in capital structure.

Theorem 5.3 shows that either of two conditions leads to debt-induced collapse:

- The optimal default barrier for the post-conversion firm is too high: $V_b^{\mathsf{PC}} > V_c$.

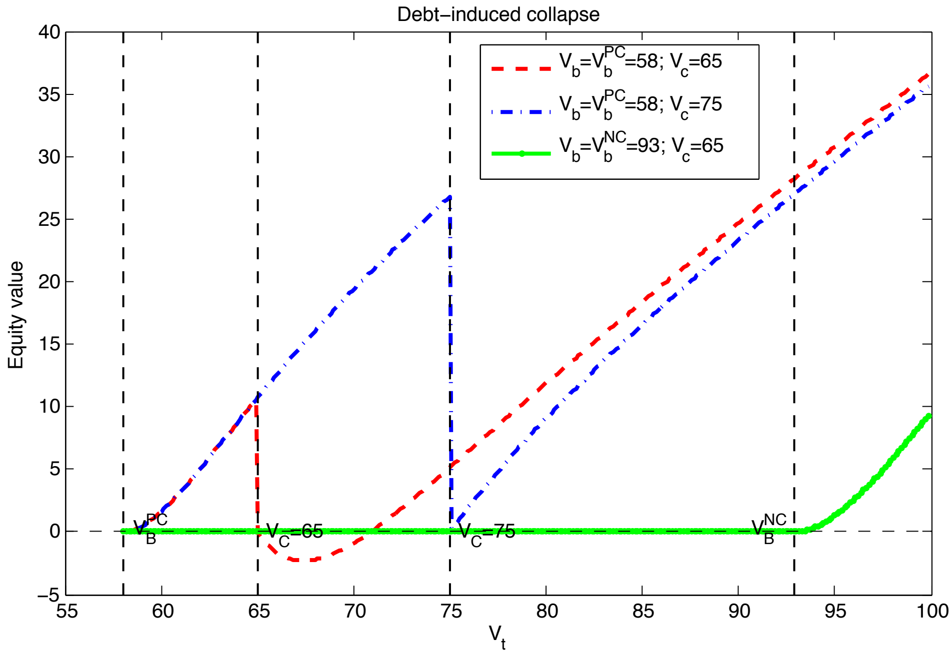


Figure 4.1: Candidate equity value as a function of asset value in three scenarios. The heavy solid (green) line reflects default at $V_b^{\mathsf{NC}} = 93$, prior to conversion. The other two lines reflect default at $V_b^{\mathsf{PC}} = 58$ with two different conversion triggers. With $V_c = 65$, equity becomes negative so V_b^{PC} is infeasible and default occurs at V_b^{NC}. With $V_c = 75$, default at V_b^{PC} is feasible, and it is optimal because it yields higher equity than V_b^{NC}.

- No default barrier lower than V_c is feasible: for any $V_b < V_c$ we can find some $V > V_c$ such that $E^{\mathsf{BC}}(V; V_b) < 0$, violating the limited liability of equity.

These two conditions provide guidance in examining when changes in capital structure result in debt-induced collapse. Theorem 4.2 makes this precise. In the theorem, we establish two critical amounts $\bar{P}_1$ and $\bar{P}_2$ such that the condition $P_1 \leq \bar{P}_1$ is equivalent to $V_b^{\mathsf{PC}} \leq V_c$, and, when this holds, $P_2 \leq \bar{P}_2$ is equivalent to $E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) \geq 0$ for all $V \geq V_c$.

Theorem 4.2. *Suppose V_c is fixed. There exist upper bounds on the amount of straight debt and CoCos above which debt-induced collapse ensues. Formally, there exist $\bar{P}_1$ and $\bar{P}_2$, where $\bar{P}_2$ depends on P_1, such that the following holds:*

- *If either $P_1 > \bar{P}_1$ or $P_2 > \bar{P}_2$, then we have debt-induced collapse.*
- *If $0 \leq P_1 \leq \bar{P}_1$ and $0 \leq P_2 \leq \bar{P}_2$, then debt-induced collapse does not occur.*

The critical levels $\bar{P}_1$ and $\bar{P}_2$ are derived in the appendix.

We illustrate these debt limits through numerical examples. We fix the parameters in Table 4.1, which are in line with our calibration results, and vary the average maturity $1/m$, and the amount of straight debt P_1.

Parameter		Value
initial asset value	V_0	100
risk free rate	r	6%
volatility	σ	8%
payout rate	δ	1%
funding benefit	κ_1, κ_2	35%
jump intensity	λ_f	0.3
firm specific jump exponent	η	4
coupon rates	(c_1, c_2)	$(r + 3\%, r + 3\%)$
bankruptcy loss	$(1 - \alpha)$	50%

Table 4.1: Base case parameters. Asset returns have a total volatility (combining jumps and diffusion) of 21%. On average every 3 years a jump costs the firm a fifth of its value. The number of shares Δ issued at conversion is set such that the market value of shares delivered is the same as the face value of the converted debt if conversion happens at exactly V_c.

Figure 4.2 shows the maximum amount of CoCos and the maximum leverage ratio that can be sustained without debt-induced collapse, with a conversion barrier $V_c = 75$. The mean maturity ranges from $1/m = 0.1$ years to $1/m = 10$ years. In the first plot we show $\bar{P}_2$ as a function of P_1. The intersection of each curve with the x-axis represents $\bar{P}_1$. For example a firm with a mean debt maturity of $1/m = 4$ years and face value $P_1 = 90$ can only add $\bar{P}_2 = 15$ CoCos to the capital structure. If the firm adds more CoCos, debt-induced collapse occurs. The second plot shows the same relationship, but now in terms of leverage. For a firm that chooses debt levels P_1 and P_2, we calculate the resulting total value of the firm F. The ratios P_1/F and P_2/F are the leverage ratios for straight debt and CoCos. A firm with debt maturity of 10 years and a straight debt leverage of 80% can increase the CoCo leverage only up to 5%. Finally, in the third plot we show the total leverage $(P_1 + P_2)/F$ as a function of straight debt leverage. A firm with a debt maturity of 1 year cannot lever up to more than 78% without triggering debt-induced collapse, regardless of how it chooses its capital structure.

As we have noted before, the optimal default barrier $V_b^{\mathsf{PC}} = P_1 \epsilon_1$ is proportional to the amount of straight debt. If V_c is far above V_b^{PC}, a large amount of CoCos can be issued. A short mean maturity $1/m$ results in a higher default barrier V_b^{PC} and hence also in a lower critical level $\bar{P}_2$. If the amount of straight debt is high, this also increases V_b^{PC} and the same effect takes place.

Figure 4.3 shows how the critical values in the top panel of Figure 4.2 change when we remove any funding benefit for CoCos by setting $\kappa_2 = 0$. The figure shows that the upper limit on CoCo issuance to avoid debt-induced collapse decreases. We interpret this effect as

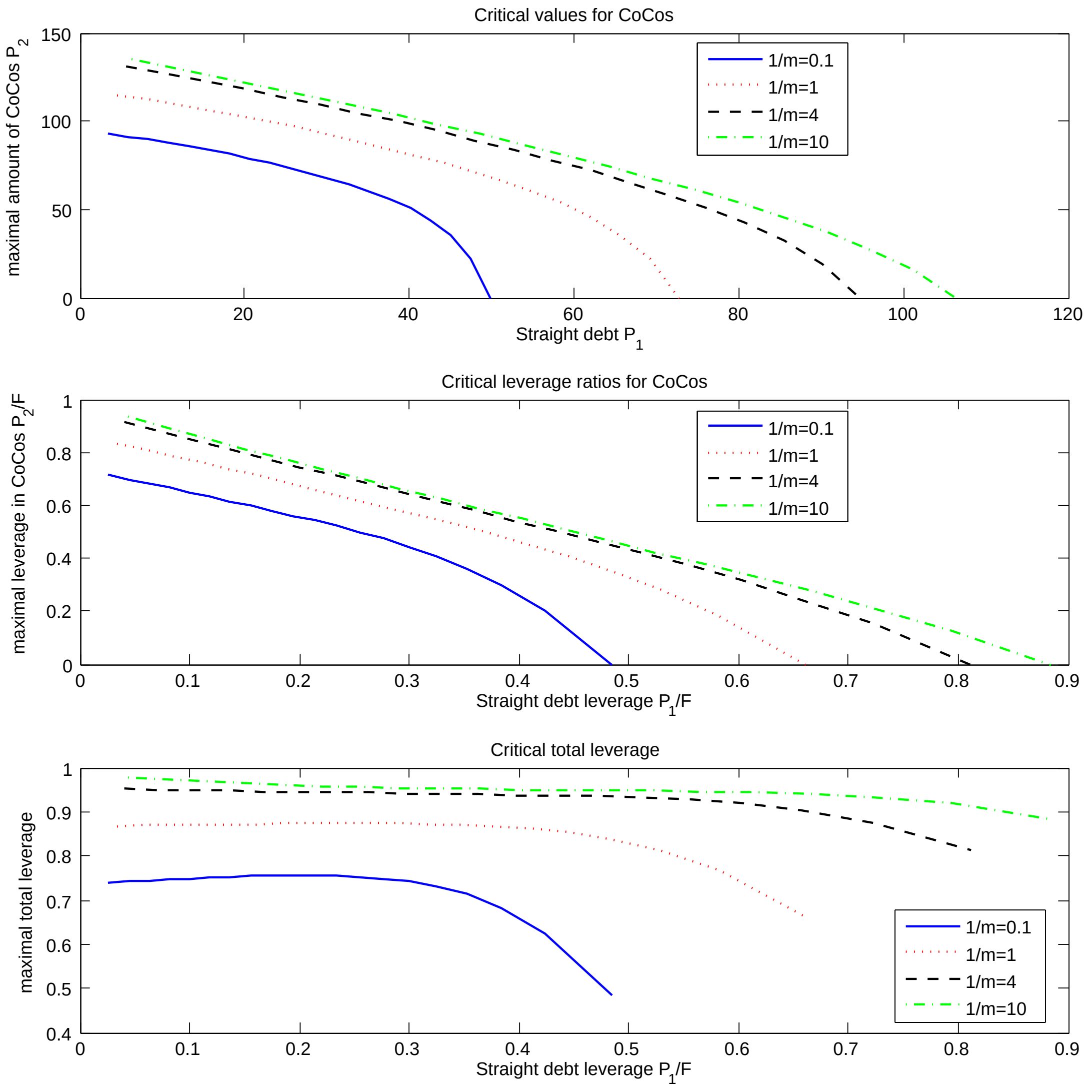


Figure 4.2: Top: Critical values of CoCos $\bar{P}_2$ as a function of straight debt P_1 for different mean maturities and $V_c = 75$. Middle: Critical leverage ratios of CoCos $\bar{P}_2/F$ as a function of straight debt leverage P_1/F. Bottom: Critical leverage $(P_1 + \bar{P}_2)/F$ as a function of straight debt leverage P_1/F.

follows. With the funding benefit reduced, shareholders have less incentive to keep the firm operating and will therefore raise the default barrier; raising the default barrier expands the scope of debt-induced collapse.

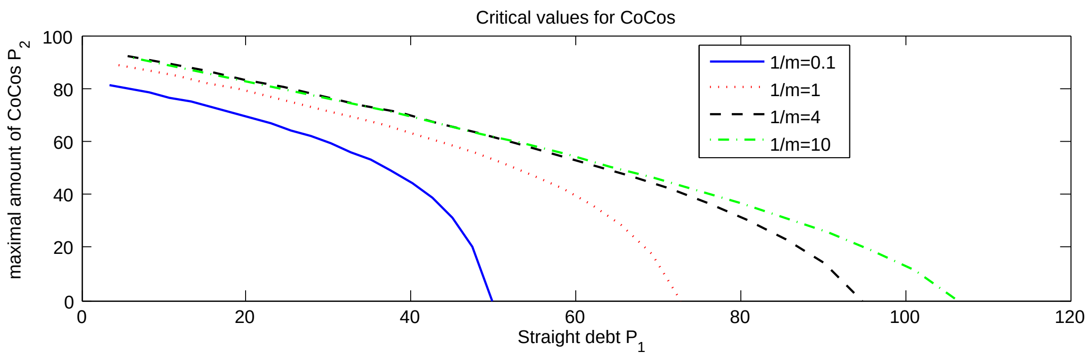


Figure 4.3: Critical values of CoCos $\bar{P}_2$ as a function of straight debt P_1 for different mean maturities and $V_c = 75$, when CoCos have no funding benefit ($\kappa_2 = 0$).

4.3.3 Constraints with a Capital Ratio Trigger

In the analysis of Theorem 4.2, we held V_c fixed while varying P_1 and P_2 to isolate individual effects. We now let V_c vary with the debt levels by setting $V_c = (P_1 + P_2)/(1 - \rho)$, following the conversion rule in (4.4) based on a minimum capital ratio ρ. According to Theorem 5.3, we have debt-induced collapse if $V_b^{\mathsf{PC}} > V_c$, which now reduces to the condition

$$(\epsilon_1(1-\rho) - 1)P_1 > P_2. \tag{4.17}$$

As before, ϵ_1 is given explicitly by equation (C.1) in the appendix. It follows from (4.17) that a sufficiently large P_1 will produce debt-induced collapse if $\epsilon(1 - \rho) > 1$. We explore when this condition[8] holds and its implications through numerical examples varying the debt rollover frequency m and coupon c.

For the numerical examples we take the baseline values in Table 4.1 and set the capital ratio and the amount of CoCos to be $\rho = 5\%$ and $P_2 = 5$, respectively. Figure 4.4 plots the critical levels of P_1 that lead to (4.17). For example, a firm with mean maturity of four months ($1/m = 0.3$) and coupon of $c = 0.11$ on its straight debt will experience debt-induced collapse at any P_1 larger than 80. Note that this condition is completely independent of the parameters of the CoCos other than the amount P_2. Figure 4.4 reveals an important interaction between debt maturity and debt-induced collapse: rolling over debt more frequently lowers the threshold of P_1 for debt-induced collapse. Directly from (4.17), it is also clear that lowering the required capital ratio ρ also widens the scope of parameters leading to debt-induced collapse.

The threshold for P_1 depicted in Figure 4.4 gives a sufficient condition based on (4.17): setting P_1 above the critical value guarantees debt-induced collapse. It is actually possible to have debt-induced collapse at an even lower value of P_1 if the feasibility condition $\min_{V \geq V_c} E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) > 0$ is violated. This condition is more complicated because the minimum is not necessarily monotonic in P_1. However, in numerical experiments we have found

[8]This condition also yields debt-induced collapse with the Tier 1 trigger in the footnote just before (4.4).

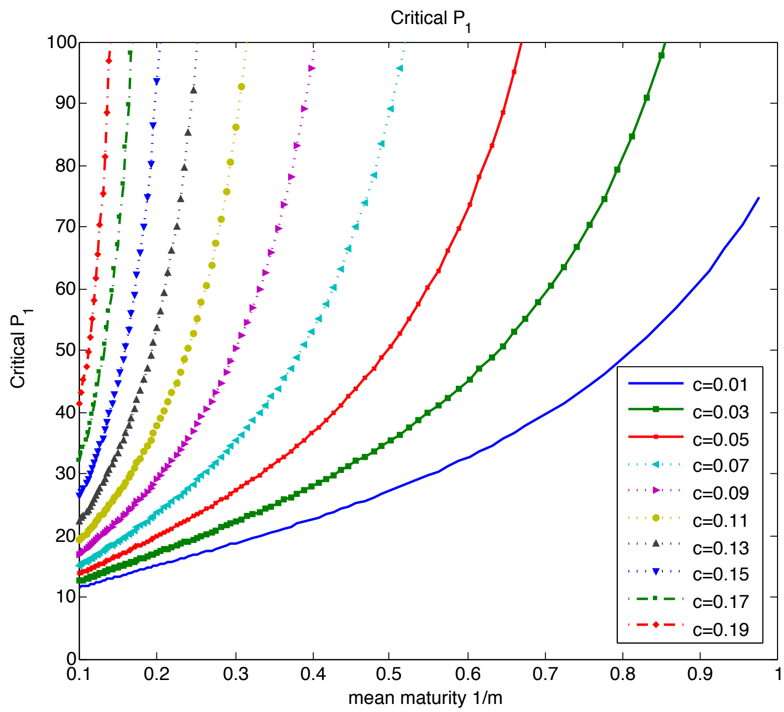


Figure 4.4: Critical values of straight debt P_1, that lead to $V_b^{\mathsf{PC}} > V_c$ and hence debt-induced collapse for a capital ratio trigger. We set $\rho = 0.05$ and $P_2 = 5$.

that debt-induced collapse typically occurs only when P_1 is larger than the threshold determined by (4.17) and illustrated in Figure 4.4, so (4.17) appears to be the more important of the two possible conditions leading to debt-induced collapse.

4.3.4 A Too-Big-To-Fail Firm

As a final illustration, we consider a simple model of a firm that is too big to fail (TBTF) and show that debt-induced collapse becomes even more of a concern in this setting. As in Albul et al. (2010), we take the key feature of a TBTF firm to be an implicit government guarantee on senior debt. We assume that at bankruptcy the government steps in and makes the senior bond holders whole. Bond holders anticipate this guarantee, but the firm does not pay for it. The net effect is to reduce the firm's cost of issuing debt: it issues P_1 in senior debt at a market value of $P_1(c_1 + m)/(r + m)$ rather than (4.7). The rest of our analysis extends accordingly.

The difference between the market value of riskless and risky debt functions like a government subsidy to the firm. With debt rollover, part of this subsidy is captured by shareholders.[9] Shareholders maximize their benefit by raising the default boundary, thus increasing

[9]This is not case in Albul et al. (2010) because in their model all debt is perpetual so all benefits of

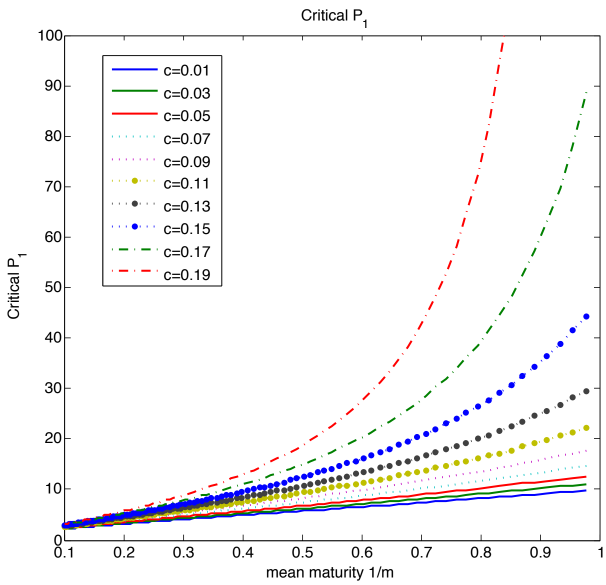


Figure 4.5: Critical values of straight debt P_1 for a TBTF firm, that lead to $V_b^{\mathsf{PC}} > V_c$ for a CET1 conversion trigger. We set $\rho = 0.05$ and $P_2 = 5$.

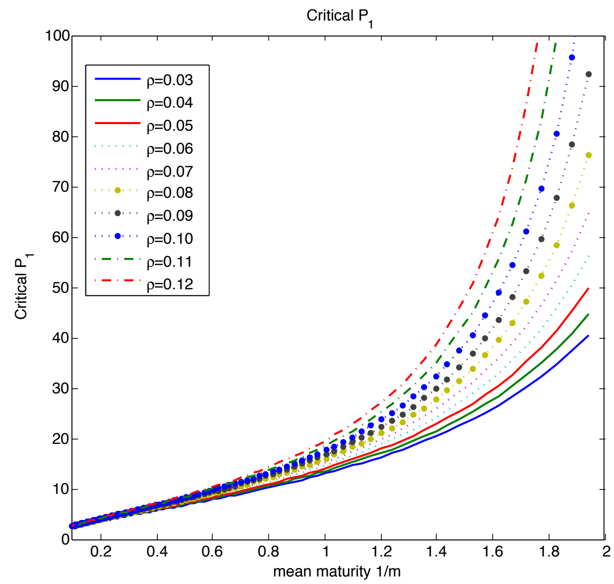


Figure 4.6: Critical values of straight debt P_1 for a TBTF firm, that lead to $V_b^{\mathsf{PC}} > V_c$ for a CET1 conversion trigger. We set $c_1 = r$ and $P_2 = 5$.

the value of the subsidy and increasing the propensity for debt-induced collapse.

Figure 4.5 displays the same comparison as Figure 4.4 but for a TBTF firm. Debt-induced collapse is now almost unavoidable at maturities shorter than one year. As the straight debt of a TBTF firm is basically risk-free, we set the coupon c_1 equal to the risk-free rate r and plot different capital ratio triggers ρ in Figure 4.6. Given the current leverage ratios of banks, a TBTF firm with a mean maturity of 2 years for its straight debt needs a trigger ratio ρ larger than 10% to avoid debt-induced collapse, which would be very high. The main point of this example is to illustrate a directional effect: an implicit guarantee widens the scope of parameters at which debt-induced collapse occurs.

4.4 The Impact of Debt Rollover

In this section, we investigate the impact of debt rollover on equity value under various changes in capital structure.

The process of rolling debt is important to our analysis, so we briefly describe this feature of our model. Under our exponential maturity assumption, old debt is continuously maturing and new debt is continuously issued. Within each debt category, the coupon and the total

reduced credit risk accrue to bond holders. This important feature of debt rollover is further explored in the next section.

par value outstanding remain constant; but while debt matures at par value, it is issued at *market* value. If the par value is greater, the difference is a cash shortfall that needs to be paid out by the firm; if the market value is greater, the difference generates additional cash for the firm. We refer to these as rollover costs — a positive cost in the first case, a negative cost in the second — and treat them the same way we treat coupon payments. Rollover costs will change as the firm's asset value changes, becoming larger as asset value declines, the firm gets closer to default, and the market value of its debt decreases. Rollover costs thus capture the increased yield demanded of riskier firms.[10]

The comparisons in this section are based on combinations of qualitative properties and numerical examples. For the numerical examples, we enrich the base model, expanding the capital structure through additional layers of straight debt and allowing two types of jumps in asset value. Details of these extensions and parameter values for the numerical illustrations are discussed in Appendix C.3. We use the parameters given there in Table C.1. The firm initially funds 100 in assets with a total par value of 85 in non-convertible debt and 15 in equity or a combination of equity and CoCos. Under any change in capital structure, we recompute the optimal default barrier and recompute the value of the firm and its liabilities. Throughout this section, we limit ourselves to changes that keep the firm within the no-collapse region so that the CoCos do not degenerate to straight debt.

4.4.1 Replacing Straight Debt with CoCos

We begin by replacing some straight debt with CoCos. The consequences of the substitution are as follows.

- If coupon payments on CoCos are not tax deductible, then replacing straight debt with CoCos has the immediate effect of reducing firm value by reducing the value of the tax shield. Even if CoCo coupons are tax deductible or enjoy other funding benefits, these benefits end at conversion, so, other things being equal, the substitution still has the immediate impact of reducing firm value; see (4.10). The reduction in firm value has the direct effect of lowering the value of equity.

- However after conversion the firm will have less debt outstanding and lower debt service payments (coupons and rollover costs) than it would without the substitution of CoCos for straight debt. With lower debt service, more of the cash generated by the firm's assets flows to equity holders in dividends. This reduces the default barrier V_b^*, which extends the life of the firm, reduces the bankruptcy cost and thus increases firm value in (4.10).

- We thus have two opposite effects on firm value: the reduced funding benefit from CoCos reduces firm value, but the reduced default probability and bankruptcy cost

[10]Debt rollover also have important implications for asymmetric information and monitoring, as in Calomiris and Kahn (1991), and liquidity risk, as in He and Xiong (2012), but these features are outside the scope of our model.

increases firm value. In our numerical examples, we find that the second effect dominates over a wide range of parameter values, so that the net effect of replacing straight debt with CoCos is to increase firm value.

- Part of this increase in firm value is captured by debt holders because the reduced bankruptcy risk increases the value of the debt. Part of the increase is also captured by equity holders: the increased debt value reduces rollover costs which increases the flow of dividends. Thus, *equity holders have a positive incentive to issue CoCos.*

This conclusion contrasts with that of Albul et al. (2010), who find that equity holders would never voluntarily replace straight debt with contingent convertibles. In their model, straight debt has infinite maturity and is never rolled. As a result, all of the benefit of reduced bankruptcy costs from CoCos is captured by debt holders. This difference highlights the importance of debt rollover in influencing incentives for equity holders, an effect we return to at several points.

The line marked with crosses in the left panel of Figure 4.7 shows the increase in equity value resulting from a substitution of one unit (market value) of CoCos for one unit (market value) of straight debt, plotted against the value of the firm's asset value. The conversion level V_c is 75. Despite the dilutive effect of conversion, the benefit to equity holders of the substitution is greatest just above the conversion level and decreases as asset level increases. This follows from the fact that the benefit to equity holders derives from the reduction in bankruptcy costs, which is greater at lower asset values. We will discuss the other curves in the left panel shortly.

The right panel of Figure 4.7 incorporates a friction in the conversion of debt to equity. To this point, we have valued each security as the expected present value of its cash flows. In practice, the markets for debt and equity are segmented, and some bond investors may be unwilling (or unable under an investment mandate) to own equity. Such investors would value CoCos at less than their present value, and this effect could well move the price at which the market clears, given the comparatively small pool of investors focused on hybrid securities.

To capture this effect, we suppose that the equity received by CoCo investors at conversion is valued at 80% of market value. For example, we can think of CoCo investors as dumping their shares at a discount, with the discount reflecting a market impact that is only temporary and therefore does not affect the original equity holders. CoCo investors anticipate that they will not receive the full value of equity at conversion and thus discount the price of CoCos up front. This makes CoCos more expensive for the firm as a source of funding. The line marked with crosses in the right panel shows the benefit to equity holders of the same substitution examined in the left panel. As one would expect, the benefit is substantially reduced near the conversion trigger of 75 (comparing the two panels); at higher asset values, the difference between the cases vanishes, with the crossed lines in both panels near 0.3 at an asset level of 100. To summarize: *Segmentation between debt and equity investors creates a friction in conversion that reduces the benefit of issuing CoCos; this effect is especially pronounced near the conversion trigger.*

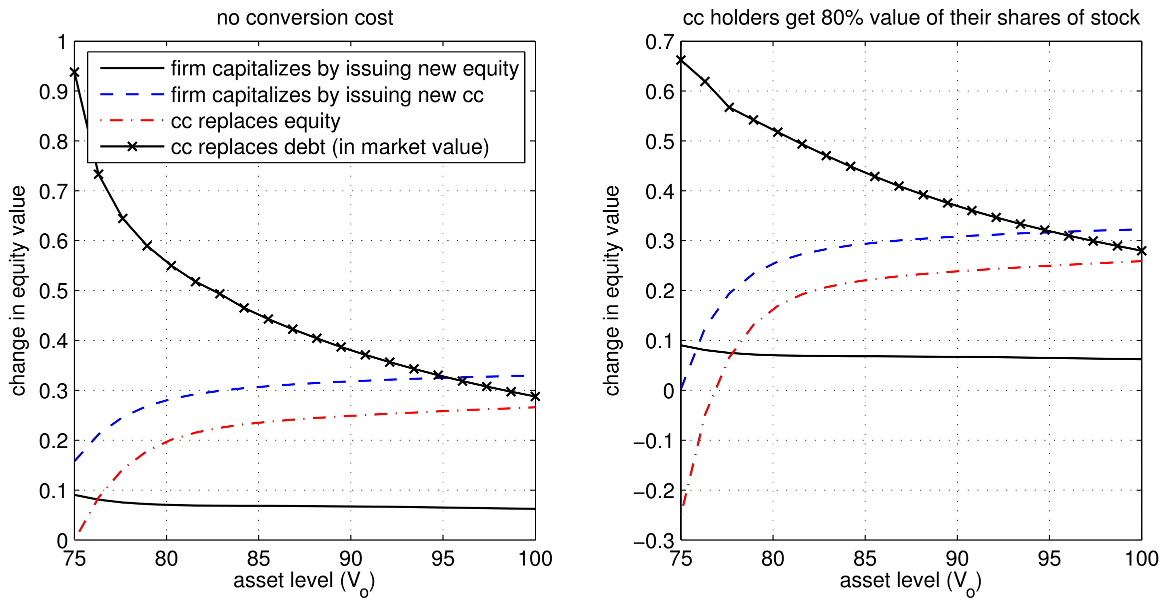


Figure 4.7: Change in equity value resulting from various changes in capital structure. In the right hand figure the CoCo holders dump their shares in the market following the conversion and as a result lose 20% value of their shares due price impact and transaction fees.

4.4.2 Increasing the Balance Sheet with CoCos

We now consider the effects of issuing CoCos without an offsetting reduction in any other liabilities. The proceeds from issuing CoCos are used to scale up the firm's investments. The consequences of this change are as follows:

- Because the post-conversion debt outstanding is unchanged, the endogenous default barrier V_b^* is unchanged, so long as the firm stay within the no-collapse region of debt levels.
- In this case, the risk of default decreases because an increase in assets moves the firm farther from the default barrier. The reduction in bankruptcy costs increases firm value and the value of straight debt. The additional funding benefit from issuing CoCos (assuming, for example, that their coupons are tax-deductible) further increases firm value.
- Shareholders benefit from the increase in firm value combined with the decrease in rollover costs for straight debt and the increase in cash generated from the larger asset base. These benefits work in the opposite direction of the increase in coupon payments required for the new CoCos.
- With a sufficiently large CoCo issue, the firm faces debt-induced collapse: the value of equity drops, the firm's default probability and bankruptcy costs jump up.

The dashed line in each panel of Figure 4.7 shows the benefit to shareholders of issuing a unit of new CoCos within the no-collapse region of debt levels. The benefit is lower on the right in the presence of a conversion friction. Whereas the incentive for debt substitution decreases with asset value, the incentive for issuing new CoCos increases with asset value. For completeness the figures also include the impact of replacing some equity with CoCos, which is roughly parallel to the effect of issuing new CoCos.

4.4.3 The Bail-In Case

Figure 4.8 illustrates the same comparisons made in the left panel of Figure 4.7, but now for the bail-in case. The main observation is that the incentive (for shareholders) to issue convertible debt is greater in Figure 4.8 than in Figure 4.7. This is primarily due to the lowering of the conversion threshold — the trigger is 75 in Figure 4.7 whereas the bail-in point is a bit below 70 in Figure 4.8. As long as conversion occurs before bankruptcy, the level of the conversion threshold has no effect on firm value or the value of straight debt. It does affect how value is apportioned between equity holders and CoCo investors.

4.5 Debt Overhang and Investment Incentives

In most capital structure models, equity holders are least motivated to invest in a firm precisely when the firm most needs additional equity. For a firm near bankruptcy, much of

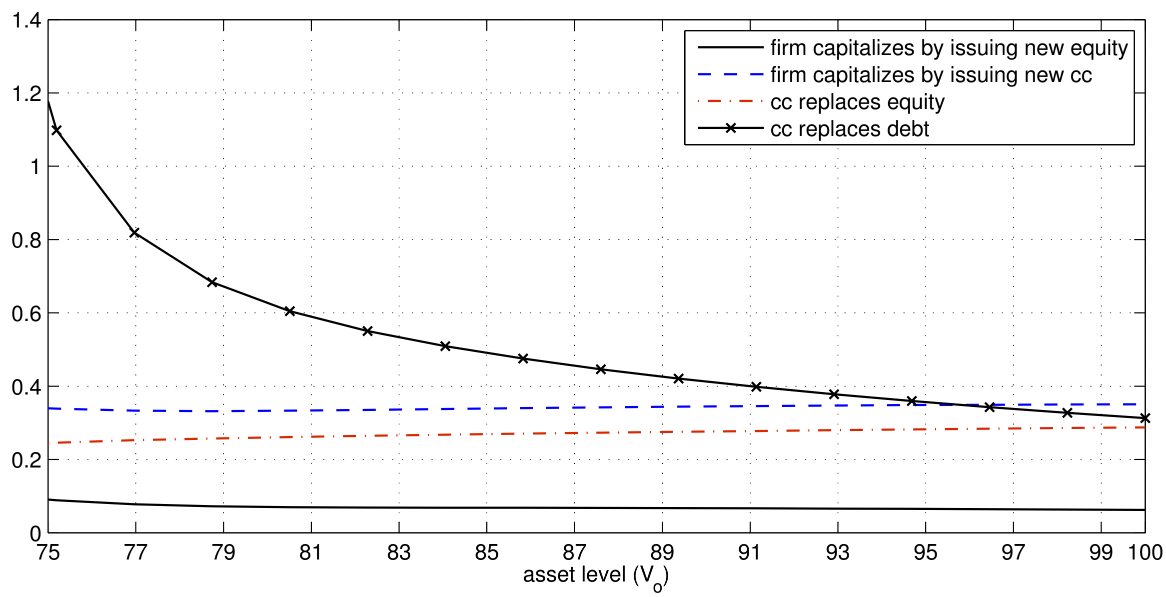


Figure 4.8: Change in equity value resulting from changes in capital structure with bail-in debt.

the value of an additional equity investment is captured by debt holders as the additional equity increases the market value of the debt by reducing the chances of bankruptcy. This is a problem of debt overhang (Myers (1977)), and it presents a significant obstacle to recapitalizing ailing banks. Duffie (2010) has proposed mandatory rights offerings as a mechanism to compel investment. Here we examine the effect of CoCos on investment incentives.

The phenomenon of debt overhang is easiest to see in a static model, viewing equity as a call option on the assets of a firm with a strike price equal to the face value of debt, as in Merton (1974). At a low asset value, where the option is deep out-of-the-money, the option delta is close to zero: a unit increase in asset value produces much less than a unit increase in option value, so equity holders have no incentive to invest. Indeed, in this static model, the net benefit of investment is always negative.

At least three features distinguish our setting from the simple static model. First, the reduction in rollover costs that follows from safer debt means that equity holders have the potential to derive some benefit from an increase in their investment. Second, the dilutive effects of CoCo conversion creates an incentive for shareholders to invest to prevent conversion. Third, if CoCo coupons are tax deductible, shareholders have an added incentive to invest in the firm near the conversion trigger to avoid the loss of this tax benefit.

Figure 4.9 shows the cost to equity holders of an additional investment of 1 in various scenarios. Negative costs are benefits. For this example, we use the longer maturities for debt in Table C.1, as the overhang problem is more acute in this case. This is illustrated by the solid black line in the left panel, which shows the overhang cost is positive throughout the range of asset values displayed.

The solid blue line and the dashed line show the overhang cost after the firm has issued CoCos. The blue line corresponds to replacing equity with CoCos, and the dashed line corresponds to replacing straight debt with CoCos. As we move from right to left, tracing a decline in asset value toward the conversion threshold $V_c = 75$, we see a dramatic increase in the benefit (negative cost) to equity holders of an additional investment. In other words, *the presence of CoCos creates a strong incentive for equity holders to invest in the firm to avoid conversion.* After conversion (below an asset level of 75), the overhang cost reverts to its level in a firm without CoCos.

The right panel of Figure 4.9 provides further insight into the investment incentive illustrated in the left panel. If we lower the conversion trigger from 75 to 70, we see from the solid black line that the investment incentive becomes greatest at 70, as expected, where it is a bit greater than the greatest value in the left figure. Removing the tax-deductibility (and any other funding benefit) of CoCo coupons yields the dashed black line, which shows that the investment incentive is reduced but not eliminated. In the solid red line, we have returned the conversion trigger to 75 but removed the jumps from the asset process. This eliminates close to half the incentive for investment, compared to the left panel. Removing both the tax shield on CoCos and jumps in asset value eliminates almost all the investment incentive, as indicated by the dashed red line.

The tax effect is evident: the tax shield increases the value to shareholders of avoiding

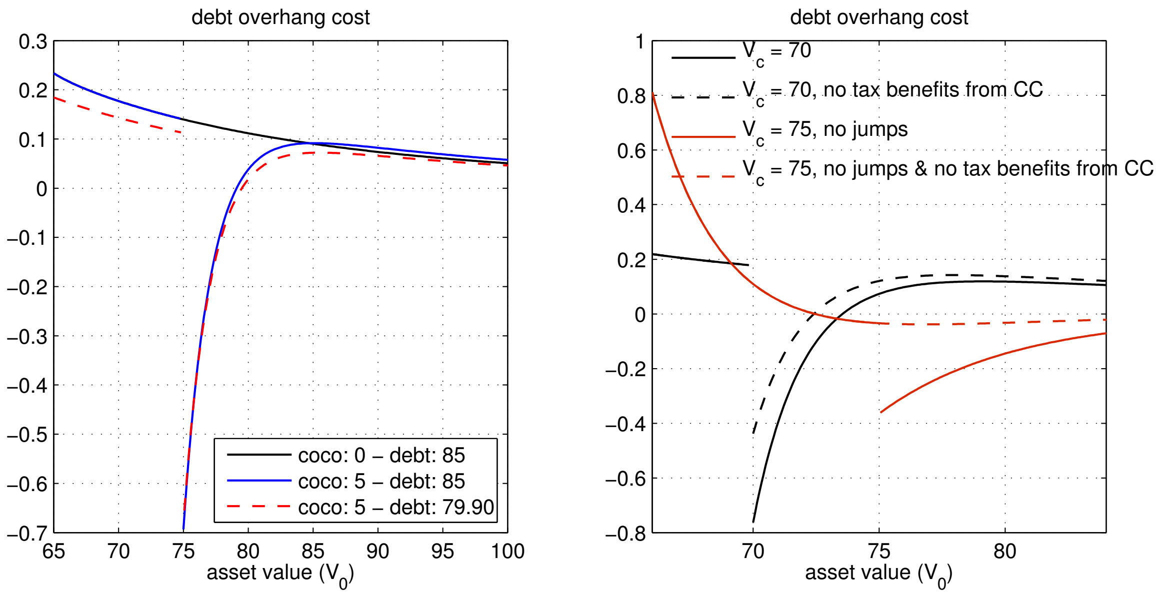


Figure 4.9: Net cost to shareholders of increasing the firm's asset by 1. Negative costs are gains. The figures show that CoCos and tail risk create a strong incentive for additional investment by equity holders near the conversion trigger.

the conversion of CoCos and thus creates a greater incentive for investment. The jump effect requires some explanation. Recall that the conversion ratio Δ is set so that the market value of the shares into which the CoCos convert would equal the face value of the converted debt *if conversion were to occur at an asset level of* V_c. If a downward jump takes V_t from a level above the trigger V_c to a level below it, then conversion occurs at an asset level lower than V_c, and the market value of the equity granted to CoCo investors is less than the face value of the debt. Equity holders thus prefer conversion following a jump to conversion at the trigger; indeed, conversion right at the trigger is the worst conversion outcome for equity holders, and this creates an incentive for investment as asset value approaches the trigger. The equity holders would prefer to delay conversion and, in effect, bet on converting at a jump rather than right at the trigger. This suggests that CoCos may create an incentive for equity investors to take on further tail risk, an issue we investigate in the next section.

4.6 Asset Substitution and Risk Sensitivity

We reviewed the problem of debt overhang in the previous section in Merton's (1974) model, which views equity as a call option on the firm's assets. The same model predicts that equity value increases with the volatility of the firm's assets, giving equity holders an incentive to increase the riskiness of the firm's investments after they have secured funding from creditors. In this section, we examine this phenomenon in our dynamic model, focusing on how CoCos change the incentives.[11]

We can summarize our main observations as follows. Because of the need to roll maturing debt, equity holders do not necessarily prefer more volatile assets in a dynamic model; longer debt maturity makes riskier assets more attractive to equity holders. Even when equity value does increase with asset volatility, CoCos can mitigate or entirely offset this effect, in part because equity holders are motivated to avoid conversion. In some cases, CoCos can make tail risk more attractive to equity holders even while making diffusive risk less attractive.

To illustrate these points, we start with the lower panel of Figure 4.10, which shows the sensitivity of equity to diffusive volatility as a function of asset value. The solid black line corresponds to a firm with no contingent capital — the sensitivity of equity to σ is positive throughout the range and peaks just above the default barrier. As the firm nears bankruptcy, the equity holders are motivated to take on extra risk in a last-ditch effort at recovery.

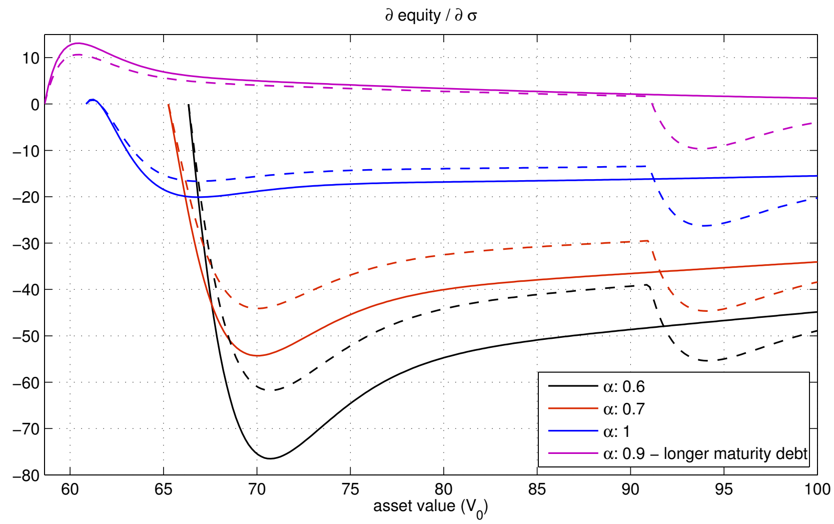
We see a very different pattern in the two blue lines, corresponding to a firm in which some straight debt has been replaced with CoCos, and the two red lines, based on replacing some equity with CoCos. In both cases, the solid line is based on a conversion trigger of 85, and the dashed line uses a trigger of 70. This gives us four combinations of capital structure and trigger level. In all four, the sensitivity is *negative* at high asset values and turns sharply negative as asset value decreases toward the conversion boundary before becoming slightly positive just above the trigger, where equity holders would prefer to gamble to avoid conversion. After conversion, the pattern naturally follows that of a firm without CoCos. The key implication of the figure is that CoCos decrease, and even reverse, the incentive for the shareholders to increase the riskiness of the firm's assets.

The top half of Figure 4.10 illustrates the effect of debt maturity and bankruptcy costs on the risk-shifting incentive. In each pair of lines, the dashed line has the same level of straight debt as the solid line but it also has CoCos. Considering first the solid lines, we see that with long-maturity debt, the risk-shifting incentive is positive, even at a rather high recovery rate of $\alpha = 90\%$. In contrast, with shorter maturity debt, the sensitivity is nearly always negative, even with a recovery rate of 100% — i.e., with no bankruptcy costs. Thus, debt maturity and not bankruptcy cost is the main driver of the sign of the risk-sensitivity. CoCos therefore have a greater effect on the risk-shifting incentive when the rest of the firm's debt has longer average maturity. The impact of CoCos is not very sensitive to the recovery rate α.

Figures 4.11 and 4.12 illustrate similar comparisons but with the sensitivity at each asset level normalized by the value of equity at that asset level; we interpret this as measuring the risk-shifting incentive per dollar of equity. Also, the figures compare sensitivities to diffusive volatility on the left with sensitivity to tail risk, as measured by $1/\eta_f$, on the right. Figure 4.12 uses a longer average maturity of debt than Figure 4.11.

The left panels of Figures 4.11 and 4.12 are consistent with what we saw in Figure 4.10 for

[11] Related questions of risk-shifting incentives are studied in Albul et al. (2010), Hilscher and Raviv (2011), Koziol and Lawrenz (2012), and Pennacchi (2010) with contingent capital and in Bhanot and Mello (2006) for debt with rating triggers.



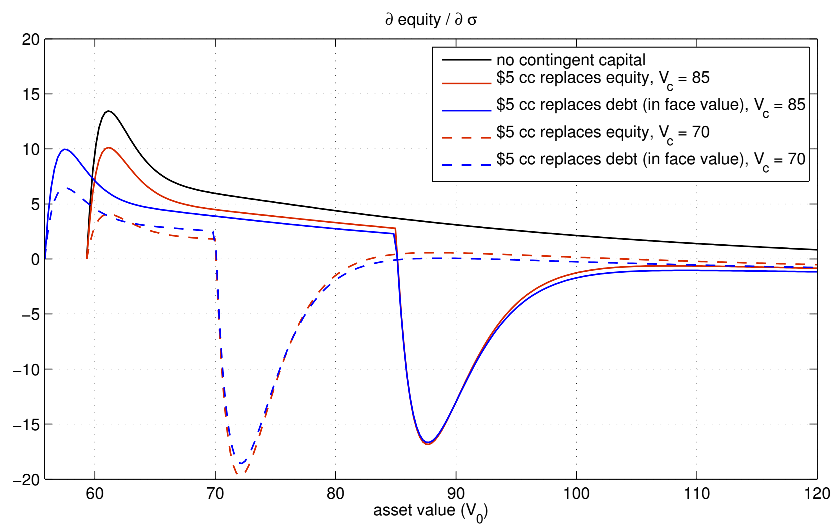


Figure 4.10: Sensitivity of equity value to diffusive volatility σ. With longer maturity debt, equity holders have a positive risk-shifting incentive. CoCos tend to reverse this incentive.

the unnormalized sensitivities: with longer maturity debt, CoCos reverse the risk-shifting incentive; with shorter maturity debt, equity holders already have an incentive to reduce risk, particularly at low asset values, and CoCos make the risk sensitivity more negative.

The right panels add new information by showing sensitivity to tail risk. In both Figures 4.11 and 4.12, equity holders have a positive incentive to add tail risk, particularly with long maturity debt, but also with short maturity debt at low asset levels. Indeed, the incentive becomes very large in both cases as asset value falls. Increasing the size of the firm's balance sheet by adding CoCos leads to a modest increase in this incentive above

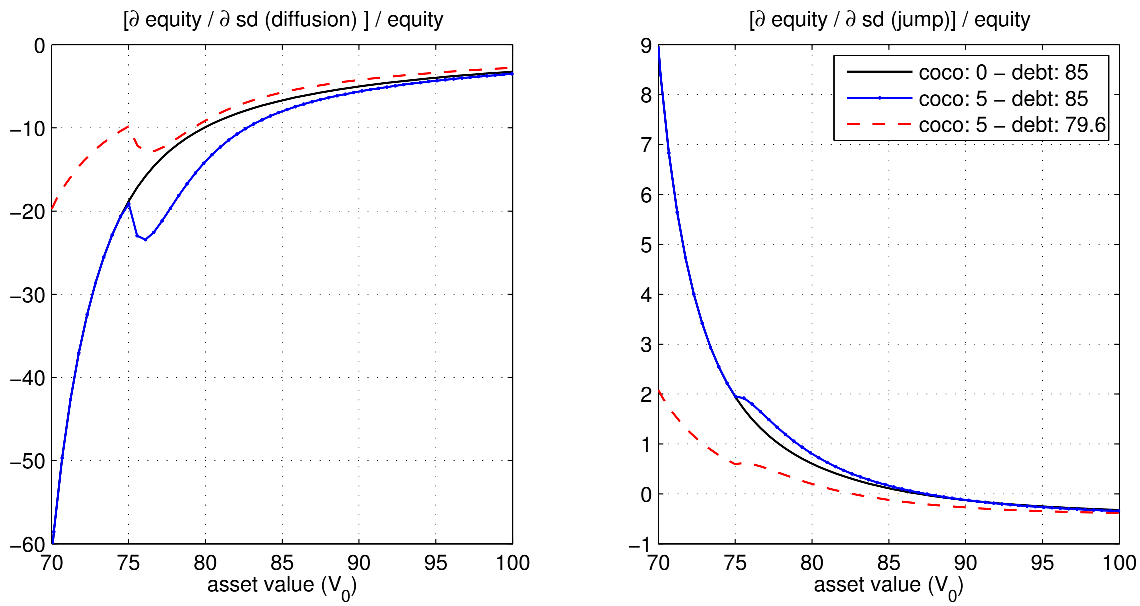


Figure 4.11: Sensitivity of equity value to diffusive volatility and jump risk in assets.

the conversion trigger. Replacing some straight debt with CoCos reduces the incentive to take on tail risk but does not reverse it. Related comparisons are examined in Albul et al. (2010) and Pennacchi (2010). Pennacchi's (2010) conclusions appear to be consistent with ours, though modeling differences make a direct comparison difficult; the conclusions in Albul et al. (2010) are quite different, given the absence of jumps and debt rollover in their framework.

The patterns in our results can be understood, in part, from the asset dynamics in (4.2); in particular, whereas the diffusive volatility σ plays no role in the (risk-neutral) drift, increasing the mean jump size increases the drift. In effect, the firm earns a higher continuous yield on its assets by taking on greater tail risk. This has the potential to generate additional dividends for shareholders, though the additional yield needs to be balanced against increased rollover costs resulting from increased default risk. In addition to generating a higher yield, jump risk is attractive to shareholders because the cost of conversion is lower if it takes place at a lower asset value than at the conversion trigger. Moreover, shareholders are indifferent between bankruptcy at an asset value below their default barrier or right at their barrier, so they are motivated to earn the higher yield from tail risk without bearing all of the downside consequences.

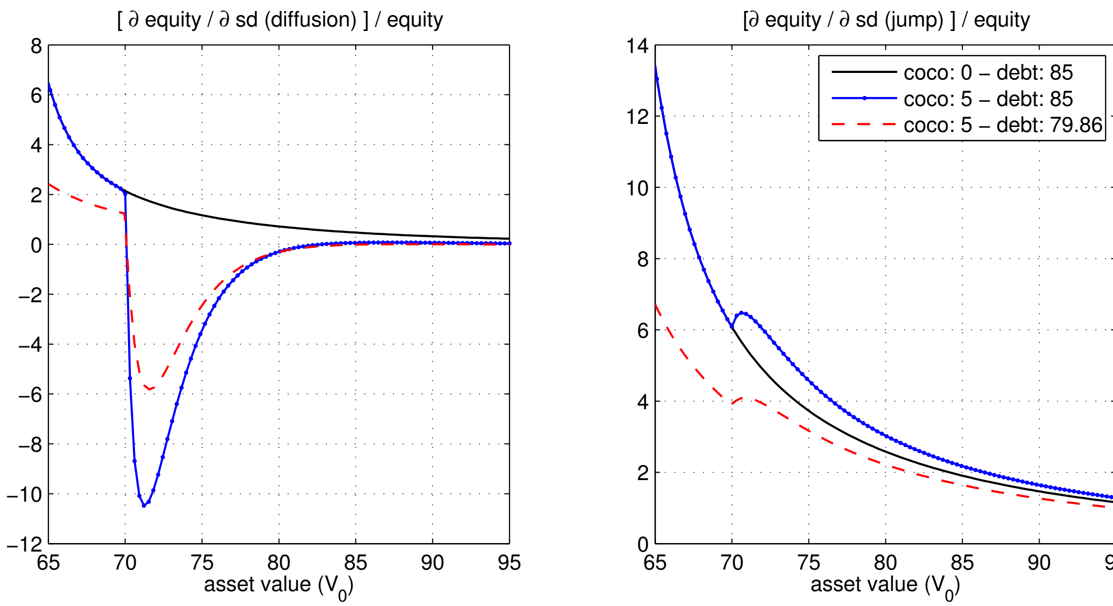


Figure 4.12: Same comparisons as Figure 4.11 but with longer average maturity. In all plots, at the same asset level the dashed line corresponds to a larger distance to default due to less outstanding regular debt.

4.7 Calibration to Bank Data Through the Crisis

In this section, we calibrate our model to specific banks. We focus on the years leading up to and during the financial crisis, with the objective of gauging what impact CoCos might have had, had they been issued in advance of the crisis. We examine the increase in the banks' ability to absorb losses, relative to the amount of straight debt replaced with CoCos, and we calculate the reduction in debt overhang costs as an indication of whether CoCos would have created greater incentives for equity holders to inject private capital at various points in time.

As candidates for our calibration, we chose the 19 bank holding companies (the largest 19 at the time) that underwent the Supervisory Capital Assessment Program (SCAP) in 2009. From this list, we removed MetLife because banking is a small part of its overall business, and we removed GMAC (now Ally) because it is privately held. The banks are listed in Table 4.2, in order of asset value in 2009.

We obtain quarterly balance sheet information from each bank holding company's quarterly 10-Q/10-K S.E.C. filings from 2004 through the third quarter of 2011, except in the case of American Express, for which we begin in 2006 because of a large spin-off in 2005.

Several of the firms became bank holding companies late in our time window, so Y-9 reports would not be available throughout the period. Also, the Y-9 reports contain less information about debt maturities and interest expenses than the quarterly reports. We group all debt into three categories — deposits, short-term debt, and long-term debt — in this order of seniority. We do not separate subordinated debt from other long-term debt because of difficulties in doing so consistently and reliably. The distinction would not have much effect on our calculations. We calculate average debt maturity within each category using information provided in annual reports. We calculate total dividends and interest payments to get a total payout rate.

We linearly interpolate values within each quarter, using values from the beginning of the quarter and the beginning of the subsequent quarter; this gives us values at a weekly frequency and avoids abrupt changes at the end of each quarter. For debt maturities, we interpolate between annual reports.

Our model is driven by asset value, but asset value is not observable. So, we fit our model using balance sheet and market information and then use the model to infer asset value or a model-defined proxy for asset value. In more detail, at each week we use the linearly interpolated values to determine the bank's debt profile, dividends, and interest. As the risk-free rate, we use the Treasury yield corresponding to the weighted average maturity of each bank's debt.

Jump parameters are difficult to estimate, particularly for rare jumps as contemplated by our model. For the calibrations, we limit the model to a single type of jump and choose from a finite set of values for the jump rate λ and the mean jump size $1/\eta$. For each (λ, η), we calibrate a value for the diffusive volatility σ iteratively as follows. Given a starting value for σ, we can numerically invert our model's formula for equity at each point in time (using the market value of equity at each point in time) to get an implied market value for the assets. We then calculate the annualized sample standard deviation of the implied asset log returns, excluding returns of magnitude greater than 3.3σ, which we treat as jumps, and compare it with σ. We adjust σ up or down depending on whether the standard deviation is larger or smaller than σ, proceeding iteratively until the values match. At that point, we have found a path of underlying assets that reproduces the market value of equity with an internally consistent level of asset volatility, for a fixed (λ, η).

We repeat this procedure over a grid of (λ, η) values. We limit λ to 0.1 or 0.3; for η, we consider integer values between 5 and 10, but if the best fit occurs at the boundary we extend the range to ensure that does not improve the fit. We choose from the set of (λ, η, σ) values by comparing model implied debt prices with market data of traded debt from the Fixed Income Securities Database and TRACE databases. We add up the total principal of traded debt and total market price paid in those transactions. Their ratio gives an average discount rate that the market applies to the debt. We calculate the corresponding model implied average discount for each (λ, η, σ) using quarterly balance sheet data for the principal of debt outstanding and the model implied prices. The interest payments are already matched through our choice of coupon rates, so we choose the (λ, η, σ) that comes closest to matching the discount on the principal as our calibrated parameters. The parameters for the 17 banks

	Parameters			Conversion Date	
Bank Holding Company	λ	η	σ	50%	75%
Bank of America Corp	0.1	5	4.1%	Jan-09	
JPMorgan Chase & Co.	0.1	8	4.4%		
Citigroup Inc.	0.1	9	3.9%	Nov-08	
Wells Fargo & Company	0.1	5	4.7%		
Goldman Sachs Group, Inc.	0.1	5	3.8%	Nov-08	
Morgan Stanley	0.1	8	4.2%	Sep-08	Dec-08
PNC Financial Services	0.3	8	7.0%	Nov-08	Jan-09
U.S. Bancorp	0.3	5	5.5%	Jan-09	
Bank of New York Mellon Corp.	0.3	6	7.3%	Oct-08	
SunTrust Banks, Inc.	0.3	9	4.1%	Apr-08	Jan-09
Capital One Financial Corp.	0.3	7	7.9%	Jun-08	Jan-09
BB&T Corporation	0.3	6	5.3%	Jun-08	
Regions Financial Corporation	0.3	8	4.7%	Jun-08	Jan-09
State Street Corporation	0.3	5	7.4%	Oct-08	
American Express Company	0.3	8	8.6%		
Fifth Third Bancorp	0.3	5	6.3%	Jan-08	Jun-08
KeyCorp	0.3	8	4.2%	Nov-07	Nov-08

Table 4.2: The table shows the calibrated parameter values (λ, η, σ) for each bank holding company. The last two columns show the months in which CoCo conversion would have been triggered, according to the calibration, assuming CoCos made up 10% of debt. The 50% and 75% dilution ratios correspond to higher and lower triggers, respectively.

are reported in Table 4.2.

Given the path of asset value and all the other model parameters, we can calculate model-implied quantities. As a first step, we calculate the endogenous bankruptcy level V_b^* based on the bank's debt profile at each point in time. We can also undertake a counterfactual experiment in which part of the debt is replaced with CoCos and recalculate the default boundary. We take CoCos to be 10% of total debt, keeping the relative proportions of other types of debt unchanged. Recall that the default boundary does not depend on the CoCo conversion trigger or conversion ratio, as long as the trigger is above the default boundary, so we do not need to specify values for these features to determine V_b^*. In other words, we assume that the conversion trigger is set to prevent debt-induced collapse.

Table 4.3 provides more detailed information at four points in time. Under each date, the value on the left is the ratio of increased loss absorption to the market value of CoCos, where the increased loss absorption is the change in the default barrier resulting from the CoCos. A ratio of 1 indicates that a dollar of CoCos absorbs a dollar of additional losses; a ratio greater or smaller than 1 indicates a greater or smaller degree of loss absorption. The second entry under each date is the distance to default as a percentage of asset value. Comparing a single

	Jan-2006		Jan-2007		Jan-2008		Jan-2009	
Bank of America Corp	1.47	7%	1.43	8%	1.63	5%	1.54	3%
JPMorgan Chase & Co.	1.29	6%	1.29	6%	1.49	5%	1.50	5%
Citigroup Inc.	1.34	7%	1.32	6%	1.42	4%	-	2%
Wells Fargo & Company	1.11	19%	1.06	22%	1.44	9%	1.60	5%
Goldman Sachs Group, Inc.	1.35	4%	1.41	5%	1.52	4%	-	4%
Morgan Stanley	1.43	4%	1.38	4%	1.50	5%	-	5%
PNC Financial Services	1.17	19%	1.11	21%	1.29	14%	-	8%
U.S. Bancorp	0.95	32%	0.98	32%	1.11	24%	1.17	18%
Bank of New York Mellon	1.15	24%	1.06	28%	1.04	28%	0.80	17%
SunTrust Banks, Inc.	0.91	21%	0.87	22%	0.91	16%	-	8%
Capital One Financial Corp.	0.93	29%	0.92	26%	0.97	16%	-	12%
BB&T Corporation	1.03	25%	1.03	23%	0.97	14%	-	9%
Regions Financial Corp.	0.90	24%	0.89	19%	0.87	12%	-	4%
State Street Corporation	1.33	18%	1.25	20%	1.07	24%	-	11%
American Express Company	1.15	38%	1.13	36%	1.26	28%	1.50	18%
Fifth Third Bancorp	0.89	26%	0.77	31%	-	17%	-	6%
KeyCorp	1.11	17%	1.01	20%	-	10%	-	5%
mean	1.15	18.81%	1.11	19.23%	1.23	13.73%	1.35	8.15%
median	1.15	19.32%	1.06	20.52%	1.26	13.80%	1.50	5.81%

Table 4.3: Under each date the left column shows the ratio of the increase in loss absorption (the change in the default boundary after CoCo issuance) to CoCo size (as measured by market value). The right column is the distance to default (without CoCos) as a percentage of asset level. The dilution ratio is 50%.

institution at different points in time, the pattern that emerges is that the loss absorption ratio tends to be greater when the firm is closer to default. The pattern does not hold across institutions because there are too many other differences in their balance sheets besides the distance to default.

The design and market value of the CoCos depends on two contractual features, the trigger V_c and the conversion price Δ. By the definition of Δ, the fraction of total equity held by CoCo investors just after conversion is $\Delta P_2/(1+\Delta P_2)$, where P_2 is the face value of CoCos issued. We choose Δ so that this ratio is either 50% or 75%, and we refer to this as the dilution ratio. We then set the conversion level V_c so that if conversion were to occur exactly at $V_t = V_c$, the market value of the equity CoCo investors would receive would equal the face value P_2 of the CoCos: conversion at $V_t = V_c$ implies neither a premium nor a discount. In order that the equity value received be equal to P_2 at both 50% and 75% dilution ratios, the higher dilution ratio must coincide with a lower conversion trigger. The results in Table 4.3 are based on a 50% dilution ratio, but the corresponding results with 75% dilution are virtually identical.

The last two columns of Table 4.2 report the month in which the model calibrations predict each of the banks would have triggered conversion of CoCos with a high trigger (50% dilution ratio) and a low trigger (75% dilution ratio). In each case, the CoCo size is equal to 10% of the bank's total debt. The calibrations predict that all the banks except JPMorgan Chase, Wells Fargo, and American Express would have crossed the high conversion trigger sometime between November 2007 and January 2009; seven of the banks would have crossed the lower conversion trigger as well.

Next, we consider debt overhang costs. For each bank in each week, we calculate the size of the equity investment required to increase assets by 1%. From this we subtract the net increase in equity value, which we calculate by taking the value of equity just after the investment (as calculated by the model) and subtracting the value of equity just before the investment (as observed in the data). This is our measure of debt overhang cost: if it is positive, it measures how much less equity holders get from their investment than they put in. A negative cost indicates a net benefit to investment.

Table 4.4 presents more detailed information at three dates prior to key points in the financial crisis: one month before the announcement of JP Morgan's acquisition of Bear Stearns; one month before final approval of the acquisition; and one month before the Lehman bankruptcy. For each date, the table shows the debt overhang cost without CoCos and with high-trigger CoCos; the third column under each date shows the distance to the conversion boundary as a percentage of asset value. Interestingly, several of the largest banks show significantly negative debt overhang costs even without CoCos. Recall from Section 4.5 that this is possible in a model with debt rollover, though not with a single (finite or infinite) debt maturity. Greater asset value implies greater bankruptcy costs and reducing these costs may partly explain the motivation for shareholders to increase their investments in the largest firms. Also, if the market perceives a too-big-to-fail guarantee for the largest banks that is absent from our model, then the model's shareholders may see the largest banks as overly leveraged relative to the market's perception.

We focus on comparisons between columns of the table — a single firm under different conditions — rather than comparisons across rows. With few exceptions, the effect of the CoCos is to lower the debt overhang cost, and the impact is often substantial. The effect depends on the interaction of several factors, including leverage, debt maturity, and the risk-free rate, which enters into the risk-neutral drift. The largest reductions in debt overhang cost generally coincide with a small distance to conversion, and, in most cases in which a bank draws closer to the conversion boundary over time, the resulting reduction in debt overhang cost becomes greater. The values in the table are for 50% dilution. The pattern with 75% is similar, but the decrease in the debt overhang cost is smaller in that case because the distance from the conversion trigger is greater.

The magnitudes of the quantities reported in these tables and figures are subject to the many limitations and simplifications of our model and calibration. We see these results as providing a useful additional perspective on the comparative statics of earlier sections of the paper; the directional effects and the comparisons over time should be more informative than the precise numerical values. These calibrations and our exploration of counterfactual

	Feb-2008			Apr-2008			Aug-2008		
Bank of America Corp	-29%	-32%	6%	-26%	-30%	5%	-28%	-42%	3%
JPMorgan Chase & Co.	-75%	-51%	5%	-43%	-41%	5%	-93%	-60%	3%
Citigroup Inc.	-42%	-53%	3%	-24%	-45%	2%	-54%	-65%	2%
Wells Fargo & Company	-35%	-23%	8%	-33%	-20%	8%	-33%	-21%	7%
Goldman Sachs Group	-51%	-45%	2%	-33%	-42%	2%	-53%	-54%	2%
Morgan Stanley	21%	-42%	1%	21%	-36%	1%	-20%	-58%	2%
PNC Financial Services	-11%	-16%	7%	-7%	-12%	8%	-10%	-12%	8%
U.S. Bancorp	4%	4%	13%	5%	5%	13%	5%	5%	11%
Bank of New York Mellon	-3%	-2%	17%	-1%	0%	14%	6%	4%	8%
SunTrust Banks, Inc.	-2%	-20%	2%	5%	-	-	9%	-	-
Capital One Financial	-4%	-28%	3%	4%	-34%	2%	6%	-	-
BB&T Corporation	2%	-11%	4%	4%	-12%	4%	6%	-60%	1%
Regions Financial Corp.	-7%	-24%	3%	-8%	-42%	2%	-9%	-	-
State Street Corporation	2%	2%	11%	5%	-1%	6%	0%	-11%	5%
American Express Co.	-12%	-13%	20%	-7%	-10%	20%	-10%	-12%	17%
Fifth Third Bancorp	12%	-79%	0%	17%	-	-	19%	-	-
KeyCorp	-6%	-137%	0%	-1%	-	-	5%	-	-

Table 4.4: Under each date, the first column is the debt overhang cost as a percentage of the increase in assets with no CoCos. The second column quotes the same value when 10% of debt is replaced with CoCos and CoCo investors receive 50% of equity at conversion. The third column is the distance to conversion as the percentage of assets. The dates correspond to one month before announcement and final approval of acquisition of Bear Stearns by JPMorgan and one month before the Lehman bankruptcy. A table entry is blank if the corresponding date is later than the CoCo conversion date for the corresponding bank.

scenarios, though hypothetical, shed light on how CoCo issuance in advance of the financial crisis might have affected loss absorption capacity, incentives for additional equity investment, and how the choice of conversion trigger and dilution ratio might have determined the timing of conversion.

4.8 Concluding Remarks

The key contribution of this paper lies in combining endogenous default, debt rollover, and jumps and diffusion in income and asset value to analyze the incentive effects of contingent convertibles and bail-in debt. Through debt rollover, shareholders capture some of the benefits (in the form of lower bankruptcy costs) from reduced asset riskiness and lower leverage — benefits that would otherwise accrue solely to creditors. These features shape many of the incentives we consider, as do the tax treatment of CoCos and tail risk. The phenomenon of debt-induced collapse, which is observable only when CoCos are combined with endogenous default, points to the need to set the conversion trigger sufficiently high so that conversion unambiguously precedes bankruptcy. Our calibrations suggest that CoCos could have had a significant impact on the largest U.S. bank holding companies in the lead up to the financial crisis.

Our analysis does not include asymmetric information, nor does it directly incorporate agency issues; both considerations are potentially relevant to the incentives questions we investigate. Some important practical considerations, such as the size of the investor base for CoCos, the behavior of stock and bond prices near the trigger, and the complexity of these instruments are also outside the model. The analysis provided here should nevertheless help inform the discussion of the merits and potential shortcomings of CoCos and other hybrid capital instruments.

Chapter 5

Contingent Convertible Bonds: Modeling and Evaluation

5.1 Introduction

This chapter presents a formal model for a new regulatory hybrid security for financial firms, so-called contingent convertible bonds. This instrument has the features of normal debt in normal times, but converts to equity when the issuing firms are under financial stress. I develop and compare different modeling approaches for contingent capital with tail risk, debt rollover and endogenous default. In order to apply contingent convertible capital in practice it is desirable to base the conversion on observable market prices that can constantly adjust to new information in contrast to accounting triggers. I show how to use credit spreads and the risk premium of credit default swaps to construct the conversion trigger and to evaluate the contracts under this specification.

Although contingent bonds could in principle be used by any firm, the focus here is to analyze their potential as a regulatory instrument for banks. As the ongoing financial crisis has illustrated, banks play an important role in the economy. When they are healthy, banks channel savings into productive assets. But when they are distressed, this role is compromised and banks lend less with adverse effects on investment, output and employment. In this situation governments often intervene, but as we could see during the past crisis, the measures are costly to taxpayers and may be limited in effectiveness. There are several reasons why banks may inadequately recapitalize on their own in the first place.

First, there is the so-called debt overhang problem. If a bank suffers substantial losses, the managers, who should act in the interest of the shareholders, may prefer not to issue new equity. If a distressed bank issues new equity, the bank's bondholders profit from this as the new capital increases the likelihood that they will get repaid. On the other hand, existing shareholders bear costs as their claims on the firm are diluted. In this sense issuing new equity creates a transfer from existing shareholders to bondholders. Hence, in order to satisfy capital requirements shareholders may prefer the bank to sell risky assets or to reduce new lending instead of issuing new equity. If during a financial crisis other banks are in trouble, too, they will also cut lending and thus the economy as a whole suffers.

Second, there is the moral hazard problem of a government bailout. A government bailout represents an implicit guarantee for the bondholders that their debt will be repaid. If bondholders believe that the government will not allow a bank to fail, the bondholders may be more willing to lend money to a bank that pursues more risky strategies and have less incentives to control the bank. This problem is particularly severe for the most important financial institutes, that are "too big to fail".

Furthermore, bankruptcy reorganization for banks is different than default restructuring for other firms. Banking business relies on confidence. If a bank is in trouble, there is the danger of a bank run as clients and short-term creditors may withdraw their capital. As in the case of Lehman Brothers, distress for a financial firm often leads to partial or complete liquidation in contrast to a restructuring according to Chapter 11 which can help a "normal" company to return to economic viability.

In summary, the debt overhang problem can make banks reduce lending or sell assets instead of recapitalizing themselves and maintaining their lending capacity. If restructuring takes place, it is usually ineffective and disruptive and can affect other institutions. The possibility of a government bailout can increase the riskiness of the strategies of a bank. For this reason the discussion in the aftermath of the financial crisis has focussed on a resolution mechanism that can allow quick and less disruptive recapitalization of distressed banks, but does not shift the costs of risky activities to the government. A possible solution is contingent convertible bonds.

Contingent convertible bonds (CCB) are instruments that convert into equity if the bank is financially distressed. The bank would issue these bonds before a crisis and if a certain trigger is reached, conversion takes place automatically. The automatic conversion of debt into equity would transform an undercapitalized bank into a well capitalized bank at no cost to the taxpayer.

The key issues for specifying CCBs are: When does conversion take place and how many shares are given to the bondholders at conversion? The automatic conversion should be triggered by the same mechanism that triggers default. From our understanding bankruptcy is caused if the value of the firm's assets is below a default barrier. Hence, conversion should take place if the value of the firm's assets reaches another barrier, namely the conversion barrier. In view of the application as a regulation instrument it is sensible to require this conversion barrier to be higher than the default barrier. In addition to the trigger, the rate at which the debt converts into equity has to be specified. There are basically two different approaches. In the first approach each dollar of debt converts into a fixed quantity of equity shares. In this case the total value of the portfolio of shares granted at conversion depends on the stock price at conversion. In the second approach the conversion is specified in terms of the market value of equity. In this case the number of shares granted at conversion depends on the stock price at the time of conversion. This work considers both types of converting debt into equity.

CBBs have been traded only very recently. In 2009 Lloyd's bank issued the first £7 billion CCBs. The first and only time so far that contingent convertible bonds were used as a regulatory instrument was in Switzerland, when the Credit Suisse Group AG issued $2 billion of these new bonds on February 14th, 2011. The coupon payments of the contingent convertible bonds were substantially higher than for normal debt: 7.875% vs. 4% on average.

The idea of contingent convertible bonds has been a very vivid area of research in the last years. However, the literature on formal models of contingent convertible bonds is still limited. Qualitative discussions can be found in Flannery (2009a+b), Squam Lake Working Group on Financial Regulation (2009), McDonald (2010) and Calomiris and Herring (2011). The first structural formal model is presented in Albul, Jaffee and Tchistyi (2010). Their model is based on Leland's (1994b) structural credit risk model with optimal default barrier. The firm's value process follows a geometric Brownian motion and the bonds are of infinite maturity. Conversion is triggered when the firm's value process reaches an exogenous conversion barrier and the conversion value is expressed in terms of the market value of equity. Their paper provides many very interesting insights concerning regulation and capital structure decisions. Our paper is closely related to Albul, Jaffee and Tchistyi's (2010) work, but our model is based on Hilberink and Rogers' (2002) and Chen and Kou (2009)'s jump process framework. As in Albul, Jaffee and Tchistyi we work with a structural credit risk model, in which the optimal default barrier is chosen endogenously by the shareholders by trading off tax benefits and bankruptcy costs. However, the bonds in our model have finite maturity and are issued such that we obtain a stationary debt structure. The firm's value in our model follows a specific jump diffusion process, namely a Kou process. This particular choice of process allows us to obtain a non-zero credit spread limit for a maturity approaching zero. The valuation of CCBs becomes considerable more challenging under a jump diffusion process. A jump that triggers conversion can be sufficiently large to trigger bankruptcy as well. Hence, the conversion value of CCBs explicitly depends on the features of the straight debt. In Albul, Jaffee and Tchistyi the valuation of straight debt and contingent convertible debt could be separated. In our framework we can show that the valuation of the two different debt instruments is interlinked. Furthermore, we do not only treat the conversion barrier as exogenously given, but we also consider the case of it being chosen optimally by the shareholders or the firm.

Specifying the conversion trigger in terms of the firm's value process is conceptionally appealing and allows us to derive analytical solutions for all prices. However, the firm's value process is unobservable and it would be desirable to base the conversion event on observable prices. Some literature, e.g. De Spiegeleer and Schoutens (2011), propose to trigger conversion when the stock price process first crosses a barrier level. As we will show the stock price is in general not a sufficient statistic for the firm's value process. Therefore, conversion based on the stock price cannot be incorporated into our modeling framework. Moreover, it is possible that more than one stock price and CCB price are consistent with our equilibrium conditions if conversion is based on the stock price. These shortcomings are not appealing. However, we can show that the unobservability of the firm's value process can be circumvented by using credit spreads or the risk premiums of credit default swaps (CDS). Credit spreads and CDS risk premiums have the same advantages as stock prices as they constantly adjust to new information in contrast to accounting triggers. We will prove that they are a sufficient statistic for the firm's value process. Thus, defining the conversion event in terms of credit spreads or CDS risk premiums is equivalent to using the firm's value process. In summary, our evaluation formulas can be applied in practice with a trigger event based on observable market prices.

We extend our model into several directions. First, we introduce a general approach which allows all parameters of the firm's value process to change after the conversion. In particular we can model the special case, that the drift of the firm's value process increases after conversion as interest payments decrease. Second, we introduce exogenous noise trading into the stock price process and can show that under certain conditions our pricing formulas are not affected. Third, we show that contracts, for which the number of shares granted at conversion is fixed a priori, are more robust against manipulation by the contingent convertible bondholders than contracts with a fixed value at conversion. It is possible to design a contract that is robust against manipulation by the equity holders and contingent convertible bondholders.

In the last part of the chapter we analyze whether contingent convertible bonds can be used as a regulation instrument. We can show that under a technical assumption, the no-early-default condition, a regulation that combines a restriction on the maximal leverage ratio and the requirement of issuing a certain fraction of contingent convertible bonds as part of the whole debt, can efficiently lower the default probability without reducing the total value of the firm. However, if the no-early-default condition is violated, a regulation based on contingent convertible debt can actually increase the risk. Therefore, it is crucial to make sure that this condition is satisfied. If the contingent convertible bonds are issued in a smaller amount than the straight debt, have a long maturity and the conversion barrier is sufficiently high, the no-early-default condition is generally satisfied.

The chapter is organized as follows. In Section 5.2 we present the formal model of Chen and Kou. In Section 5.3 we add the contingent convertible bonds to the model. In Section 5.4 we derive closed-form prices for CCBs. Section 5.5 discusses the choice of the default and conversion barrier. In Section 5.6 we consider the case where conversion is based on observable market prices instead of the unobservable firm's value process. In Section 5.7 we present some numerical simulations. Section 5.8 discusses the optimal design of CCBs. In Section 5.9 we consider extensions to our model. Section 5.10 focusses on contingent convertible bonds as a regulation instrument. Section 5.11 concludes. Most of the proofs and the special case of a pure diffusion process are collected in the Appendix.

5.2 Model for Normal Debt

In this section we review Chen and Kou's (2009) model, in which the firm's value process follows a particular jump-diffusion process. It is based on Leland's (1994a) diffusion and Hilberink and Rogers' (2002) jump diffusion structural credit risk models with optimal default barrier. As the model presented here applies to any firm and as we will only consider the special case of banks in the section about regulation, we use the generic terminology "firm" instead of "bank".

The value of the firm's assets at time t is denoted by V_t and under the risk-neutral measure $\mathbb{P}$ [1] it evolves as

$$dV_t = V_t(dZ_t + (r - \delta)dt) \tag{5.1}$$

where Z is some martingale, r is the constant riskless interest rate and δ represents the proportional rate at which a part of the assets is disbursed to investors.[2]

We will follow Chen and Kou (2009) and specify Z as a jump diffusion process with double exponentially distributed jumps. Note that as the firm has bondholders and shareholders, δ cannot be seen as a dividend rate. First the coupons and principal repayments have to be paid before the residual can be paid out as dividends. By assuming a constant riskless interest rate we neglect the interest rate risk.

The firm is partly financed by debt which has two features: Its time structure and its riskiness. In order to obtain a stationary debt structure, debt is constantly retired and reissued. Assume that at every point in time the firm issues new debt in the amount of p_D, i.e. in the time interval $(t, t+dt)$ new bonds with face value $p_D dt$ are issued. The debt has the maturity profile φ, where φ can be any non-negative function with $\int_0^\infty \varphi(s)ds = 1$ and can be interpreted as a density function. We will choose the maturity profile as $\varphi(t) = me^{-mt}$, i.e. the maturity of a specific bond is chosen randomly according to an exponentially distributed random variable. Of all the debt issued in $(t, t+dt)$ the debt with face value $p_D\varphi(s)dtds$ will mature in the time interval $(t+s, t+s+ds)$. If we also consider all the debt that was issued before $t = 0$ the face value of debt maturing in $(s, s+ds)$ is

$$\int_{-\infty}^{0} p_D\varphi(s-x)dxds = \int_{s}^{\infty} p_D\varphi(y)dyds = p_D\Psi(s)ds \quad , \quad \Psi(s) \equiv \int_{s}^{\infty} \varphi(y)dy.$$

For our maturity profile this equals

$$\int_{-\infty}^{0} p_D m e^{-m(s-x)}dxds = p_D ds.$$

The face value of all the newly issued debt is $p_D ds$. Hence, the face value of all the debt maturing in $(s, s+ds)$ is equal to the face value of the newly issued debt. Thus, the face value of debt stays constant and at every point in time equals

$$P_D = p_D \int_0^\infty \Psi(s)ds = \frac{p_D}{m}.$$

Note, that for our maturity profile $\varphi(t) = me^{-mt}$ the parameter m is a measure of maturity. As m increases a higher fraction of the debt matures earlier. If default never

[1] As we consider only one martingale measure in this section, $\mathbb{E}$ will denote the expected value with respect to $\mathbb{P}$.

[2] We use the drift $r - \delta$ because we work under the risk-neutral measure $\mathbb{P}$. Under the real world probability measure the drift would be $\mu - \delta$, where μ is some constant.

occurs, the average maturity of debt is:

$$\int_0^\infty t\varphi(t)dt = \int_0^\infty t(me^{-mt})dt = \frac{1}{m}.$$

For $m = 0$ only consol bonds are issued as in Leland (1994b). The choice of the exponential maturity is needed in our model to express the debt in terms of some Laplace transform. This enables us to derive an explicit solution for double exponential jump diffusion processes.

We assume that all debt is of equal seniority and is paid by coupons at the fixed rate $c_D dt$ for the time interval $(t, t + dt)$ until either maturity or default occurs. The first time the value of the firm falls to some level V_B or lower and thus default happens is denoted by τ. The default barrier V_B will be chosen optimally by the shareholders. In the case of default, we assume that a fraction α of the value of the firm's assets is lost. The value of a bond issued at time 0 with face value 1 and maturity t is therefore

$$d_D(V, V_B, t) = \mathbb{E}\left[\int_0^{t\wedge\tau} c_D e^{-rs} ds\right] + \mathbb{E}\left[e^{-rt}\mathbb{1}_{\{t<\tau\}}\right] + \frac{1}{P_D}(1-\alpha)\mathbb{E}\left[V(\tau)e^{-r\tau}\mathbb{1}_{\{\tau\leq t\}}\right].$$

The first term represents the net present value of all coupons up to the minimum of t and τ. The second term can be interpreted as the net present value of the firm's repayment when default does not happen. The last term is the net present value of the assets if bankruptcy occurs. We assume that the face value of all debt is P_D and thus a bondholder with a bond with face value 1 gets the fraction $1/P_D$ of the value $(1 - \alpha)V(\tau)$ that remains after bankruptcy. Note that if V were continuous, $V(\tau)$ would simply be V_B, but for a process with jumps this need not be the case.

The total value of all debt outstanding given our assumptions about the maturity profile was derived by Chen and Kou (2009) and is given in the next proposition.

Proposition 5.1. *The total value of all outstanding debt for the maturity profile* $\varphi(t) = me^{-mt}$ *is*

$$\begin{aligned} D(V, V_B) &= \int_0^\infty p_D\Psi(t)d_D(V, V_B, t)dt \\ &= \frac{c_D P_D + mP_D}{m+r}\mathbb{E}\left[1 - e^{-(m+r)\tau}\right] + (1-\alpha)\mathbb{E}\left[V(\tau)e^{-(m+r)\tau}\mathbb{1}_{\{\tau<\infty\}}\right]. \end{aligned}$$

The main problem now is to compute the two expectations. For V following a geometric Brownian motion an explicit solution is easily available. Note that the expectation $\mathbb{E}\left[V(\tau)e^{-(r+m)\tau}\mathbb{1}_{\{\tau<\infty\}}\right]$ is bounded from above by V_B and below by 0. In the two extreme cases for the maturity rate m, Lebesgue's dominated convergence theorem yields the following corollary:

Corollary 5.1. *The value of the debt for* $m \to \infty$ *and* $m \to 0$ *is given by*

$$\begin{aligned} \lim_{m\to 0} D(V, V_B) &= \frac{c_D P_D}{r}\mathbb{E}\left[1 - e^{-r\tau}\right] + (1-\alpha)\mathbb{E}\left[V(\tau)e^{-r\tau}\mathbb{1}_{\{\tau<\infty\}}\right] \\ \lim_{m\to\infty} D(V, V_B) &= P_D. \end{aligned}$$

The limit for $m \to 0$ corresponds to the case of consol bonds as in Leland's (1994b) paper.

The total coupon rate of all the debt equals $C_D = c_D P_D$. As by the choice of our maturity profile, P_D is constant over time and c_D is assumed to be fixed, we get a stationary debt structure: A unit of bonds issued one year ago will look exactly the same as a unit of bonds issued today. If the value of the firm's assets does not change they will also have the same price.

According to Modigliani and Miller the total value of the firm is expressed as the sum of the asset value plus tax benefits minus bankruptcy costs. We assume that there is a proportional corporate tax rate $\bar{c}$ and coupon payments can be offset against tax. Thus, for the total coupon rate $C_D = c_D P_D$ the firm receives an additional income stream of $\bar{c} C_D dt$.

Definition 5.1. *The tax benefits associated to the debt are denoted by TB_D and the bankruptcy costs by BC. The total value of the firm is defined as*

$$G_{debt}(V, V_B) = V + TB_D(V, V_B) - BC(V, V_B).$$

Proposition 5.2. *The total value of the firm equals*

$$G_{debt}(V, V_B) = V + \frac{\bar{c} C_D}{r} \mathbb{E}\left[1 - e^{-r\tau}\right] - \alpha \mathbb{E}\left[V(\tau) e^{-r\tau} \mathbb{1}_{\{\tau < \infty\}}\right]. \tag{5.2}$$

The value of the firm consists of the value of its assets plus the net present value of the tax rebates minus the net present value of the losses on default. Now, we can express the value of the firm's equity as

$$EQ_{debt}(V, V_B) = G_{debt}(V, V_B) - D(V, V_B).$$

Optimal capital structure and optimal endogenous default are two interlinked problems. The optimal debt level P_D and the optimal bankruptcy trigger V_B have to be chosen simultaneously. When a firm chooses P_D in order to maximize the total value of the firm at time 0, the decision depends on V_B. Vice versa, the optimal default trigger V_B is a function of the amount of debt P_D. Leland (1994a+b) and Leland and Toft (1996) have shown how to choose P_D and V_B according to a two-stage optimization problem. In the first stage, for a fixed P_D, equity holders choose the optimal default barrier by maximizing the equity value subject to the limited liability constraint. In a second stage, the firm determines the amount of debt P_D that maximizes the total value of the firm. More precisely, the first stage problem is

$$\max_{V_B} EQ_{debt}(V, V_B) \qquad \text{such that } EQ_{debt}(V', V_B) > 0 \text{ for all } V' > V_B$$

The "smooth pasting" condition as derived in Leland and Toft (1996) delivers an optimality criterion:

$$\frac{\partial EQ_{debt}}{\partial V}(V, V_B)|_{V=V_B} = 0.$$

In the case of two-sided jumps Chen and Kou prove that the solution to the smooth pasting condition indeed maximizes the equity, respecting the constraint that the value of equity must remain non-negative at all times. The optimal default barrier will be denoted by V_B^* and is clearly a function of P_D. The second stage optimization is formulated as

$$\max_{P_D} G_{debt}(V, V_B^*(P_D)).$$

In our model bankruptcy occurs at an endogenously determined asset value V_B. For all asset values larger than V_B equity has a positive value. Note, that this does not mean that bankruptcy occurs when debt service payments exceed the cash flow δV. At any point in time the firm issues bonds which are worth D, but has to make a debt service of P_D and after-tax coupon payment of$(1-\bar{c})C_D$. Hence, $(\delta V-(1-\bar{c})C_D-P_D+D)$ is the payout rate to the shareholders. As V falls, the cash flow δV declines and the price of the debt $D(V, VB)$ will fall as well, which can result in a negative payout rate to the shareholders. As long as the equity value is positive, new stock can be issued to meet debt service requirements. Hence, bankruptcy can only occur when the equity value becomes zero.

The initial total coupon rate C_D will be chosen, such that debt sells at par, i.e. the price of the debt equals its face value. Therefore, if the value of the firm's assets does not change, the firm can use the newly issued debt to repay the face value of the old debt. As we will see later the optimal value of V_B depends on the coupon rate C_D. Hence, for a fixed amount of debt P_D, the initial coupon payment C_D is computed by solving the following two equations simultaneously. First, debt has to sell at par:

$$P_D = D(V_0, V_B, P_D, C_D). \tag{5.3}$$

Second, the optimality criterion is to maximize the equity value:

$$\left(\frac{\partial EQ_{debt}(V_0, V_B, P_D, C_D)}{\partial V_0}\right)_{V_0=V_B} = 0. \tag{5.4}$$

5.3 A Model for Contingent Convertible Debt

5.3.1 Modeling Conversion

The special property of contingent convertible bonds is that the debt automatically converts to equity if the firm or bank reaches a specified level of financial distress. We will model this by introducing a barrier V_C. The first time the value of the firm falls to or below this level, the convertible bond fully converts into equity. Hence, the conversion time is defined as

$$\tau_C = \inf(t \in (0, \infty) : V(t) \leq V_C).$$

The challenge of modeling convertible debt lies in the specification of the conversion value. First, we present a model where the conversion value is based on a fixed number of shares. In this case the stock price processes has to be modeled as well. We label them fixed share contingent convertible bonds (FSC). Second, we consider a model where the conversion value is based on the market value of equity, i.e. the number of equity shares depends on the stock price at conversion. We label them as fixed value convertibles (FVC):

1. **FSC (Fixed share convertibles):** Conversion value in terms of a fixed number of shares.

   Here, the number of shares granted in exchange for a contingent convertible bond are fixed in the contract. Therefore, the conversion value depends on the stock market price at the time of conversion. We will model the stock price endogenously as a fraction of the equity. There are basically two ways to choose the fixed number of shares granted at conversion. Either this number is fixed at time zero without any reference to other prices; in this case we obtain a unique price for the FSCs. The alternative is to express this number in terms of the stock price $S(t)$ at time $t = 0$. The number of shares granted at conversion for a single CCB will then equal $\frac{\ell}{S(0)}$. The coefficient ℓ is a contract term. The value of the corresponding shares at time τ_C is $S(\tau_C)$. Hence, at the time of conversion bondholders receive equity valued at its market price $\ell \frac{S(\tau_C)}{S(0)}$. However, the stock price at time 0 depends on the features of the CCBs, while the price of the CCBs also depends on the stock price $S(0)$. Hence, the stock price $S(0)$ and the price of CCBs have to be determined in an equilibrium. We will show that for this case there exist in general two equilibrium prices for FSCs.

2. **FVC (Fixed value convertibles):** Conversion value in terms of the market value of equity.

   Bondholders receive equity valued at its market price in the amount of ℓ at the time of conversion τ_C. The coefficient ℓ is a contract term that determines the fraction of the conversion value to the face value of the convertible bond at the time of conversion. In the following we will assume that the value of equity for $V_t = V_C$ is sufficient to pay the conversion value. This makes sense as the bondholders would only agree on a contract, where it is known a priori that it is possible to fulfill the contractual obligations. There are basically two ways of how the payments at conversion can be specified. In a model without jumps both approaches coincide. Suppose that conversion is triggered by a jump in V_t that crosses the conversion barrier V_C. In the first approach, we assume that the number of shares granted to the contingent convertible bondholders is determined as if the firm's value process first touches V_C, conversion takes place at this time and then the firm's value process jumps to the value V_{τ_C}. This means the number of shares granted at conversion for a single bond with face value 1 is $n' = \frac{\ell}{S(V_C)}$, where $S(V_C)$ denotes the stock price given the firm's value process is equal to V_C. The value of the payment is then $n' \cdot S(\tau_C)$. We can think of $S(V_C)$ as a hypothetical stock price and show that it is known at time zero when the contract is written. Hence, the number n'

is a constant and this kind of FVC contract is actually a FSC contract. We will label it FVC1.

In the second approach the number of shares granted is determined based on the stock price $S(\tau_C) = S(V_{\tau_C})$, i.e. $n' = \frac{\ell}{S(V_{\tau_C})}$. In this case conversion takes place after the firm's value process has crossed the conversion barrier. If the crossing happens by a jump, $S(V_{\tau_C})$ is smaller than $S(V_C)$. This implies that the number n' is a random variable, which makes the contract different from the above FSCs. We use the name FVC2 for this second specification. In this paper we have solved the model for both cases. The first approach has the advantage that FVCs and FSCs can be incorporated into the same framework. Hence, in the main part of the text we focus on this contract specification. In the Appendix we present the detailed solution to the second approach.

In analogy to the normal debt case we denote by P_C the total value of the convertible debt. The fixed coupon paid by a unit contingent convertible debt is c_C and the total amount of the coupon payments equals C_C. The maturity profile of the contingent convertible and the normal debt is the same, but it is straightforward to relax this assumption. In summary we have the following equations for a bond d_D with face value 1 and a contingent convertible bond d_C with face value 1:

$$\tau = \inf(t \in (0,\infty) : V_t \le V_B) \qquad \tau_C = \inf(t \in (0,\infty) : V_t \le V_C)$$

$$P_D = p_D \int_0^\infty \Psi(s)ds = \frac{p_D}{m} \qquad P_C = p_c \int_0^\infty \Psi(s)ds = \frac{p_C}{m}$$

$$C_D = c_D P_D \qquad C_C = c_C P_C.$$

Implicitly, we make the following assumption:

Assumption 5.1. *The conversion level is always equal or larger than the bankruptcy level:*

$$V_C \ge V_B.$$

If contingent convertible debt is to be used as a regulation instrument, it is sensible to make the even stronger assumption that $V_C > V_B$. In the case where $V_C \le V_B$ the contingent convertible debt degenerates to straight debt without any recovery payment.

The pricing structure of contingent convertible bonds is similar to that of straight debt bonds: The price consists of the net present value of the coupon payments until conversion, the net present value of the firm's repayment if conversion does not occur and finally the conversion value if conversion happens before maturity. The two different types of CCBs considered in this paper distinguish themselves only in the conversion value. The value of a single contingent convertible bond with face value 1 and maturity t equals

$$d_C(V, V_B, V_C, t) = \mathbb{E}\left[\int_0^{t\wedge\tau_C} c_C e^{-rs} ds\right] + \mathbb{E}\left[e^{-rt}\mathbb{1}_{\{t<\tau_C\}}\right] + conv(V, V_B, V_C, t)$$

where $conv(V, V_B, V_C, t)$ is the conversion value of the respective bond. Following the same argument as for straight debt we obtain the following proposition:

Proposition 5.3. *The total value of all outstanding convertible debt equals*

$$\begin{aligned} CB(V, V_B, V_C) &= \int_0^\infty p_C \Psi(t) d_C(V, V_B, V_C, t) dt \\ &= \left(\frac{c_C P_C + m P_C}{m + r} \right) E\left[1 - e^{-(m+r)\tau_C}\right] + CONV(V, V_B, V_C) \end{aligned}$$

where CONV is the total conversion value

$$CONV(V, V_B, V_C) = \int_0^\infty p_C \cdot \Psi(t) \cdot conv(V, V_B, V_C, t) dt$$

We have decided to use the terminology "conversion value" instead of "conversion payment" for $CONV$. The reason is that $CONV$ is not an actual cash payment like the coupon payments or the repayment of the principal value if conversion does not take place. The conversion value represents a redistribution of the equity value among shareholders in the event of conversion. This perspective is important when evaluating the value of the equity. A more appropriate but less practicable name for $CONV$ would be "the total value of the shares granted to contingent convertible bondholders at conversion".

5.3.2 Consistency and Equilibrium Requirements

The total tax benefits are the sum of the tax benefits of the straight debt and the tax benefits of the contingent convertible debt:

$$\begin{aligned} TB(V_t, V_B, V_C) &= TB_D(V_t, V_B, V_C) + TB_C(V_t, V_B, V_C) \\ &= \frac{\bar{c} C_D}{r} \mathbb{E}\left[1 - e^{-r\tau}\right] + \frac{\bar{c} C_C}{r} \mathbb{E}\left[1 - e^{-r\tau_C}\right] \end{aligned}$$

where $C_D = c_D P$ and $C_C = c_C P_C$ denote the total values of the coupon payments and $\bar{c}$ is the tax rate. As before coupon payments are tax deductible. The bankruptcy cost are

$$BC(V_t, V_B) = \alpha E\left[V(\tau) e^{-r\tau} \mathbb{1}_{\{\tau < \infty\}}\right].$$

The total value of the firm equals

$$G = V_t + TB - BC.$$

The value of the equity consists of the total value of the firm minus the payments which the equity holders have to make to the bondholders. The payments to the holders of straight debt have a different structure than the payments to contingent convertible bondholders. The value of a contingent convertible bond could be split into two parts: First, the value of the coupon payments and the repayment of the principal value if conversion does not take place. These are actual cash payments. Second, the value of the shares granted at conversion. The conversion shares given to the contingent convertible bondholders are not a cash payment but represent a redistribution among shareholders. The value of the conversion shares depends on the value of the equity as a whole. For the old shareholders it represents an actual cost, but it does not change the value of the equity as a whole. The total equity will be only affected by the value of actual cash payments. Hence, the conversion value for the various contingent convertible bonds $CONV(V, V_B, V_C)$ does not directly enter the valuation formula for the total equity. The term $CB - CONV$ equals the principal value paid in the case of no conversion and the value of the coupon payments of the contingent convertible bonds. We define the total equity as

$$\begin{aligned} &EQ(V_t, V_B, V_C) \\ =&V_t + TB(V_t, V_B, V_C) - D(V_t, V_B) - (CB(V_t, V_B, V_C) - CONV(V_t, V_B, V_C)) - BC(V_t, V_B). \end{aligned}$$

Note that at any time t the following budget equation has to hold:

$$V_t + TB = EQ + D + CB - CONV + BC.$$

In Section 5.3.3 we introduce dilution costs $DC(V_t, V_B, V_C)$. These are the costs that the old shareholders have to bear because the claim on equity will be distributed among more shareholders after the conversion. At the time of conversion there is only one value transfer: The contingent convertible bondholders receive the conversion value $CONV$ and the old equity holders suffer from a loss in value equal to the dilution costs DC. As there is no other value created or destroyed the budget equation requires the dilution costs to equal the conversion value.

Lemma 5.1. *The dilution costs DC coincide with the conversion value $CONV$.*

After having specified the concrete form of the dilution costs and the conversion value, we will verify in Section 5.3.3 that the above lemma holds in our model. In this sense our model is consistent. Denote by

$$EQ_{old}(V_t, V_B, V_C) = EQ(V_t, V_B, V_C) - DC(V_t, V_B, V_C)$$

the value of the equity for the old shareholders.[3] By Lemma 5.1 this can be written as

$$EQ_{old}(V_t, V_B, V_C) = V_t + TB(V_t, V_B, V_C) - BC(V_t, V_B) - D(V_t, V_B) - CB(V_t, V_B, V_C)$$

In a model without contingent convertible debt the equity holders choose the conversion barrier V_B such that it maximizes the value of the equity subject to the constraint that the value of the equity is strictly positive for a firm's value process larger than V_B. In a model including contingent convertible debt the default decision before conversion is made by the old shareholders. However, if default does not happen before conversion, the contingent convertible bondholders become equity holders as well. The optimal default barrier for the

[3] The labeling "total equity" for EQ and "equity for the old shareholders" for EQ_{old} is our own notation. In other papers, e.g. Albul, Jaffee, Tchistyi (2010), the equity for the old shareholders is just called equity.

old and new equity holders after conversion can be different than for the old shareholders before conversion. The fact, that the old shareholders cannot commit to their optimal choice before conversion, will be called the commitment problem. Hence, for a given amount of debt P_D and P_C the old shareholders will choose a default barrier that maximizes the value of their equity subject to the limited liability constraint and the commitment problem:

$$\max_{V_B} EQ_{old}(V_t, V_B)$$
$$\text{s.t. } EQ_{old}(V', V_B) > 0 \text{ for all } V' > V_B \text{ and s.t. the commitment problem.}$$

In Section 5.5 we will formulate the problem formally and derive a general solution to it. For a given level of debt P_D and P_C, the coupon values will be determined at time $t = 0$ such that all the debt sells at par:

$$P = D(V_0, V_B, P_D, P_C, C_D, C_C) \tag{5.5}$$
$$P_C = CB(V_0, V_B, V_C, P_D, P_C, C_D, C_C). \tag{5.6}$$

As the variables V_B, C_D and C_C are determined endogenously, the remaining choice variables are $V_C, P_D, P_C, m, \ell, \bar{c}, V_0$ and r. In the following we will usually suppress the dependence of the functions on all the parameters and use a short-hand notation where we only implicitly write the dependence on the variables of interest.

5.3.3 Modeling the Stock Price Process

Stock Price Process without CCBs

A single stock is a claim on a fixed portion of the equity of a firm. Hence, we can define the stock price process as the value of the equity for the old shareholders divided by the number of shares.

Definition 5.2. *If the capital structure of a firm includes only straight debt, but no contingent convertible debt, the stock price is defined as*

$$S_t = S(V_t) = \frac{EQ_{debt}(V_t)}{n}$$

where n is the number of shares $n = EQ(V_0)_{debt}/S(0)$.

Dilution Costs

Shareholders do not only profit from the additional tax benefits from issuing contingent convertible bonds, but also face the risk of dilution. This creates a tradeoff. In more detail, at the time of conversion all the cash payments of the contingent convertible bonds, i.e. the coupon payments and the repayment of the face value, fall to zero. Hence, the total value of the equity of a firm at conversion is the same as the total value of the equity of an identical firm that did not issue any CCBs. At the time of conversion holders of contingent convertible bonds become equity holders. As this implies that new shares are issued, the value of the shares of the old shareholders decreases: "The size of the cake stays the same, but is divided among more people." Here we want to model the dilution costs for the old shareholders. Denote by n the number of shares of the old shareholders and by n' the number of shares of all the new shareholders after conversion. Hence the costs of dilution DC are defined as

$$DC(V_t, V_B, V_C) = \mathbb{E}\left[\frac{n'}{n+n'} EQ(V_{\tau_C}) e^{-(r+m)(\tau_C - t)} \mathbb{1}_{\{\tau_C < \infty\}} \mathbb{1}_{\{\tau_C < \tau\}} | \mathfrak{F}_t\right].$$

DC corresponds to the value of the shares of the new shareholders. It is calculated as the present value of their fraction of the total equity weighted by the maturity profile. Hence, the value of the equity for the old shareholders equals

$$EQ_{old}(V_t, V_B, V_C) = EQ(V_t, V_B, V_C) - DC(V_t, V_B, V_C).$$

The main difference between our two different contingent convertible bonds is the number of shares granted to the bondholders.

Stock Price Process with CCBs

If the capital structure of a firm includes debt and CCBs we define the stock price in the following way:

Definition 5.3. *If Assumption 1 is satisfied, the endogenous stock price process is defined as*

$$S_t = S(V_t) = \begin{cases} \frac{EQ(V_t) - DC(V_t)}{n} & \text{if } t < \tau_C \\ \frac{EQ(V_t)}{n+n'} & \text{if } \tau_C \leq t < \tau \\ 0 & \text{if } t \geq \tau \end{cases}$$

where n is the number of "old" shares

$$n = \frac{EQ(V_0) - DC(V_0)}{S_0}$$

and n' is the number of "new" shares issued at conversion.[4]

Note that $EQ_{old}(V_t) = EQ(V_t) - DC(V_t)$ is just the value of the equity for the old shareholders.

[4]The definition of the stock price implicitly assumes that no new shares can be issued before conversion, i.e. the number of old shares n at time $t = 0$ is the same as the number of old shares n at conversion $t = \tau_C$. This restrictive assumption is only needed for the evaluation of FSCs. For the evaluation of FVC1 and FVC2 contracts and all the other results in this paper, this assumption is not necessary and can be relaxed.

5.3.4 Pricing CCBs

Pricing FSCs

The old shareholders own a number of shares that is equal to the value of equity to them divided by the price of the stock at time $t = 0$:

$$n = \frac{EQ(V_0) - DC(V_0)}{S_0}.$$

The number n is fixed at time zero. Note, that the stock price S_0 has to be determined endogenously and will depend on the features of the CCBs. Assume first, that n' is fixed and does not depend on S_0. The contingent convertible bondholders receive a fixed number of shares at conversion if and only if conversion and bankruptcy do not happen at the same time. This condition is captured by $\tau_C < \tau$ or equivalently $V_{\tau_C} > V_B$.

Proposition 5.4. *If the value of the shares, that holders of a single contingent convertible bond with face value 1 receive at conversion, is $n'S(\tau)/P_C$, then the value of the individual bond satisfies*

$$d_C(V, V_B, V_C, t) = E\left[\int_0^{t\wedge\tau_C} c_C e^{-rs} ds\right] + E\left[e^{-rt}\mathbb{1}_{\{t<\tau_C\}}\right] + \frac{n'}{P_C}\mathbb{E}\left[S(\tau_C)e^{-r\tau_C}\mathbb{1}_{\{\tau_C\leq t\}}\mathbb{1}_{\{V_{\tau_C}>V_B\}}\right].$$

Under the assumption of an exponential maturity profile $\varphi(t) = me^{-mt}$ the total value of the convertible debt CB is given by:

$$CB(V, V_B, V_C) = \left(\frac{c_C P_C + mP_C}{m+r}\right) E\left[1 - e^{-(m+r)\tau_C}\right] + n'\mathbb{E}\left[S(\tau_C)e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{V_{\tau_C}>V_B\}}\right].$$

In Lemma 5.1 we have already noted that $CONV(V_0) = DC(V_0)$. Hence, instead of evaluating the conversion value, we will focus on the dilution costs, which are equal to

$$DC(V_t) = \frac{n'}{n+n'}\mathbb{E}\left[EQ(V_{\tau_C})e^{-(r+m)(\tau_C - t)}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{\tau_C<\tau\}}|\mathfrak{F}_t\right] \tag{5.7}$$

We will evaluate this expression analytically in Section 5.4.3.

The value of the equity $EQ(V_{\tau_C})$ is independent of the contingent convertible bonds as conversion has already taken place. We confirm that the consistency result in Lemma 5.1 is satisfied for our choice of the stock price process.

Corollary 5.2. *The conversion value equals the dilution costs:*

$$CONV(V_t) = DC(V_t)$$

Proof. Plugging in the definition of $S(t)$ yields

$$\begin{aligned} CONV(V_0) =& n'\mathbb{E}\left[S(\tau_C)e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{\tau_C<\tau\}}\right] \\ =& \frac{n'}{n+n'}\mathbb{E}\left[EQ(V_{\tau_C})e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\tau\}}\mathbb{1}_{\{\tau_C<\infty\}}\right] = DC(V_0) \end{aligned}$$

□

Assume that the number n' of total shares granted at conversion is expressed in terms of the stock price. If a single FSCs gives a bondholder ℓ/S_0 shares, then the total number of shares equals

$$n' = \frac{P_C \ell}{S_0} = \frac{n\ell P_C}{EQ(V_0) - DC(V_0)}.$$

This is an intuitive way of specifying the contract as it just says how much of the face value of debt are convertible bondholders going to get in terms of the current stock price if conversion takes place. However, introducing CCBs into the capital structure changes the stock price. As soon as the agents in the economy anticipate that contingent convertible capital will be issued they will discount the current stock price by the dilution costs. As the dilution costs and the stock price at time 0 are interlinked variables, specifying n' in terms of S_0 will in general lead to two possible equilibrium prices for FSCs. Therefore, it is undesirable to specify the conversion value of FSCs in terms of the stock price S_0, and one should avoid such a contract design.

Proposition 5.5. *If* $n' = \frac{P_C \ell}{S_0}$*, then there exist two different combinations of prices for* $\{S_0, DC(0)\}$*, which satisfy the consistency and equilibrium conditions for FSCs. The dilution costs* $DC(V_0)$ *at time* $t = 0$ *for a contingent convertible bond with such a fixed number of shares equal:*

$$\begin{aligned} DC(V_0) =& \frac{EQ(V_0) + n'S_0}{2} \\ &\pm \sqrt{\left(\frac{EQ(V_0) + n'S_0}{2}\right)^2 - \mathbb{E}\left[EQ(V_{\tau_C}) e^{-(r+m)\tau_C} \mathbb{1}_{\{\tau_C < \infty\}} \mathbb{1}_{\{\tau_C < \tau\}}\right] n'S_0}. \end{aligned}$$

In order to rule out multiple equilibrium prices, the FSC contract has to be specified such that n and n' are fixed a priori and chosen independently of S_0. After the equilibrium stock price is realized, we can determine a posteriori for which contract parameter ℓ the condition $n' = \frac{\ell P_C}{S_0}$ is satisfied. In our numerical analysis we will refer to n' as a fraction of the face value of debt in terms of the current market value of equity, because it is easier to interpret. However, this will not be the contract definition but an a posteriori consequence of the contract specification.

Pricing FVCs

The conversion value of FSCs depends on the stock market price at conversion. FVCs are designed to avoid this uncertainty and the conversion value is defined as a fraction ℓ of the face value of convertible debt. There are basically two ways of how to conceptualize the conversion value for FVCs. In the first case, which we will label FVC1, we introduce the hypothetical stock price $S(V_C)$, which would be realized if the firm's value process only touches but does not jump over the conversion barrier V_C. The total number of shares granted to the bondholders equals $n' = \frac{\ell P_C}{S(V_C)}$. Given certain assumptions about the dynamics of the firm's value process the value $S(V_C)$ is uniquely defined and known a priori at time 0. Hence, FVC1s are actually a specific version of FSC contracts. Although the stock price appears in the definition of n' we will show that there exists a unique equilibrium price. FVC1 have the advantage that they are relatively easy to implement.

In the second approach, labeled as FVC2, we try to take into consideration the fact that if conversion is triggered by a jump in the firm's value process, the stock price at conversion $S(V_{\tau_C}) = S_{\tau_C}$ is lower than the hypothetical stock price $S(V_C)$. In particular, it can be so low that the value of the equity is not sufficient to make the promised payment. If the value of the equity after conversion is sufficiently large, the bondholder get $n' = \frac{\ell P_C}{S_{\tau_C}}$ shares, otherwise they take possession of the whole firm. Note, that n' is a random variable in this case, which makes FVC2 different from FSC contracts. Because of the possibility of complete dilution at conversion the stock price process has to be modeled differently for FVC2s than for the other contracts. Most of the derivations for FVC2s are explained in Appendix D.1.

FVC1: In the case of FVC1s we make the assumption that the value of the equity at conversion is sufficient to make the promised payment. Relaxing this assumption is straightforward as the agents would just price in the additional risk. However, it seems sensible that the agents would usually only agree on a contract where it is known a priori that the contractual obligations can be fulfilled.

Assumption 5.2. *The parameters of the FVC1 contract are chosen such that*

$$EQ(V_C) \geq \ell P_C,$$

i.e. the equity value at conversion is sufficient to give shares to the bondholders with a value equal to the promised payment.

Proposition 5.6. *If the value of the shares given to holders of contingent convertible bonds at conversion is ℓ, the values of the individual bonds of FVC1 under Assumption 5.2 satisfy*

$$d_C(V, V_B, V_C, t) = E\left[\int_0^{t\wedge\tau_C} c_C e^{-rs} ds\right] + E\left[e^{-rt}\mathbb{1}_{\{t<\tau_C\}}\right] + \ell\mathbb{E}\left[\frac{S(\tau_C)}{S(V_C)} e^{-r\tau_C}\mathbb{1}_{\{\tau_C\leq t\}}\mathbb{1}_{\{V_{\tau_C}>V_B\}}\right].$$

The total value of the convertible debt FVC1 for an exponential maturity profile $\varphi(t) = me^{-mt}$ under Assumption 5.2 is given by:

$$CB(V, V_B, V_C) = \left(\frac{c_C P_C + mP_C}{m+r}\right) E\left[1 - e^{-(m+r)\tau_C}\right] + \ell P_C \mathbb{E}\left[\frac{EQ(V_{\tau_C})}{EQ(V_C)} e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{V_{\tau_C}>V_B\}}\right].$$

Note, that the value of the equity after conversion is independent of any features of the contingent convertible debt. Hence, we obtain a unique equilibrium price for FVC1 contracts although the number of new shares n' is defined with respect to some stock price.

FVC2: For FVC2s we do not require Assumption 5.2. The conversion value for FVC2s requires us to distinguish several cases. If $\tau_C < \tau$, i.e. the downward movement of V_{τ_C} is not sufficient to trigger bankruptcy, the contingent convertible bondholders receive a payment. If on the one hand the value of the equity is sufficiently large, they get a number of stocks such that the value of the total payment equals ℓP_C. If on the other hand the value of the equity is insufficient to make the promised payment to the contingent convertible bond holders, they take possession of the whole equity and the old shareholders are completely diluted out. We assume that the face value of all contingent convertible debt is P_C and thus a bondholder with a bond with face value 1 gets a fraction $1/P_C$ of the value of the equity $EQ(V_{\tau_C})$ after conversion in this case.

Proposition 5.7. *If the payment to holders of contingent convertible bonds at conversion is ℓ, the values of the individual bonds of FVC2 satisfy*

$$d_C(V,V_B,V_C,t) = E\left[\int_0^{t\wedge\tau_C} c_C e^{-rs} ds\right] + E\left[e^{-rt}\mathbb{1}_{\{t<\tau_C\}}\right] + \ell\mathbb{E}\left[e^{-r\tau_C}\mathbb{1}_{\{\tau_C\le t\}}\mathbb{1}_{\{\ell P_C \le EQ(V_{\tau_C})\}}\mathbb{1}_{\{\tau_C<\tau\}}\right]$$
$$+ \frac{1}{P_C}\mathbb{E}\left[e^{-r\tau_C} EQ(V_{\tau_C})\mathbb{1}_{\{\tau_C\le t\}}\mathbb{1}_{\{\tau_C<\tau\}}\mathbb{1}_{\{\ell P_C > EQ(V_{\tau_C})\}}\right].$$

For an exponential maturity profile $\varphi(t) = me^{-mt}$ the total value of the convertible debt FVC2 is given by:

$$CB(V,V_B,V_C) = \left(\frac{c_C P_C + mP_C}{m+r}\right) E\left[1 - e^{-(m+r)\tau_C}\right] + \ell P_C \mathbb{E}\left[e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\tau\}}\mathbb{1}_{\{\ell P_C\le EQ(V_{\tau_C})\}}\right]$$
$$+ \mathbb{E}\left[EQ(V_{\tau_C}) e^{-(m+r)\tau_C}\mathbb{1}_{\{V(\tau_C)>V_B\}}\mathbb{1}_{\{\ell P_C > EQ(V_{\tau_C})\}}\mathbb{1}_{\{\tau_C<\infty\}}\right].$$

A detailed treatment of FVC2s is provided in Appendix D.1. In Appendix D.3 we show that the two contracts FVC1 and FVC2 are identical if no jumps are included in the firm's value process. The idea is that without jumps the firm's value process has to touch the conversion barrier at the time of conversion, i.e. $V_{\tau_C} = V_C$. As a consequence $EQ(V_{\tau_C}) = EQ(V_C)$ and thus $S(V_{\tau_C}) = S(V_C)$. Under Assumption 5.2 the number of shares granted at conversion is $\frac{\ell P_C}{S(V_{\tau_C})}$ for FVC2s and $\frac{\ell P_C}{S(V_C)}$ for FVC1s. As both numbers coincide, the two contracts are the same.

In Appendix D.2 we compare FVC and FSC contracts in terms of the number of shares n' granted at conversion for different conversion parameters ℓ.

5.4 Evaluating the Model

5.4.1 Dynamics of the Firm's Value Process

Now we make some explicit assumptions about the martingale in (5.1) and assume that it is a Lévy process. Thus V can be expressed as

$$V_t = V_0 \exp(X_t) \tag{5.8}$$

where X is a Lévy process. The key tool for our analysis will be the Laplace exponent of X. The moment generating function of a Lévy process is of the form

$$E[\exp(zX_t)] = \exp(t\psi(z)) \tag{5.9}$$

for some function ψ being analytic in the interior of its domain of definition. The Lévy-Khintchine representation theorem characterizes a Lévy process, identifying a drift term, a Brownian motion component and a jump component:

$$\psi(z) = bz + \frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}} (e^{zy} - 1 - zy\mathbb{1}_{\{|y|<1\}})\nu(dy)$$

where b is the drift, σ corresponds to the Brownian motion component and ν is the Lévy measure identifying the jumps. The roots of $\psi(z) = \lambda$ will be of importance in the following. We can now define the first passage times in terms of X: τ is the first hitting time defined as $\tau = \tau_x = \inf(t \geq 0 : X(t) \leq x)$ with $x = \log(V_B/V)$ and $\tau_C = \tau_{x_C} = \inf(t \geq 0 : X(t) \leq x_C)$ with $x_C = \log(V_C/V)$.

For a general Lévy process it is very difficult to characterize the distribution of the first passage times. Following Chen and Kou (2009) we propose a two-sided jump model for the evolution of the firm's assets with a double exponential jump diffusion process. The main advantage of the double exponential distribution is that it leads to an analytical solution for various Laplace transforms of the first passage times. Due to the conditional memoryless property of the exponential distribution we can also analytically evaluate the Laplace transform of the conversion time for jumps that are too small to trigger default. We assume that under the risk-neutral measure the value of the firm's assets V follows

$$dV_t = V_t\left((r-\delta)dt + \sigma dW_t + d\left(\sum_{i=1}^{N_t}(Z_i - 1)\right) - \lambda\xi dt\right)$$

where N is a Poisson process with constant intensity rate λ. Z_i are i.i.d. random variables and the $Y_i = \log(Z_i)$ possess a double exponential density:

$$f_Y(y) = p\eta_1 e^{-\eta_1 y}\mathbb{1}_{\{y\geq 0\}} + q\eta_2 e^{\eta_2 y}\mathbb{1}_{\{y<0\}}$$

where η_1, η_2, p and q are positive numbers and $p+q = 1$. The parameters p and η_1 correspond to the upward jumps and q and η_2 to the downward jumps respectively. The mean percentage jump size ξ is given by

$$\xi = E[Z-1] = E[e^Y - 1] = \frac{p\eta_1}{\eta_1 - 1} + \frac{q\eta_2}{\eta_2 + 1} - 1.$$

The sources of randomness N, W and Y are assumed to be independent. In order to ensure that $\xi < \infty$, we assume that $\eta_1 > 1$. This condition implies that the average upward jump cannot exceed 100%, which is reasonable in reality. Applying Itô's lemma for jump diffusions yields

$$V_t = V_0 \exp(X_t) = V_0 \exp\left(bt + \sigma W_t + \sum_{i=1}^{N_t} Y_i\right).$$

where $b = r - \delta - \frac{1}{2}\sigma^2 - \lambda\xi$ and for $z \in (-\eta_2, \eta_1)$ the Lévy-Khintchine formula is given by

$$\begin{aligned} \psi(z) &= bz + \frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}} (e^{zy} - 1)\lambda f_Y(y)dy \\ &= bz + \frac{1}{2}\sigma^2 z^2 + \lambda\left(\frac{p\eta_1}{\eta_1 - z} + \frac{q\eta_2}{\eta_2 + z} - 1\right). \end{aligned}$$

Note that under the above risk-neutral measure V_t is a martingale after proper discounting, i.e.

$$V_t = \mathbb{E}\left[e^{-(r-\delta)(T-t)} V_T | \mathfrak{F}_t\right]$$

where $\mathfrak{F}_t$ is the information up to time t. Kou and Wang (2003) prove the following results:

Lemma 5.2. *For any $\rho > 0$ it holds that*

$$E\left[e^{-\rho\tau}\right] = \frac{\eta_2 - \beta_{3,\rho}}{\eta_2} \frac{\beta_{4,\rho}}{\beta_{4,\rho} - \beta_{3,\rho}} e^{x\beta_{3,\rho}} + \frac{\beta_{4,\rho} - \eta_2}{\eta_2} \frac{\beta_{3,\rho}}{\beta_{4,\rho} - \beta_{3,\rho}} e^{x\beta_{4,\rho}}$$

where τ denotes the first passage time of X_t to x and $-\beta_{3,\rho} > -\beta_{4,\rho}$ are the two negative roots of the equation $\psi(\beta) = \rho$.

Lemma 5.3. *For any $\rho > 0$ and $\theta > -\eta_2$ it holds that*

$$\mathbb{E}\left[e^{-\rho\tau + \theta X_\tau} \mathbb{1}_{\{\tau_x < \infty\}}\right] = e^{\theta}\left(\frac{\beta_{4,\rho} + \theta}{\eta_2 + \theta} \frac{\eta_2 - \beta_{3,\rho}}{\beta_{4,\rho} - \beta_{3,\rho}} e^{x\beta_{3,\rho}} + \frac{\beta_{3,\rho} + \theta}{\eta_2 + \theta} \frac{\beta_{4,\rho} - \eta_2}{\beta_{4,\rho} - \beta_{3,\rho}} e^{x\beta_{4,\rho}}\right)$$

where τ denotes the first passage time of X_t to x and $-\beta_{3,\rho} > -\beta_{4,\rho}$ are the two negative roots of the equation $\psi(\beta) = \rho$.

For the evaluation of contingent convertible bonds we need to consider the case where conversion occurs but the jumps are not large enough to trigger bankruptcy. For this reason we show the following proposition.

Proposition 5.8. *Assume that X_t follows a Kou process and τ denotes the first passage time to $x < 0$, i.e. $\tau = \inf(0 \le t : X_t \le x)$. It holds that for $y > 0$, $\theta > -\eta_2$ and $\rho > 0$:*

$$
\begin{aligned}
&\mathbb{E}\left[e^{-\rho\tau+\theta X_\tau}\mathbb{1}_{\{\tau<\infty,-(X_\tau-x)<y\}}\right] \\
&= \left(\frac{\eta_2-\beta_{3,\rho}}{\beta_{4,\rho}-\beta_{3,\rho}}e^{x\beta_{3,\rho}}+\frac{\beta_{4,\rho}-\eta_2}{\beta_{4,\rho}-\beta_{3,\rho}}e^{x\beta_{4,\rho}}\right)e^{\theta x}+e^{\theta x}\frac{\eta_2}{\theta+\eta_2}\frac{\left(1-e^{-(\theta+\eta_2)y}\right)}{(1-e^{-\eta_2 y})} \\
&\quad\cdot\left(\frac{e^{x\beta_{3,\rho}}}{\beta_{4,\rho}-\beta_{3\rho}}\left(\frac{\eta_2-\beta_{3,\rho}}{\eta_2}\beta_{4,\rho}-(\eta_2-\beta_{3,\rho})-e^{-\eta_2 y}\frac{(\eta_2-\beta_{3,\rho})(\beta_{4,\rho}-\eta_2)}{\eta_2}\right)\right. \\
&\quad\left.+\frac{e^{x\beta_{4,\rho}}}{\beta_{4,\rho}-\beta_{3,\rho}}\left(\frac{\beta_{4,\rho}-\eta_2}{\eta_2}\beta_{3,\rho}-(\beta_{4,\rho}-\eta_2)+e^{-\eta_2 y}\frac{\eta_2-\beta_{3,\rho}}{\eta_2}(\beta_{4,\rho}-\eta_2)\right)\right)
\end{aligned}
$$

where $-\beta_{3,\rho} > -\beta_{4,\rho}$ are the two negative roots of the equation $\psi(\beta) = \rho$.

Definition 5.4. *The function J is defined as*

$$
J(x,\theta,y,\rho) = \mathbb{E}\left[e^{-\rho\tau+\theta X_\tau}\mathbb{1}_{\{\tau<\infty,-(X_\tau-x)<y\}}\right],
$$

where $x < 0$, $\theta > -\eta_2$, $y > 0$, $\rho > 0$ and $\tau = \inf(t \ge 0 : X_t \le x)$. The explicit form of $J(x, \theta, y, \rho)$ is given in Proposition 5.8.

Corollary 5.3. *Assume that X_t follows a Kou process and τ denotes the first passage time to $x < 0$, i.e. $\tau = \inf(t \ge 0 : X_t \le x)$. It holds that for $y > 0$ and $\rho > 0$*

$$
\begin{aligned}
\mathbb{E}\left[e^{-\rho\tau}\mathbb{1}_{\{-(X_\tau-x)<y\}}\right] =& \frac{e^{x\beta_{3,\rho}}}{\beta_{4,\rho}-\beta_{3\rho}}\left(\frac{\eta_2-\beta_{3,\rho}}{\eta_2}\beta_{4,\rho}-e^{-\eta_2 y}\frac{(\eta_2-\beta_{3,\rho})(\beta_{4,\rho}-\eta_2)}{\eta_2}\right) \\
&+\frac{e^{x\beta_{4,\rho}}}{\beta_{4,\rho}-\beta_{3,\rho}}\left(\frac{\beta_{4,\rho}-\eta_2}{\eta_2}\beta_{3,\rho}+e^{-\eta_2 y}\frac{\eta_2-\beta_{3,\rho}}{\eta_2}(\beta_{4,\rho}-\eta_2)\right).
\end{aligned}
$$

Definition 5.5. *The function G is defined as*

$$
G(x,y,\rho) = \mathbb{E}\left[e^{-\rho\tau}\mathbb{1}_{\{-(X_\tau-x)<y\}}\right],
$$

where $x < 0$, $y > 0$, $\rho > 0$ and $\tau = \inf(t \ge 0 : X_t \le x)$. The explicit form of $G(x, y, \rho)$ is given in Corollary 5.3.

5.4.2 Evaluation of the Debt

Now it is straightforward to derive an expression for the firm's debt D. The following three propositions are shown in Chen and Kou (2009).

Proposition 5.9. *The value of the firm's debt equals*

$$D = \frac{C_D + mP_D}{r+m}\left(1 - \frac{\beta_{4,r+m}}{\eta_2}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_B}{V}\right)^{\beta_{3,r+m}} - \frac{\beta_{3,r+m}}{\eta_2}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_B}{V}\right)^{\beta_{4,r+m}}\right)$$
$$+ (1-\alpha)V_B\left(\frac{\beta_{4,r+m}+1}{\eta_2+1}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_B}{V}\right)^{\beta_{3,r+m}} + \frac{\beta_{3,r+m}+1}{\eta_2+1}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_B}{V}\right)^{\beta_{4,r+m}}\right)$$

where $-\beta_{3,\rho} > -\beta_{4,\rho}$ are the only two negative roots of $\psi(\beta) = \rho$ and the value is independent of the specification of the contingent convertible bonds.

Proposition 5.10. *If the capital structure does not include any contingent convertible debt the total value of the firm G_{debt} equals*

$$\begin{aligned} G_{debt} &= V + \frac{\bar{c}C}{r}\left(1 - \frac{\beta_{4,r}}{\eta_2}\frac{\eta_2 - \beta_{3,r}}{\beta_{4,r} - \beta_{3,r}}\left(\frac{V_B}{V}\right)^{\beta_{3,r}} - \frac{\beta_{3,r}}{\eta_2}\frac{\beta_{4,r} - \eta_2}{\beta_{4,r} - \beta_{3,r}}\left(\frac{V_B}{V}\right)^{\beta_{4,r}}\right) \\ &\quad -\alpha V_B\left(\frac{\beta_{4,r}+1}{\eta_2+1}\frac{\eta_2 - \beta_{3,r}}{\beta_{4,r} - \beta_{3,r}}\left(\frac{V_B}{V}\right)^{\beta_{3,r}} + \frac{\beta_{3,r}+1}{\eta_2+1}\frac{\beta_{4,r} - \eta_2}{\beta_{4,r} - \beta_{3,r}}\left(\frac{V_B}{V}\right)^{\beta_{4,r}}\right). \end{aligned}$$

Recall that if the capital structure does not include any contingent convertible debt the value of the equity of the firm EQ_{debt} is the difference between the total value of the firm and the value of its debt:

$$EQ_{debt}(V, V_B) = G_{debt}(V, V_B) - D(V, V_B).$$

Proposition 5.11. *For all $V > V_B$ and $V_B > V_B^*$ the function $EQ_{debt}(V, V_B)$ is a strictly increasing function in V*

$$\frac{\partial EQ_{debt}(V, V_B)}{\partial V} > 0$$

where V_B^* is the optimal default barrier as defined in Section 5.5.

As we will show in Section 5.5 later, the smallest possible default barrier, that we need to take into consideration is V_B^*. Therefore, Proposition 5.11 is general enough for our purposes. In order to evaluate the conversion value, we need to calculate the value of the equity at the time of conversion $EQ(V_{\tau_C})$.

Lemma 5.4. *The value of the equity at conversion $EQ(V_{\tau_C})$ satisfies*

$$EQ(V_{\tau_C}) = EQ_{debt}(V_{\tau_C}) = \sum_i \alpha_i V_{\tau_C}^{\theta_i} = \sum_i V_0^{\theta_i} \alpha_i e^{X(\tau_C)\theta_i}$$

The coefficients α_i and θ_i are defined in Lemma D.5.

Definition 5.6. *The function $T : (V_B, \infty) \to (0, \infty)$ is defined as*

$$T(V_t) = EQ_{debt}(V_t, V_B)$$

Corollary 5.4. *The condition $EQ(V_{\tau_C}) \geq \ell P_C$ is equivalent to $V_{\tau_C} \geq T^{-1}(\ell P_C)$.*

5.4.3 Evaluation of the CCBs

Evaluation of FSCs

The only difficulty is to evaluate the dilution costs given by equation 5.7.

Lemma 5.5.

$$\mathbb{E}\left[EQ(V_{\tau_C})e^{-(r+m)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{\tau_C<\tau\}}\right] = \sum_i \alpha_i V_0^{\theta_i} J\left(\log\left(\frac{V_C}{V_0}\right), \theta_i, \log\left(\frac{V_C}{V_B}\right), r+m\right)$$

where α_i and θ_i are as in Lemma 5.4.

Proof. The total equity value is the difference between the total value of the firm and the value of actual debt payments: In Lemma 5.4 we have shown that $EQ(V_{\tau_C})$ has a structure of the form

$$EQ(V_{\tau_C}) = \sum_i \alpha_i V_{\tau_C}^{\theta_i}.$$

Therefore, calculating $\mathbb{E}\left[EQ(V_{\tau_C})e^{-(r+m)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{\tau_C<\tau\}}\right]$ boils down to

$$\begin{aligned}&\sum_i \mathbb{E}\left[\alpha_i V_{\tau_C}^{\theta_i} e^{-(r+m)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{\tau_C<\tau\}}\right] \\ =&\sum_i \alpha_i V_0^{\theta_i}\mathbb{E}\left[e^{\theta_i X_{\tau_C}-(r+m)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{-(X_{\tau_C}-x_C)<\log(V_C/V_B)\}}\right].\end{aligned}$$

□

As a consequence the price of FSCs is given by the following theorem:

Theorem 5.1. *The price of FSCs for $t<\tau_C$ equals*

$$\begin{aligned}CB(V_t) =& \frac{C_C+mP_C}{r+m}\left(1-\frac{\beta_{4,r+m}}{\eta_2}\frac{\eta_2-\beta_{3,r+m}}{\beta_{4,r+m}-\beta_{3,r+m}}\left(\frac{V_C}{V_t}\right)^{\beta_{3,r+m}} - \frac{\beta_{3,r+m}}{\eta_2}\frac{\beta_{4,r+m}-\eta_2}{\beta_{4,r+m}-\beta_{3,r+m}}\left(\frac{V_C}{V_t}\right)^{\beta_{4,r+m}}\right) \\ &+\frac{n'}{n+n'}\left(\sum_i \alpha_i V_t^{\theta_i} J\left(\log\left(\frac{V_C}{V_t}\right),\theta_i,\log\left(\frac{V_C}{V_B}\right), r+m\right)\right).\end{aligned}$$

where α_i and θ_i are as in Lemma 5.4.

Proof. By Proposition 5.4, Lemma 5.1 and equation 5.7 the price of the FSC is given by

$$\begin{aligned}CB(V,V_B,V_C) =& \left(\frac{c_C P_C + mP_C}{m+r}\right) E\left[1-e^{-(m+r)\tau_C}\right] \\ &+\frac{n'}{n+n'}\mathbb{E}\left[EQ(V_{\tau_C})e^{-(r+m)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{\tau_C<\tau\}}\right].\end{aligned}$$

The first summand corresponds to the value of the coupon payments and the repayment of the face value if conversion does not take place. The second summand is the value of the dilution costs. Applying Lemma 5.2, we can explicitly calculate the Laplace transformation of the conversion time appearing in the first summand. Lemma 5.5 gives us an explicit expression for the expectation in the second summand. □

Evaluation of FVCs

We start with FVC1:

Proposition 5.12. *The price of FVC1s for $t < \tau_C$ equals*

$$CB(V_t) = \frac{C_C + mP_C}{r+m}\left(1 - \frac{\beta_{4,r+m}}{\eta_2}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_C}{V_t}\right)^{\beta_{3,r+m}} - \frac{\beta_{3,r+m}}{\eta_2}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_C}{V_t}\right)^{\beta_{4,r+m}}\right)$$
$$+ \frac{\ell P_C}{\sum_i \alpha_i V_C^{\theta_i}}\left(\sum_i \alpha_i V_t^{\theta_i} J\left(\log\left(\frac{V_C}{V_t}\right), \theta_i, \log\left(\frac{V_C}{V_B}\right), r+m\right)\right).$$

where α_i and θ_i are as in Lemma 5.4.

Proof. By Proposition 5.6 the price of FVC1s is given by

$$CB(V, V_B, V_C) = \left(\frac{c_C P_C + mP_C}{m+r}\right) E\left[1 - e^{-(m+r)\tau_C}\right]$$
$$+ \ell P_C \mathbb{E}\left[\frac{EQ(V_{\tau_C})}{EQ(V_C)} e^{-(m+r)\tau_C} \mathbb{1}_{\{\tau_C < \infty\}} \mathbb{1}_{\{V_{\tau_C} > V_B\}}\right].$$

The first summand is the value of the coupon payments and the repayment of the face value if conversion does not happen. The second term corresponds to the conversion value. Lemma 5.2 allows us to calculate the expectation in the first term. We combine Lemma 5.4 and Lemma 5.5 to derive an expression for the second summand. Note, that $EQ(V_C)$ is nonrandom and thus can be taken out of the expectation. □

Next, we deal with FVC2:

Theorem 5.2. *The price of FVC2s for $t < \tau_C$ equals*

$$CB = \frac{C_C + mP_C}{r+m}\left(1 - \frac{\beta_{4,r+m}}{\eta_2}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_C}{V_t}\right)^{\beta_{3,r+m}} - \frac{\beta_{3,r+m}}{\eta_2}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_C}{V_t}\right)^{\beta_{4,r+m}}\right)$$
$$+ \ell P_C \cdot G\left(\log\left(\frac{V_C}{V_t}\right), \log\left(\frac{V_C}{\max(T^{-1}(\ell P_C), V_B)}\right), m+r\right) \mathbb{1}_{\{V_C > T^{-1}(\ell P_C)\}}$$
$$+ \sum \alpha_i V_t^{\theta_i}\left(J\left(\log\left(\frac{V_C}{V_t}\right), \theta_i, \log\left(\frac{V_C}{V_B}\right), m+r\right)\right.$$
$$\left. - J\left(\log\left(\frac{V_C}{V_t}\right), \theta_i, \log\left(\frac{V_C}{T^{-1}(\ell P_C)}\right), m+r\right) \mathbb{1}_{\{V_C > T^{-1}(\ell P_C)\}}\right) \mathbb{1}_{\{V_B < T^{-1}(\ell P_C)\}}$$

5.5 Optimal Default Barrier

Only Straight Debt without Contingent Convertible Debt

Choosing the optimal debt level P_D and the optimal bankruptcy trigger V_B are two entangled problems. When a firm chooses P in order to maximize the total value of the firm at time 0, the decision depends on V_B. Vice versa, the optimal default trigger V_B is a function of the amount of debt P_D. Following Leland (1994a+b) P_D and V_B are chosen according to a two-stage optimization problem. In the first stage, for a fixed P_D, equity holders choose the optimal default barrier by maximizing the equity value subject to the limited liability constraint. In a second stage, the firm determines the amount of debt P_D that maximizes the total value of the firm. In this section, we will focus on the first stage problem. The solution to the whole problem is presented in Section 5.10. First, we summarize how the first stage problem is solved in the case with only straight debt and in the next subsection we extend the analysis to a firm that issues contingent convertible debt and straight debt. The maximization problem is

$$\max_{V_B} EQ_{debt}(V_t, V_B) \qquad \text{such that } EQ_{debt}(V', V_B) > 0 \qquad \forall V' > V_B.$$

The case of only straight debt was already considered in Chen and Kou (2009):

Proposition 5.13. *The optimal default barrier without CCBs solves the smooth pasting condition*

$$\left(\frac{\partial(V + TB_D(V,V_B) + BC(V,V_B) - D(V,V_B))}{\partial V} \;|_{V=V_B}\right) = 0$$

and equals:

$$V_B^* = \frac{\frac{C_D+mP_D}{r+m}\beta_{3,r+m}\beta_{4,r+m} - \frac{\bar{c}C_D}{r}\beta_{3,r}\beta_{4,r}}{\alpha(\beta_{3,r}+1)(\beta_{4,r}+1) + (1-\alpha)(\beta_{3,r+m}+1)(\beta_{4,r+m}+1)}\frac{\eta_2+1}{\eta_2}.$$

In the following we want to outline the arguments presented in Chen and Kou (2009).

Proof. Denote by V_B^* the solution to

$$\frac{\partial EQ_{debt}(V,V_B)}{\partial V} \;|_{V=V_B} = 0.$$

By the formula for the equity without contingent convertible debt we can easily verify that

$$V_B^* = \frac{\frac{C_D+mP_D}{r+m}\beta_{3,r+m}\beta_{4,r+m} - \frac{\bar{c}C_D}{r}\beta_{3,r}\beta_{4,r}}{\alpha(\beta_{3,r}+1)(\beta_{4,r}+1) + (1-\alpha)(\beta_{3,r+m}+1)(\beta_{4,r+m}+1)}\frac{\eta_2+1}{\eta_2}.$$

Define $H(V, V_B)$ by

$$H(V,V_B) = \frac{\partial}{\partial V_B} EQ(V,V_B).$$

The proof consists of the following steps:

1. The optimal V_B satisfies $V_B \geq V_B^*$.
2. It holds that $\frac{\partial EQ_{debt}(V,V_B)}{\partial V} \geq 0$ for all $V \geq V_B \geq V_B^*$, i.e. the equity value is increasing in the firm's value.
3. It holds that $H(V, V_B) \leq 0$ for all $V \geq V_B \geq V_B^*$. Hence
$$EQ_{debt}(V, y_1) \geq EQ(V, y_2) \qquad \text{for all } V_B^* \leq y_1 \leq y_2 \leq V,$$
i.e. the firm will choose the lowest default barrier that satisfies the non-negativity constraint

First, by definition $EQ_{debt}(V_B^*, V_B^*) = 0$ and by step 2 $EQ_{debt}(V, V_B^*)$ is nondecreasing in V. Thus V_B^* satisfies the non-negativity constraint $EQ_{debt}(V', V_B^*) \geq 0$ for all $V' \geq V_B^*$. Second, any $V_B \in (V_B^*, V]$ cannot yield a higher equity value because of step 3. □

Case 1: Optimal default barrier V_B for exogenous conversion barrier V_C

The optimization problem for the old shareholders changes, when contingent convertible bonds are included in the capital structure. We have to consider two different "notions of equity" here: The value of the equity for the old shareholders equals

$$EQ_{old}(V, V_B, V_C) = V + TB_D + TB_C - BC - D - CCB.$$

The value of the debt excluding any features of the contingent convertible bonds is

$$EQ_{debt}(V, V_B) = V + TB_D - BC - D.$$

The old shareholders will choose the default barrier V_B such the value of their equity EQ_{old} is maximized subject to the constraint, that EQ_{old} has to be nonnegative. There are basically two different solutions to this optimization problem. Either the resulting default barrier V_B is larger than the conversion barrier V_C or smaller. After conversion, the contingent convertible bondholders will become equity holders and the optimization problem of the equity holders is the same as in the case of only straight debt. If the old shareholders decide on a default barrier V_B that is smaller than V_C, they will not be able to commit to it as after conversion V_B^* (see last subsection) will be chosen. However, it is possible that the value of the equity for the old shareholders $EQ_{old}(V, V_B^*, V_C)$ becomes negative for $V > V_C$. As the old shareholders anticipate this, they will choose a default barrier larger than V_C in this case.

More formally, the optimization problem is formulated as follows:

Definition 5.7. *If the capital structure includes straight debt and contingent convertible bonds, the old shareholders choose V_B to maximize*

$$\max_{V_B} EQ_{old}(V, V_B, V_C) \qquad \text{such that } EQ_{old}(V', V_B, V_C) > 0 \text{ for all } V' > V_B.$$

subject to the commitment problem that $V_B = V_B^*$ if $V_B < V_C$.

We need to clarify what happens in the case where the default and conversion barrier are crossed at the same time for $V_B < V_C$. Passing both barriers simultaneously can occur because a jump that crosses the conversion barrier is large enough to cross the default barrier as well. We treat the crossing as if it happened sequentially. First the conversion barrier is passed and the contingent convertible bondholders become equity holders. For the equity holders, consisting of the old and new shareholders, the optimal default barrier will be V_B^*, but not the former barrier V_B. Therefore the crossing of V_B will not trigger default. Default happens only, if the new barrier V_B^* is passed.

Next, we need to clarify how to treat the case $V_B \geq V_C$. If default happens before conversion, all the payments linked to the contingent convertible bonds are nil as well. Hence, we will assume that value of the payments of a contingent convertible bond are the same as if $V_B = V_C$:

$$CCB(V, V_B, V_C) = CCB(V, V_B, V_B) \qquad \text{for } V_B > V_C$$

How does the commitment problem affect the optimal choice of V_B in this case? If $V_B \geq V_C$ and a jump crosses both barriers at the same time, we treat this case as if the crossing had happened sequentially. First, the default barrier is passed and the firm defaults. Second, the conversion barrier is passed. But as default has already taken place, the value of the equity is zero and the contingent convertible bondholders cannot be compensated with stocks. In particular, the contingent convertible bondholders cannot change the default barrier to V_B^* as in the previous case. If the default barrier and conversion barrier coincide, i.e. $V_B = V_C$, we can either assume that default happens first or that conversion takes place first. We have decided, that first default should take place and after that we deal with the conversion. This is a purely technical convention, which does not affect any of our main results qualitatively, but simplifies the exposition.

We will now show, that there are only two possible solutions to the optimization problem.

Theorem 5.3. *There are only two possible solutions for the optimal default barrier. Either the optimal default barrier coincides with the optimal default barrier with only straight debt or it equals the maximum of the conversion barrier and* $V_B^{**}: V_B = V_B^*$ *or* $V_B = \max(V_C, V_B^{**})$, *where*

$$V_B^{**} = \frac{\frac{C_D + C_C + m(P_D + P_C)}{r+m}\beta_{3,r+m}\beta_{4,r+m} - \frac{\bar{c}(C_D + C_C)}{r}\beta_{3,r}\beta_{4,r}}{\alpha(\beta_{3,r}+1)(\beta_{4,r}+1) + (1-\alpha)(\beta_{3,r+m}+1)(\beta_{4,r+m}+1)} \frac{\eta_2 + 1}{\eta_2}.$$

V_B^{**} *equals the optimal default barrier of a firm with only straight debt with face value* $P_D + P_C$ *and coupon* $C_D + C_C$. *If*

$$EQ_{old}(V, V_B^*, V_C) \geq EQ_{debt}(V, V_B^{**}, P_D + P_C, C_D + C_C) \qquad \textit{for all } V \geq V_C$$

for $V_B^{**} > V_C > V_B^*$ *or*

$$EQ_{old}(V, V_B^*, V_C) \geq 0 \qquad \textit{for all } V \geq V_C.$$

for $V_B^* < V_B^{**} \leq V_C$ *then*

$$V_B = V_B^*$$

otherwise

$$V_B = \max(V_C, V_B^{**}).$$

The intuition behind the proof is the following. V_B^* can only be the optimal default barrier, when it is feasible, i.e. the equity value of the old shareholders is always positive before conversion $EQ_{old}(V, V_B^*, V_C) > 0$ for all $V > V_C$. Even, when it is feasible to choose V_B^*, it may be optimal for the old shareholders to default before conversion. We show that the equity value of the old shareholders for $V_B > V_C$ is the same as for a firm that issues only straight debt in the amount $P_D + P_C$ with coupon $C_D + C_C$: $EQ_{old}(V, V_B, V_C, P_D, P_C, C_D, C_C) = EQ_{debt}(V, V_B, P_D + P_C, C_D + C_C)$. The optimal default barrier for this amount of straight debt is V_B^{**}. If $V_B^{**} > V_C$ and $EQ_{debt}(V, V_B^{**}, P_D + P_C, C_D + C_C) > EQ_{old}(V, V_B^*, V_C)$, then the old shareholders will prefer to default before conversion. However, if $V_B^{**} < V_C$, the old shareholder can get at most $EQ_{debt}(V, V_C, P_D + P_C, C_D + C_C)$ if they decide to default before conversion, i.e. the optimal default barrier before conversion is V_C itself. However, we can show that in this case the old shareholders will always prefer to default after conversion, i.e. take $V_B = V_B^*$ if it is feasible.

We are particularly interested in the case, where the default and conversion barrier are not the same. Hence, we assume

Assumption 5.3. *The default barrier* $V_B^* < V_C$ *satisfies the no-early-default condition: For* $V_B^{**} \leq V_C$

$$EQ_{old}(V, V_B^*, V_C) \geq 0 \qquad \text{for all } V \geq V_C$$

and for $V_B^{**} > V_C$

$$EQ_{old}(V, V_B^*, V_C) \geq EQ_{debt}(V, V_B^{**}, P_D + P_C, C_D + C_C) \qquad \text{for all } V \geq V_C.$$

The no-early-default condition is composed of two statements: First, V_B^* is a feasible default barrier, i.e. it satisfies the limited liability constraint. Second, it is never profitable for the old shareholders to default before conversion, i.e. the value of the equity for the old shareholders for V_B^* is always larger than the corresponding value for the optimal default barrier larger than the conversion barrier. Assumption 5.3 is easily testable for a given amount of debt.

Proposition 5.14. *If Assumption 5.3 is satisfied, then the firm chooses the same default barrier as in the case without contingent convertible capital:*

$$V_B = V_B^*.$$

Lemma 5.6. *For a fixed amount of debt* P_D *and* P_C *and fixed coupon values* C_D *and* C_C *the default barrier* V_B^{**} *is always larger than* V_B^**.*

Case 2: V_B and V_C chosen optimally

In the previous subsection we assumed that V_C is exogenously given. In this subsection we treat V_C as a choice variable. From a decision theoretical point of view, V_B and V_C are different. V_B is not agreed on explicitly, when the debt is issued. By the very nature of debt, default happens when the cash payments cannot be made anymore. As long as the equity value is positive ($EQ_{old} > 0$), shareholders can (and will) always issue more equity to avoid default. When $EQ_{old} < 0$, the firm defaults. Hence, V_B is agreed on only implicitly as all agents anticipate the shareholders' actions. If the parameters change, e.g. more debt is issued, the default barrier V_B changes as well, as it is not a contract term. In contrast, V_C is specified in the contract a priori and cannot be changed a posteriori.

We will analyze two cases: In the first case the shareholders choose V_C to maximize the value of their equity and in the second case V_C is determined to maximize the total value of the firm. The main result of this section is that contingent convertible debt can degenerate to straight debt without recovery payment. In this case the conversion barrier will coincide with the default barrier V_B^{**}. If conversion takes place before default, the optimal conversion barrier of the shareholders is strictly higher than the optimal conversion barrier for the firm as a whole.

First, we consider the two-dimensional optimization problem of the shareholders. We will simplify it to a two-stage optimization problem. The first stage is to choose V_B optimally for a given V_C subject to the commitment problem, i.e. we have the same problem as in the previous subsection. The second stage is to choose a V_C, that maximizes the equity value for the old shareholders. However, the solution may not be unique. We adopt the convention that the smallest possible conversion barrier is chosen by the old shareholders if the solution is not unique. This is motivated by the fact, that the shareholders have also additional costs from dilution which are not explicitly modeled here, e.g. less control over the company. Hence, a lower conversion barrier should be preferred. In summary, the second stage optimization problem is

$$V_C = \inf\{\arg\max_{V_C} EQ_{old}(V, V_B(V_C), V_C) \text{ s.t. } V \geq V_C \geq V_B(V_C)\}$$

Define $\bar{V}_C$ as the smallest conversion barrier, such that the limited liability constraint for V_B^* is satisfied:

$$\bar{V}_C = \inf\{V_C \geq V_B^* : EQ_{old}(V, V_B^*, V_C) \geq 0 \text{ for all } V \geq V_C\}.$$

This infimum exists as the above set is non-empty (e.g. $V_C = V$ is included in the set). Next, we define $V_C^* \geq \bar{V}_C$ as the smallest conversion barrier that maximizes the value of the equity of the old shareholders for the default barrier V_B^*:

$$V_C^* = \inf\{\arg \max_{V_C: V \geq V_C \geq \bar{V}_C} EQ_{old}(V, V_B^*, V_C)\}.$$

Note, that $EQ_{old}(V, V_B^*, V_C)$ is a continuous function in V_C and the set over which we are maximizing is compact. Hence, a maximum exists. The infimum of the nonempty set is well-defined and unique.

Proposition 5.15. *Assume that the shareholders choose $\{V_B, V_C\}$ according to the two stage optimization problem in order to maximize EQ_{old}. If*

$$EQ_{old}(V, V_B^*, V_C^*) \geq EQ_{debt}(V, V_B^{**}, V_B^{**}) \text{ for all } V > V_B^{**}$$

then the optimal solution is

$$V_B = V_B^* \text{ and } V_C = V_C^*,$$

otherwise

$$V_B = V_B^{**} \text{ and } V_C = V_B^{**}$$

Now we change the optimization problem and consider the case where V_C is chosen to maximize the total value of the firm. This affects only the second stage, while the first stage remains unaffected:

$$V_C = \inf\{\arg\max_{V_C} G(V, V_B, V_C) \text{ s.t. } V \geq V_C \geq V_B(V_C)\}$$

Proposition 5.16. *Assume that V_C is chosen to maximize the total value of the firm in a second stage. If $G(V, V_B^*, \bar{V}_C) > G(V, V_B^{**}, V_B^{**})$ for all $V > V_B^{**}$, then the optimal solution is*

$$V_B = V_B^* \text{ and } V_C = \bar{V}_C,$$

otherwise

$$V_B = V_B^{**} \text{ and } V_C = V_B^{**}$$

Note, that in the case where $V_B = V_B^*$, the optimal conversion barrier for the old equity holders is higher than the conversion barrier that is optimal for the firm as a whole. The reason is that the firm and the old shareholders face different tradeoffs. The total value of the firm is strictly increasing in a lower conversion barrier as a late conversion means more tax benefits. The old shareholders also profit from a late conversion as it implies more tax benefits and a lower conversion value. However, there is also a cost to the old shareholders if conversion takes place later, as the coupon and face value payments for the contingent convertible bonds increase. Therefore, it is in general not optimal for the old shareholders to choose the lowest possible conversion barrier.

The key result of this section is the following: If the conversion barrier is chosen endogenously by the firm, the contingent convertible bonds could degenerate to straight debt without recovery payment. In this case the optimal default barrier V_B^{**} will be larger than V_B^* for the same face value of debt and the same coupon payments. As we will discuss in Section 5.10, a higher default barrier implies a higher default risk. Thus, a regulator prefers a lower default barrier. Therefore, this section gives a strong argument for V_C being fixed exogenously by the regulator such that the contingent convertible debt does not degenerate to debt without any conversion payment. Hence, in the following we focus on an exogenously given V_C which satisfies the no-early-default condition from Assumption 5.3.

5.6 Conversion Triggered by Observable Market Prices

5.6.1 Stock Price as a Sufficient Condition for Conversion

The firm's value process is in general not observable. Our model so far has specified the event of conversion in terms of the firm's value process. In this subsection we want to analyze whether conversion could also be specified in terms of the observable stock price.

The stock price process S_t can be expressed as a function of V_t. Recall that in the case of a firm that has not issued any contingent convertible bonds, but only straight debt, the relationship is very simple:

$$S_t = \frac{EQ(V_t)}{n}.$$

As $\partial EQ(V_t)/\partial V_t > 0$ (see Proposition 5.11), the stock price is a strictly increasing function in the value of the firm's assets. If the firm also issues contingent convertible bonds, the situation becomes more complicated. In Section 5.3.3 we have defined the stock price $S_t = S(V_t)$ as a function of V_t:

$$S_t = S(V_t) = \begin{cases} \frac{EQ_{old}(V_t)}{n} = \frac{EQ(V_t) - DC(V_t)}{n} & \text{if } t < \tau_C \\ \frac{EQ(V_t)}{n+n'} & \text{if } \tau_C \leq t < \tau \end{cases}$$

We have shown that the dilution costs equal the conversion value for all specifications of CCBs. Hence, it follows for all types of CCBs that

$$EQ(V_{\tau_C}) - DC(V_{\tau_C}) = \frac{n}{n+n'} EQ(\tau_C).$$

Hence, the stock price is a continuous function in V_t, even when we include contingent convertible bonds. Given our closed form solutions for the conversion value and the dilution costs we can give an explicit formula for the stock price as a function of V_t. However, if we include contingent convertible debt, $EQ_{old}(V_t)$ and therefore also $S(V_t)$ are not necessarily strictly increasing functions in V_t any more. First, we consider the special case, where $S(.)$ is still a strictly increasing function.

Assumption 5.4. *Assume that the mapping between S_t and V_t is strictly increasing for $V_t > V_B$. This implies that the mapping $S(.)$ is invertible for $V > V_B$ and its inverse $S(.)^{-1}$ is strictly increasing as well.*

Proposition 5.17. *Under Assumption 5.4 the stock price is a sufficient statistic for conversion, i.e.*

$$\tau_C = \inf\{t \geq 0 : V_t \leq V_C\} = \inf\{t \geq 0 : S_t \leq S_C\} \qquad \textit{with probability 1}$$

with $S_C = S(V_C)$.

Proof. Obviously, $V_t \leq V_C$ implies $S_t \leq S_C$. As $S(.)$ is invertible with a strictly increasing inverse, $S_t \leq S_C$ also implies $V_t \leq V_C$. □

Corollary 5.5. *If Assumption 5.4 is satisfied, we can base the conversion event on the stock price and obtain a unique pricing equilibrium. The prices for the different CCBs, which are presented in Section 5.4.3, are still valid.*

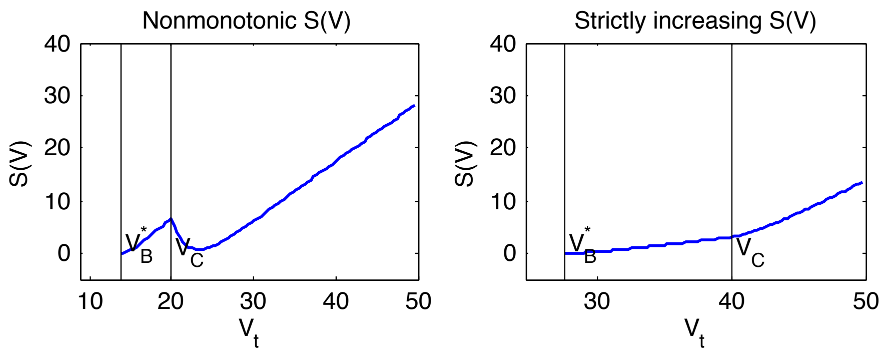


Figure 5.1: Stock price as a function of the firm's value process.

The situation changes if Assumption 5.4 is violated.[5] In Figure 5.1 we plot the stock price as a function of the firm's value process for two different FVC1 contracts. In the right plot Assumption 5.4 is satisfied. However, as we can see from the left plot, there exists parameter values such that $S^{-1}(S_t) = \{V_t^L, V_t^M, V_t^H\}$ contains three elements, i.e. the stock price S_t is the equilibrium price for the firm's values $V_t^L < V_t^M < V_t^H$. Particularly problematic is that $V_t^L < V_C < V_t^H$, i.e. the event of conversion and not conversion are consistent under a certain stock price. If conversion is based on the stock price and the stock price is a nonmonotonic function of the firm's value process, then the prices cannot be evaluated within our modeling framework.

More formally, define a conversion triggering stock price S_C. Next, we define the conversion time based on observables as

$$\tilde{\tau}_C = \inf\{t \geq 0 : \tilde{S}_t \leq S_C\}$$

[5]Sundaresan and Wang (2010) present a structural model for CoCo Bonds, in which conversion is based on the stock price process. They claim that, in order to obtain a unique equilibrium, mandatory conversion must not result in any value transfer between equity and CoCo holders (Theorem 1). However, their claim is wrong. A pricing equilibrium can exist and be unique if the trigger price and conversion ratio are chosen independently. The only crucial assumption is that the mapping between the stock price and the firm's assets is strictly monotonic. Sundaresan and Wang also illustrate their argument in a two period model. However, they are just stating, that we need a monotonic relationship between the stock price and the firm's value to obtain a unique pricing equilibrium. In a two period model this condition coincides with their stronger condition that there is no value transfer between shareholders and CoCo holders.

where $\tilde{S}_t$ is the stock price process. This can be used to define a new CCB contract $\widetilde{CB}$, which is identical to the former one except for the conversion time. Hence, the equity $\widetilde{EQ}$ and stock price function $\tilde{S}(V)$ under this new conversion time change as well. Note, that as the CCB price now explicitly depends on the stock price process (through $\tilde{\tau}$), an equilibrium stock price has to be the solution to the equation $\tilde{S}_t = \tilde{S}(V_t)$. In more detail, an equilibrium stock price is a function $\tilde{S}(.) : (V_B, \infty) \to \mathbb{R}_0^+$, that solves the following equation for all $V_t \in (V_B, \infty)$:

$$\begin{aligned}
n \cdot \tilde{S}(V_t) =& EQ_{debt} + \widetilde{TB}_C(V_T) - \widetilde{CB} \qquad \Leftrightarrow \\
n \cdot \tilde{S}(V_t) =& EQ_{debt} + \frac{\bar{c}C_C}{r}\mathbb{E}\left[1 - e^{-r\tilde{\tau}_C}\right] - \left(\frac{c_C P_C + mP_C}{m+r}\right)\mathbb{E}\left[1 - e^{-(m+r)\tilde{\tau}_C}\right] \\
& - \mathbb{E}\left[n'\tilde{S}(V_{\tilde{\tau}_C})e^{-(m+r)\tilde{\tau}_C}\mathbb{1}_{\{\tilde{\tau}_C < \tau < \infty\}}\right]
\end{aligned}$$

This is a fix point problem. However, the fix point in this case is a function. In order to prove existence and uniqueness of a solution we need to apply fix point theorems for infinitely dimensional Banach spaces. We can show, that if we restrict the set of possible solutions to functions $\tilde{S}(V_t)$ for which the conversion time can be expressed in the form $\inf\{t \geq 0 : V_t \leq V_C\}$ for some V_C, then it is possible that no equilibrium exists. If we simplify the model to discrete time, we can show that multiple solutions can exist. It seems to be a nontrivial problem to make a general statement about the existence and uniqueness of a solution. For this reason we will propose different observable prices as conversion triggers in the next subsection, which do not suffer from this shortcoming.

5.6.2 Conversion based on credit spreads and credit default swaps

In this paper we have developed a consistent and complete model for CCBs where the event of conversion is based on the unobserved firm's value. For practical purposes we need to specify conversion in terms of an observable variable. As we have seen in the last subsection, defining the event of conversion in terms of the stock price will only lead under Assumption 5.4 to the same pricing formulas as a model where conversion is based on the firm's value process. The reason is that the stock price implicitly depends on the features of the CCBs. As a consequence it is possible that a particular stock price is consistent with two different firm's values. Hence, the stock price is in general not a sufficient statistic for the firm's value process. Our goal is it to find an observable variable that could fully reveal the firm's value process. We propose the credit spread and the risk premium of a portfolio of credit default swaps (CDS). As we will show, both, the credit spread and the risk premium for CDSs, are not affected by the features of CCBs and this will allow us to use them as a sufficient statistic for the firm's value process.

First, we prove that the credit spread fully reveals the stock price. The credit spread is defined as the risk premium between a risky and an identical risk-free bond.

Definition 5.8. *The price of a unit default-free coupon bond with face value 1, maturity t and coupon c is denoted by*

$$b(c,t) = \int_0^t ce^{-rs}ds + e^{-rt}.$$

Lemma 5.7. *The aggregated total value of default-free bonds for a firm that issues p default-free unit bonds with maturity profile $\varphi(t) = me^{-mt}$ equals*

$$B(C) = \frac{C + Pm}{m + r}$$

with $C = Pc$ and $P = p/m$.

Proof.

$$\begin{aligned} B(C) &= \int_0^{-\infty} p\Psi(t)b(c,t)dt = \int_0^{-\infty} pc\left(\frac{1 - e^{-rt}}{r}\right)e^{-mt}dt + \int_0^{\infty} pe^{-rt}e^{-mt}dt \\ &= pc\left(\frac{1}{mr} - \frac{1}{r(m+r)}\right) + \frac{p}{m+r} = pc\frac{1}{m(m+r)} + \frac{Pm}{m+r}. \end{aligned}$$

□

The credit spread π is defined as the difference in the coupon payments of a risky and an identical risk-free unit bond, that trade at the same price:

$$b(c,t) = d_D(V, V_B, c + \pi, t)$$

We will first look at an aggregated credit spread Π, i.e.

$$B(C) = D(V, V_B, C + \Pi)$$

where we assume that the portfolio of risky and risk-free bonds have the same face value P_D and maturity profile.

Lemma 5.8. *The aggregated credit spread Π equals*

$$\Pi = C_D - \tilde{C}$$

where C_D is the total coupon value of the risky debt and

$$\tilde{C} = D(V, V_B, C_D)(m + r) - P_D \cdot m$$

Proof. The result follows from

$$D(V, V_B, C_D) = B(\tilde{C}) = \frac{\tilde{C} + P_D m}{m + r}.$$

□

Recall that $C_D = P_D \cdot c_D$. We conclude: The credit spread of a single bond equals

$$\pi = c_D - D(V, V_B, C_D)\frac{m + r}{P_D} + m$$

Economically, it only makes sense to consider positive spreads. A necessary condition is that the recovery payment in the case of default has a lower present value than the repayment of the face value.

Assumption 5.5. *In the following we assume that the value of the total straight debt, if it was risk-free, is larger than total value of the risky debt:*

$$B(C_D) > D(V, V_B).$$

Assumption 5.5 has an important implication:

Lemma 5.9. *Assumption 5.5 implies that the value of the total straight debt, if it was risk-free, is larger than the value of the largest possible recovery payment:*

$$B(C_D) > D(V, V_B) \qquad \Rightarrow \qquad \frac{C_D + mP_D}{r + m} > (1 - \alpha)V_B.$$

If we want to make a statement about the relationship between the credit spread π and the firm's value process, we need to analyze the dependency of $D(V, V_B)$ on V.

Lemma 5.10. *If the condition $\frac{C_D + mP_D}{r+m} \geq \frac{\eta_2}{\eta_2+1}\frac{\beta_{3,r+m}+1}{\beta_{3,r+m}}(1 - \alpha)V_B$ is satisfied, then the value of the straight debt is an increasing function in the firm's value:*

$$\frac{\partial D(V)}{\partial V} > 0 \qquad \textit{for all } V \geq V_B.$$

Note that $\frac{\eta_2}{\eta_2+1}\frac{\beta_{3,r+m}+1}{\beta_{3,r+m}} > 1$. This means, that the condition in the above lemma is stronger than Assumption 5.5. However, as the next lemma shows, it will always be satisfied in our case.

Lemma 5.11. *If the default barrier is chosen optimally as $V_B = V_B^*$ (i.e. V_B is chosen optimally by the shareholders as in the case with only straight debt), then the condition*

$$\frac{C_D + mP_D}{r + m} \geq \frac{\eta_2}{\eta_2 + 1}\frac{\beta_{3,r+m} + 1}{\beta_{3,r+m}}(1 - \alpha)V_B$$

is always satisfied.

Now we can completely characterize the relationship between π and V.

Corollary 5.6. *If the default barrier is chosen as $V_B = V_B^*$, the credit spread $\pi(V)$ as a function of the firm's value process is strictly decreasing:*

$$\frac{\partial \pi}{\partial V} < 0 \qquad \text{for all } V \geq V_B.$$

Our goal was it to express the conversion trigger in terms of an observable process. The next theorem shows that the credit spread is a suitable candidate.

Theorem 5.4. *Assume that $V_B = V_B^*$. There exists a unique value π_C such that $\pi(V_C) = \pi_C$. If the conversion time is defined as*

$$\tau_C^* = \inf(t \in [0, \infty) : \pi(V_t) \geq \pi_C)$$

we get exactly the same evaluation formulas for CCBs as in the case where the conversion time is defined as

$$\tau_C = \inf(t \in [0, \infty) : V_t \leq V_C).$$

Proof. As $\pi(V)$ is strictly decreasing on $[V_B, \infty)$ and $V_C > V_B$, existence and uniqueness of π_C follow. Furthermore, the strict monotonicity implies that with probability 1

$$\{t \in [0, \infty) : V_t \leq V_C\} = \{t \in [0, \infty) : \pi(V_t) \geq \pi(V_C)\} = \{t \in [0, \infty) : \pi(V_t) \geq \pi_C\}$$

holds. □

Another sufficient statistic for the firm's value process is the risk premium of CDSs. A credit default swap is an agreement that the seller of the CDS will compensate the buyer in the event of a loan default. The buyer of the CDS makes a series of payments to the seller and, in exchange, receives a payoff if the loan defaults. We define the CDS fee as $\tilde{\pi}$. For a unit straight debt bond with face value 1 and maturity t the CDS risk premium has to satisfy

$$\int_0^t \tilde{\pi} e^{-rs} ds = b(t, c_D) - d_D(V, V_B, t).$$

This means that a CDS together with a defaultable bond has the same value as an otherwise identical default-free bond.

In the following analysis we construct an index, which is a strictly monotonic function in the firm's value process. Our index is a portfolio of CDS contracts such that the whole debt is "insured". For this purpose, we have to make the weak assumption that a CDS for a risky bond with every possible maturity is issued. It is important to note that this portfolio, that fully insures the aggregated debt, does not need to actually exist. As long as we observe market prices for CDS contracts for every possible maturity, we can calculate the price of our artificial index. The price of the portfolio is a weighted average of the CDS prices, where the different maturities have the same weights as in the debt portfolio.

Proposition 5.18. *The risk premium for credit default swaps on the aggregated debt satisfies the following equation:*

$$\tilde{\pi}\frac{P_D}{m+r} = (B(C_D) - D(V, V_B)) .$$

Proof. Aggregation yields:

$$\int_0^\infty p_D \int_0^t \tilde{\pi} e^{-rs} ds \Psi(t) dt = \int_0^\infty p_D(b(t, c_D) - d_D(V, V_B, t))\Psi(t)dt.$$

We only need to show the statement for the LHS.

$$\int_0^\infty p_D \frac{\tilde{\pi}}{r}\left(1 - e^{-rt}\right) e^{-mt} dt = \tilde{\pi} p_D \frac{1}{m(m+r)} = \tilde{\pi}\frac{P_D}{m+r}.$$

□

This allows us to completely characterize the relationship between $\tilde{\pi}$ and V:

Proposition 5.19. *If the default barrier is chosen as $V_B = V_B^*$, the CDS risk premium $\tilde{\pi}(V)$ as a function of the firm's value process is strictly decreasing:*

$$\frac{\partial \tilde{\pi}}{\partial V} < 0 \qquad \text{for all } V \geq V_B.$$

By the same argument as in Theorem 5.4 we conclude that conversion can be based on the CDS risk premium.

One of the arguments of critics of CoCo-Bonds was that the conversion event cannot be based on the observable stock price process. Indeed, there exist parameter values, for which our evaluation formulas based on the firm's value process and a model where conversion is triggered by movements in the stock price, differ. However, we have shown that the unobservability of the firm's value process can be circumvented by using credit spreads or the CDS risk premium. Credit spreads have the same advantages as stock prices as they constantly adjust to new information in contrast to accounting triggers. As credit spreads are not affected by the features of CCBs, they are a sufficient statistic for the firm's value process. Thus, defining the conversion event in terms of credit spreads is equivalent to using the firm's value process. The same holds for CDS risk premiums.[6]

[6] A special case are firms that are "too big to fail" (TBTF). As debt of these firms is implicitly protected by a government guarantee, the credit spread should be zero, i.e. the debt should be considered to be risk free. However, even the big banks, that enjoyed this government guarantee, had to pay a risk premium on their debt during the past crisis. This implies that the debt of TBTF firms is not completely insured. Either there is some uncertainty about the government bail-out taking place or debt holders fear a hair cut after a bail-out. In either case, the credit spread as a sufficient statistic for the firm's value process works. Assume for example that the loss after default for unprotected debt is $\alpha = 0.5$. If the probability of a bail-out is 80%, then the expected bankruptcy loss of the debt is $\tilde{\alpha} = 0.8 \cdot 0 + 0.2 \cdot 0.5 = 0.1$. If on the other hand a bailout is certain, but a hair cut of 10% is to be expected, the expected bankruptcy loss of the debt is $\tilde{\alpha} = 0.1$. Either way, we could apply the methods of this section, where we replace α with $\tilde{\alpha}$ in the formula for the debt.

5.7 Numerical Examples

In this subsection we will calculate several scenarios numerically and discuss their interpretations. For the computations the values of the following parameters are fixed:

$$V_0 = 100, r = 7.5\%, \delta = 7\%, \alpha = 50\%, \bar{c} = 35\%.$$

These parameter values are similar to those used by Leland (1994a), Leland and Toft (1996) and Hilberink and Rogers (2002) and are chosen to be consistent with the U.S. environment.

We display the spread of normal debt and contingent convertible bonds as a function of log-maturity. By letting $\log(m^{-1})$ vary between -4 to 10 we receive mean maturity profiles from about a week to 1000 years. For example a log-maturity $\log(m^{-1})$ of 1 corresponds to an average maturity of $\frac{1}{m} = 2.7$ years. The spread here is defined as an aggregate spread, i.e.

$$spread = \frac{C_D}{P_D} - r$$

for traditional debt and respectively

$$spread_C = \frac{C_C}{P_C} - r$$

for contingent convertible debt.

We consider three different firms. The parameters of the firm's value process are chosen such that the amount of "uncertainty" for all three firms is the same, i.e. the quadratic variation is kept constant: $\langle \log\left(\frac{V_t}{V_0}\right)\rangle = \sigma^2 + 2\lambda\left(\frac{p}{\eta_1^2} + \frac{1-p}{\eta_2^2}\right) = 0.25$

1. "No jumps": $\sigma = 0.25$

2. "Infrequent large jumps": $\sigma = 0.15, \eta_1 = 2, \eta_2 = 2, p = 0.5, \lambda = 0.2$.
   On average every five years the firm's value jumps. With 50% probability the firm losses one third of its value, while with 50 % probability it gains one third.

3. "Frequent moderate jumps": $\sigma = 0.15, \eta_1 = 10, \eta_2 = 10, p = 0.25, \lambda = 0.5$.
   On average every two years the firm's value jumps. With 75% probability the firm losses 1/11 of its value, while with 25% probability it gains 1/11.

We analyze the aggregate credit spreads and the dilution costs for different choices of the conversion value parameter ℓ, the conversion barrier V_C and the amount of straight debt and contingent convertible debt. For the straight debt we consider three different levels of debt $P_D = 10, 30$ and 40 and for the contingent convertible debt we vary $P_C = 10$ and $P_C = 40$. The value of the firm's assets is hold constant, which means that we swap debt respectively CCBs for equity. For Figure 5.2 to 5.10 we assume that for every maturity the coupon values are determined such that the debt sells at par. This can be interpreted as the case of a firm that creates a new capital structure. In Figure 5.11 to 5.16 we fix the coupon payments such that the debt sells at par at time 0 for $V_0 = 100$. Here we think of a firm that has set up its capital structure at time $t = 0$ and we follow the dynamics of the value of its assets over time.

5.7.1 Comparing FVC and FSC contracts

In Figure 5.2 we plot the spread and the default and conversion barrier for a firm without jumps. The conversion ratio ℓ of the FVC1 contract is set to 1, i.e. the Coco bondholders receive equity at conversion which has the same market value as the face value of the CCBs. The most striking result is that the spread of the CCBs is completely independent of the capital structure and equal to zero. This result illustrates that a model without jumps produces unrealistic results.

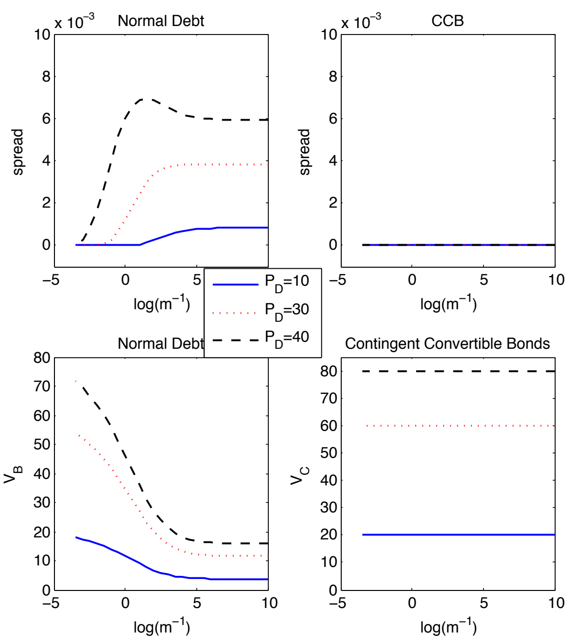


Figure 5.2: "No jumps": FVC1 spread as a function of maturity for different levels of debt and different conversion barriers. The parameters are $\ell = 1, P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 60, 80$.

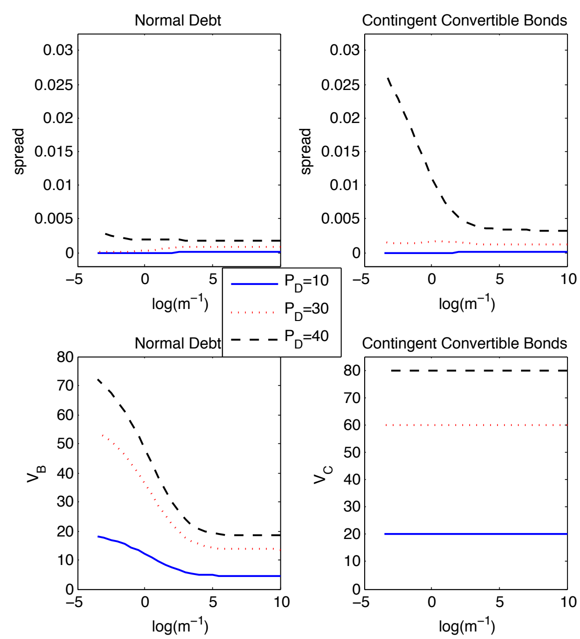


Figure 5.3: "Frequent, moderate jumps": FVC1 spread as a function of maturity for different levels of debt and different conversion barriers. The parameters are $\ell = 1, P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 60, 80$.

The humped-shaped form for normal debt was already found by Leland and Toft. A higher curve corresponds to a higher leverage in terms of normal debt. This makes sense as for a larger leverage the default barrier is higher which in turn implies a higher default risk. Therefore the spread as a risk premium is also higher. Firms with low levels of debt have a small spread, which increases with maturity. These firms are far away from the bankruptcy level V_B and thus the credit spread as a measurement of risk is low. As maturity is growing the firm has more time to approach the critical level V_B and thus the spread increases. As the leverage increases the spread curve becomes more humped. Why is the spread of the highly levered firm falling for a certain level of maturity. This can be explained by the argument that if a firm has survived for a long period of time it is very likely that its value has gone up. Thus conditioning on survival for a long time the firm's value has to be on average far away from V_B and hence the lower spread indicates the decrease in riskiness. Note, that the credit spreads for normal debt are equal to zero for short maturity.

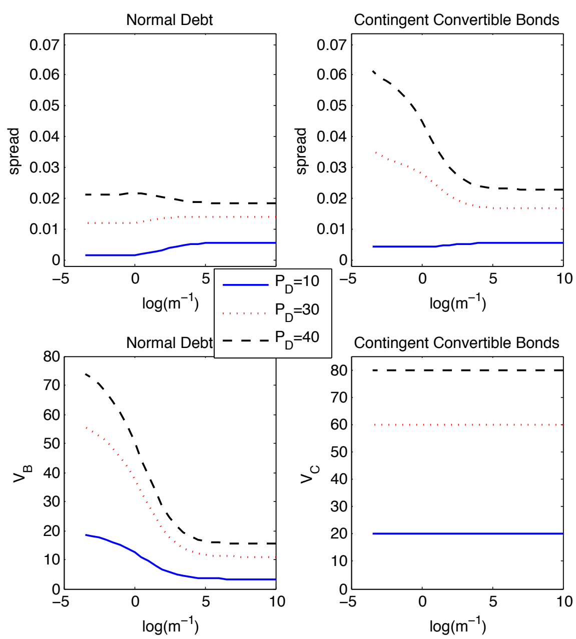


Figure 5.4: "Infrequent, large jumps": FVC1 spread as a function of maturity for different levels of debt and different conversion barriers. The parameters are $\ell = 1, P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 60, 80$.

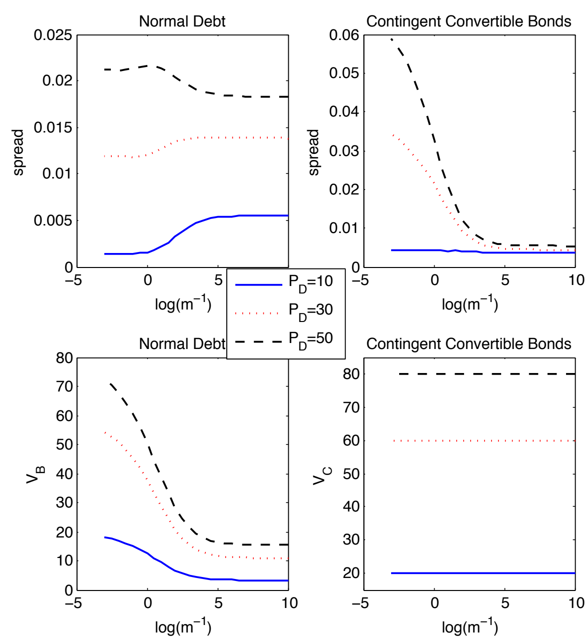


Figure 5.5: "Infrequent, large jumps": FVC2 spread as a function of maturity for different levels of debt and different conversion barriers. The parameters are $\ell = 1, P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 60, 80$.

In Figure 5.3 and 5.4 we consider the two firms with jumps. In both cases we observe substantial credit spreads which depend strongly on the amount of straight debt. Figure 5.2 has shown, that if no jumps are included in the firm's value process, if ℓ is equal to 1 and if the equity value at conversion is sufficiently high, then FVCs are risk-free. The contingent convertible debt holders will always receive their full payments, but the time when this happens may be random. In the case of FVC1 contracts with jumps the expected value of the equity at the conversion time τ_C is in general lower than the value of the equity for $V_t = V_C$. Hence, the conversion value is lower than the face value of the CCBs, which results in the positive credit spreads. The larger the jumps, the smaller is the expected value of the equity at conversion and therefore the higher the spread. The curves of contingent convertible bonds have a similar shape as the curves for the straight debt. Note that the limiting credit spreads are nonzero for both bonds. The spreads for the contingent convertible bonds are higher than for straight debt in Figures 5.3 and 5.4. This is mainly due to the fact that conversion happens substantially earlier than default for most maturities. For short maturities, where the default and conversion barrier are relatively close, conversion is most likely to occur by a jump. If the conversion barrier is crossed by a jump the equity value after conversion is lower than if it is passed by a continuous movement, which results in a higher credit spread for the contingent convertible bonds.

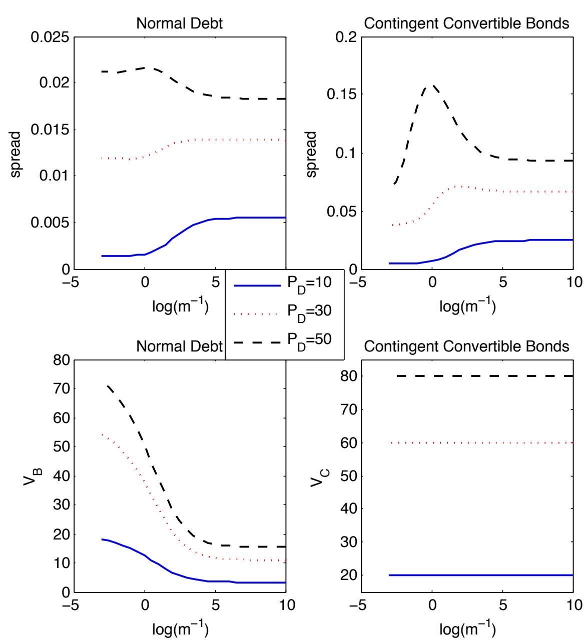


Figure 5.6: "Infrequent, large jumps": FSC spread as a function of maturity for different levels of debt and different conversion barriers. The parameters are $\ell = 1, P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 60, 80$.

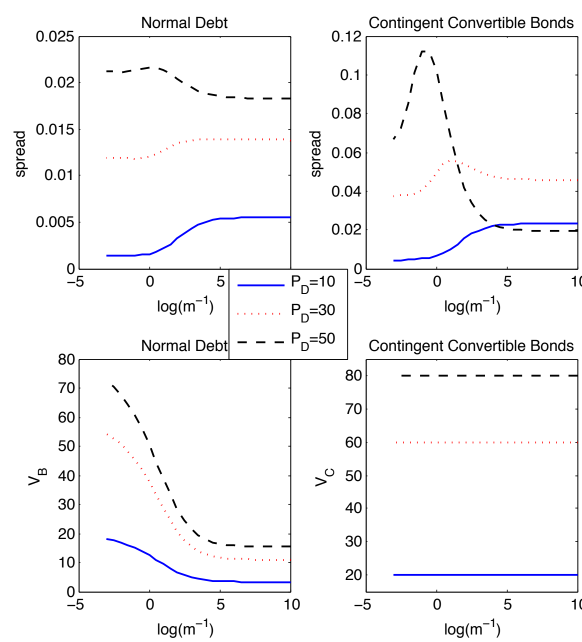


Figure 5.7: "Infrequent, large jumps": FSC spread as a function of maturity for different levels of debt and different conversion barriers. The parameters are $\ell = 1.5, P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 60, 80$.

In Figure 5.5 we consider a firm with infrequent, moderate jumps for FVC2 contracts.

As we can see the magnitude of the spreads is very similar to the corresponding FVC1 contract, however the spreads for long maturities for FVC2 contracts are lower than for FVC1 contracts. If the value of the equity at conversion is sufficient to make the promised payment, than the face value of FVC2s equals the conversion value for $\ell = 1$. For long maturities the default barrier is relatively low, which makes it more likely that the equity value at conversion is high. Hence, for long maturities FVC2s are almost risk-free.

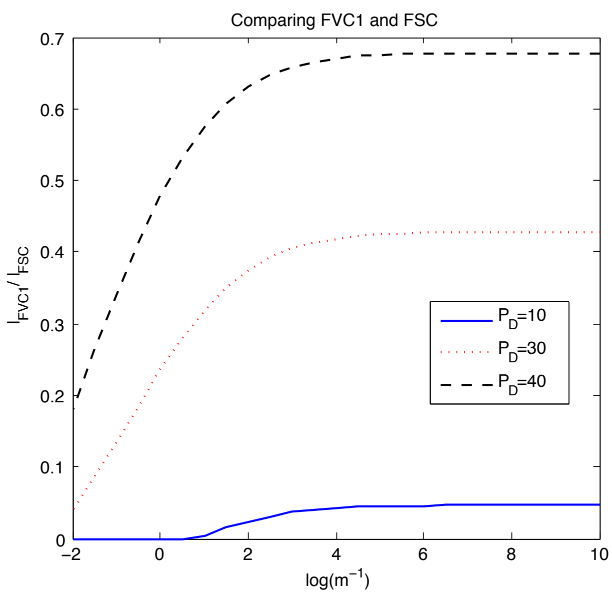


Figure 5.8: "Infrequent, large jumps": Comparing FVC1s with FSCs. We plot ℓ_{FVC1}/ℓ_{FSC} such that the two contracts are equal. The parameters are $P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 60, 80$.

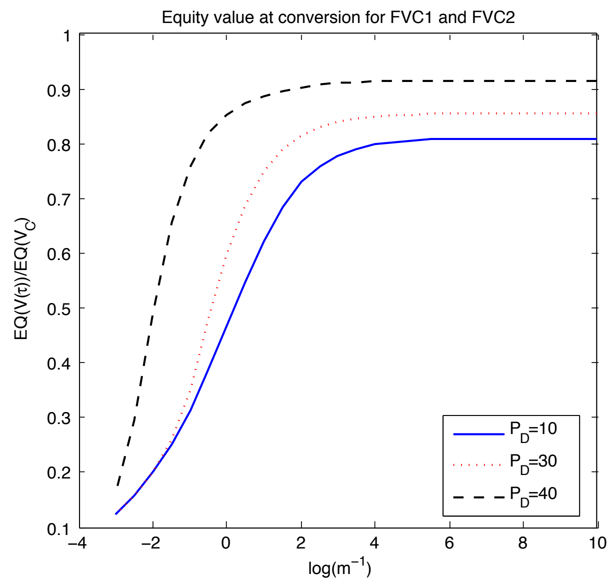


Figure 5.9: "Infrequent, large jumps": Comparison of the total equity value at conversion for FVC1 and FVC2 for different constant conversion barriers. The parameters are $\ell = 1$, $P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 60, 80$.

In Figure 5.6 we plot the spread and the default and conversion barrier for FSCs. The conversion barrier is $\ell = 1$ which means that for a contingent convertible bond with face value 1, the debt holders will get $\frac{1}{S_0}$ shares at conversion. Without loss of generality we can normalize $S_0 = 1$ and hence think of ℓ as the number of shares granted at conversion. Of course, one share at time $t = 0$ has a substantially higher value than a share at time $t = \tau_C$. Hence, we expect the conversion value to be relatively low. This is exactly, what we observe: The spreads for FSCs are substantially higher than for FVCs, due to the lower conversion value. In Figure 5.7 we increase ℓ to 1.5. The higher conversion value dramatically lowers the spread. As we have discussed before, the specification of the number of shares at conversion in terms of the stock price S_0 at time 0 is problematic as it leads to multiple equilibria. Here, we have focussed on the equilibrium with the lower conversion value.

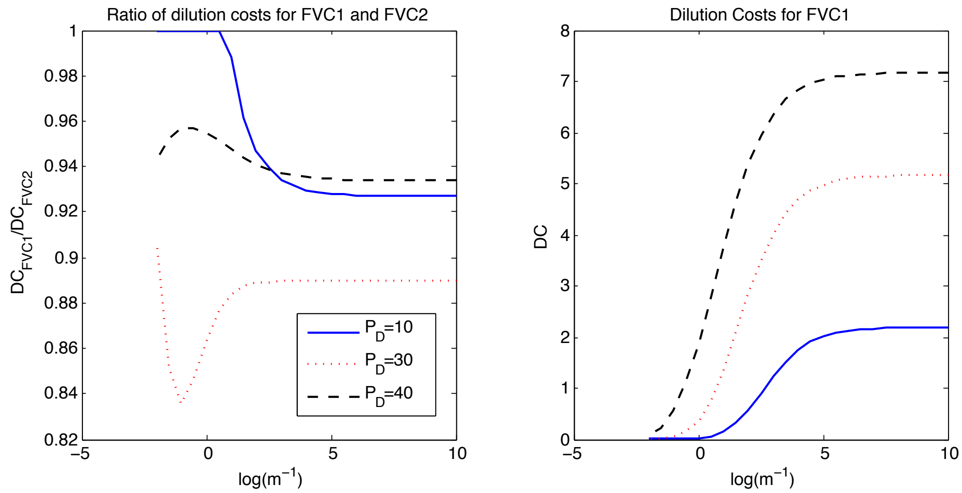


Figure 5.10: "Infrequent, large jumps": Ratio of dilution costs between FVC1 and FVC2. The parameters are $P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 60, 80$.

We already know, that a FVC1 is a special version of a FSC contract. How do we have to choose the parameter ℓ_{FVC1} such that the contract coincides with a given FSC contract with parameter ℓ_{FSC}? Simple calculations show that

$$\frac{\ell_{FVC1}}{\ell_{FSC}} = \frac{S(V_C)}{S(0)}.$$

In Figure 5.8 we plot the corresponding ratio. For long-term maturities with $P_D = 30$ a FSC contract that promises three times the face value of debt in terms of the current market value of equity is equivalent to an FVC1 contract that pays the exact face value of debt in terms of shares with value $S(V_C)$.

In Figures 5.4 and 5.5 we have seen that the spreads for FVC1 and FVC2 contracts are very similar. However, the number of shares granted at conversion differs. The main difference between the two contracts is that the conversion value of FVC1s is based on $EQ(V_C)$, while the conversion value of FVC2s is primarily based on $EQ(V_{\tau_C})$. If these two values coincide, there is no difference between the two FVC contracts. In Figure 5.9 we plot the ratio

$$\mathbb{E}\left[EQ(V_{\tau_C})e^{-(r+m)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{\tau_C<\tau\}}\right]/EQ(V_C)\mathbb{E}\left[e^{-(r+m)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{\tau_C<\tau\}}\right].$$

For long-term contracts the ratio takes values from 0.8 to 0.9, i.e. if all other parameters of the contract stay the same, the number of shares granted under FVC2s should be 10% to 20% higher than under FVC1s. Figure 5.10 plots the ratio of the dilution costs for FVC2s and FVC1s. We observe that the dilution costs for FVC2s are actually around 5% to 12 %

higher than for FVC1s in the long run, i.e. less than the expected 10% to 20%. This can be explained by the argument that for a low value of the firm's assets V_{τ_C} at conversion it is likely that the old shareholders lose (almost) any claim on the company and in this case the dilution costs for FVC1 and FVC2 contracts are the (almost) same.

5.7.2 Assumption 5.3

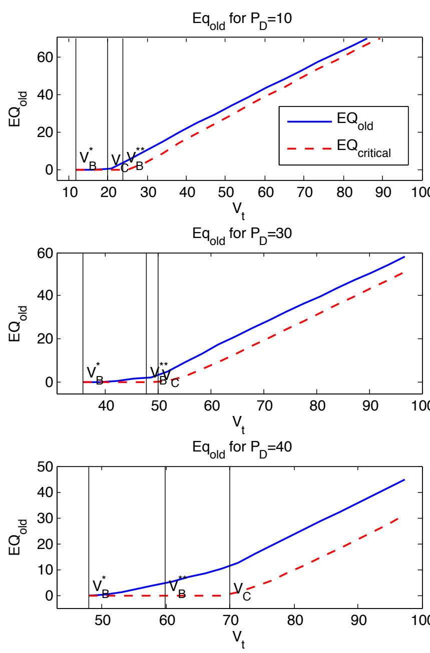


Figure 5.11: Testing Assumption 5.3 for FVC1s. The parameters are $\ell = 1, m = 1, P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 50, 70$.

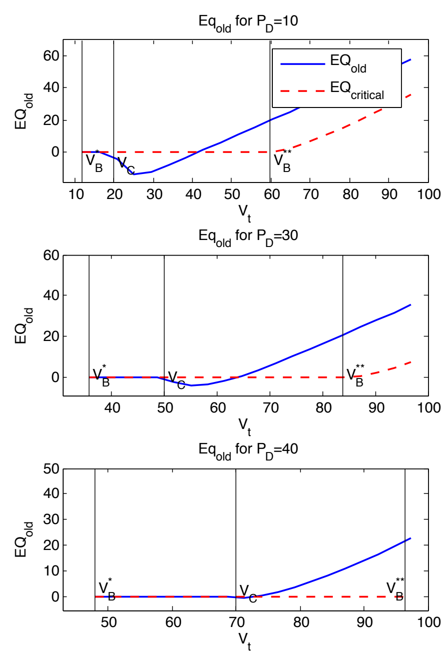


Figure 5.12: Testing Assumption 5.3 for FVC1s. The parameters are $\ell = 1, m = 1, P_C = 40, P_D = 10, 30, 40$ and $V_C = 20, 50, 70$.

If Assumption 5.3 is satisfied, the optimal default barrier is $V_B^* < V_C$ and we can apply the solution formulas developed in this paper. If the assumption is violated, the contingent convertible debt degenerates to straight debt without any recovery payment. From the perspective of a regulator, the contingent convertible bonds become unattractive. Hence, it is important to know under which parameter constellations Assumption 5.3 holds. It requires that V_B^* satisfies the limited liability constraint and that no default barrier larger than V_C yields a higher value for the old shareholders. We define a "critical" equity value:

$$EQ_{critical}(V) = \max(0, EQ_{debt}(V, \max(V_B^{**}, V_C), P_D + P_C, C_D + C_C)).$$

As long as $EQ_{old}(V, V_B^*, V_C) > EQ_{critical}(V)$ for $V > V_C$, Assumption 5.3 is satisfied. In Figure 5.11 we see that for an average maturity of 1 year and an amount of contingent convertible debt (FVC1) that does not exceed the straight debt, the assumption holds. However, if the amount of contingent convertible debt is large as in Figure 5.12, the assumption is violated. Having very short maturity debt can also create problems. In Figure 5.13 we set the average maturity to 1/10 year and the equity value of the old shareholders crosses the critical barrier. On the other hand, long term debt seems to elevate the chances of satisfying the condition. In Figure 5.14 we combine a large amount of contingent convertible debt ($P_C = 40$) with a long maturity (10 years) and the assumption holds.

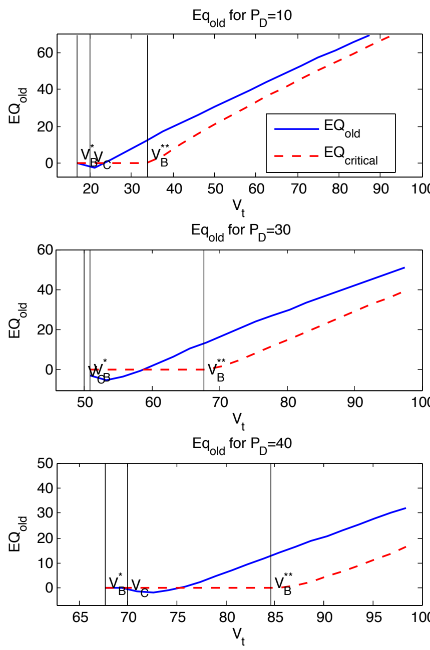


Figure 5.13: Testing Assumption 5.3 for FVC1s. The parameters are $\ell = 1, m = 10, P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 50, 70$.

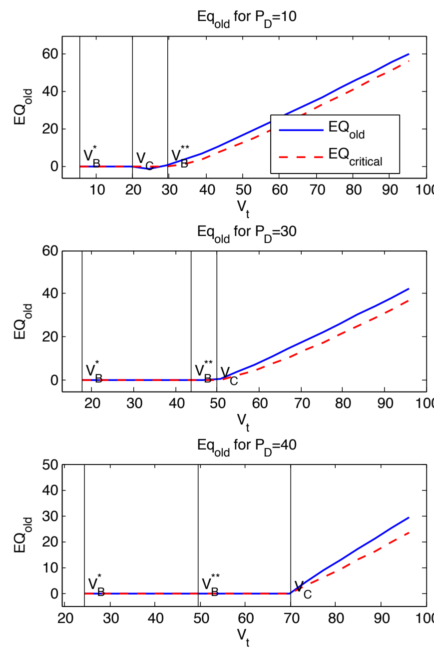


Figure 5.14: Testing Assumption 5.3 for FVC1s. The parameters are $\ell = 1, m = 0.1, P_C = 40, P_D = 10, 30, 40$ and $V_C = 20, 50, 70$.

Why can $EQ_{old}(V, V_B^*, V_C)$ be negative and why does it eventually become positive again? The equity value for the old shareholders can only be negative if the cash payments related to the contingent convertible debt (i.e. coupon payments C_C and face value P_C) are very high. If conversion takes place, the old shareholders are freed from all the cash payments of the convertible debt. The total value of the equity that remains after conversion has by definition a non-negative value. Thus, independently of how small the share of the old equity holders is after conversion, it will have a non-negative value. Therefore, if the firm's value process falls sufficiently low and the cash payments for the contingent convertible bonds are high, the equity value EQ_{old} can become negative. However, the prospect of conversion will eventually lead to a positive price, if V_t falls further.

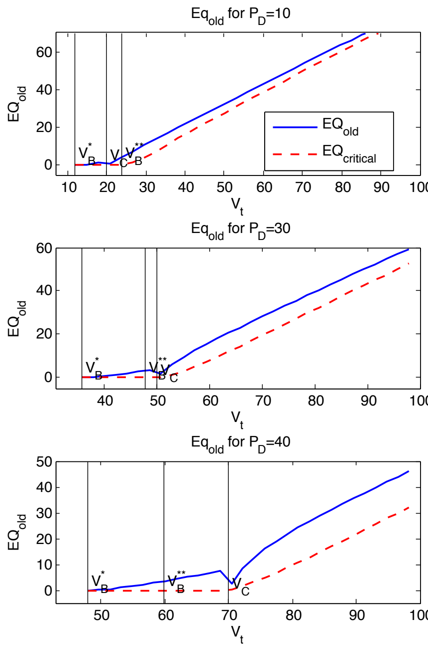


Figure 5.15: Testing Assumption 5.3 for FSCs. The parameters are $\ell = 1, m = 1, P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 50, 70$.

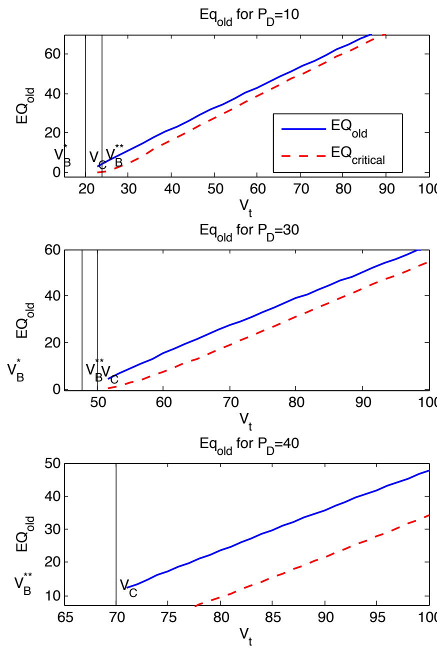


Figure 5.16: Testing Assumption 5.3 for FVC2s. The parameters are $\ell = 1, m = 1, P_C = 10, P_D = 10, 30, 40$ and $V_C = 20, 50, 70$.

In the case of FVC2s and FSCs the findings are similar as depicted in Figures 5.15 and 5.16. The key results from these simulations are that as long as the conversion ratio is sufficiently high (e.g. $\ell = 1$ for FVCs), the amount of contingent convertible debt is smaller than the amount of straight debt ($P_C < P_D$) and the maturity of the debt is sufficiently long (e.g. $\frac{1}{m} > 1$), Assumption 5.3 holds.

5.7.3 Agency Costs

In the basic Merton (1974) model, equity can be regarded as a Call option on the firm's assets. Therefore, equity holders always want to increase the risk (volatility), while debt holders would like to decrease the risk. By including a rolling debt structure and tax benefits, the risk incentives of equity holders and debt holders are in general not opposing any more. As Leland (1994b) has pointed out, for short maturity debt, equity holders and debt holders do both prefer not to scale up the risk. In our model with jumps and CCBs, the risk incentives are more complex.

We consider three different firms, that have the same amount of total risk as measured by the quadratic variation. The default barrier V_B is chosen optimally and the conversion barrier is set 20% higher than the default barrier. Following the second stage optimization

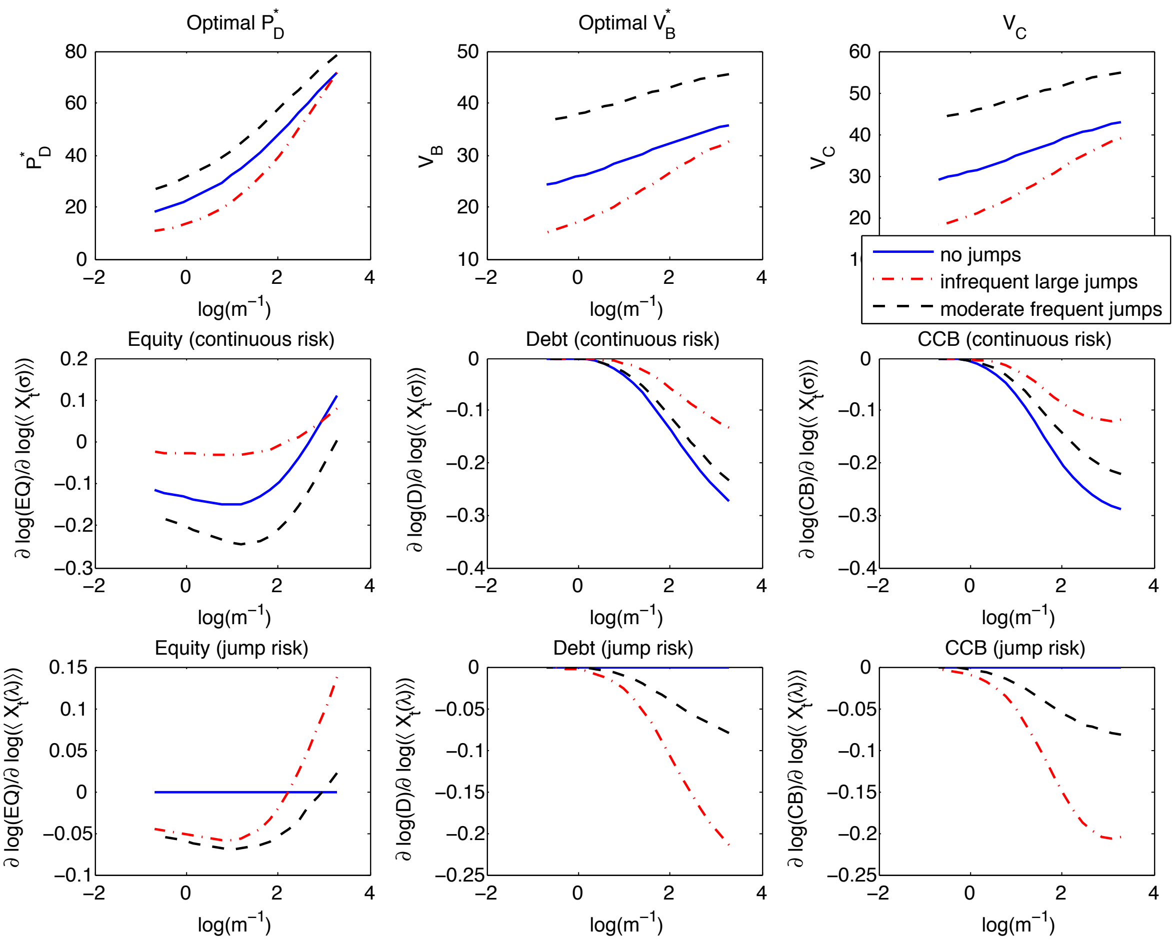


Figure 5.17: Risk incentives for FVC1 contracts with optimal debt. The parameters are $\ell = 1, P_C = 10, c_D = c_C = 0.8, V_C = 1.2 \cdot V_B$. We plot the relative change in equity and debt values for a relative change in total risk (quadratic variation).

as described in Section 5.10, the amount of straight debt P_D is chosen to maximize the total value of the firm for a fixed amount of contingent convertible bonds $P_C = 10$. Keeping the amount of debt constant, we want to analyze which agents will profit or suffer from increasing the risk. In our model we have two types of risk: continuous risk (measured by the volatility σ) and jump risk (measured by the jump intensity λ). We will change the total amount of risk (the quadratic variation) by either increasing the volatility or the jump intensity. Figure 5.17 shows by how many percentage points the equity value, debt value or CCB value changes, if we increase the total risk by 1 percentage point. There are three main observations:

1. Contingent convertible bonds have very similar risk incentives as straight debt. The

agency costs between equity holders and CCB holders are slightly larger than for debt holders.

2. For short-term debt, the incentives of equity holders have the same sign as the incentives for debt and CCB holders. This can be interpreted as lower agency costs for short maturities.

3. The agency costs related to jump risk for a firm that is mainly exposed to continuous risk are very small. Vice versa, the agency costs for continuous risk for a firm with higher jump risk are also much less pronounced than for a firm with mainly continuous risk.

5.8 How should CCBs be designed?

We have presented and completely characterized two different CCB contracts: FSCs and FVC2s. The contract FVC1 is a special case of FSC. The important question is which should be used in practice. We will analyze the different contracts with respect to manipulation, noise trading and multiple equilibria.

5.8.1 Manipulation

We think of manipulation as spreading "good" or "bad" news that will influence the stock price and the credit spread. Furthermore, we assume that conversion is based on the credit spread as described in Section 5.6. Spreading "good" news will temporarily increase the stock price S_t and lower the credit spread π_t. Spreading "bad" news results in the opposite movements. However, after the manipulation the prices will return to their former level. Equivalently, we can think of manipulation as directly affecting the firm's value process V_t. The only interesting case is when the firm's value process V_t is lowered to a level $V_{manip} < V_C$ which triggers conversion and then returns to the former level V_t.

First we consider manipulation by contingent convertible bondholders. In the case of FVC2 contracts spreading "bad" news can trigger conversion and lead to a temporary underevaluation of the stock price S_t. As the number of shares granted to contingent convertible bondholders depends on the stock price at conversion, the temporary underevaluation has a permanent effect. The lower the bondholders can press down the price, the more shares they receive. After the price correction, the contingent convertible bondholders make a profit. More formally, if the contingent convertible bondholders can temporarily manipulate the firm's value process to any arbitrary level V_{manip} with $V_C \geq V_{manip} > V_B$ and the tax benefits are sufficiently low, the will always do so independently of the parameters of the FVC2 contract:

Proposition 5.20. *Assume that a firm issues FVC2s and contingent convertible bondholders can temporarily lower the firm's value to an arbitrary V_{manip} with $V_C \geq V_{manip} > V_B$. If the* *equity value without CoCo bonds is sufficiently high, i.e.* $EQ_{debt}(V_t) > CCB(V_t)$*, they will always manipulate the market for any* $\ell \in (0, \infty)$.

We show in the proof, that $EQ_{debt}(V_t) + TB_C(V_t) > CCB(V_t)$ is always satisfied. Hence, if the tax benefits are sufficiently low, $EQ_{debt}(V_t) > CCB(V_t)$ will also hold.

The same mechanism does not work with FSC (and hence FVC1) contracts. Here the interests of the contingent convertible bondholders are more aligned with those of the shareholders. Spreading "bad" news can trigger conversion, but will not affect the number of shares granted to the bondholders. Spreading "good" news can increase the value of the stocks, but cannot trigger conversion. Hence, FSC contracts offer less incentives for manipulation to the contingent convertible bondholders. More formally, there exists always a conversion parameter ℓ such that the contingent convertible bondholders do not want to manipulate the market:

Lemma 5.12. *Assume that a firm issues FVC1s and contingent convertible bondholders can temporarily lower the firm's value to an arbitrary* $V_{manip} > V_B$*. If* ℓ *is small enough such that*

$$CCB(V_t) - \ell P_C \frac{EQ_{debt}(V_t)}{EQ_{debt}(V_C)} \geq 0$$

they will not manipulate the market at time t.

Proof. If contingent convertible bondholders do not manipulate the market, they get CCB(V_t). If the manipulate the market, they obtain shares that have a value of $\ell P_C \frac{EQ_{debt}(V_t)}{EQ_{debt}(V_C)}$ after the price correction. □

Hence, FVC1 contracts can always be designed such that the bondholders do not want to manipulate the market at time t.

Second, we consider manipulation by the equity holders. We start with a FVC1 contract. Equity holders will not manipulate the market at time t if

$$EQ_{old}(V_t, V_B, V_C) - \left(EQ_{debt}(V_t) - \ell P_C \frac{EQ_{debt}(V_t)}{EQ_{debt}(V_C)} \right) \geq 0$$

The first term is the value of their equity if they do not manipulate the market. The second term is the value of their equity in the case of manipulation after conversion and after the price correction. Plugging in the definitions, the inequality is equivalent to

$$TB_C(V_t) - CCB(V_t, V_B, V_C) + \ell P_C \frac{EQ_{debt}(V_t)}{EQ_{debt}(V_C)} \geq 0$$

If $\ell = 1$, i.e. the contingent convertible bondholders receive equity at conversion that has the same market value as the face value of the CCBs, then the inequality will always be satisfied.

Proposition 5.21. *Assume that a firm has issued straight debt and FVC1s. For $V_t \leq V_0$ and $\ell = 1$, equity holders will never manipulate the market to trigger conversion.*

As a result, if the conversion value is sufficiently high, the equity holders will not manipulate the market to enforce conversion. A similar reasoning applies to FVC2 contracts. Intuitively, the less likely manipulation by contingent convertible bondholders, the more likely manipulation by the equity holders and vice versa. However, there exist parameters, such that neither of them wants to trigger conversion. After conversion, the tax benefits of the contingent convertible bonds are lost. If these benefits are sufficiently high and the conversion parameter ℓ is chosen accordingly, neither of them will manipulate the market.

Lemma 5.13. *Assume a firm issues FVC1s. If the conversion parameter ℓ is chosen such that*

$$TB_C(V_t) \geq CCV(V_t) - \ell P_C \frac{EQ_{debt}(V_t)}{EQ_{debt}(V_C)} \geq 0$$

then neither equity holders nor contingent convertible bondholder will manipulate the market.

We have seen that FVC1 contracts are more robust against manipulation than FVC2 contracts. We favor FVC1 contracts with $\ell = 1$. Equity holders will never manipulate such a contract and for sufficiently high coupon payments c_C the contingent convertible bondholders will not do it neither.

5.8.2 Noise trading

So far we have defined the stock price process as a function of V_t, i.e.

$$S_t = \frac{1}{n}\left(EQ(V_t) - DC(V_t)\right)$$

for $t < \tau_C$ and the only source of risk was the process V_t. We will now suppose that S_t is driven by the firm's value process V_t and an additional independent process. This makes economically sense as changes in the stock price do not necessarily solely reflect changes in the fundamental value. Additional factors, e.g. noise trading, can be captured by including a noise process.

Definition 5.9. *The endogenous stock price process with noise trading before the time of conversion ($t < \tau_C$) is defined as*

$$S_t = S(V_t) = \frac{EQ(V_t) - DC(V_t)}{n}\left(1 + \tilde{X}_t\right)$$

where n is the number of "old" shares and $\tilde{X}_t$ is an arbitrary martingale process, which has expectation zero, i.e. $E[\tilde{X}_t - \tilde{X}_s|\tilde{X}_s] = 0$ for $s \leq t$ and $\tilde{X}_0 = 0$, and is independent of X_t, which is driving V_t.

The intuition behind this modeling approach is that the stock price should reflect the value of the shareholders' claim on the firm's productive assets. As it is shown empirically stock prices can be more volatile than the fundamental value of the underlying assets. Hence, the noise process $\tilde{X}_t$ should capture this additional source of uncertainty.

Proposition 5.22. *The conversion value for FSCs and FVC1s under a stock price with noise trading equals the corresponding payment without exogenous shocks.*

Proof. The conversion value equals

$$\begin{aligned} & n'\mathbb{E}\left[S(\tau_C)e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{\tau_C<\tau\}}\right] \\ =&\frac{n'}{n}\mathbb{E}\left[(EQ(V_{\tau_C})-DC(V_{\tau_C})e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{\tau_C<\tau\}}\right]E\left[\left(1+\tilde{X}_{\tau_C}\right)\right] \\ =&\frac{n'}{n+n'}\mathbb{E}\left[EQ(V_{\tau_C})e^{-(r+m)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{\tau_C<\tau\}}\right] \end{aligned}$$

by the assumption that $E[\tilde{X}_t]=0$ and the independenc of $\tilde{X}$ and X. □

The situation is different for FVC2s. In this case fluctuations in the stock price affect the number of shares granted to the contingent convertible bondholders at conversion. If for example the stock price falls due to a downward shock in $\tilde{X}$, the bondholders get a higher number of shares although the fundamental value did not change. In more detail, the value of the equity used for the redistribution at the time of conversion is $n\cdot S(\tau_C)=EQ_{debt}(\tau_C)\cdot(1+\tilde{X}_{\tau_C})$. The condition if the equity is sufficient to fully pay the promised conversion value changes to $V_{\tau_C}>T^{-1}\left(\ell P_C\left(1+\tilde{X}_{\tau_C}\right)^{-1}\right)$.

Proposition 5.23. *The conversion value for FVC2s under a stock price with noise trading does in general not equal the corresponding payment without exogenous shocks.*

Proof.

$$\begin{aligned} CONV =&\ell P_C\mathbb{E}\left[e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\tau\}}\mathbb{1}_{\{V_{\tau_C}>T^{-1}\left(\ell P_C\left(1+\tilde{X}_{\tau_C}\right)^{-1}\right)\}}\right] \\ &+\mathbb{E}\left[EQ_{debt}(V_{\tau_C})\left(1+\tilde{X}_{\tau_C}\right)e^{-(m+r)\tau_C}\mathbb{1}_{\{V(\tau_C)>V_B\}}\mathbb{1}_{\{V_{\tau_C}\leq T^{-1}\left(\ell P_C\left(1+\tilde{X}_{\tau_C}\right)^{-1}\right)\}}\mathbb{1}_{\{\tau_C<\infty\}}\right] \end{aligned}$$

The problem is that $\tilde{X}(\tau_C)$ appears in a nonlinear way in the above formula and a closed-form solution is not available. However, it is obvious to see that in general the expected value above does not coincide with the corresponding expectation without exogenous shocks. □

This is an argument in favor of FSCs and FVC1s over FVC2s.

5.8.3 Multiple equilibria

In Section 5.3.4 we have shown for FSCs, that defining the number of shares n' granted to the contingent convertible bondholders at conversion in terms of $S(0)$ will lead to multiple equilibria. This result can be extended to show, that if n' is a function of any S_t with $t < \tau_C$, there can be multiple equilibria. One way to circumvent this problem is simply to avoid linking n' to any stock price. However, when writing a contract it is natural to relate n' to some market price. A very appealing alternative are FVC1s. Here, the number n' is calculated using the model to predict $S(V_C)$. In this setup, all prices are unique. Similarly, for FVC2 contracts, we also obtain unique prices. This can be seen as an argument for FVC1 and FVC2 contracts.

5.8.4 Optimal design of CCBs

In a perfect market environment, the different CCBs should be equivalent as all information is correctly priced. However, if we take into account manipulation and noise trading, FVC1 contracts with $\ell = 1$ are more robust against this market imperfections than the other contracts. In addition, FVC1 will always have a unique equilibrium price. In order to avoid that CCBs degenerate to a straight debt contract without recovery payment, we have to ensure that Assumption 5.3 holds. CCB contracts with a high maturity are more likely to satisfy this assumption. For this reason we propose FVC1 with $\ell = 1$ and a high average maturity as the "best" contract.

5.9 Extensions

5.9.1 Time-Varying Firm's Value Process

So far we have assumed that the firm's value process does not change after conversion. In particular, the proportional rate at which profit is disbursed to investors δ is constant before and after conversion. Remember, that as the firm has bondholders and shareholders, δ cannot be seen as a dividend rate. First, the coupons and principal payments have to be paid before the residual is paid out as dividends. However, one of the arguments of introducing contingent convertible debt was that after the conversion the coupon payments are lower than before, allowing the firm to recover from financial distress. A constant δ implies that after conversion the total dividend payments equal the former dividend payments plus the payments for the contingent convertible debt. This high dividend payment could be justified economically by the argument, that after conversion the number of shareholders is larger than before. Nonetheless, it seems that a high dividend payment during times of financial distress is not very common. Hence, a more realistic model should take into account that the payout rate δ decreases after conversion. In this section, we will introduce a general approach which allows all parameters of the firm's value process to change after the conversion. Notwithstanding, the focus will be on different δs.

Assume two different payout ratios:

$$\delta_1 \text{ for } [0, \tau_C] \qquad \delta_2 \text{ for } (\tau_c, \tau]$$

and

$$dV_t = V_t\left((r-\delta_1)dt + \sigma dW_t^* + d\left(\sum_{i=1}^{N_t}(Z_i - 1)\right) - \lambda\xi dt\right) \qquad \text{for } t \le \tau_C$$

$$dV_t = V_t\left((r-\delta_2)dt + \sigma dW_t^* + d\left(\sum_{i=1}^{N_t}(Z_i - 1)\right) - \lambda\xi dt\right) \qquad \text{for } t > \tau_C.$$

By assumption we have $V_C \ge V_B$, which implies

$$\tau \ge \tau_C.$$

Note, that the probability law of τ_C does not change. However, the probability law of τ is not the same anymore. As the value of the coupon payments and the principal repayment of contingent convertible bonds depends only on certain Laplace transforms of τ_C, introducing the time-varying firm's value process does not affect these values. But the prices of straight debt coupons and the conversion values will change. In order to calculate the price of normal debt coupons we need to calculate $\mathbb{E}\left[e^{-\rho\tau}\right]$ and $\mathbb{E}\left[e^{X(\tau)-\rho\tau}\mathbb{1}_{\{\tau<\infty\}}\right]$, where X_t relates to V_t by $V_t = V_0\exp(X_t)$.

Theorem 5.5. *The Laplace transform of the default time for a firm, whose payout ratio changes at conversion, is given by*

$$\begin{aligned}
\mathbb{E}\left[e^{-\tau\rho}\right] =& \bar{c}_1\left(\frac{V_B}{V_0}\right)^{\bar{\beta}_{3,\rho}} J(\log(V_C/V_0), \bar{\beta}_{3,\rho}, \log(V_C/V_B), \rho) \\
&+ \bar{c}_2\left(\frac{V_B}{V_0}\right)^{\bar{\beta}_{4,\rho}} J(\log(V_C/V_0), \bar{\beta}_{4,\rho}, \log(V_C/V_B), \rho) \\
&+ \left(\frac{V_B}{V_C}\right)^{\eta_2} \frac{\eta_2 - \beta_{3,\rho}}{\eta_2} \frac{\beta_{4,\rho} - \eta_2}{\beta_{4,\rho} - \beta_{3,\rho}} \left(\left(\frac{V_C}{V_0}\right)^{\beta_{3,\rho}} - \left(\frac{V_C}{V_B}\right)^{\beta_{4,\rho}}\right)
\end{aligned}$$

with

$$\bar{c}_1 = \frac{\eta_2 - \bar{\beta}_{3,\rho}}{\eta_2} \frac{\bar{\beta}_{4,\rho}}{\bar{\beta}_{4,\rho} - \bar{\beta}_{3,\rho}}$$

$$\bar{c}_2 = \frac{\bar{\beta}_{4,\rho} - \eta_2}{\eta_2} \frac{\bar{\beta}_{3,\rho}}{\bar{\beta}_{4,\rho} - \bar{\beta}_{3,\rho}}$$

and $-\bar{\beta}_{3,\rho} > -\bar{\beta}_{4,\rho}$ *are the two negative roots of the equation*

$$\bar{\psi}(\beta) = \rho$$

with $\bar{\psi}$ being the Lévy exponent of $\bar{X}_t = (r-\delta_2)t + \sigma W_t^* + \sum_{i=1}^{N_t} Y_i$. The functions $-\beta_{3,\rho} > -\beta_{4,\rho}$ are the two negative roots of the equation $\psi(\beta) = \rho$, where ψ is the Lévy exponent of $X_t = (r-\delta_1)t + \sigma W_t^* + \sum_{i=1}^{N_t} Y_i$ The function J is defined as

$$J(x,\theta,y,\rho) = \mathbb{E}\left[e^{-\rho\tau+\theta X_\tau}\mathbb{1}_{\{\tau<\infty,-(X_\tau-x)<y\}}\right]$$

The explicit form of $J(x,\theta,y,\rho)$ is given in Proposition 5.8.

Theorem 5.6. *The default time for a firm, whose payout ratio changes at conversion, satisfies the following equality for $\theta > -\eta_2$:*

$$\begin{aligned}
\mathbb{E}\left[e^{-\tau\rho+\theta X_\tau}\mathbb{1}_{\{\tau<\infty\}}\right] =& \bar{d}_1\left(\frac{V_B}{V_0}\right)^{-\theta-\bar{\beta}_{3,\rho}} J(\log(V_C/V_0), -\bar{\beta}_{3,\rho}, \log(V_C/V_B), \rho) \\
&+ \bar{d}_2\left(\frac{V_B}{V_0}\right)^{-\theta-\bar{\beta}_{4,\rho}} J(\log(V_C/V_0), -\bar{\beta}_{4,\rho}, \log(V_C/V_B), \rho) \\
&+ \frac{\eta_2-\beta_{3,\rho}}{\beta_{4,\rho}-\beta_{3,\rho}}\frac{\beta_{4,\rho}+\theta}{\eta_2+\theta}\left(\frac{V_C}{V_0}\right)^{\theta+\beta_{3,\rho}} + \frac{\beta_{4,\rho}-\eta_2}{\beta_{4,\rho}-\beta_{3,\rho}}\frac{\beta_{3,\rho}+\theta}{\eta_2+\theta}\left(\frac{V_C}{V_0}\right)^{\theta+\beta_{4,\rho}} \\
&- J(\log(V_C/V_0), \theta, \log(V_C/V_B), \rho)
\end{aligned}$$

where

$$\bar{d}_1 = \frac{\eta_2-\bar{\beta}_{3,\rho}}{\bar{\beta}_{4,\rho}-\bar{\beta}_{3,\rho}}\frac{\bar{\beta}_{4,\rho}+\theta}{\eta_2+\theta}$$
$$\bar{d}_2 = \frac{\bar{\beta}_{4,\rho}-\eta_2}{\bar{\beta}_{4,\rho}-\bar{\beta}_{3,\rho}}\frac{\bar{\beta}_{3,\rho}+\theta}{\eta_2+\theta}$$

and with the same notation as in Theorem 5.5 for the rest.

This allows us to calculate the price of straight debt. For the conversion value we need to make only a small change in the evaluation formulas. The value of equity after conversion $EQ(V_{\tau_C})$ has to be determined using the second process, i.e. we replace all $\beta_{3,\rho}$ and $\beta_{4,\rho}$ with $\bar{\beta}_{3,\rho}$ and $\bar{\beta}_{4,\rho}$ in the corresponding formula. The choice of the optimal default barrier is analogous to the case of a constant δ.

5.10 Finding an Optimal Regulation Scheme

The main question is if contingent convertible bonds can be used as a regulation instrument for banks. A "good" regulation instrument would reduce the default probability of a bank without imposing to high costs on the bank. Intuitively, the higher the amount of debt of a firm, the higher the default probability. A simple way to limit the default probability is to limit the amount of debt that a firm is allowed to have. This is equivalent to requiring the firm to hold a minimum amount of equity. However, there are a cost to this regulation, as the firm would lose tax benefits. Instead of limiting the amount of debt, the regulator could require the bank to replace part of its debt with contingent convertible bonds. For example, Flannery (2009a) proposes a scenario, in which banks can choose between holding equity equal to 6% of an asset aggregate, or holding equity equal to 4% of the asset aggregate and CCBs equal to 4% of the asset aggregate.

In the following we will first analyze the optimal capital structure of a firm without regulation. Then we show that, if Assumption 5.3 is satisfied, a regulation schemes that restricts the amount of straight debt and requires mandatory issuing of CCBs, strictly dominates a regulation that only limits the amount of straight debt. The CCB regulation scheme will achieve the same upper bound on the level of risk, but the total value of the firm will be the same as under no regulation. In addition, under the CCB regulation scheme the costs to the government in terms of tax benefits are lower than if no regulation is imposed. However, if Assumption 5.3 is violated, these results do not hold any more .

Throughout the section we make the following two assumptions. First, the conversion barrier V_C is exogenously given. Second:

Assumption 5.6. *The coupon payments of the contingent convertible bonds are positive:*

$$c_C > 0.$$

Note, that as long as there are bankruptcy costs the coupon of straight debt is always positive, when debt is issued at par at time 0. Assumption 5.6 is satisfied in all practically relevant situations. For example, for all FVC1 and FVC2 with $\ell \leq 1$, which are issued at par at time 0, the assumption is satisfied.

5.10.1 Optimal Capital Structure without Regulation

First we consider a firm that issues only normal debt bonds. The parameter V_B is determined endogenously while the parameters m, λ, θ, V_0, c_D and r can be assumed to be given exogenously. Hence, the only remaining choice parameter is the amount of debt P_D. This will be chosen to maximize the total value of the firm, i.e.

$$\max_{P_D} G_{debt} = \max_{P_D}(V + TB_D - BC).$$

Chen and Kou (2009) show that $G_{debt}(P_D)$ is a strictly concave function in P_D. Hence, for any given V, there exists a unique P_D that maximizes $G_{debt}(P_D)$. As the initial value of the firm V is given, the optimal choice of P_D is a tradeoff between tax benefits TB_D and bankruptcy costs BC. Based on our endogenously determined parameters we solve

$$\frac{\partial TB_D}{\partial P_D} = \frac{\partial BC}{\partial P_D}.$$

If we allow the firm to issue CCBs in addition to normal debt, the optimization problem changes to

$$\max_{P_D, P_C} (V + TB_D + TB_C - BC).$$

We consider two cases: Either Assumption 5.3 is satisfied or not. The tricky part is, that this assumption depends on the amount of debt issued by a firm. Hence, we will first analyze the optimal capital structure conditioned on satisfying this constraint. Second, we determine the optimal amount of debt under the restriction that the assumption is violated. Finally, the firm picks the one of the two combinations $\{P_D, P_C\}$, which yields a higher total value of the firm.

Case 1: Assumption 5.3 satisfied:

Under Assumption 5.3, the optimal barrier level V_B equals V_B^* which is independent of any features of CCBs. The optimization problem becomes

$$\max_{P_D, P_C} (V + TB_D(P_D) + TB_C(P_C) - BC(P_D)).$$

The FOC for P_D is then

$$\frac{\partial TB_D(V, V_B^*)}{\partial P_D} = \frac{\partial BC(V, V_B^*)}{\partial P_D},$$

which coincides with the case without CCBs. The optimal level of straight debt does not depend on any characteristics of the CCBs. The next lemma implies, that there exists a unique value of P_D that maximizes the total value of the firm.

Lemma 5.14. *The total value of the firm $G(P_D)$ is a strictly concave function in P_D, if $V_B = V_B^*$.*

Proof. The only difference between $G(P_D)$ and $G_{debt}(P_D)$ are the tax benefits $T_C(P_C)$, which do not depend on P_D. Chen and Kou (2009) have proven the strict concavity of $G_{debt}(P_D)$. □

The total value of the firm depends on P_C only through the tax benefits TB_C, which are monotonically increasing in P_C.

Corollary 5.7. *If the coupon payments c_C are positive, then the total value of the firm $G(P_C)$ is increasing in the value of contingent convertible debt P_C.*

Proof. The total value of the firm is defined as

$$G = V + TB_D + TB_C - BC.$$

As the amount of contingent convertible bonds only affects the tax benefits TB_C which are defined by $\frac{\bar{c}c_C P_C}{r} E[1 - e^{-r\tau_C}]$ and the coupons c_C are not negative, the statement follows. □

As a consequence of Corollary 5.7, equity will be crowded out by CCBs one-to-one as long as Assumption 5.3 is satisfied. As we have seen the firm as a whole will always profit from issuing CCBs, while the taxpayer pays the cost of the additional tax shield. We will in the following assume that a firm can only issue CCBs as certain fraction of its debt. This assumption is implicit or explicit in various proposals to use CBBs for banking regulation. Hence, there will be only tax benefits for CBBs issued as part of a regulation requirement.

Case 2: Assumption 5.3 violated:

If Assumption 5.3 is violated, the optimal default barrier is the maximum of V_B^{**} and V_C. First, we will focus on the case $V_B = V_B^{**}$, which depends on P_D and P_C. The optimization problem becomes

$$\max_{P_D,P_C}(V + TB_D(P_D, P_C) + TB_C(P_D, P_C) - BC(P_D, P_C)).$$

The FOC for P_D and P_C are then

$$\frac{\partial TB_D(V, V_B^{**}) + TB_C(V, V_B^{**})}{\partial P_D} = \frac{\partial BC(V, V_B^{**})}{\partial P_D}$$
$$\frac{\partial TB_D(V, V_B^{**}) + TB_C(V, V_B^{**})}{\partial P_C} = \frac{\partial BC(V, V_B^{**})}{\partial P_C}.$$

We start with the special case $c_D = c_C$, i.e. the coupon payments for straight debt and contingent convertible debt are the same. In this case, P_D and P_C are perfect "substitutes" for the firm as the total value of the firm will only be influenced by $P_D + P_C$. The total value of the firm is the same as for a firm that issues only straight debt in the amount of $P_D + P_C$.

Corollary 5.8. *If $c_D = c_C$, then the firm would like to choose an amount of debt such the default barrier is the same as in case 1.*

Proof. Define $\tilde{P} = P_D + P_C$ and $\tilde{c} = c_D = c_C$. Obviously it holds $\tilde{C} = \tilde{c}\tilde{P} = c_D P_D + c_C P_C = C_D + C_C$ which implies that $EQ_{debt}(V, V_B, P_D + P_C, C_D + C_C) = EQ(V, V_B, \tilde{P}, \tilde{C})$ and $G_{debt}(\tilde{P}, \tilde{C}) = G_{debt}(P_D + P_C, C_D + C_C)$, i.e. the total value of the firm and the equity value of the old shareholders is only influenced by the sum $P_D + P_C$. Hence, the optimal default barrier $V_B^{**}(\tilde{P}, \tilde{C})$ will be the same as in the case where only straight debt is issued. □

The hypothetically optimal V_B^{**} is not feasible, as it would be smaller than V_C. This leads to the following conclusion:

Corollary 5.9. *Assume that $c_D = c_C$. The optimal debt choice $\{P_D, P_C\}$ is any combination of P_D and P_C such that $V_B^*(P_D + P_C) = V_C$.*

If $c_D \neq c_C$, straight debt and the degenerated contingent convertible debt are not perfect substitutes any more. We assume that at time zero all the debt is issued at par. At conversion, which coincides with default, the contingent convertible bondholders receive nothing, while the straight debt bondholders get the recovery payment. Thus, a higher coupon $c_C > c_D$ is needed to compensate the contingent convertible bondholders. Hence, we will assume that the coupon c_C for the contingent convertible debt has to be higher than c_D. The optimization problem of the firm is then

$$\max_{P_D,P_C} G_{debt}(\tilde{P},\tilde{C}) \qquad \text{subject to } \tilde{P} = P_D + P_C \text{ and } \tilde{C} = c_D P_D + c_C P_C.$$

This problem is equivalent to

$$\max_{\tilde{P},\tilde{c}} G_{debt}(\tilde{P},\tilde{c}\tilde{P}) \qquad \text{subject to } \tilde{c} \in [c_D, c_C].$$

We split this two-dimensional problem into a two-stage optimization problem. In the first stage, for any $\tilde{c} \in [c_D, c_C]$ we solve the problem and obtain a unique optimal amount of debt $\tilde{P}(\tilde{c})$ and optimal default barrier $V_B^*(\tilde{c})$ (for simplicity we express the optimal default barrier only in terms of the remaining choice variable $\tilde{c}$). In a second stage, the coupon $\tilde{c}$ is chosen that maximizes the total value of the firm. We denote the optimal coupon by c^*:

$$c^* = \arg \max_{\tilde{c}\in[c_D,c_C]} G_{debt}(\tilde{P}(\tilde{c}))$$

Proposition 5.24. *Assume that $c_D < c_C$. If $V_B^*(c^*) \geq V_C$, then the optimal debt choice $\{P_D, P_C\}$ is the combination of P_D and P_C that satisfies*

$$P_D + P_C = \tilde{P}(c^*) \qquad \textit{and} \qquad c^*(P_D + P_C) = P_D c_D + P_C c_C$$

If $V_B^*(c^*) < V_C$, then the optimal debt choice $\{P_D, P_C\}$ is the highest amount of P_C such that two conditions are satisfied: 1. $V_B^{**}(P_D, P_C) = V_C$ and 2. Assumption 5.3 is violated.

The key result of this section is the following. If Assumption 5.3 is satisfied, the default barrier will be strictly smaller, than in the other case. In addition, if Assumption 5.3 is violated, the default barrier is to some extent unresponsive to restrictions in the maximal amount of straight debt, as straight debt and the degenerated contingent convertible debt become (perfect or imperfect) substitutes. In the next section we will discuss regulation. Intuitively speaking, a regulator wants to enforce a small default barrier, because this will imply a lower default probability. Based on this section we will conclude that the regulator wants to require that only contingent convertible bonds satisfying Assumption 5.3 are issued.

5.10.2 Optimal capital structure with regulation

In this section we discuss different regulation schemes. The regulator will impose restrictions on the capital structure of a bank, such that the “risk” does not exceed a pre-specified level. We will formalize the concept of “risk” from the perspective of a regulator, but intuitively the regulator wants to enforce a low default barrier. The lower the default barrier, the lower the probability of default.

Definition 5.10. *The probability that default happens before t as a function of P_D is defined as*

$$\Upsilon(t, P_D) = \mathbb{P}(\tau \leq t).$$

We assume that the regulator uses a specific risk measure:

Definition 5.11. *Denote the parameter space of the choice variables as Θ. We are only interested in $\Theta = \{(P_D, P_C) \in [0, \infty)^2\}$. A risk measure $\chi_i(P_D, P_C)$ is a mapping from Θ to $[0, \infty)$. A regulation scheme is defined as a restriction of the parameters of our model to the set $\tilde{\Theta}$ such that $\chi_i(P_D, P_C) \leq \omega$ for all $(P_D, P_C) \in \tilde{\Theta}$ and a fixed risk level ω. If we consider only one choice variable we suppress the other in the notation of χ_i.*

We make the additional assumption that the risk measure has the property that if the default probability $\mathbb{P}(\tau \leq t)$ is higher under (P_D, P_C) than under $(\tilde{P}_D, \tilde{P}_C)$ for all t, then $\chi_i(P_D, P_C) \geq \chi_i(\tilde{P}_D, \tilde{P}_C)$.

Definition 5.12. *A capital requirement ρ_i is defined as the maximum amount of debt P_D that a firm is allowed to include in its capital structure such that risk measure $\chi_i(P_D)$ is always smaller than some critical value ω:*

$$\rho_i = \sup\{P_D \in [0, \infty) : \chi_i(\tilde{P}_D) \leq \omega \ \forall \tilde{P}_D \leq P_D\}.$$

Example 5.1. *We define*

$$\rho_1(\omega) = \sup\{P_D \in [0, \infty) : \Upsilon(1, P_D) \leq \omega\}$$

The measure ρ_2 is not restricted to the time period 1:

$$\rho_2(\omega) = \sup\left\{P_D \in [0, \infty) : \int_0^\infty s(t)\Upsilon(t, P_D)dt \leq \omega\right\}$$

where $s(t)$ is a weighting function satisfying $\int_0^\infty s(t)dt = 1$.

The intuition behind the two capital requirements ρ_i is that by setting ω sufficiently low, the default probability is restricted from above in a certain sense. Our model allows us to calculate the Laplace transform of the default time in closed form. Applying Laplace inversion we can numerically calculate the two capital requirement regulation schemes.

If we want to compare different regulation schemes, we have to specify what a "good" regulation means. Regulation can be costly to the firm and the taxpayer. The taxpayers are affected by the amount of tax benefits that they are granting to the firm, while a not "optimal" amount of debt can lower the total value of the firm G.

Definition 5.13. *If two regulation schemes have the same maximum amount of risk ω as specified by the risk measure χ_i, the first regulation scheme is said to be more efficient if the total value of the firm is strictly higher than under the second scheme.*

We want to show that requiring a firm to replace a certain amount of its straight debt by contingent convertible debt can be a more efficient regulation scheme than using only maximal capital requirements.

Throughout this section we assume that Assumption 5.3 is satisfied. This has the following consequences:

1. $G(P_D)$ is strictly concave and there exists a unique optimal amount of debt, which we will denote by P_D^*.
2. The optimal default barrier is V_B^*.

We can conclude the following:

Corollary 5.10. *1. The optimal default barrier V_B^* is a function of P_D but not P_C.*

2. The optimal amount of debt P_D is independent of any features of CCBs.

3. The default probability $\Upsilon(t, P_D)$ is strictly increasing in P_D.

4. The default probability $\Upsilon(t, P_D)$ is independent of P_C.

5. Any risk measure χ_i is increasing in P_D.

6. The risk measures χ_1 and χ_2 are independent of P_C.

Now we want to compare a capital requirement regulation scheme ρ_i with a regulation scheme that requires the mandatory issuing of CCBs.

Definition 5.14. *A CCB regulation scheme is a tuple $\phi_i = (\phi_i^D, \phi_i^C)$ of an upper bound on the amount of straight debt ϕ_i^D and a fixed amount of CCBs ϕ_i^C such that*

$$\chi_i(\tilde{P}_D, \tilde{P}_C) \leq \omega \quad \forall \tilde{P}_D \leq \phi_i^D, \quad \tilde{P}_C = \phi_i^C.$$

In order to prove rigorously that a CCB regulation scheme is more efficient than a capital requirement regulation scheme, we have required Assumption 5.3, which can be tested. The economic intuition behind our regulation approach is straightforward. First, we assume that there exists an optimal level of leverage of straight debt. This makes sense as a firm issuing straight debt faces the tradeoff between tax benefits and bankruptcy costs and the optimal leverage should set the marginal gains of tax benefits equal to the marginal costs of bankruptcy costs. Next, it is also intuitive to assume that a higher amount of straight debt increases the default probability. Hence, if the optimal leverage of a firm implies a too high default probability from the point of view of the regulator, one way to reduce it is to require the firm to lower its level of debt. This would also lower the tax benefits associated with the straight debt. As the new level of leverage is not optimal for the firm any more the total value of the firm will be lower under such a regulation. However, the firm as a whole would benefit from issuing CCBs as it profits from the tax benefits. The amount of CCBs can be chosen such that its tax benefits exactly compensate for the loss due to the capital requirement.

Proposition 5.25. *Consider first a firm without any regulation. Its optimal amount of debt is P_D^* and the maximal total value of the firm is $G(P_D^*)$. Second consider a capital requirement ρ_i. The risk measured by χ_i for $i = 1, 2$ under this scheme is limited to ω and the loss in total value to the firm is $G(P_D^*) - G(\rho_i)$. Third, we define a CCB regulation scheme as (ϕ_i^D, ϕ_i^C), where $\phi_i^D = \rho_i$ and ϕ_i^C is such that $TB(\phi_i^C) = G(P_D^*) - G(\rho_i)$. The risk under the CCB regulation scheme is bounded by ω and the total value of the firm is equal to the value under no regulation, i.e. it is efficient compared to the capital requirement and the firm is indifferent between the CCB regulation and no regulation.*

Note that the above regulation scheme is in a certain sense equivalent to a regulation where existing straight debt is partly replaced by CCBs. Hence, if the optimal amount of debt P_D^* without regulation is known, requiring the firm to replace a certain fraction of the optimal debt amount by CCBs and hence ending up with a lower level of straight debt, will yield exactly the same outcome.

We make an additional assumption:

Assumption 5.7. *The value of the contingent convertible debt is larger than the related tax benefits: $CB > TB_C$.*

For a realistic tax rebate rate, this assumption will always be satisfied.

Lemma 5.15. *The maximal total tax benefits under the CCB regulation scheme are lower than the maximal tax benefits under no regulation.*

Definition 5.15. *The total leverage is defined as*

$$TL = \frac{P_D + P_C}{G}.$$

Lemma 5.16. *The maximal possible total leverage under the CCB regulation scheme is higher than the maximal possible total leverage under a pure capital requirement regulation scheme, if Assumption 5.7 is satisfied.*

The above CCB regulation scheme yields the same total value for the firm as the case where no regulation is imposed. Hence, the firm as a whole does not suffer. Next, the tax deduction costs for the taxpayer are lower compared to the case without regulation. Therefore, the taxpayer is better off. Most importantly, the default probability is lower than in the case without regulation.

The above results hold only if Assumption 5.3 is satisfied. If this assumption is violated, the optimal default barrier is at least V_C, which is by definition larger than V_B^*. Furthermore, if Assumption 5.3 is violated, the default barrier can increase in the amount of contingent convertible bonds P_C. Hence, requiring a bank to issue CoCo bonds can actually increase its risk. Next, as in this case straight debt and degenerated contingent convertible bonds become (perfect or imperfect) substitutes, any capital requirement on the straight debt can be circumvented by issuing more CoCos. The traditional capital requirement regulation would become ineffective. Hence, a regulator would always impose restrictions such that Assumption 5.3 holds.

5.10.3 TBTF Firms

In this section we analyze firms that are "too big to fail" (TBTF). As bankruptcy of such firms might result in a crisis of the overall financial system, the government will not let them fail. At the time of default of a TBTF firm the government will take over its assets and its obligations to make payments to debt holders. Hence, the debt holders of TBTF firms have an implicit government guarantee on their debt contract, which makes their debt basically risk-free.

We will first model formally a TBTF firm, that issues only straight debt. As its debt is risk-free, the value equals:

$$D_{debt}^{TBTF} = \frac{C_D + mP_D}{m + r}.$$

This comes at a cost to the government. At the time of default, the government steps in and obtains assets worth $V(\tau)$. In return, the government takes over the obligation to make the coupon payments and repayments of the face value of debt forever. Therefore, the value of the government subsidy for the firm is

$$SUB^{TBTF}(V, V_B) = \frac{C_D + mP_D}{m + r}\mathbb{E}\left[e^{-(m+r)\tau}\right] - \mathbb{E}\left[V(\tau)e^{-r\tau}\right].$$

The total value of the firm equals the value of the firm's assets plus the tax benefits and the government subsidy. Because of the potential government bailout, the bankruptcy costs do not appear in total value of the firm.

$$\begin{aligned} G_{debt}^{TBTF}(V, V_B) &= V + TB_D(V, V_B) + SUB^{TBTF}(V, V_B) \\ &= V + \frac{\bar{c}C_D}{r}\mathbb{E}\left[1 - e^{-r\tau}\right] + \frac{C_D + mP_D}{m + r}\mathbb{E}\left[e^{-(m+r)\tau}\right] - \mathbb{E}\left[V(\tau)e^{-r\tau}\right] \end{aligned}$$

The equity value is the residual claim of the total value of the firm after the value of the debt is subtracted:

$$EQ_{debt}^{TBTF}(V, V_B) = G_{debt}^{TBTF}(V, V_B) - D_{debt}^{TBTF}$$

Albul, Jaffee and Tchistyi (2010) consider TBTF firms in their model with infinite maturity bonds. They show, that if only consol bonds are issued, the value of the equity and the optimal default barrier are the same as for a normal firm. However, in our model with a rolling debt structure this result does not hold any more.

Proposition 5.26. *The optimal default barrier of a TBTF firm equals*

$$V_B^{***} = \frac{\frac{C_D + mP_D}{r+m}\beta_{3,r+m}\beta_{4,r+m} - \frac{\bar{c}C_D}{r}\beta_{3,r}\beta_{4,r}}{(\beta_{3,r} + 1)(\beta_{4,r} + 1)}\frac{\eta_2 + 1}{\eta_2}.$$

Proof. V_B^{***} is simply the solution to the smooth pasting condition:

$$\left(\frac{\partial(EQ_{debt}^{TBTF})(V,V_B)}{\partial V}\,|_{V=V_B}\right) = 0$$

□

A TBTF firm does not face the tradeoff between tax benefits and bankruptcy costs. As long as $V_B = V_B^{***}$ the firm will issue as much debt as possible.

Proposition 5.27. *Assume that $V_B = V_B^{***}$. The total value of a TBTF firm is strictly increasing in the amount of straight debt:*

$$\frac{\partial G_{debt}^{TBTF}}{\partial P_D} > 0$$

Hence, the regulator should restrict the total amount of debt, that a TBTF firm is allowed to issue. Assume that the regulator wants to limit the risk of all banks to specific level. According to our definition of risk, this is equivalent to imposing an upper bound on the default barrier V_B. Let's denote this target default barrier by $\bar{V}_B$. Assume, that we have two firms that are identical, but one is considered TBTF, while the other does not profit from an implicit government guarantee. What is the upper bound on the amount of straight debt for these two firms, that ensures that the default barrier is below $\bar{V}_B$? Proposition 5.13 and Proposition 5.26 imply that the optimal default barriers are proportional to P_D:

Proposition 5.28. *The optimal default barriers for a normal firm and a TBTF firm can be written as*

$$V_B^* = \kappa^* P_D \qquad V_B^{***} = \kappa^{***} P_D.$$

It holds $\kappa^* \leq \kappa^{***}$. Therefore, in order to enforce that the default barrier is below the critical level $\bar{V}_B$, the regulator has to use a stricter capital requirement for TBTF firms ($P_D \leq \bar{V}_B/\kappa^{***}$), than for a normal firm ($P_D \leq \bar{V}_B/\kappa^*$).

This proposition says that the default risk is increasing faster in the amount of straight debt P_D for a TBTF firm than for a normal firm. Extending the evaluation formulas for CCBs to a TBTF firm is straightforward. We just need to replace EQ_{debt} by EQ_{debt}^{TBTF} and V_B^* by V_B^{***}. The regulator can apply a similar CCB regulation scheme to a TBTF firm as described in the last subsection. The CCBs can be used to compensate the firm for its loss in the total value due to the capital requirement. As the TBTF already profits from the government subsidy SUB^{TBTF}, it will usually need less tax benefits from the CCBs to obtain the same total value as a firm without this subsidy. The main takeaway of this subsection is that a TBTF firm will always have a lower amount of straight debt than a comparable normal firm under regulation.

5.11 Conclusion

In the aftermath of the financial crisis of 2008 contingent convertible bonds were discussed as regulation instruments for banks. CCBs are new debt instruments that automatically convert to equity when the issuing firm or bank reaches a specified level of financial distress. We conceptualize the modeling of CCBs and present a formal model for this new hybrid security, which incorporates jumps in the firm's value process and allows for a rolling debt structure. We extend Chen and Kou's model to incorporate contingent convertible debt by introducing a second barrier which triggers conversion. We are able to completely characterize two different types of CCBs: In the first case the number of shares granted at conversion is fixed a priori. In the second specification the number of shares granted at conversion is chosen a posteriori such that the value of the shares equals a specified value. We determine the dilution costs to the old shareholders for the two types of CCBs. Our analysis shows that CCBs behave similarly to straight debt in many ways: The credit spread as a function of maturity is humped-shaped and the limiting credit spread for a maturity approaching zero is generally non-zero.

However, the specification of the conversion payment has huge effects on the features of CCBs. In order to obtain a unique equilibrium price for FSCs, certain restrictions have to be imposed on the design of this debt contract. There are two different conceptional approaches to modeling FVCs. In one approach FVCs and FSCs can be incorporated into the same unified framework. In the other approach, FSCs and FVCs have very distinct properties. We also explain how to evaluate the model if the parameters of the firm's value process change after conversion.

We discuss whether conversion can be based on observable market prices. We show that the conversion event can be specified in terms of credit spreads or risk premiums for CDSs, leading to the same pricing formulas that we have obtained when conversion was triggered by movements in the firm's value process. Hence, our evaluation formulas can be applied in practice with a trigger event based on observable market prices.

An important question concerns the optimal design of CCBs. We show that FSCs and FVC1s are not affected by noise in the stock price process. Furthermore, FSC and FVC1 contracts are more robust against manipulation by the contingent convertible bondholders. If in the case of conversion bond holders of FVC1s get shares that have the same value as the face value of debt (i.e. the conversion parameter ℓ is equal to 1), then equity holders will never manipulate the market. Therefore, we favor FVC1 contracts with $\ell = 1$.

Last but not least we analyze the potential of CCBs as a regulation instrument. If the no-early-default condition is satisfied a regulation combining leverage restrictions and the requirement of issuing a certain fraction of CCBs can efficiently lower the default probability without reducing the total value of the firm. However, if the no-early-default condition is violated, a CCB regulation can actually increase the risk. In order to ensure that this condition holds, only CCBs with a long maturity should be issued.

References

Ahn, S. C., and A. R. Horenstein, 2013, Eigenvalue ratio test for the number of factors, *Econometrica* 81, 1203–1227.

Aït-Sahalia, J. Fan, Y., and D. Xiu, 2010, High-frequency covariance estimates with noisy and asynchronous data, *Journal of the American Statistical Association* 105, 1504–1517.

Aït-Sahalia, P. A. Mykland, Y., and L. Zhang, 2005a, How often to sample a continuous-time process in the presence of market microstructure noise, *Review of Financial Studies* 18, 351–416.

Aït-Sahalia, P. A. Mykland, Y., and L. Zhang, 2005b, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* 100, 1394–1411.

Aït-Sahalia, Y., 2004, Disentangling diffusion from jumps, *Journal of Financial Economics* 74, 487–528.

Aït-Sahalia, Y., J. Fan, and Y. Li, 2013, The leverage effect puzzle: Disentangling sources of bias at high frequency, *Journal of Financial Economics* .

Aït-Sahalia, Y., J. Fan, and D Xiu, 2010, High-frequency estimates with noisy and asynchronous financial data, *Journal of the American Statistical Association* 105, 1504–1516.

Aït-Sahalia, Y., and J. Jacod, 2009a, Estimating the degree of activity of jumps in high frequency data, *Annals of Statistics* 37, 2202–2244.

Aït-Sahalia, Y., and J. Jacod, 2009b, Testing for jumps in a discretely observed process, *Annals of Statistics* 37, 184–222.

Aït-Sahalia, Y., and J. Jacod, 2012, Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data, *Journal of Economic Literature* 50, 1005–1048.

Aït-Sahalia, Y., and J. Jacod, 2014, *High-Frequency Financial Econometrics* (New Jersey: Princeton University Press).

Aït-Sahalia, Y., and D. Xiu, 2015a, Principal component analysis of high frequency data, *Working paper* .

Aït-Sahalia, Y., and D. Xiu, 2015b, Principal component estimation of a large covariance matrix with high-frequency data, *Working paper* .

Albul, B., D.M. Jaffee, and A. Tchistyi, 2010, Contingent convertible bonds and capital structure decisions, *Working paper* .

Aldous, D. G., and G. K. Eagleson, 1978, On mixing and stability of limit theorems, *Annal of Probability* 6, 325–331.

Allen, F., E. Carletti, and R. Marquez, 2013, Deposits and bank capital structure, *SSRN working paper 2238048* .

Andersen, T.G., T. Bollerslev, and F. X. Diebold, 2010, Parametric and non-parametric volatility measurement, *In: Hansen, L.P., Ait-Sahalia, Y. (Eds.), Handbook of Financial Econometrics, North-Holland, Amsterdam* 67–138.

Andersen, T.G., T. Bollerslev, F. X. Diebold, and P. Labys, 2001, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* 42, 42–55.

Andersen, T.G., D. Dobrev, and E. Schaumburg, 2012, Jump-robust volatility estimation using nearest neighbour truncation, *Journal of Econometrics* 169, 75–93.

Anderson, T. W., 1963, Asymptotic theory for principal component analysis, *Annals of Mathematical Statistics* 34, 122–148.

Andrews, D. W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 34, 122–148.

Ansel, J. P., and C. Stricker, 1991, Lois de martingale, densities et dècomposition de Föllmer-Schweizer, *Preprint, Universitè de Franche-Comtè, Besancon* .

Auh, J. K., and S. M. Sundaresan, 2014, Repo runs and the bankruptcy code, *SSRN working paper 2217669* .

Back, K., 1991, Asset prices for general processes, *Journal of Mathematical Economics* 20, 371–395.

Bai, J., 2003, Inferential theory for factor models of large dimensions, *Econometrica* 71, 135–171.

Bai, J., and S. Ng, 2002, Determining the number of factors in approximate factor models, *Econometrica* 70, 191–221.

Bai, J., and S. Ng, 2006, Evaluating latent and observed factors in macroeconomics and finance, *Journal of Econometrics* 507–537.

Bai, Z. D., and Y. Q. Yin, 1993, Limit of the smallest eigenvalue of a large dimensional sample covariance matrix, *The Annals of Probability* 21, 1275–1294.

Bakshi, G.S., N. Ju, and H. Ou-Yang, 2006, Estimation of continuous-time models with an application to equity volatility, *Journal of Financial Economics* 82, 227–249.

Bali, T. T., R. F. Engle, and Y. Tang, 2014, Dynamic conditional beta is alive and well in the cross- section of daily stock returns., *Working paper* .

Bandi, F., and R. Reno, 2012, Time-varying leverage effects, *Journal of Econometrics* 12, 94–113.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 2008, Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise, *Econometrica* 76, 1481–1536.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 2011, Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, *Journal of Econometrics* 162, 149–169.

Barndorff-Nielsen, O.E., and N. Shephard, 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society* 253–280.

Barndorff-Nielsen, O.E., and N. Shephard, 2004, Power and bipower variation with stochastic volatility and jumps, *Journal of Financial Econometrics* 2, 1–48.

Barndorff-Nielsen, O.E., and N. Shephard, 2006, Econometrics of testing for jumps in financial economics using bipower variation, *Journal of Financial Econometrics* 4, 1–30.

Barndorff-Nielsen, O.E., and N. Shephard, 2007, Variation, jumps, market frictions and high frequency data in financial econometrics, *In: Blundell, R., Torsten, P., Newey, W.K. (Eds.), Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, vol. III. , Cambridge University Press, Cambridge* 328–372.

Barndorff-Nielsen, O.E., N. Shephard, and M Winkel, 2006, Limit theorems for multipower variation in the presence of jumps, *Stochastic Processes and their Applications* 116, 796–806.

Bhanot, K., and A. Mello, 2006, Should corporate debt include a rating trigger?, *Journal of Financial Economics* 79, 69–98.

Bielecki, T.R., and M. Rutkowski, 2002, *Credit risk: Modeling, valuation and hedging* (Springer).

Black, F., 1976, Studies of stock price volatility changes, *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics* 6, 177–181.

Bohn, J.R., 1999, Characterizing credit spreads, Working paper: University of California, Berkeley.

Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen, 2009, A discrete-time model for daily S&P500 returns and realized variations: Jumps and leverage effects, *Journal of Financial Econometrics* 150, 151–166.

Bollerslev, T., T.H. Law, and G. Tauchen, 2008, Risk, jumps, and diversification, *Journal of Financial Econometrics* 144, 234–256.

Bollerslev, T., S. Z. Li, and V. Todorov, 2013, Jump tails, extreme dependencies and the distribution of stock returns, *Journal of Econometrics* 172, 307–324.

Bollerslev, T., S. Z. Li, and V. Todorov, 2015a, Roughing up beta: Continuous vs. discontinuous betas, and the cross section of expected stock returns, *Working paper* .

Bollerslev, T., and V. Todorov, 2010, Jumps and betas: A new theoretical framework for disentangling and estimating systematic risks, *Journal of Econometrics* 157, 220–235.

Bollerslev, T., and V. Todorov, 2011a, Estimation of jump tails, *Econometrica* 79, 1727–1783.

Bollerslev, T., and V. Todorov, 2011b, Tails, fears and risk premia, *Journal of Finance* 66, 2165–2211.

Bollerslev, T., V. Todorov, and L. Xu, 2015b, Tail risk premia and return predictability., *Journal of Financial Economics* Forthcoming.

Bollerslev, T., and H. Zhou, 2002, Estimating stochastic volatility diffusion using conditional moments of integrated volatility, *Journal of Financial Econometrics* 109, 33–65.

Bolton, P., and F. Samama, 2012, Capital access bonds: Contingent capital with an option to convert, *Economic Policy* 27, 275–317.

Calomiris, C., and C. Kahn, 1991, The role of demandable debt in structuring optimal banking arrangements, *American Economic Review* 81, 497–513.

Calomiris, C.W., and R.J. Herring, 2011, Why and how to design a contingent convertible debt requirement., *Working paper* .

Campbell, A. W. Lo, C.J., and A.C. Mackinlay, 1997, *The Econometrics of Financial Markets.* (New Jersey: Princeton University Press).

Campbell, J. Y., 2000, Asset pricing at the millenium, *Journal of Finance* 1515–1567.

Campbell, J. Y., and L. Hentschel, 1992, No news is good news: An asymmetric model of changing volatility in stock returns, *Journal of Financial Economics* 31, 281–318.

Carhart, M. M., 1997, On persistence in mutual fund performance, *Journal of Finance* 1, 57–82.

Carr, P., and L. Wu, 2009, Variance risk premiums, *Review of Financial Studies* 22, 1311–1341.

Chamberlain, G., 1988, Asset pricing in multiperiod securities markets, *Econometrica* 56, 1283–1300.

Chamberlain, G., and M. Rothschild, 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* 51, 1281–1304.

Chen, N., P. Glasserman, B. Nouri, and M. Pelger, 2012, Cocos, bail-in, and tail risk, *Working Paper 0004, Office of Financial Research. Available at www.treasury.gov/ofr.* .

Chen, N., and S. G. Kou., 2009, Credit spreads, optimal capital structure, and implied volatility with endogenous default and jump risk, *Mathematical Finance* 19, 343–378.

Chernov, M., A.R. Gallant, E. Ghysels, and G Tauchen, 2003, Alternative models for stock price dynamics, *Journal of Financial Econometrics* 116, 225–257.

Christensen, K., R.C.A. Oomen, and M. Podolskij, 2010, Realised quantile-based estimation of the integrated variance, *Journal of Financial Econometrics* 159, 74–98.

Christie, A.A., 1982, The stochastic behavior of common stock variances: Value, leverage and interest rate effects, *Journal of Financial Economics* 10, 407–432.

Cochrane, 2011, Discount rates, *Journal of Finance* 4, 1047–1108.

Connor, G., and R. Korajczyk, 1988, Risk and return in an equilibrium APT: Application to a new test methodology, *Journal of Financial Economics* 21, 255–289.

Connor, G., and R. Korajczyk, 1993, A test for the number of factors in an approximate factor model, *Journal of Finance* 58, 1263–1291.

Cont, R., and P. Tankov, 2004, *Financial Modelling with Jump Processes.* (Chapman & Hall/CRC.).

Dao, T., and M. Jeanblanc, 2006, A double exponential structural jump diffusion model with endogenous default barrier, *Working Paper* .

De Spiegeleer, J., and W. Schoutens, 2011, Pricing contingent convertibles: A derivatives approach, Working paper.

DeAngelo, H., and R. Stulz, 2013, Why high leverage is optimal for banks., *Working paper w19139, National Bureau of Economic Research.* .

Décamps, J., and S. Villeneuve, 2014, Rethinking dynamic capital structure models with roll-over debt, *Mathematical Finance* 24, 66–96.

Dellacherie, C., and P. A. Meyer, 1982, *Probabilities and Potential* (North-Holland).

Diamond, D., and Z. He, 2014, A theory of debt maturity: The long and short of debt overhang, *Journal of Finance* 69, 719–762.

Doherty, N., and S. Harrington, 1995, Investment incentives, bankruptcies and reverse convertible debt, Working paper, Wharton School.

Dudley, W., 2009, Some lessons from the crisis, Remarks at the Institute of International Bankers Membership Luncheon, NYC.

Duffie, D., 2001, *Dynamic Asset Pricing Theory*, third edition (Princeton University Press).

Duffie, D., 2010, *A Contractual Approach to Restructuring Financial Institutions, in G. Schultz, K. Scott, and J. Taylor, eds.* (G. Schultz, K. Scott, and J. Taylor, eds., Ending Government Bailouts as we Know Them).

Engle, R. F., and V. K. Ng, 1993, Measuring and testing the impact of news on volatility, *Journal of Finance* 48, 1749–1778.

Eraker, B., 2004, Do stock prices and volatility jump? Reconciling evidence from spot and option prices,, *Journal of Finance* 59, 1367–1404.

Eraker, B., M. Johannes, and N. Polson, 2003, The impact of jumps in volatility and returns, *Journal of Finance* 58.

Fama, E. F., and K. R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, E. F., and R. French, K., 2004, The capital asset pricing model: Theory and evidence, *Journal of Economic Perspective* 25–46.

Fan, J., Y. Fan, and J. Lv, 2008, High dimensional covariance matrix estimation using a factor model, *Journal of Econometrics* 147, 186–197.

Fan, J., A. Furger, and D. Xiu, 2014, Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data, *Working paper* .

Fan, J., Y. Li, and K. Yu, 2012, Vast volatility matrix estimation using high-frequency data for portfolio selection, *Journal of the American Statistical Association* 107, 412–428.

Fan, L., Y. Liao, and M. Mincheva, 2013, Large covariance estimation by thresholding principal orthogonal complements, *Journal of the Royal Statistical Society* 75, 603–680.

Feigin, P. D., 1985, Stable convergence of semimartingales, *Stochstic Processes and their Applications* 19, 125–134.

Figleskwi, S., and X. Wang, 2000, Is the "leverage effect" a leverage effect, *Working paper* .

Flannery, M. J., 2005, *"No Pain, No Gain?" Effecting Market Discipline via "Reverse Convertible Debentures,"* (Capital Adequacy Beyond Basel: Banking, Securities, and Insurance, Oxford University Press).

Flannery, M. J., 2009a, Market-valued triggers will work for contingent capital instruments, *Solicited Submission to U.S. treasury working group on bank capital* .

Flannery, M. J., 2009b, Stabilizing large financial institutions with contingent capital certificates, *Working paper* .

Forni, M., M. Hallin, M. Lippi, and L. Reichlin, 2000, The generalized dynamic-factor model: Identification and estimation,, *REVIEW* 82, 540–554.

French, R., K., G. W. Schwert, and R. F. Stambaugh, 1987, Expected stock returns and volatility, *Journal of Financial Economics* 19, 3–29.

Gabaix, X., 2012, Variable rare disasters: An exactly solved framework for ten puzzles in macrofinance, *Quarterly Journal of Economics* 645–700.

Glasserman, P., and B. Nouri, 2012, Contingent capital with a capital-ratio trigger, *Management Science* .

Goldstein, R., N. Ju, and H. Leland, 2001, An ebit-based model of dynamic capital structure, *Journal of Business* 74, 483–512.

Goyal, A., 2012, Empirical cross-sectional asset pricing: a survey, *Finance Market Portfolio Management* 3–38.

Group, Squam Lake Working, 2009, An expedited resolution mechanism for distressed financial firms: Regulatory hybrid securities, *Council on Foreign Relations* .

Hall, P., and C.C. Heyde, 1980, *Martingale Limit Theory and its Application* (Academic Press).

Hallin, M., and R. Liska, 2007, The generalized dynamic factor model: Determining the number of factors, *Journal of the American Statistical Association* 102, 603–617.

Hansen, P., and A. Lunde, 2006, Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* 24, 127–161.

Hanson, S. G., A.K. Kashyap, and J.C. Stein, 2011, A macroprudential approach to financial regulation, *Journal of Economic Perspectives* 25, 3–28.

Harding, J.P., X. Liang, and S. Ross, 2013, The optimal capital structure of banks: Balancing deposit insurance, capital requirements and tax-advantaged debt, *Journal of Financial Services Research* 43, 127–148.

Hayashi, T., and N. Yoshida, 2005, On covariance estimation of non-synchronously observed diffusion processes, *Bernoulli* 11, 359–379.

He, Z., and W. Xiong, 2012, Rollover risk and credit risk, *Journal of Finance* 67, 391–429.

Hilberink, B., and L.C.G. Rogers, 2002, Optimal capital structure and endogenous default, *Finance and Stochastics* 6, 237–263.

Hilscher, J., and A. Raviv, 2011, Bank stability and market discipline: Debt-for-equity swap versus subordinated notes, *Working paper* .

Himmelberg, C.P., and S. Tsyplakov, 2012, Pricing contingent capital bonds: Incentives matter, *Working paper* .

Horn, R. A., and C. R. Johnson, 1991, *Topics in Matrix Analysis* (Cambridge: Cambridge University Press).

Huang, X., and G. Tauchen, 2005, The relative contribution of jumps to total price variance, *Journal of Financial Econometrics* 3, 456–499.

Jacod, J., Y. Li, P.A. Mykland, M. Podolskij, and M. Vetter, 2009, Microstructure noise in the continuous case: The pre-averaging approach, *Stochastic Processes and their Applications* 119, 2249–2276.

Jacod, J., and P. Protter, 2012, *Discretization of Processes* (Heidelberg: Springer).

Jacod, J., and V. Todorov, 2010, Do price and volatility jump together?, *Annals of Applied Probability* 20, 1425–1469.

Jagannathan, R., E. Schaumburg, and G. Zhou, 2010a, Cross-sectional asset pricing test, *Annals of Review of Financial Economics* 2, 49–74.

Jagannathan, R., G. Skoulakis, and Z. Wang, 2010b, The analysis of cross section of security returns, *Handbook of Financial Econometrics* 2, 73–134.

Jagannathan, R., and Z. Wang, 1996, The conditional CAPM and the cross-section of expected stock returns, *Journal of Finance* 3–53.

Jarrow, R.A., and E.R. Rosenfeld, 1984, Jump risks and the intertemporal capital asset pricing model, *Journal of Business* 57, 337–351.

Johannes, M., 2004, The statistical and economic role of jumps in continuous-time interest rate models, *Journal of Finance* 59, 227–260.

Kallenberg, O., 1997, *Foundations of Modern Probability* (Springer).

Kou, S. G., 2002, A jump diffusion model for option pricing, *Management Science* 48, 1086–1101.

Kou, S. G., and H. Wang, 2003, First passage times of a jump diffusion process, *Advances in Applied Probability* 35, 504–531.

Kou, S.G., and H. Wang, 2004, Option pricing under a double exponential jump diffusion model, *Management Science* 50, 1178–1192.

Koziol, C., and J. Lawrenz, 2012, Contingent convertibles: Solving or seeding the next banking crisis?, *Journal of Banking and Finance* 36, 90–104.

Kreps, D. M., 1981, Arbitrage and equilibrium in economies with infinitely many commodities, *Journal of Mathematical Economics* 8, 15–35.

Le Courtois, O., and F. Quittard-Pinon, 2006, Risk-neutral and actual default probabilities with an endogenous bankruptcy jump-diffusion model, *Asia-Pacific Financial Markets* 13, 11–39.

Lee, S.S., and P.A. Mykland, 2008, Jumps in financial markets: A new nonparametric test and jump dynamics, *Review of Financial Studies* 21, 2535–2563.

Leland, H. E., 1994a, Corporate debt value, bond covenants, and optimal capital structure, *Journal of Finance* 49, 1213–1252.

Leland, H. E., and K. B. Toft, 1996, Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads, *Journal of Finance* 51, 987–1019.

Leland, H.E., 1994b, Bond prices, yield spreads, and optimal capital structure with default risk, IBER Working paper: University of California, Berkeley.

Leland, H.E., 2004, Predictions of default probabilities in structural models, *Journal of Investment Management* 1–16.

Lettau, M., and S. Ludvigson, 2011, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 1238–1287.

Lintner, J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economic and Statistics* 1, 13–37.

Liu, J., F. Longstaff, and J. Pan, 2003, Dynamic asset allocation with event risk, *Journal of Finance* 59, 755–793.

Lunde, A., N. Shephard, and K. Sheppard, 2014, Econometric analysis of vast covariance matrices using composite realized kernels, *Working paper* .

Madan, D., and W. Schoutens, 2010, Conic coconuts: The pricing of contingent capital notes using conic finance, *Working paper* .

Mancini, C, 2009, Non parametric threshold estimation for models with stochastic diffusion coefficient and jumps, *Scandinavian Journal of Statistics* 42–52.

Marcenko, V. A., and L. A. Pastur, 1967, Distribution of eigenvalues for some sets of random matrices,, *Math. USSR-Sbornik* 457–483.

McDonald, R., 2013, Contingent capital with a dual price trigger, *Journal of Financial Stability* 9, 230–241.

Medvedev, A., and O. Scaillet, 2007, Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility, *Review of Financial Studies* 20, 427–459.

Merton, R. C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* 29, 449–470.

Modigliani, F., and M. Miller, 1958, The cost of capital, corporation finance and the theory of investment, *American Economic Review* 48, 261–297.

Murphy, G., M. Walsh, and M. Willison, 2012, Precautionary contingent capital, *Financial Stability Paper 16, Bank of England* .

Myers, S. C., 1977, Determinants of corporate borrowing, *Journal of Financial Economics* 5, 147–175.

Naik, V., and M. Lee, 1990, General equilibrium pricing of options on the market portfolio with discontinuous returns, *Review of Financial Studies* 3, 493–521.

Nelson, D. B., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.

Onatski, A., 2010, Determining the number of factors from empirical distribution of eigenvalues, *Review of Economic and Statistics* 92, 1004–1016.

Pan, J., 2002, The jump risk premium implicit in options: Evidence from an integrated time-series study, *Journal of Financial Economics* 3–50.

Paul, D., and J. Silverstein, 2008, No eigenvalues outside the support of limiting empirical spectral distribution of a separable covariance matrix, *Working paper* .

Pazarbasioglu, C., J. Zhou, V. Le Leslé, and M. Moore, 2011, Contingent capital: Economic rationale and design features, *Staff Discussion Note 11/010, International Monetary Fund, Washington, D.C.* .

Pelger, M., 2015a, Large-dimensional factor modeling based on high-frequency observations, *Working paper* .

Pelger, M., 2015b, Understanding systematic risk: A high-frequency approach, *Working paper* .

Pennacchi, G., 2010, A structural model of contingent bank capital, *Federal Reserve Bank of Cleveland Working Paper 10-04* .

Pennacchi, G., T. Vermaelen, and C.C.P. Wolff, 2010, Contingent capital: The case for COERCs, *Working paper* .

Prigent, J., 2003, *Weak Convergence of Financial Markets* (Heidelberg: Springer).

Protter, P.E, 2004, *Stochastic Integration and Differential Equations* (Springer-Verlag).

Rogers, C., and D. Williams, 2000, *Diffusions, Markov Processes and Martingales*, volume 2, second edition (Cambridge: Cambridge University Press).

Ross, S. A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Schweizer, M., 1992, Martingale densities for general asset prices, *Journal of Mathematical Economics* 21, 363–378.

Sharpe, W., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 3, 425–442.

Silverstein, J., and S. Choi, 1995, Analysis of the limiting spectral distribution of large dimensional random matrices, *Journal of Multivariate Analysis* 54, 295–309.

Song, Z., and D. Xiu, 2012, A tale of two option markets: State-price densities and volatility risk, *Working paper* .

Stock, J., and M. Watson, 2002a, Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economic Statistics* 20, 147–162.

Stock, J. H., and M. W. Watson, 2002b, Forecasting using principal components from a large number of predictors, *Journal of American Statistical Association* 97, 1167–1179.

Stricker, C., 1990, Arbitrage et lois de martingale, *Ann. Inst. Henri Poincaré* 26, 451–460.

Subrahmanyam, A., 2010, The cross section of expected stock returns: What have we learnt from the past twenty-five years of research?, *European Financial Management* 1, 27–42.

Sundaresan, S., and Z. Wang, 2014a, Bank liability structure, *Working paper* .

Sundaresan, S., and Z. Wang, 2014b, Design of contingent capital with a stock price trigger for conversion, *Journal of Finance* .

Tao, M., Y. Wang, and X. Chen, 2013a, Fast convergence rates in estimating large volatility matrices using high-frequency financial data, *Econometric Theory* 29, 838–856.

Tao, M., Y. Wang, and H. H. Zhou, 2013b, Optimal sparse volatility matrix estimation for high dimensional Itô processes with measurement errors, *Annals of Statistics* 41, 1816–1864.

Tauchen, G.E., and H. Zhou, 2011, Realized jumps on financial markets and predicting credit spreads, *Journal of Econometrics* 160, 102–118.

Todorov, V., 2009, Estimation of continuous-time stochastic volatility models with jumps using high-frequency data, *Journal of Econometrics* 148, 131–148.

Todorov, V., 2010, Variance risk premium dynamics: The role of jumps, *Review of Financial Studies* 23, 345–383.

Todorov, V., 2011, Econometric analysis of jump-driven stochastic volatility models, *Journal of Financial Econometrics* 160, 12–21.

Wang, D. C., and P. A. Mykland, 2014, The estimation of leverage effect with high frequency data, *Journal of the American Statistical Association* 109, 197–215.

Wang, Y., and J. Zhou, 2010, Vast volatility matrix estimation for high-frequency financial data, *Annals of Statistics* 38, 943–978.

Xiu, D., and I. Kalnina, 2014, Nonparametric estimation of the leverage effect using information from derivatives markets, *Working paper* .

Yu, J., 2005, On leverage in a stochastic volatility model, *Journal of Econometrics* 127, 165–178.

Zhang, L., 2011, Estimating covariation: Epps effect, microstructure noise, *Journal of Econometrics* 160, 33–47.

Appendix A

Appendix to Chapter 1

A.1 Structure of Appendix

The appendix of Chapter 1 is structured as follows. Appendix A.2 specifies the class of stochastic processes used in this paper. In Appendix A.3 I collect some intermediate asymptotic results, which will be used in the subsequent proofs. Appendix A.4 proves the results for the loading estimator. Appendix A.5 treats the estimation of the factors. In Appendix A.6 I show the results for the common components. In Appendix A.7 I derive consistent estimators for the covariance matrices of the estimators. Appendix A.8 deals with separating the continuous and jump factors. The estimation of the number of factors is in Appendix A.9. Appendix A.10 proves the test for identifying the factors. Last but not least I discuss the proofs for microstructure noise in Appendix A.11. Finally, for convenience Appendix A.12 contains a collection of limit theorems. In the proofs C is a generic constant that may vary from line to line.

A.2 Assumptions on Stochastic Processes

Definition A.1. ***Locally bounded special Itô semimartingales***
The stochastic process Y is a locally bounded special Itô semimartingale if it satisfies the following conditions. Y is a d-dimensional special Itô semimartingale on some filtered space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$, which means it can be written as

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_E \delta(s,x)(\mu - \nu)(ds, dx)$$

where W is a d-dimensional Brownian motion and μ is a Poisson random measure on $\mathbb{R}_+ \times E$ with $(E, \mathbb{E})$ an auxiliary measurable space on the space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$. The predictable compensator (or intensity measure) of μ is $\nu(ds, dx) = ds \times v(dx)$ for some given finite or sigma-finite measure on $(E, \mathbb{E})$. This definition is the same as for an Itô semimartingale with the additional assumption that $\| \int_0^t \int_E \|\delta(s,x)\| \mathbb{1}_{\{\|\delta\|>1\}} \nu(ds, dx)\| < \infty$ for all t. Special *semimartingales have a unique decomposition into a predictable finite variation part and a local martingale part.*

The coefficients $b_t(\omega), \sigma_t(\omega)$ *and* $\delta(\omega, t, x)$ *are such that the various integrals make sense (see Jacod and Protter (2012) for a precise definition) and in particular* b_t *and* σ_t *are optional processes and* δ *is a predictable function.*

The volatility σ_t *is also a d-dimensional Itô semimartingale of the form*

$$\begin{aligned}\sigma_t =&\sigma_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \tilde{\sigma}'_s dW'_s + \int_0^t \int_E \mathbb{1}_{\{\|\tilde{\delta}\| \leq 1\}} \tilde{\delta}(s,x)(\mu - \nu)(ds, dx) \\ &+ \int_0^t \int_E \mathbb{1}_{\{\|\tilde{\delta}\| > 1\}} \tilde{\delta}(s,x)\mu(ds, dx)\end{aligned}$$

where W' *is another Wiener process independent of* (W, μ). *Denote the predictable quadratic covariation process of the martingale part by* $\int_0^t a_s ds$ *and the compensator of* $\int_0^t \int_E \mathbb{1}_{\{\|\tilde{\delta}\| > 1\}} \tilde{\delta}(s,x)\mu(ds, dx)$ *by* $\int_0^t \tilde{a}_s ds$.

1. *I assume a local boundedness condition holds for* Y*:*
   - *The process* b *is locally bounded and càdlàg.*
   - *The process* σ *is càdlàg.*
   - *There is a localizing sequence* τ_n *of stopping times and, for each* n*, a deterministic nonnegative function* Γ_n *on* E *satisfying* $\int \Gamma_n(z)^2 v(dz) < \infty$ *and such that* $\|\delta(\omega, t, z)\| \wedge 1 \leq \Gamma_n(z)$ *for all* (ω, t, z) *with* $t \leq \tau_n(\omega)$.
2. *The volatility process also satisfy a local boundedness condition:*
   - *The processes* $\tilde{b}$ a *and* $\tilde{a}$ *are locally bounded and progressively measurable*
   - *The processes* $\tilde{\sigma}$ *and* $\tilde{b}$ *are càdlàg or càglàd*
3. *Furthermore both processes* $\sigma\sigma^\top$ *and* $\sigma_{t-}\sigma_{t-}^\top$ *take their values in the set of all symmetric positive definite* $d \times d$ *matrices.*

Remark A.1. ***Interpretation of Definition A.1****:*
Definition A.1 accommodates almost all models for stochastic volatility, including those with jumps and allows for correlation between the volatility and asset price processes. Condition (1) is very mild and for example

$$X_t = X_0 + \int_0^t H_s dZ_s$$

where Z *is a multidimensional Lévy process and* H *is a predictable and locally bounded process automatically satisfies condition (1). Condition (2) is stronger, but nevertheless very often* *satisfied. For example when X is the solution (weak or strong, when it exists) of a stochastic differential equation of the form*

$$X_t = X_0 + \int_0^t f(s, X_{s-}) dZ_s$$

with Z again a Lévy process and f is a $C^{1,2}$ function on $\mathbb{R}_+ \times \mathbb{R}^N$.

The nondegeneracy condition (3) says that almost surely the continuous martingale part of X is not identically 0 on any interval $[0, t]$. Most results will hold without (3), but for almost any application this condition is satisfied.

The assumption of a special semimartingale essentially requires that the process has a finite first moment. For any practical purpose in finance this assumption is satisfied.

The local boundedness assumptions allow us to apply a localization procedure. There exist stopping times such that the stopped processes are bounded. The derivation of most of my results requires the processes to be bounded. However all relevant results hold under stopping, which means a local boundedness condition is sufficient to treat the processes as bounded in the proofs. For more details see Theorem A.1.

Examples of processes satisfying Definition A.1

1. **CIR model:** It can be shown that the Cox-Ingersoll-Ross model satisfies the assumption. It is defined as

$$dX_t = a(b - X_t)dt + \sigma\sqrt{X_t}dW_t$$

with $2ab > \sigma^2$ and $X_0 > 0$.

2. **Heston model:** Also the Heston model satisfies Definition A.1:

$$\begin{aligned} dX_t &= cX_t dt + \sqrt{\sigma_t^2} X_t dW_t \\ d\sigma_t^2 &= a(b - \sigma_t^2)dt + \tilde{\sigma}\sqrt{\sigma_t^2}d\tilde{W}_t \\ dW_t d\tilde{W}_t &= \rho dt \end{aligned}$$

with $2ab > \tilde{\sigma}^2$.

3. **Barndorff-Nielsen and Shephard Ornstein-Uhlenbeck stochastic volatility model:**
In the Barndorff-Nielsen and Shephard (2002) model the stochastic volatility follows a positive Ornstein-Uhlenbeck process:

$$\begin{aligned} X_t &= X_0 \exp(Y_t) \\ dY_t &= (a + \beta\sigma_t^2)dt + \sigma_t dW_t + \rho dZ_t \\ d\sigma_t^2 &= -b\sigma_t^2 dt + dZ_t\,, \qquad \sigma_0^2 > 0 \end{aligned}$$

where $\rho \leq 0$, $b > 0$ and Z is a Lévy process.

4. **Itô semimartingale models for price data:** Assume that Y satisfies Definition A.1. Then we can model asset prices as:

$$X_t = X_0 \exp(Y_t)$$

5. **Andersen, Benzoni and Lund (2002)**: Stochastic volatility with log-normal jumps generated by a non-homogeneous Poisson process.

6. **Affine models** as defined in Duffie, Pan and Singleton (2000).

A.3 Some Intermediate Asymptotic Results

A.3.1 Convergence Rate Results

Proposition A.1. *Assume Y is a d-dimensional Itô-semimartingale satisfying Definition A.1:*

$$Y_t = Y_0 + \int_0^t b_Y(s)ds + \int_0^t \sigma_Y(s)dW_Y(s) + \int_0^t \delta_Y \star (\mu - \nu)_t$$

Assume further that Y is square integrable. Assume $\bar{Z}_N = \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i$, where each Z_i is a local Itô-martingale satisfying Definition A.1:

$$Z_i(t) = \int_0^t \sigma_{Z_i}(s)dW_i(s) + \delta_{Z_i} \star (\mu_{Z_i} - \nu_{Z_i})_t$$

and each Z_i is square integrable. Assume that $[\bar{Z}_N, \bar{Z}_N]_T$ and $\langle \bar{Z}_N, \bar{Z}_N \rangle_T$ are bounded for all N. Divide the interval $[0, T]$ into M subintervals. Assume further that Y is either independent of Z_N or a square integrable martingale.

Then, it holds that

$$\sqrt{M} \left(\sum_{j=1}^M \Delta_j Y \Delta_j Z_N - [Y, Z_N]_T \right) = O_p(1)$$

Proof. **Step 1: Localization**
Using Theorem A.1 and following the same reasoning as in Section 4.4.1 of Jacod (2012), we can replace the local boundedness conditions with a bound on the whole time interval. I.e. without loss of generality, we can assume that there exists a constant C and a non-negative function Γ such that

$$\|\sigma_{Z_i}\| \leq C, \quad \|Z_i(t)\| \leq C, \quad \|\delta_{Z_i}\|^2 \leq \Gamma, \quad \int \Gamma(z)\nu_{Z_i}(dz) \leq C$$
$$\|\sigma_Y\| \leq C, \quad \|Y(t)\| \leq C, \quad \|\delta_Y\|^2 \leq \Gamma, \quad \int \Gamma(z)\nu_Y(dz) \leq C$$
$$\|b_Y\| \leq C$$

σ_{Z_N}, $\delta_{\bar{Z}_N}$ and $\nu_{\bar{Z}_N}$ are defined by

$$\langle \bar{Z}_N \rangle_t = \int_0^t \left(\sigma^2_{\bar{Z}_N}(s) + \int \delta^2_{\bar{Z}_N}(z,s)\nu_{\bar{Z}_N}(dz) \right) ds$$

Given our assumptions, we can use wlog that

$$\|\sigma_{\bar{Z}_N}\| \leq C, \qquad \|\bar{Z}_N(t)\| \leq C, \qquad \|\delta^2_{\bar{Z}_N}\| \leq \Gamma, \qquad \int \Gamma(z)\nu_{\bar{Z}_N}(dz) \leq C$$

Step 2: Bounds on increments
Denote the time increments by $\Delta_M = T/M$. Lemmas A.32, A.33 and A.34 together with the bounds on the characteristics of Y and Z_N imply that

$$\begin{aligned}\mathbb{E}\left[\sup_{0\leq s\leq \Delta_M} \|Y_{t+s} - Y_t\|^2\right] \leq & C\Delta_M \mathbb{E}\left[\int_t^{t+\Delta_M} \|b_Y(s)\|^2 ds\right] + C\mathbb{E}\left[\int_t^{t+\Delta_M} \|\sigma_Y(s)\|^2 ds\right] \\ & + C\mathbb{E}\left[\int_t^{t+\Delta_M}\int \|\delta_Y(s,z)\|^2 \nu_Y(dz)ds\right] \leq \frac{C}{M}\end{aligned}$$

and similarly

$$\mathbb{E}\left[\sup_{0\leq s\leq \Delta_M} \|\bar{Z}_N(s+t) - \bar{Z}_N(t)\|^2\right] \leq \frac{C}{M}$$

Step 3: Joint convergence
Define $G_{MN} = \sqrt{M}\left(\sum_{j=1}^M \Delta_j Y \Delta_j \bar{Z}_N - [Y, \bar{Z}_N]_T\right)$. We need to show, that $\forall \epsilon > 0$ there exists an n and a finite constant C such that

$$\mathbb{P}\left(\|G_{MN}\| > C\right) \leq \epsilon \qquad \forall M, N > n$$

By Markov' s inequality, if $\mathbb{E}\left[\|G_{MN}\|^2\right] < \infty$

$$\mathbb{P}\left(\|G_{MN}\| > C\right) \leq \frac{1}{C^2}\mathbb{E}\left[\|G_{MN}\|^2\right]$$

Hence it remains to show that $\mathbb{E}\left[\|G_{MN}\|^2\right] < \infty$ for $M, N \to \infty$.

Step 4: Bounds on sum of squared increments
By Itô's lemma, we have on each subinterval

$$\Delta_j Y \Delta_j \bar{Z}_N - \Delta_j[Y, \bar{Z}_N] = \int_{t_j}^{t_{j+1}} (Y(s) - Y(t_j))d\bar{Z}_N(s) + \int_{t_j}^{t_{j+1}} (\bar{Z}_N(s) - \bar{Z}_N(t_j))dY(s)$$

As $\bar{Z}_N$ is square integrable and a local martingale, it is a martingale. By assumption Y is either independent of $\bar{Z}_N$ or a martingale as well. In the first case it holds that

$$\mathbb{E}\left[\Delta_j Y \Delta_j \bar{Z}_N - \Delta_j[Y, \bar{Z}_N] | \mathfrak{F}_{t_j}\right] = \mathbb{E}\left[\Delta_j Y | \mathfrak{F}_{t_j}\right]\mathbb{E}\left[\Delta_j \bar{Z}_N | \mathfrak{F}_{t_j}\right] = 0$$

In the second case both stochastic integrals $\int_0^t Y(s)d\bar{Z}_N(s)$ and $\int_0^t \bar{Z}_N(s)dY(s)$ are martingales. Hence in either case, $\Delta_j Y \Delta_j \bar{Z}_N - \Delta_j[Y, Z_N]$ forms a sequence of martingale differences and we can apply Burkholder's inequality for discrete time martingales (Lemma A.30):

$$\begin{aligned}
\mathbb{E}\left[\|G_{MN}\|^2\right] \leq & M\sum_{j=1}^{M}\mathbb{E}\left[\|\Delta_j Y\Delta_j\bar{Z}_N - \Delta_j[Y,\bar{Z}_N]\|^2\right] \\
\leq & M\sum_{j=1}^{M}\mathbb{E}\left[\|\int_{t_j}^{t_{j+1}}(Y(s)-Y(t_j))d\bar{Z}_N(s) + \int_{t_j}^{t_{j+1}}(\bar{Z}_N(s)-\bar{Z}_N(t_j))dY(s)\|^2\right] \\
\leq & M\sum_{j=1}^{M}\mathbb{E}\left[\|\int_{t_j}^{t_{j+1}}(Y(s)-Y(t_j))d\bar{Z}_N(s)\|^2\right] \\
& + M\sum_{j=1}^{M}\mathbb{E}\left[\|\int_{t_j}^{t_{j+1}}(\bar{Z}_N(s)-\bar{Z}_N(t_j))dY(s)\|^2\right]
\end{aligned}$$

It is sufficient to show that $\mathbb{E}\left[\|\int_{t_j}^{t_{j+1}}(Y(s)-Y(t_j))d\bar{Z}_N\|^2\right] = \frac{C}{M^2}$ and $\mathbb{E}\left[\|\int_{t_j}^{t_{j+1}}(\bar{Z}_N(s)-\bar{Z}_N(t_j))dY\|^2\right] = \frac{C}{M^2}$. By Lemma A.31 and step 1 and 2:

$$\begin{aligned}
E\left[\|\int_{t_j}^{t_{j+1}}(Y(t)-Y(t_j))d\bar{Z}_N\|^2\right] & \leq \mathbb{E}\left[\int_{t_j}^{t_{j+1}}\|Y(t)-Y(t_j)\|^2 d\langle\bar{Z}_N\rangle\right] \\
& \leq \mathbb{E}\left[\int_0^T \|Y(t)-Y(t_j)\|^2\left(\sigma^2_{\bar{Z}_N}(t) + \int\delta^2_{\bar{Z}_N}(z,t)\nu_{\bar{Z}_N}(z)\right)dt\right] \\
& \leq C\mathbb{E}\left[\int_{t_j}^{t_{j+1}}\|Y(t)-Y(t_j)\|^2 dt\right] \\
& \leq C\mathbb{E}\left[\sup_{t_j\leq t\leq t_{j+1}}\|Y(t)-Y(t_j)\|^2\right]\frac{1}{M} \\
& \leq \frac{C}{M^2}.
\end{aligned}$$

Similarly using Lemma A.32 for the drift of Y and A.31 for the martingale part, we can bound the second integral:

$$
\begin{aligned}
\mathbb{E}\left[\| \int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j)) dY \|^2\right] \leq & \mathbb{E}\left[\| \int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j)) b_Y dt \|^2\right] \\
& + \mathbb{E}\left[\| \int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - Z_N(t_j)) \left(\sigma_Y dW_Y + \delta_Y d(\mu - \nu)\right) \|^2\right] \\
\leq & \frac{1}{M} C \mathbb{E}\left[\int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j))^2 \|b_Y(t)\|^2 dt\right] \\
& + CE\Bigg[\int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j))^2 \\
& \cdot \left(\|\sigma_Y(t)\|^2 + \int \|\delta_Y\|^2(z,t)\nu_Y(z)\right) dt\Bigg] \\
\leq & \frac{1}{M} C \mathbb{E}\left[\int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j))^2 dt\right] \\
& + C\mathbb{E}\left[\int_{t_j}^{t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j))^2 (t) dt\right] \\
\leq & C\mathbb{E}\left[\sup_{t_j \leq t \leq t_{j+1}} (\bar{Z}_N(t) - \bar{Z}_N(t_j))^2\right] \frac{1}{M} \\
\leq & \frac{C}{M^2}
\end{aligned}
$$

Putting things together, we obtain:

$$
\mathbb{E}\left[\|G_{MN}\|^2\right] \leq M \sum_{j=1}^{M} \frac{C}{M^2} \leq C
$$

which proves the statement. □

Lemma A.1. *Assumption 1.1 holds. Then*

$$
\frac{1}{N} F e \Lambda = O_p\left(\frac{1}{\sqrt{MN}}\right)
$$

Proof. Apply Proposition A.1 with $Y = F$ and $\bar{Z}_N = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \Lambda_k e_k$. □

Lemma A.2. *Assumption 1.1 holds. Then*

$$
\frac{1}{N} \sum_{k=1}^{N} \left(\sum_{j=1}^{M} e_{ji} e_{jk} - [e_i, e_k]\right) \Lambda_k = O_p\left(\frac{1}{\sqrt{MN}}\right)
$$

Proof. Apply Proposition A.1 with $Y = e_i$ and $\bar{Z}_N = \frac{1}{\sqrt{N}} \sum_{k=1}^N \Lambda_k e_k$. □

Lemma A.3. *Assume Assumption 1.1 holds. Then*

$$\frac{1}{N}\sum_{i=1}^N \Lambda_i e_i(T) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

Proof. By Burkholder's inequality in Lemma A.31 we can bound

$$\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^N \Lambda_i e_i(T)\right)^2\right] \leq \mathbb{E}\left[\frac{1}{N^2}\Lambda^\top \langle e, e\rangle \Lambda\right] \leq \frac{C}{N}$$

based on Assumption 1.1. □

Lemma A.4. *Assume Assumption 1.1 holds. Then*

$$\sum_{j=1}^M e_{ji}e_{jk} - [e_i, e_k]_T = O_p\left(\frac{1}{\sqrt{M}}\right)$$

Proof. Apply Theorem A.2. □

Proof of Lemma 1.1:

Proof. If e_i has independent increments it trivially satisfies weak serial dependence. The harder part is to show that the second and third condition imply weak cross-sectional dependence. We need to show

$$\sum_{j=1}^M \frac{1}{N^2}\sum_{i=1}^N \sum_{r=1}^N \mathbb{E}\left[e_{j,i}^2 e_{j,r}^2\right] = O\left(\frac{1}{\delta}\right)$$

Step 1: Decompose the residuals into their continuous and jump component respectively:

$$\begin{aligned}
&\sum_{j=1}^M \frac{1}{N^2}\sum_{i=1}^N \sum_{r=1}^N \mathbb{E}\left[\left(e_{j,i}^C + e_{j,i}^D\right)^2\left(e_{j,r}^C + e_{j,r}^D\right)^2\right] \\
\leq & C\sum_{j=1}^M \frac{1}{N^2}\sum_{i=1}^N \sum_{r=1}^N \Bigg(\mathbb{E}\left[{e_{j,i}^C}^2 {e_{j,r}^C}^2\right] + \mathbb{E}\left[{e_{j,i}^D}^2 {e_{j,r}^D}^2\right] + \mathbb{E}\left[{e_{j,i}^C}^2 {e_{j,r}^D}^2\right] \\
& + \mathbb{E}\left[e_{j,i}^C e_{j,i}^D {e_{j,r}^C}^2\right] + \mathbb{E}\left[e_{j,i}^C e_{j,i}^D {e_{j,i}^D}^2\right] + \mathbb{E}\left[e_{j,i}^C e_{j,i}^D e_{j,r}^C e_{j,r}^D\right]\Bigg).
\end{aligned}$$

Step 2: To show: $\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[{e_{j,i}^{C}}^2 {e_{j,r}^{C}}^2 \right] = O_p\left(\frac{1}{\delta}\right)$
This is a a consequence the Cauchy-Schwartz inequality and Burkholder's inequality in Lemma A.31:

$$\mathbb{E}\left[{e_{j,i}^{C}}^2 {e_{j,r}^{C}}^2 \right] \leq C \mathbb{E}\left[{e_{j,i}^{C}}^4 \right]^{1/2} \mathbb{E}\left[{e_{j,r}^{C}}^4 \right]^{1/2} \leq \frac{C}{M^2}$$

Step 3: To show: $\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[{e_{j,i}^{D}}^2 {e_{j,r}^{D}}^2 \right] = O_p\left(\frac{1}{\delta}\right)$

$$\begin{aligned} &\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[{e_{j,i}^{D}}^2 {e_{j,r}^{D}}^2 \right] \leq \max_{j,r} |{e_{j,r}^{D}}^2| \cdot \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbb{E}\left[{e_{j,i}^{D}}^2 \right] \\ \leq & C \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbb{E}\left[\Delta_j \langle e_i^D, e_i^D \rangle \right] \leq \frac{C}{N} \mathbb{E}\left[\sum_{i=1}^{N} \langle e_i^D, e_i^D \rangle \right] \leq O\left(\frac{1}{\delta}\right) \end{aligned}$$

where we have used the second and third condition.
Step 4: To show: $\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[e_{j,i}^{C} e_{j,i}^{D} e_{j,r}^{C} e_{j,r}^{D} \right] = O_p\left(\frac{1}{\delta}\right)$

$$\begin{aligned} \sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[e_{j,i}^{C} e_{j,i}^{D} e_{j,r}^{C} e_{j,r}^{D} \right] &\leq \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[\sum_{j=1}^{M} |e_{j,i}^{D}||e_{j,r}^{D}| \sup_{j,i,r} \left(|e_{j,i}^{C}||e_{j,r}^{C}| \right) \right] \\ &\leq C \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[\left(\sum_{j=1}^{M} {e_{j,i}^{D}}^2 \right)^{1/2} \left(\sum_{j=1}^{M} {e_{j,r}^{D}}^2 \right)^{1/2} \sup_{j,i} ({e_{j,i}^{C}}^2) \right] \\ &\leq C \mathbb{E}\left[\sup_{j,i} ({e_{j,i}^{C}}^2) \right] \leq \frac{C}{M}. \end{aligned}$$

Step 5: The other moments can be treated similarly as in step 2 to 4. □

Proposition A.2. ***Consequence of weak dependence***
Assume Assumption 1.1 holds. If additionally Assumption 1.5, i.e. weak serial dependence and weak cross-sectional dependence, holds then we have:

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} F_j e_{ji} \sum_{l=1}^{M} e_{li} = O_p\left(\frac{1}{\delta}\right)$$

Proof. By the localization procedure in Theorem A.1, we can assume without loss of generality that there exists a constant C such that

$$\begin{aligned} &\|b_F(t)\| \leq C \qquad \|\sigma_F(t)\| \leq C \qquad \|F(t)\| \leq C \qquad \|\delta_F(t,z)\|^2 \leq \Gamma(z) \qquad \int \Gamma(z) v_F(dz) \leq C \\ &\qquad\qquad\qquad \|\sigma_{e_i}(t)\| \leq C \qquad \|e_i(t)\| \leq C \qquad \|\delta_{e_i}(t,z)\|^2 \leq \Gamma(z) \qquad \int \Gamma(z) v_{e_i}(dz) \leq C \end{aligned}$$

We want to show

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}F_j e_{ji}e_i(T)=O_p\left(\frac{1}{\delta}\right)$$

where $e_i(T)=\sum_{l=1}^{M}e_{li}$. I proceed in several steps: First, I define

$$\tilde{Z}=\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(F_j e_{ji}e_i(T)-\mathbb{E}_j\left[b_j^F\Delta_j\langle e_i,e_i\rangle\right]\right)$$

with the notation $\mathbb{E}_j[.]=\mathbb{E}[.|\mathfrak{F}_{t_j}]$ as the conditional expectation and $b_j^F=\int_{t_j}^{t_{j+1}}b^F(s)ds$ as the increment of the drift term of F. The proof relies on the repeated use of different Burkholder inequalities, in particular that $b_j^F=O_p\left(\frac{1}{M}\right),\Delta_j\langle e_i,e_i\rangle=O_p\left(\frac{1}{M}\right)$ and $\mathbb{E}[F_j]=O_p\left(\frac{1}{\sqrt{M}}\right)$.

Step 1: To show $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathbb{E}_j\left[b_j^F\Delta_j\langle e_i,e_i\rangle\right]=O_p\left(\frac{1}{\delta}\right)$

$$\left|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathbb{E}_j\left[b_j^F\Delta_j\langle e_i,e_i\rangle\right]\right|\leq\sup|\mathbb{E}_j[b_j^F]|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}|\mathbb{E}_j\left[\Delta_j\langle e_i,e_i\rangle\right]|\leq O_p\left(\frac{1}{M}\right)O_p(1)$$

Step 2: To show: $\tilde{Z}=O_p\left(\frac{1}{\delta}\right)$

Note that by the independence assumption between F and e, the summands in $\tilde{Z}$ follow a martingale difference sequence. Thus, by Burkholder's inequality for discrete time martingales:

$$\begin{aligned}\mathbb{E}\left[\tilde{Z}^2\right]\leq&C\mathbb{E}\left[\sum_{j=1}^{M}\left(\frac{1}{N}\sum_{i=1}^{N}\left(F_j e_{ji}e_i(T)-\mathbb{E}_j[b_j^F\Delta_j\langle e_i,e_i\rangle]\right)\right)^2\right]\\ \leq&C\mathbb{E}\Bigg[\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}F_j^2 e_{ji}e_{jr}e_i(T)e_r(T)+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\left(\mathbb{E}_j[b_j^F]^2\mathbb{E}_j[\Delta_j\langle e_i,e_i\rangle]\mathbb{E}_j[\Delta_j\langle e_r,e_r\rangle]\ \right)\\ &-\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\left(F_j e_{ji}e_i(T)\mathbb{E}_j[b_j^F]\mathbb{E}_j[\Delta_j\langle e_r,e_r\rangle]+F_j e_{jr}e_r(T)\mathbb{E}_j[b_j^F]\mathbb{E}_j[\Delta_j\langle e_i,e_i\rangle]\right)\Bigg]\end{aligned}$$

The first term can be written as

$$\begin{aligned}&\mathbb{E}\left[\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}F_j^2 e_{ji}e_{jr}e_i(T)e_r(T)\right]\\ =&\mathbb{E}\left[\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}F_j^2 e_{ji}e_{jr}\sum_{l\neq j}e_{li}\sum_{s\neq j}e_{sr}\right]+\mathbb{E}\left[\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}F_j^2 e_{ji}^2 e_{jr}^2\right]\end{aligned}$$

Under the assumption of weak serial dependence in Assumption 1.5 the first sum is bounded by

$$\begin{aligned}
&\mathbb{E}\left[\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}F_j^2 e_{ji}e_{jr}\sum_{l\neq j}e_{li}\sum_{s\neq j}e_{sr}\right]\\
\leq & C\left(\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}[F_j^2]|\mathbb{E}[e_{ji}e_{jr}]|\left|\mathbb{E}\left[\sum_{l\neq j}e_{li}\sum_{s\neq j}e_{sr}\right]\right|\right)\\
\leq & C\left(\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}[F_j^2]|\mathbb{E}[e_{ji}e_{jr}]|\left|\mathbb{E}\left[\sum_{l\neq j}e_{li}e_{lr}\right]\right|\right)\\
\leq & C\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}|\mathbb{E}[\Delta_j\langle e_i,e_r\rangle]|\\
\leq & C\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}|\mathbb{E}\left[\int_{t_j}^{t_{j+1}}G_{i,r}(s)ds\right]|\\
\leq & C\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N^2}\sum_{r=1}^{N}\mathbb{E}\left[\int_{t_j}^{t_{j+1}}\sum_{i=1}^{N}|G_{i,r}(s)|ds\right]\\
\leq & C\frac{1}{MN}
\end{aligned}$$

Next, we turn to the second sum of the first term:

$$\begin{aligned}
&\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}\left[F_j^2\right]\mathbb{E}\left[e_{ji}^2e_{jr}^2\right]\\
\leq & \frac{C}{M}\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}\left[e_{ji}^2e_{jr}^2\right]\\
\leq & \frac{C}{M\delta}
\end{aligned}$$

In the last line, we have used weak cross-sectional dependence in Assumption 1.5. The third term can be bounded as follows

$$\sum_{j=1}^{M}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\mathbb{E}\left[\mathbb{E}_j[b_j^F]^2\mathbb{E}_j[\Delta_j\langle e_i,e_i\rangle]\mathbb{E}_j[\Delta_j\langle e_r,e_r\rangle]\right]\leq\frac{C}{M^2}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{r=1}^{N}\sum_{j=1}^{M}\frac{C}{M^2}\leq\frac{C}{M^3}$$

The final two terms can be treated the same way:

$$\begin{aligned}
&\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[F_j e_{ji} e_i(T) \mathbb{E}_j \left[b_j^F \Delta_j \langle e_i, e_i \rangle\right]\right] \\
\leq &\sum_{j=1}^{M} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{r=1}^{N} \mathbb{E}\left[F_j \mathbb{E}_j [b_j^F]\right] \mathbb{E}\left[e_{ji} e_i(T) \mathbb{E}_j \left[\Delta_j \langle e_r, e_r \rangle\right]\right] \\
\leq &\sum_{j=1}^{M} \mathbb{E}\left[F_j \mathbb{E}_j [b_j^F]\right] \mathbb{E}\left[\left|\frac{1}{N} \sum_{i=1}^{N} e_{ji} e_i(T)\right| \mathbb{E}_j \left[\frac{1}{N} \sum_{r=1}^{N} \Delta_j \langle e_r, e_r \rangle\right]\right] \\
\leq &\frac{C}{M^{3/2}} \sum_{j=1}^{M} \mathbb{E}\left[\left|\frac{1}{N} \sum_{i=1}^{N} e_{ji} e_i(T)\right|\right] \frac{C}{M} \\
\leq &\frac{C}{M^{3/2}} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[|e_{ji}|\right] \leq \frac{C}{M^2}
\end{aligned}$$

□

Lemma A.5. ***Convergence rate of sum of residual increments:*** *Under Assumptions 1.1 and 1.2 it follows that*

$$\frac{1}{N} \sum_{i=1}^{N} \Lambda_i e_{j,i} = O_p\left(\frac{1}{\delta}\right)$$

Proof. We apply Burkholder's inequality from Lemma A.31 together with Theorem A.1:

$$\mathbb{E}\left[\left(\frac{1}{N} \sum_{i=1}^{N} \Lambda_i e_{j,i}\right)^2\right] \leq C \mathbb{E}\left[\frac{1}{N^2} \Lambda^\top \Delta_j \langle e, e \rangle \Lambda\right] \leq C \mathbb{E}\left[\frac{1}{N^2} \Lambda^\top \int_{t_j}^{t_{j+1}} G(s) ds \Lambda\right] \leq \frac{C}{NM}$$

which implies

$$\frac{1}{N} \sum_{i=1}^{N} \Lambda_i e_{j,i} = O_p\left(\frac{1}{\sqrt{NM}}\right).$$

□

A.3.2 Central Limit Theorems

Lemma A.6. ***Central limit theorem for covariation between F and e_i*** *Assume that Assumptions 1.1 and 1.2 hold. Then*

$$\sqrt{M} \sum_{j=1}^{M} F_j e_{ji} \stackrel{L-s}{\to} N(0, \Gamma_i)$$

where the entry $\{l, g\}$ of the $K \times K$ matrix Γ_i is given by

$$\Gamma_{i,l,g} = \int_0^T \sigma_{F^l,F^g} \sigma_{e_i}^2 ds + \sum_{s \le T} \Delta F^l(s) \Delta F^g(s) \sigma_{e_i}^2(s) + \sum_{s' \le T} \Delta e_i^2(s') \sigma_{F^g,F^l}(s')$$

F^l denotes the l-th component of the the K dimensional process F and σ_{F^l,F^g} are the entries of its $K \times K$ dimensional volatility matrix.

Proof. Apply Theorem A.2 using that independence of F and e_i implies $[F, e_i] = 0$. □

Lemma A.7. ***Martingale central limit theorem with stable convergence to Gaussian martingale***
Assume $Z^n(t)$ is a sequence of local square integrable martingales and Z is a Gaussian martingale with quadratic characteristic $\langle Z, Z \rangle$. Assume that for any $t > 0$

1. $\int_0^t \int_{|z|>\epsilon} z^2 \nu^n(ds, dx) \xrightarrow{p} 0 \; \forall \epsilon \in (0, 1]$

2. $[Z^n, Z^n]_t \xrightarrow{p} [Z, Z]_t$

Then $Z^n \xrightarrow{L-s} Z$.

Proof. The convergence in distribution follows immediately from Lemma A.29. In order to show the stable weak convergence in Theorem A.4, we need to show that the nesting condition for the filtration holds. We construct a triangular array sequence $X^n(t) = Z^n([tk_n])$ for $0 \le t \le 1$ and some $k_n \to \infty$. The sequence of histories is $\mathfrak{F}_t^n = \mathfrak{H}_{[tk_n]}^n; 0 \le t \le 1$, where $\mathfrak{H}^n$ is the history of Z^n. Now, $t_n = \frac{1}{\sqrt{k_n}}$ is a sequence that satisfies the nesting condition. □

Lemma A.8. ***Martingale central limit theorem for sum or residuals***
Assume that Assumption 1.1 is satisfied and hence, in particular $e_i(t)$ are square integrable martingales. Define $Z_N = \frac{1}{\sqrt{N}} \sum_{i=1}^N \Lambda_i e(t)$. Assume that for any $t > 0$

1. $\frac{1}{N} \Lambda^\top \langle e, e \rangle_t^D \Lambda \xrightarrow{p} 0$

2. $\frac{1}{N} \Lambda^\top [e, e]_t^D \Lambda \xrightarrow{p} 0$

3. $\frac{1}{N} \Lambda^\top [e, e]_t \Lambda \xrightarrow{p} \Phi_t$

Then, conditioned on its quadratic covariation Z_N converges stably in law to a normal distribution.

$$Z_N \xrightarrow{L-s} N(0, \Phi_t).$$

Proof. By Lemma A.7 $Z_N \xrightarrow{L-s} Z$, where Z is a Gaussian process with $\langle Z, Z \rangle_t = \Phi_t$. Conditioned on its quadratic variation, the stochastic process evaluated at time t has a normal distribution. □

A.4 Estimation of the Loadings

Lemma A.9. ***A decomposition of the loadings estimator***
Let V_{MN} be the $K \times K$ matrix of the first K largest eigenvalues of $\frac{1}{N}X^\top X$. Define $H = \frac{1}{N}\left(F^\top F\right)\Lambda^\top\hat{\Lambda}V_{MN}^{-1}$. Then we have the decomposition

$$V_{MN}\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) = \frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k[e_i, e_k]_T + \frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\phi_{ki} + \frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\eta_{ki} + \frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{ki}$$

with

$$\phi_{ki} = \sum_{j=1}^{M} e_{ji}e_{jk} - [e_i, e_k]_T$$

$$\eta_{ki} = \Lambda_k^\top \sum_{j=1}^{M} F_j e_{ji}$$

$$\xi_{ki} = \Lambda_i^\top \sum_{j=1}^{M} F_j e_{jk}$$

Proof. This is essentially the identity in the proof of Theorem 1 in Bai and Ng (2002). From

$$\left(\frac{1}{N}X^\top X\right)\hat{\Lambda} = \hat{\Lambda}V_{MN}$$

it follws that $\frac{1}{N}X^\top X\hat{\Lambda}V_{MN}^{-1} = \hat{\Lambda}$. Substituting the definition of X, we obtain

$$\left(\hat{\Lambda} - \Lambda H\right)V_{MN} = \frac{1}{N}e^\top e\hat{\Lambda} + \frac{1}{N}\Lambda F^\top F\Lambda^\top\hat{\Lambda} + \frac{1}{N}e^\top F\Lambda^\top\hat{\Lambda} + \frac{1}{N}\Lambda F^\top e\hat{\Lambda} - \Lambda H V_{MN}$$

H is chosen to set

$$\frac{1}{N}\Lambda F^\top F\Lambda^\top\hat{\Lambda} - \Lambda H V_{MN} = 0.$$

□

Lemma A.10. ***Mean square convergence of loadings estimator*** *Assume Assumption 1.1 holds. Then*

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - H^\top\Lambda_i\|^2 = O_p\left(\frac{1}{\delta}\right).$$

Proof. This is essentially Theorem 1 in Bai and Ng (2002) reformulated for the quadratic variation and the proof is very similar. In Lemma A.12 it is shown that $\|V_{MN}\| = O_p(1)$. As $(a+b+c+d)^2 \leq 4(a^2+b^2+c^2+d^2)$, we have $\|\hat{\Lambda}_i - \Lambda_i H\|^2 \leq (a_i + b_i + c_i + d_i) \cdot O_p(1)$ with

$$\begin{aligned} a_i &= \frac{1}{N^2} \| \sum_{k=1}^N \hat{\Lambda}_k [e_k, e_i] \|^2 \\ b_i &= \frac{1}{N^2} \| \sum_{k=1}^N \hat{\Lambda}_k \phi_{ki} \|^2 \\ c_i &= \frac{1}{N^2} \| \sum_{k=1}^N \hat{\Lambda}_k \eta_{ki} \|^2 \\ d_i &= \frac{1}{N^2} \| \sum_{k=1}^N \hat{\Lambda}_k \xi_{kI} \|^2 \end{aligned}$$

Step 1: To show: $\frac{1}{N}\sum_{i=1}^N a_i = O_p\left(\frac{1}{N}\right)$

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^N a_i &\leq \frac{1}{N}\sum_{i=1}^N \left(\frac{1}{N^2} \| \sum_{k=1}^N \hat{\Lambda}_k [e_k, e_i] \|^2 \right) \\ &\leq \frac{1}{N} \left(\frac{1}{N} \sum_{k=1}^N \|\hat{\Lambda}_k\|^2 \right) \left(\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^N [e_k, e_i]_T^2 \right) \\ &= O_p\left(\frac{1}{N}\right) \end{aligned}$$

The first term is $\frac{1}{N}\sum_{i=1}^N \|\hat{\Lambda}_k\|^2 = O_p(1)$. The second term can be bounded by using the norm equivalence between the Frobenius and the spectral norm. Note that $\sum_{i=1}^N \sum_{k=1}^N [e_k, e_i]_T^2$ is simply the squared Frobenius norm of the matrix $[e, e]$. It is well-known that any $N \times N$ matrix A with rank N satisfies $\|A\|_F \leq \sqrt{N}\|A\|_2$. Therefore

$$\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^N [e_k, e_i]_T^2 \leq \|[e, e]\|_2^2 = O_p(1).$$

Step 2: To show: $\frac{1}{N}\sum_{i=1}^N b_i = O_p\left(\frac{1}{M}\right)$

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} b_i &\leq \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N^2}\|\sum_{k=1}^{N}\hat{\Lambda}_k\phi_{ki}\|^2\right) \\
&\leq \frac{1}{N}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\hat{\Lambda}_k^\top\hat{\Lambda}_l\phi_{ki}\phi_{li} \\
&\leq \frac{1}{N}\left(\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left(\hat{\Lambda}_k^\top\hat{\Lambda}_l\right)^2\right)^{1/2}\left(\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left(\sum_{i=1}^{N}\phi_{ki}\phi_{li}\right)^2\right)^{1/2} \\
&\leq \frac{1}{N}\left(\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\hat{\Lambda}_k^\top\hat{\Lambda}_l\right)^{1/2}\left(\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left(\sum_{i=1}^{N}\phi_{ki}\phi_{li}\right)^2\right)^{1/2}
\end{aligned}
$$

The second term is bounded by

$$
\left(\sum_{i=1}^{N}\phi_{ki}\phi_{li}\right)^2 \leq N^2\max_{k,l}\phi_{kl}^4
$$

As $\phi_{kl}^4 = \left(\sum_{j=1}^{M} e_{jk}e_{jl} - [e_k, e_l]\right)^4 = O_p\left(\frac{1}{M^2}\right)$, we conclude

$$
\frac{1}{N}\sum_{i=1}^{N} b_i \leq \frac{1}{N}O_p\left(\frac{N}{M}\right) = O_p\left(\frac{1}{M}\right)
$$

Step 3: To show: $\frac{1}{N}\sum_{i=1}^{N} c_i = O_p\left(\frac{1}{M}\right)$

$$
\begin{aligned}
\frac{1}{N^3}\sum_{i=1}^{N}\|\sum_{k=1}^{N}\hat{\Lambda}_k\eta_{ki}\|^2 &\leq \frac{1}{N}\sum_{i=1}^{N}\|F^\top e_i\|^2\left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k\|^2\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|\Lambda_k\|^2\right) \\
&\leq \frac{1}{N}\left(\sum_{i=1}^{N}\|F^\top e_i\|^2\right)O_p(1) \leq O_p\left(\frac{1}{M}\right)
\end{aligned}
$$

The statement is a consequence of Lemma A.6.
Step 4: To show: $\frac{1}{N}\sum_{i=1}^{N} d_i = O_p\left(\frac{1}{M}\right)$

$$
\begin{aligned}
\frac{1}{N^2}\|\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{ki}\|^2 &= \frac{1}{N^2}\|\sum_{k=1}^{N}\sum_{j=1}^{M}\hat{\Lambda}_k\Lambda_i^\top F_j e_{jk}\|^2 \\
&\leq \|\Lambda_i\|^2\left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k\|^2\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|\sum_{j=1}^{M}F_j e_{jk}\|^2\right)
\end{aligned}
$$

The statement follows again from Lemma A.6.
Step 5: From the previous four steps we conclude

$$\frac{1}{N}\sum_{i=1}^{N}(a_i+b_i+c_i+d_i)=O_p\left(\frac{1}{\delta}\right)$$

□

Lemma A.11. ***Convergence rates for components of loadings estimator***
Under Assumptions 1.1 and 1.2, it follows that

1. $\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k[e_k,e_i]_T=O_p\left(\frac{1}{\sqrt{N\delta}}\right)$
2. $\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\phi_{ki}=O_p\left(\frac{1}{\sqrt{M\delta}}\right)$
3. $\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\eta_{ki}=O_p\left(\frac{1}{\sqrt{\delta}}\right)$
4. $\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{ki}=O_p\left(\frac{1}{\sqrt{M\delta}}\right)$

Proof. This is essentially Lemma A.2 in Bai (2003). The proof follows a similar logic to derive a set of inequalities. The main difference is that we use Lemmas A.1, A.2, A.4 and A.6 for determining the rates.
Proof of (1.):

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k[e_k,e_i]=\frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k-H^\top\Lambda_k\right)[e_k,e_i]+\frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k[e_k,e_i]$$

The second term can be bounded using Assumption 1.2

$$\frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k[e_k,e_i]\leq\max_k\|\Lambda_k\|\|H\|\frac{1}{N}\sum_{k=1}^{N}\|[e_k,e_i]\|=O_p\left(\frac{1}{N}\right)$$

For the first term we use Lemma A.10:

$$\begin{aligned}\left\|\frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k-H^\top\Lambda_k\right)[e_k,e_i]\right\|&\leq\left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k-H^\top\Lambda_k\|^2\right)^{1/2}\frac{1}{\sqrt{N}}\left(\sum_{k=1}^{N}[e_k,e_i]^2\right)^{1/2}\\&=O_p\left(\frac{1}{\sqrt{\delta}}\right)O_p\left(\frac{1}{\sqrt{N}}\right)=O_p\left(\frac{1}{\sqrt{N\delta}}\right)\end{aligned}$$

The local boundedness of every entry of $[e,e]$ and Assumption 1.2 imply that

$$\sum_{k=1}^{N}\|[e_k,e_i]\|^2\leq max_{l=1,\ldots N}\|[e_l,e_i]\|\sum_{k=1}^{N}\|[e_k,e_i]\|=O_p(1)$$

Proof of (2.):

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\phi_{ki} = \frac{1}{N}\sum_{k=1}^{N}\phi_{ki}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right) + \frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k\phi_{ki}$$

Using Lemma A.4 we conclude that the first term is bounded by

$$\left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\sum_{j=1}^{M}e_{ji}e_{jk} - [e_i,e_k]_T\|^2\right)^{1/2} = O_p\left(\frac{1}{\sqrt{\delta}}\right)O_p\left(\frac{1}{\sqrt{M}}\right)$$

The second term is $O_p\left(\frac{1}{\sqrt{M\delta}}\right)$ by Lemma A.4.

Proof of (3.):

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\eta_{ki} = \frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\Lambda_k{}^\top F^\top e_i + \frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k\Lambda_k{}^\top F^\top e_i$$

Applying the Cauchy-Schwartz inequality to the first term yields

$$\begin{aligned}\frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\eta_{ki} &\le \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\eta_{ki}^2\right)^{1/2}\\ &\le O_p\left(\frac{1}{\sqrt{\delta}}\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|\Lambda_k\|^2\|F^\top e_i\|^2\right)^{1/2}\\ &\le O_p\left(\frac{1}{\sqrt{\delta}}\right)\left(\|F^\top e_i\|^2\right)^{1/2} \le O_p\left(\frac{1}{\sqrt{\delta M}}\right).\end{aligned}$$

For the second term we obtain the following bound based on Lemma A.6:

$$\frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k\Lambda_k{}^\top F^\top e_i = H^\top\left(\frac{1}{N}\sum_{k=1}^{N}\Lambda_k\Lambda_k{}^\top\right)\left(F^\top e_i\right) \le O_p\left(\frac{1}{\sqrt{M}}\right)$$

Proof of (4.): We start with the familiar decomposition

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{ki} = \frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\xi_{ki} + \frac{1}{N}\sum_{k=1}^{N}H^\top\Lambda_k\xi_{ki}$$

The first term is bounded by

$$\begin{aligned}\|\frac{1}{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\Lambda_i{}^\top F^\top e_k\| &\le \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|F^\top e_k\|^2\right)^{1/2}\|\Lambda_i\|\\ &\le O_p\left(\frac{1}{\sqrt{\delta}}\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|F^\top e_k\|^2\right)^{1/2}\\ &\le O_p\left(\frac{1}{\sqrt{\delta M}}\right)\end{aligned}$$

The rate of the second term is a direct consequence of Proposition A.1:

$$\frac{1}{N}\sum_{k=1}^{N} H^\top \Lambda_k e_k^\top F \Lambda_i = O_p\left(\frac{1}{\sqrt{MN}}\right)$$

This very last step is also different from the Bai (2003) paper. They essentially impose this last conversion rate as an assumption (Assumption F.2), while I derive explicit conditions for the stochastic processes in Proposition A.1. □

Lemma A.12. ***Limit of*** V_{MN}
Assume Assumptions 1.1 and 1.2 hold. For $M, N \to \infty$*, we have*

$$\frac{1}{N}\hat{\Lambda}^\top \left(\frac{1}{N}X^\top X\right)\hat{\Lambda} = V_{MN} \xrightarrow{p} V$$

and

$$\frac{\hat{\Lambda}^\top \Lambda}{N}\left(F^\top F\right)\frac{\Lambda^\top \hat{\Lambda}}{N} \xrightarrow{p} V$$

where V *is the diagonal matrix of the eigenvalues of* $\Sigma_\Lambda^{1/2^\top}\Sigma_F \Sigma_\Lambda^{1/2}$

Proof. See Lemma A.3 in Bai's (2003) and the paper by Stock and Watson (2002). □

Lemma A.13. ***The matrix*** Q
Under Assumptions 1.1 and 1.2

$$plim_{M,N\to\infty}\frac{\hat{\Lambda}^\top \Lambda}{N} = Q$$

where the invertible matrix Q *is given by* $V^{1/2}\Upsilon^\top \Sigma_F^{-1/2}$ *with* Υ *being the eigenvector of* $\Sigma_F^{1/2}\Sigma_\Lambda \Sigma_F^{1/2}$

Proof. The statement is essentially Proposition 1 in Bai (2003) and the proof follows the same logic. Starting with the equality $\frac{1}{N}X^\top X\hat{\Lambda} = \hat{\Lambda}V_{MN}$, we multiply both sides by $\frac{1}{N}(F^\top F)^{1/2}\Lambda^\top$ to obtain

$$(F^\top F)^{1/2}\frac{1}{N}\Lambda^\top\left(\frac{X^\top X}{N}\right)\hat{\Lambda} = (F^\top F)^{1/2}\left(\frac{\Lambda^\top \hat{\Lambda}}{N}\right)V_{MN}$$

Plugging in $X = F\Lambda^\top + e$, we get

$$(F^\top F)^{1/2}\left(\frac{\Lambda^\top \Lambda}{N}\right)(F^\top F)\left(\frac{\Lambda^\top \hat{\Lambda}}{N}\right) + d_{NM} = (F^\top F)^{1/2}\left(\frac{\Lambda^\top \hat{\Lambda}}{N}\right)V_{MN}$$

with

$$d_{NM} = (F^\top F)^{1/2} \left(\frac{\Lambda^\top e^\top F}{N} \frac{\Lambda^\top \hat{\Lambda}}{N} + \frac{\Lambda^\top \Lambda}{N} \frac{F^\top e \hat{\Lambda}}{N} + \frac{\Lambda^\top e^\top e \hat{\Lambda}}{N^2} \right)$$

Applying Lemmas A.1 and A.2, we conclude $d_{NM} = o_p(1)$. The rest of the proof is essentially identical to Bai's proof. □

Lemma A.14. *Properties of* Q *and* H
Under Assumptions 1.1 and 1.2

1. $plim_{M,N\to\infty} H = Q^{-1}$
2. $Q^\top Q = \Sigma_\Lambda$
3. $plim_{M,N\to\infty} HH^\top = \Sigma_\Lambda^{-1}$

Proof. Lemma A.13 yields $H = (F^\top F)\left(\frac{\Lambda^\top \hat{\Lambda}}{N}\right) V^{-1} \xrightarrow{p} \Sigma_F Q^\top V^{-1}$ and the definition of V is $\Upsilon V \Upsilon^\top = {\Sigma_F^{1/2}}^\top \Sigma_\Lambda \Sigma_F^{1/2}$. Hence, the first statement follows from

$$\begin{aligned} H^\top Q &= V^{-1} Q \Sigma_F Q^\top + o_p(1) \\ &= V^{-1} V^{1/2} \Upsilon^\top \Sigma_F^{-1/2} \Sigma_F {\Sigma_F^{-1/2}}^\top \Upsilon V^{1/2} + o_p(1) \\ &= V^{-1} V + o_p(1) = I + o_p(1) \end{aligned}$$

The second statement follows from the definitions:

$$\begin{aligned} Q^\top Q &= {\Sigma_F^{-1/2}}^\top \Upsilon V^{1/2} V^{1/2} \Upsilon^\top \Sigma_F^{1/2} \\ &= {\Sigma_F^{-1/2}}^\top {\Sigma_F^{1/2}}^\top \Sigma_\Lambda \Sigma_F^{1/2} \Sigma_F^{-1/2} \\ &= \Sigma_\Lambda \end{aligned}$$

The third statement is a simple combination of the first two statements. □

Proof of Theorem 1.3:

Proof. Except for the asymptotic distribution of $\sqrt{M} F^\top e_i$, the proof is the same as for Theorem 1 in Bai (2003). By Lemma A.11

$$\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) V_{MN} = O_p\left(\frac{1}{\sqrt{M\delta}}\right) + O_p\left(\frac{1}{\sqrt{N\delta}}\right) + O_p\left(\frac{1}{\sqrt{M}}\right) + O_p\left(\frac{1}{\sqrt{M\delta}}\right)$$

The dominant term is $\frac{1}{N}\sum_{k=1}^N \hat{\Lambda}_k \eta_{ki}$. Hence, we get the expansion

$$\sqrt{M}\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) = V_{MN}^{-1} \frac{1}{N} \sum_{k=1}^N \hat{\Lambda}_k \Lambda_k^\top \sqrt{M} F^\top e_i + O_p\left(\frac{\sqrt{M}}{\delta}\right)$$

If $\frac{\sqrt{M}}{N} \to 0$, then using Lemmas A.6 and A.13, we obtain

$$\sqrt{M}(\hat{\Lambda}_i - H^\top \Lambda_i) \xrightarrow{L-s} N\left(0, V^{-1} Q \Gamma_i Q^\top V^{-1}\right)$$

If $\liminf \frac{\sqrt{M}}{N} \geq \tau > 0$, then

$$N(\hat{\Lambda}_i - \Lambda_i H) = O_p\left(\frac{N}{\sqrt{M\delta}}\right) + O_p\left(\frac{\sqrt{N}}{\sqrt{\delta}}\right) + O_p\left(\frac{N}{\sqrt{M}}\right) + O_p\left(\frac{N}{\sqrt{M\delta}}\right) = O_p(1)$$

□

Lemma A.15. ***Consistency of loadings***
Assume Assumption 1.1 holds. Then

$$\hat{\Lambda}_i - H^\top \Lambda_i = O_p\left(\frac{1}{\sqrt{\delta}}\right).$$

Proof. If we impose additionally the Assumption 1.2, then this lemma is a trivial consequence of Theorem 1.3. However, even without Assumption 1.2, Lemma A.11 can be modified to show that

$$V_{MN}\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) = O_p\left(\frac{1}{\sqrt{\delta}}\right) + O_p\left(\frac{1}{\sqrt{N\delta}}\right) + O_p\left(\frac{1}{\sqrt{M}}\right) + O_p\left(\frac{1}{\sqrt{M\delta}}\right).$$

□

A.5 Estimation of the Factors

Lemma A.16. *Assume that Assumptions 1.1 and 1.2 hold. Then*

$$\sum_{j=1}^{M} \frac{1}{N} F_j (\Lambda - \hat{\Lambda} H^{-1})^\top \hat{\Lambda} = O_p\left(\frac{1}{\delta}\right)$$

Proof. The overall logic of the proof is similar to Lemma B.1 in Bai (2003), but the underlying conditions and derivations of the final bounds are different. It is sufficient to show that

$$\frac{1}{N}(\hat{\Lambda} - \Lambda H)^\top \Lambda = O_p\left(\frac{1}{\delta}\right).$$

First using Lemma A.9 we decompose this term into

$$\begin{aligned}\frac{1}{N}(\hat{\Lambda} - \Lambda H)^\top \Lambda &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k \phi_{ik} + \frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k [e_i, e_k] + \frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k \eta_{ki} + \frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k \xi_{ki}\right)\Lambda_i^\top \\ &= I + II + III + IV\end{aligned}$$

We will tackle all four terms one-by-one.
Term I: The first term can again be decomposed into

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k\phi_{ik}{\Lambda_i}^\top = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)\phi_{ik}{\Lambda_i}^\top + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k\phi_{ik}{\Lambda_i}^\top$$

Due to Lemmas A.2 and A.10 the first term of I is bounded by

$$\begin{aligned}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)\phi_{ik}{\Lambda_i}^\top &\leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\phi_{ik}{\Lambda_i}^\top\|^2\right)^{1/2} \\ &\leq O_p\left(\frac{1}{\sqrt{\delta}}\right)\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}(e_{ji}e_{jk} - [e_i, e_k]){\Lambda_i}^\top\|^2\right)^{1/2} \\ &= O_p\left(\frac{1}{\sqrt{\delta}}\right)O_p\left(\frac{1}{\sqrt{MN}}\right)\end{aligned}$$

Now we turn to the second term, which we can bound using Lemma A.2 again:

$$\begin{aligned}\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k\phi_{ik}{\Lambda_i}^\top\| &\leq \|H\|\|\frac{1}{N}\sum_{k=1}^{N}\Lambda_k\frac{1}{N}\sum_{i=1}^{N}\phi_{ik}{\Lambda_i}^\top\| \\ &\leq O_p(1)\left(\frac{1}{N}\sum_{k=1}^{N}\|\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\phi_{ik}{\Lambda_i}^\top\|^2\right)^{1/2} \\ &\leq O_p\left(\frac{1}{\sqrt{MN}}\right)\end{aligned}$$

Hence, I is bounded by the rate $O_p\left(\frac{1}{\sqrt{MN}}\right)$.
Term II: Next we deal with II:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k[e_i, e_k]{\Lambda_i}^\top = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)[e_i, e_k]{\Lambda_i}^\top + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k[e_i, e_k]{\Lambda_i}^\top$$

Cauchy-Schwartz applied to the first term yields

$$\begin{aligned}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)[e_i, e_k]{\Lambda_i}^\top &\leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}[e_i, e_k]{\Lambda_i}^\top\|^2\right)^{1/2} \\ &= O_p\left(\frac{1}{\sqrt{\delta}N}\right)\end{aligned}$$

We used Lemma A.10 for the first factor and Assumption 1.2 in addition with the boundedness of $\|\Lambda_i\|$ for the second factor. By the same argument the second term of II converges at the following rate

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k [e_i, e_k]\Lambda_i{}^\top \leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}[e_i, e_k]\Lambda_i{}^\top\|^2\right)^{1/2}$$
$$\leq O_p\left(\frac{1}{N}\right)$$

Thus, the rate of II is $O_p\left(\frac{1}{N}\right)$. Next, we address III.
Term III: We start with the familiar decomposition

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k\eta_{ki}\Lambda_i{}^\top = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\eta_{ki}\Lambda_i{}^\top + \frac{1}{N^2}\sum_{k=1}^{N}\sum_{i=1}^{N}H^\top\Lambda_k\eta_{ki}\Lambda_i{}^\top$$

We use Lemmas A.1 and A.10 and the boundedness of $\|\Lambda_k\|$. The first term is bounded by

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\eta_{ki}\Lambda_i{}^\top \leq \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\cdot$$
$$\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\Lambda_k{}^\top F_j e_{ji}\Lambda_i\|^2\right)^{1/2}$$
$$\leq O_p\left(\frac{1}{\sqrt{\delta NM}}\right)$$

The second term is bounded by

$$\frac{1}{N^2}\sum_{k=1}^{N}\sum_{i=1}^{N}H^\top\Lambda_k\eta_{ki}\Lambda_i{}^\top \leq \left(\frac{1}{N}\sum_{k=1}^{N}\|H^\top\Lambda_k\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\Lambda_k{}^\top F_j e_{ji}\Lambda_i\|^2\right)^{1/2}$$
$$\leq O_p\left(\frac{1}{\sqrt{NM}}\right)$$

This implies that III is bounded by $O_p\left(\frac{1}{\sqrt{MN}}\right)$.
Term IV: Finally, we deal with IV:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{ki}\Lambda_i{}^\top = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\xi_{ki}\Lambda_i{}^\top + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k\xi_{ki}\Lambda_i{}^\top.$$

The first term can be bounded using Lemmas A.10 and Lemma A.6:

$$\begin{aligned}\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)\xi_{ki}\Lambda_i{}^\top\| \leq & \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2} \cdot \\ & \left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}\Lambda_i{}^\top F^\top e_i\Lambda_i{}^\top\|^2\right)^{1/2} \\ \leq & O_p\left(\frac{1}{\sqrt{\delta M}}\right)\end{aligned}$$

For the second term we need the boundedness of Λ_i and a modification of Proposition A.1:

$$\begin{aligned}\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k\xi_{ki}\Lambda_i{}^\top\| &= \|\frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{M}H^\top\Lambda_k e_{jk}F_j^\top\left(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\Lambda_i{}^\top\right)\| \\ &\leq \|\left(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i{}^\top\Lambda_i\right)\|\|\frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{M}F_j e_{jk}\Lambda_k^\top H\| \\ &\leq O_p\left(\frac{1}{\sqrt{MN}}\right).\end{aligned}$$

In conclusion, IV is bounded by $O_p\left(\frac{1}{\sqrt{MN}}\right)$. Putting things together, we get

$$\frac{1}{N}(\hat{\Lambda} - \Lambda H)^\top\Lambda = O_p\left(\frac{1}{\sqrt{MN}}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{MN}}\right) + O_p\left(\frac{1}{\sqrt{MN}}\right) = O_p\left(\frac{1}{\delta}\right).$$

□

Lemma A.17. *Assume that Assumptions 1.1 and 1.2 hold. Then*

$$\sum_{j=1}^{M}\sum_{k=1}^{N}\frac{1}{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)e_{jk} = O_p\left(\frac{1}{\delta}\right) + O_p(1)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}F_j e_{ji}\sum_{l=1}^{M}e_{li}\right)$$

Without further assumptions the RHS is $O_p\left(\frac{1}{\delta}\right) + O_p\left(\frac{1}{\sqrt{M}}\right)$.

Proof. The general approach is similar to Lemma B.2 in Bai (2003), but the result is different, which has important implications for Theorem 1.4.

Note that $e_i(T) = \sum_{j=1}^{M} e_{ji}$. We want to show:

$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\Lambda}_i - H^\top\Lambda_i\right)e_i(T) = O_p\left(\frac{1}{\delta}\right) + O_p(1)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}F_j e_{ji}\sum_{l=1}^{M}e_{li}\right).$$

We substitute the expression from Lemma A.9:

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\Lambda}_i - H^\top\Lambda_i\right)e_i(T) =& \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k[e_i,e_k]e_i(T) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k\phi_{ik}e_i(T)\\ &+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k\eta_{ik}e_i(T) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k\xi_{ik}e_i(T)\\ &= I + II + III + IV\end{aligned}$$

Term I: We first decompose I into two parts:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k[e_i,e_k]e_i(T) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)[e_i,e_k]e_i(T) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k[e_i,e_k]e_i(T).$$

Lemma A.10, Assumption 1.2 and the boundedness of $e_i(T)$ yield for the first term of I:

$$\begin{aligned}\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)[e_i,e_k]e_i(T)\| \leq& \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\cdot\\ &\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}e_i(T)[e_i,e_k]\|^2\right)^{1/2}\\ \leq& O_p\left(\frac{1}{\sqrt{\delta}}\right)O_p\left(\frac{1}{N}\right).\end{aligned}$$

Using Assumption 1.2 , we bound the second term

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k[e_i,e_k]e_i(T) = O_p\left(\frac{1}{N}\right).$$

Hence, I is $O_p\left(\frac{1}{N}\right)$.

Term II: We split II into two parts:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k\phi_{ik}e_i(T) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)\phi_{ik}e_i(T) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}H^\top\Lambda_k\phi_{ik}e_i(T)$$

As before we apply the Cauchy-Schwartz inequality to the first term and then we use Lemma A.4:

$$\begin{aligned}\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)\phi_{ik}e_i(T)\| \leq& \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2}\\ &\cdot\left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N}e_i(T)\left(\sum_{j=1}^{M}e_{ji}e_{jk} - [e_i,e_k]\right)\|^2\right)^{1/2}\\ \leq& O_p\left(\frac{1}{\sqrt{\delta}}\right)O_p\left(\frac{1}{\sqrt{M}}\right)\end{aligned}$$

The second term can be bounded by using a modification of Lemma A.2 and the boundedness of $e_i(T)$:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k \left(\sum_{j=1}^{M} e_{ji}e_{jk} - [e_i, e_k]\right) e_i(T) \leq O_p\left(\frac{1}{\sqrt{MN}}\right).$$

Thus, II is $O_p\left(\frac{1}{\sqrt{\delta M}}\right)$.

Term III: This term yields a convergence rate different from the rest and is responsible for the extra summand in the statement:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k \eta_{ik} e_i(T) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top \Lambda_k\right)\eta_{ik}e_i(T) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k \eta_{ik} e_i(T)$$

The first term can be controlled using Lemma A.10 and Lemma A.6:

$$\begin{aligned}\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top \Lambda_k\right)\eta_{ik}e_i(T)\| \leq & \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2} \cdot \\ & \left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N} e_i(T)\Lambda_k{}^\top \sum_{j=1}^{M} F_j e_{ji}\|^2\right)^{1/2} \\ \leq & O_p\left(\frac{1}{\sqrt{\delta}}\right) O_p\left(\frac{1}{\sqrt{M}}\right)\end{aligned}$$

Without further assumptions, the rate of the second term is slower than of all the other summands and can be calculated using Lemma A.6:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k \Lambda_k{}^\top \sum_{j=1}^{M} F_j e_{ji} e_i(T) = O_p(1)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} F_j e_{ji}\sum_{l=1}^{M} e_{li}\right) = O_p\left(\frac{1}{\sqrt{M}}\right)$$

Term IV: We start with the usual decomposition for the last term:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}\hat{\Lambda}_k \xi_{ik} e_i(T) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)\xi_{ik}e_i(T) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k \xi_{ik} e_i(T)$$

For the first term we use Lemma A.10 and Lemmas A.6 and A.8:

$$\begin{aligned}\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{\Lambda}_k - H^\top\Lambda_k)\xi_{ik}e_i(T)\| \leq & \left(\frac{1}{N}\sum_{k=1}^{N}\|\hat{\Lambda}_k - H^\top\Lambda_k\|^2\right)^{1/2} \cdot \\ & \left(\frac{1}{N}\sum_{k=1}^{N}\|\frac{1}{N}\sum_{i=1}^{N} e_i(T)\Lambda_i{}^\top \sum_{j=1}^{M} F_j e_{jk}\|^2\right)^{1/2} \\ \leq & O_p\left(\frac{1}{\sqrt{\delta MN}}\right).\end{aligned}$$

Similarly for the second term:

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{N} H^\top \Lambda_k \xi_{ik} e_i(T) = \frac{1}{N}\sum_{k=1}^{N} H^\top \Lambda_k \left(\frac{1}{N}\sum_{i=1}^{N} e_i(T)\Lambda_i{}^\top\right)\left(\sum_{j=1}^{M} F_j e_{jk}\right)$$
$$= O_p\left(\frac{1}{\sqrt{MN}}\right)$$

In conclusion, IV is $O_p\left(\frac{1}{\sqrt{MN}}\right)$. Putting the results together, we obtain

$$I + II + III + IV = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{\delta M}}\right) + O_p\left(\frac{1}{\sqrt{M}}\right) + O_p\left(\frac{1}{\sqrt{MN}}\right)$$
$$= O_p\left(\frac{1}{\delta}\right) + O_p\left(\frac{1}{\sqrt{M}}\right).$$

Term III is responsible for the low rate of convergence. □

Proof of Theorem 1.4:

Proof.

$$\begin{aligned}
\hat{F} - F{H^{-1}}^\top &= \frac{1}{N} X\hat{\Lambda} - F{H^{-1}}^\top \\
&= (F(\Lambda - \hat{\Lambda}H^{-1} + \hat{\Lambda}H^{-1})^\top + e)\frac{1}{N}\hat{\Lambda} - F{H^{-1}}^\top \\
&= F\Lambda^\top\hat{\Lambda}\frac{1}{N} - F{H^{-1}}^\top\hat{\Lambda}^\top\hat{\Lambda}\frac{1}{N} + F{H^{-1}}^\top + e\hat{\Lambda}\frac{1}{N} - F{H^{-1}}^\top \\
&= \frac{1}{N}F(\Lambda - \hat{\Lambda}H^{-1})^\top\hat{\Lambda} + \frac{1}{N}e\hat{\Lambda} \\
&= \frac{1}{N}F(\Lambda - \hat{\Lambda}H^{-1})^\top\hat{\Lambda} + \frac{1}{N}e(\hat{\Lambda} - \Lambda H) + \frac{1}{N}e\Lambda H.
\end{aligned}$$

By Lemmas A.16 and A.17, only the last term is of interest

$$\begin{aligned}
\sum_{j=1}^{M}\left(\hat{F}_j - H^{-1}F_j\right) =& \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\hat{\Lambda}_k\left(\Lambda_k - {H^{-1}}^\top\hat{\Lambda}_k\right)^\top F_j + \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\left(\hat{\Lambda}_k - H^\top\Lambda_k\right)e_{jk} \\
&+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} H^\top\Lambda_k e_{jk} \\
=& O_p\left(\frac{1}{\delta}\right) + O_p(1)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} F_j e_{ji}\sum_{l=1}^{M} e_{li}\right) + \frac{1}{N}e(T)\Lambda H.
\end{aligned}$$

Under Assumption 1.5 Proposition A.2 implies $\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}F_j e_{ji}\sum_{l=1}^{M}e_{li}\right)=O_p\left(\frac{1}{\delta}\right)$. If $\frac{\sqrt{N}}{M}\to 0$ then

$$\sqrt{N}\sum_{j=1}^{M}\left(\hat{F}_j-H^{-1}F_j\right)=o_p(1)+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}H^{\top}\Lambda_i e_i(T)$$

By Lemma A.8, we can apply the martingale central limit theorem and the desired result about the asymptotic mixed normality follows. In the case $\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}F_j e_{ji}\sum_{l=1}^{M}e_{li}\right)=O_p\left(\frac{1}{\sqrt{M}}\right)$, the arguments are analogous. □

Lemma A.18. ***Consistency of factors***
Assumptions 1.1 and 1.2 hold. Then $\hat{F}_T-H^{-1}F_T=O_p\left(\frac{1}{\sqrt{\delta}}\right)$.

Proof. The Burkholder-Davis-Gundy inequality in Lemma A.31 implies $\frac{1}{N}e_T\Lambda H=O_p\left(\frac{1}{\sqrt{N}}\right)$. In the proof of Theorem 1.4, we have shown that Assumptions 1.1 and 1.2 are sufficient for

$$\sum_{j=1}^{M}\left(\hat{F}_j-H^{-1}F_j\right)=O_p\left(\frac{1}{\delta}\right)+O_p\left(\frac{1}{\sqrt{M}}\right)+\frac{1}{N}e_T\Lambda H.$$

□

Lemma A.19. ***Consistency of factor increments***
Under Assumptions 1.1 and 1.2 we have

$$\hat{F}_j=H^{-1}F_j+O_p\left(\frac{1}{\delta}\right)$$

Proof. Using the same arguments as in the proof of Theorem 1.4 we obtain the decomposition

$$\hat{F}_j-H^{-1}F_j=\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\left(\Lambda_k-H^{-1^{\top}}\hat{\Lambda}_k\right)^{\top}F_j+\frac{1}{N}\sum_{k=1}^{N}e_{jk}\left(\hat{\Lambda}_k-H^{\top}\Lambda_k\right)+\frac{1}{N}\sum_{k=1}^{N}H^{\top}\Lambda_k e_{jk}.$$

Lemma A.16 can easily be modified to show that

$$\frac{1}{N}\sum_{k=1}^{N}\hat{\Lambda}_k\left(\Lambda_k-H^{-1^{\top}}\hat{\Lambda}_k\right)^{\top}F_j=O_p\left(\frac{1}{\delta}\right).$$

Lemma A.17 however requires some additional care. All the arguments go through for $e_{l,i}$ instead of $\sum_{l=1}^{M}e_{l,i}$ except for the term $\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}F_j e_{ji}e_{li}\right)$. Based on our previous results we have $\sum_{j=1}^{M}F_j e_{j,i}=O_p\left(\frac{1}{\sqrt{M}}\right)$ and $e_{l,i}=O_p\left(\frac{1}{\sqrt{M}}\right)$. This yields

$$\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}F_j e_{ji}e_{li}\right)=O_p\left(\frac{1}{M}\right)=O_p\left(\frac{1}{\delta}\right)$$

Therefore

$$\frac{1}{N}\sum_{k=1}^{N} e_{jk}\left(\hat{\Lambda}_k - H^\top \Lambda_k\right) = O_p\left(\frac{1}{\delta}\right).$$

Lemma A.5 provides the desired rate for the last term $\frac{1}{N}\sum_{k=1}^{N} H^\top \Lambda_k e_{jk} = O_p\left(\frac{1}{\delta}\right)$. □

Lemma A.20. ***Consistent estimation of factor covariation***
Under Assumptions 1.1 and 1.2 we can consistently estimate the quadratic covariation of the factors if $\frac{\sqrt{M}}{N} \to 0$. *Assume* $Y(t)$ *is a stochastic process satisfying Definition A.1. Then*

$$\|\hat{F}^\top \hat{F} - H^{-1}[F,F]_T H^{-1^\top}\| = o_p(1) \qquad \|\sum_{j=1}^{M} \hat{F}_j Y_j - H^{-1}[F,Y]\| = o_p(1)$$

Proof. This is a simple application of Lemma A.19:

$$\begin{aligned}\sum_{j=1}^{M} \hat{F}_j \hat{F}_j^\top &= H^{-1}\left(\sum_{j=1}^{M} F_j F_j^\top\right) H^{-1^\top} + O_p\left(\frac{1}{\delta}\right)\sum_{j=1}^{M} |F_j| + \sum_{j=1}^{M} O_p\left(\frac{1}{\delta^2}\right) \\ &= H^{-1}\left(\sum_{j=1}^{M} F_j F_j^\top\right) H^{-1^\top} + O_p\left(\frac{\sqrt{M}}{\delta}\right) + O_p\left(\frac{M}{\delta^2}\right)\end{aligned}$$

By Theorem A.2

$$\left(\sum_{j=1}^{M} F_j F_j^\top\right) - [F,F]_T = O_p\left(\frac{1}{\sqrt{\delta}}\right)$$

The desired result follows for $\frac{\sqrt{M}}{N} \to 0$. The proof for $[F,Y]$ is analogous. □

A.6 Estimation of Common Components

Proof of Theorem 1.5:

Proof. The proof is very similar to Theorem 3 in Bai (2003). For completeness I present it here:

$$\hat{C}_{T,i} - C_{T,i} = \left(\hat{\Lambda}_i - H^\top \Lambda_i\right)^\top H^{-1} F_T + \hat{\Lambda}_i^\top \left(\hat{F}_T - H^{-1} F_T\right).$$

From Theorems 1.3 and 1.4 we have

$$\begin{aligned}\sqrt{\delta}\left(\hat{\Lambda}_i - H^\top \Lambda_i\right) &= \sqrt{\frac{\delta}{M}} V_{MN}^{-1} \frac{1}{N}\sum_{k=1}^{N} \hat{\Lambda}_k \Lambda_k^\top \sqrt{M} F^\top e_i + O_p\left(\frac{1}{\sqrt{\delta}}\right) \\ \sqrt{\delta}\left(\hat{F}_T - H^{-1} F_t\right) &= \sqrt{\frac{\delta}{M}} \sum_{i=1}^{N} H^\top \Lambda_i e_{T,i} + O_p\left(\sqrt{\frac{\delta}{M}}\right) + O_p\left(\frac{1}{\sqrt{\delta}}\right).\end{aligned}$$

If Assumption 1.5 holds, the last equation changes to

$$\sqrt{\delta}\left(\hat{F}_T - H^{-1}F_t\right) = \sqrt{\frac{\delta}{M}}\sum_{i=1}^{N} H^\top \Lambda_i e_{T,i} + O_p\left(\frac{1}{\sqrt{\delta}}\right).$$

In the following, we will assume that weak serial dependence and cross-sectional dependence holds. The modification to the case without it is obvious. Putting the limit theorems for the loadings and the factors together yields:

$$\begin{aligned}\hat{C}_{T,i} - C_{T,i} =& \sqrt{\frac{\delta}{M}} F^\top H^{-1^\top} V_{MN}^{-1}\left(\frac{1}{N}\Lambda^\top\hat{\Lambda}\right)\sqrt{M}F^\top e_i \\ &+ \sqrt{\frac{\delta}{N}}\Lambda_i^\top H H^\top\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Lambda_i e^{T,i}\right) + O_p\left(\frac{1}{\sqrt{\delta}}\right).\end{aligned}$$

We have used

$$\begin{aligned}\hat{\Lambda}_i^\top\left(\hat{F}_T - H^{-1}F_T\right) &= \Lambda_i^\top H(\hat{F}_T - H^{-1}F_T) + \left(\hat{\Lambda}_i^\top - \Lambda_i^\top H\right)\left(\hat{F}_T - H^{-1}F_T\right) \\ &= \Lambda_i^\top H(\hat{F}_T - H^{-1}F_T) + O_p\left(\frac{1}{\delta}\right).\end{aligned}$$

By the definition of H it holds that

$$H^{-1^\top} V_{MN}^{-1}\left(\frac{\hat{\Lambda}^\top\Lambda}{N}\right) = \left(F^\top F\right)^{-1}.$$

Using the reasoning behind Lemma A.14, it can easily be shown that

$$HH^\top = \left(\frac{1}{N}\Lambda^\top\Lambda\right)^{-1} + O_p\left(\frac{1}{\delta}\right).$$

Define

$$\begin{aligned}\xi_{NM} &= F_T^\top\left(F^\top F\right)^{-1}\sqrt{M}F^\top e_i \\ \phi_{NM} &= \Lambda_i^\top\left(\frac{1}{N}\Lambda^\top\Lambda\right)^{-1}\frac{1}{\sqrt{N}}\Lambda^\top e_T\end{aligned}$$

By Lemmas A.6 and A.8, we know that these terms converge stably in law to a conditional normal distribution:

$$\xi_{NM} \stackrel{s-L}{\to} N(0, V_{T,i}) \qquad , \qquad \phi_{NM} \stackrel{s-L}{\to} N(0, W_{T,i})$$

Therefore,

$$\sqrt{\delta}\left(\hat{C}_{T,i} - C_{T,i}\right) = \sqrt{\frac{\delta}{M}}\xi_{NM} + \sqrt{\frac{\delta}{N}}\phi_{NM} + O_p\left(\frac{1}{\sqrt{\delta}}\right)$$

ξ_{NM} and ϕ_{NM} are asymptotically independent, because one is the sum of cross-sectional random variables, while the other is the sum of a particular time series of increments. If $\frac{\delta}{M}$ and $\frac{\delta}{N}$ converge, then asymptotic normality follows immediately from Slutzky's theorem. $\frac{\delta}{M}$ and $\frac{\delta}{N}$ are not restricted to be convergent sequences. We can apply an almost sure representation theory argument on the extension of the probability space similar to Bai (2003). □

Lemma A.21. ***Consistency of increments of common component estimator*** *Under Assumptions 1.1 and 1.2 it follows that*

$$\hat{C}_{j,i} = C_{j,i} + O_p\left(\frac{1}{\delta}\right)$$
$$\hat{e}_{j,i} = e_{j,i} + O_p\left(\frac{1}{\delta}\right)$$

with $\hat{e}_{j,i} = X_{j,i} - \hat{C}_{j,i}$.

Proof. As in the proof for Theorem 1.5 we can separate the error into a component due to the loading estimation and one due to the factor estimation.

$$\hat{C}_{j,i} - C_{j,i} = \left(\hat{\Lambda}_i - H^\top \Lambda_i\right)^\top H^{-1}F_j + \hat{\Lambda}_i^\top \left(\hat{F}_j - H^{-1}F_j\right).$$

By Lemmas A.15 and A.19 we can bound the error by $O_p\left(\frac{1}{\delta}\right)$. □

Lemma A.22. ***Consistent estimation of residual covariation*** *Assume Assumptions 1.1 and 1.2 hold. Then if* $\frac{\sqrt{M}}{\delta} \to 0$ *we have for* $i,k = 1,...,N$ *and any stochastic process* $Y(t)$ *satisfying Definition A.1:*

$$\sum_{j=1}^M \hat{e}_{j,i}\hat{e}_{j,k} = [e_i, e_k] + o_p(1), \qquad \sum_{j=1}^M \hat{C}_{j,i}\hat{C}_{j,k} = [C_i, C_k] + o_p(1).$$
$$\sum_{j=1}^M \hat{e}_{j,i}Y_j = [e_i, Y] + o_p(1), \qquad \sum_{j=1}^M \hat{C}_{j,i}Y_j = [C_i, Y] + o_p(1).$$

Proof. Using Lemma A.21 we obtain

$$\sum_{j=1}^M \hat{e}_{j,i}\hat{e}_{j,k} = \sum_{j=1}^M e_{j,i}e_{j,k} + \sum_{j=1}^M O_p\left(\frac{1}{\delta^2}\right) + \sum_{j=1}^M |e_{j,i}| O_p\left(\frac{1}{\delta}\right) = \sum_{j=1}^M e_{j,i}e_{j,k} + o_p(1) = [e_i, e_k] + o_p(1).$$

The rest of the proof follows the same logic. □

Proof of Theorem 1.1:

Proof. This is a collection of the results in Lemmas A.15, A.18, A.20, A.21 and A.22. □

A.7 Estimating Covariance Matrices

Proposition A.3. ***Consistent unfeasible estimator of covariance matrix of loadings***
Assume Assumptions 1.1, 1.2 and 1.3 hold and $\frac{\sqrt{M}}{N} \to 0$. *By Theorem 1*

$$\sqrt{M}(\hat{\Lambda}_i - H^\top \Lambda_i) \xrightarrow{L-s} N(0, \Theta_\Lambda)$$

with

$$\Theta_{\Lambda,i} = V^{-1} Q \Gamma_i Q^\top V^{-1}$$

where the entry $\{l, g\}$ *of the* $K \times K$ *matrix* Γ_i *is given by*

$$\Gamma_{i,l,g} = \int_0^T \sigma_{F^l,F^g} \sigma_{e_i}^2 ds + \sum_{s \le T} \Delta F^l(s) \Delta F^g(s) \sigma_{e_i}^2(s) + \sum_{s' \le T} \Delta e_i^2(s') \sigma_{F^g,F^l}(s').$$

F^l *denotes the* l*-th component of the the* K *dimensional process* F *and* σ_{F^l,F^g} *are the entries of its* $K \times K$ *dimensional volatility matrix. Take any sequence of integers* $k \to \infty$, $\frac{k}{M} \to 0$. *Denote by* $I(j)$ *a local window of length* $\frac{2k}{M}$ *around* j *with some* $\alpha > 0$ *and* $\omega \in \left(0, \frac{1}{2}\right)$.

Define a consistent, but unfeasible, estimator for Γ_i *by*

$$\begin{aligned}
\bar{\Gamma}_{i,l,g} = & M \sum_{j=1}^{M} F_j^l F_j^g e_{j,i}^2 \mathbb{1}_{\{|F_j^l| \le \alpha \Delta_M^\omega, |F_j^g| \le \alpha \Delta_M^\omega, |e_{j,i}| \le \alpha \Delta_M^\omega\}} \\
& + \frac{M}{2k} \sum_{j=k+1}^{M-k} F_j^l F_j^g \mathbb{1}_{\{|F_j^l| \ge \alpha \Delta_M^\omega, |F_j^g| \ge \alpha \Delta_M^\omega\}} \left(\sum_{h \in I(j)} e_{h,i}^2 \mathbb{1}_{\{|e_{h,i}| \le \alpha \Delta_M^\omega\}} \right) \\
& + \frac{M}{2k} \sum_{j=k+1}^{M-k} e_{j,i}^2 \mathbb{1}_{\{|e_{j,i}| \ge \alpha \Delta_M^\omega\}} \left(\sum_{h \in I(j)} F_h^l F_h^g \mathbb{1}_{\{|F_h^l| \le \alpha \Delta_M^\omega, |F_h^g| \le \alpha \Delta_M^\omega\}} \right)
\end{aligned}$$

Then

$$V_{MN}^{-1} \left(\frac{\hat{\Lambda}^\top \Lambda}{N} \right) \bar{\Gamma}_i \left(\frac{\Lambda^\top \hat{\Lambda}}{N} \right) V_{MN}^{-1} \xrightarrow{p} \Theta_{\Lambda,i}$$

Proof. The Estimator for Γ_i is an application of Theorem A.3. Note that we could generalize the statement to include infinite activity jumps as long as their activity index is smaller than 1. Finite activity jumps trivially satisfy this condition. The rest follows from Lemmas A.12 and A.13. □

Proof of Theorem 1.6:

Proof. By abuse of notation the matrix $e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ has elements $e_{j,i}1_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ and the matrix $F\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^\top$ has elements $F_j\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda_i^\top$. A similar notation is applied for other combinations of vectors with a truncation indicator function.

Step 1: To show: $\frac{1}{N}\hat{X}_j^C\hat{\Lambda} - \sum_{i=1}^N \mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}H^{-1}F_j = O_p\left(\frac{1}{\delta}\right)$

We start with a similar decomposition as in Theorem 1.4:

$$\frac{\hat{X}^C\hat{\Lambda}}{N} - F\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}{H^{-1}}^\top\frac{\hat{\Lambda}^\top\hat{\Lambda}}{N} = \frac{1}{N}F\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\Lambda - \hat{\Lambda}H^{-1}\right)^\top\hat{\Lambda} + \frac{1}{N}e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\hat{\Lambda} - \Lambda H\right) + \frac{1}{N}e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda H.$$

It can be shown that

$$\frac{1}{N}F_j\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\Lambda - \hat{\Lambda}H^{-1}\right)^\top\hat{\Lambda} = O_p\left(\frac{1}{\delta}\right)$$

$$\frac{1}{N}e_j\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\hat{\Lambda} - \Lambda H\right) = O_p\left(\frac{1}{\delta}\right)$$

$$\frac{1}{N}e_j\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda H = \frac{1}{N}e_j^C\Lambda H + \frac{1}{N}\left(e_j\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}} - e_j^C\right)\Lambda H = O_p\left(\frac{1}{\delta}\right).$$

The first statement follows from Lemma A.16. The second one can be shown as in Lemma A.19. The first term of the third statement can be bounded using Lemma A.5. The rate for the second term of the third equality follows from the fact that the difference $e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}} - e_{j,i}^C$ is equal to some drift term which is of order $O_p\left(\frac{1}{M}\right)$ and to $-\frac{1}{N}e_{j,i}^C$ if there is a jump in $X_{j,i}$.

Step 2: To show: $\frac{1}{N}\hat{X}_j^D\hat{\Lambda} - \sum_{i=1}^N \mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}H^{-1}F_j = O_p\left(\frac{1}{\delta}\right)$

As in step 1 we start with a decomposition

$$\frac{\hat{X}^D\hat{\Lambda}}{N} - F\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}{H^{-1}}^\top\frac{\hat{\Lambda}^\top\hat{\Lambda}}{N} = \frac{1}{N}F\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\left(\Lambda - \hat{\Lambda}H^{-1}\right)^\top\hat{\Lambda} + \frac{1}{N}e\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\left(\hat{\Lambda} - \Lambda H\right) + \frac{1}{N}e\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\Lambda H.$$

It follows

$$\frac{1}{N}F_j\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\left(\Lambda - \hat{\Lambda}H^{-1}\right)^\top\hat{\Lambda} = O_p\left(\frac{1}{\delta}\right)$$

$$\frac{1}{N}e_j\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\left(\hat{\Lambda} - \Lambda H\right) = O_p\left(\frac{1}{\delta}\right)$$

$$\frac{1}{N}e_j\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}\Lambda H = \frac{1}{N}e_j^D\Lambda H + \frac{1}{N}\left(e_j\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}} - e_j^D\right)\Lambda H = O_p\left(\frac{1}{\delta}\right).$$

The first rate is a consequence of Lemma A.16, the second rate follows from Lemma A.15 and the third rate can be derived using similar arguments as in step 1.

Step 3: To show: $\hat{X}^C_{j,i} - \frac{\hat{X}^C_j \hat{\Lambda}}{N}\hat{\Lambda}_i - e_{j,i}\mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta^{\bar{\omega}}_M\}} = O_p\left(\frac{1}{\delta}\right)$
By a similar decomposition as in Lemma A.21 we obtain

$$\begin{aligned}\hat{X}^C_{j,i} - \frac{\hat{X}^C_j \hat{\Lambda}}{N}\hat{\Lambda}_i - e_{j,i}\mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta^{\bar{\omega}}_M\}} &= \left(\hat{\Lambda}_i - H^\top \Lambda_i\right)^\top H^{-1} F_j \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta^{\bar{\omega}}_M\}} \\ &\quad + \hat{\Lambda}_i^\top \left(\frac{\hat{\Lambda}^\top \hat{X}^{C\top}_j}{N} - H^{-1} F_j \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta^{\bar{\omega}}_M\}}\right) \\ &= O_p\left(\frac{1}{\sqrt{\delta}}\right) \|F_j \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta^{\bar{\omega}}_M\}}\| + O_p\left(\frac{1}{\delta}\right) \\ &= O_p\left(\frac{1}{\sqrt{\delta M}}\right) + O_p\left(\frac{1}{\delta}\right)\end{aligned}$$

The first rate follows from Lemma A.15 and the second rate can be deduced from step 1.
Step 4: To show $\hat{X}^D_{j,i} - \frac{\hat{X}^D_j \hat{\Lambda}}{N}\hat{\Lambda}_i - e_{j,i}\mathbb{1}_{\{|X_{j,i}| > \alpha \Delta^{\bar{\omega}}_M\}} = O_p\left(\frac{1}{\delta}\right) + O_p\left(\frac{1}{\sqrt{\delta}}\right) \|F_j \mathbb{1}_{\{|X_{j,i}| > \alpha \Delta^{\bar{\omega}}_M\}}\|$
A similar decomposition as in the previous step yields

$$\begin{aligned}\hat{X}^D_{j,i} - \frac{\hat{X}^D_j \hat{\Lambda}}{N}\hat{\Lambda}_i - e_{j,i}\mathbb{1}_{\{|X_{j,i}| > \alpha \Delta^{\bar{\omega}}_M\}} &= \left(\hat{\Lambda}_i - H^\top \Lambda_i\right)^\top H^{-1} F_j \mathbb{1}_{\{|X_{j,i}| > \alpha \Delta^{\bar{\omega}}_M\}} \\ &\quad + \hat{\Lambda}_i^\top \left(\frac{\hat{\Lambda}^\top \hat{X}^{D\top}_j}{N} - H^{-1} F_j \mathbb{1}_{\{|X_{j,i}| > \alpha \Delta^{\bar{\omega}}_M\}}\right) \\ &\leq O_p\left(\frac{1}{\sqrt{\delta}}\right) \|F_j \mathbb{1}_{\{|X_{j,i}| > \alpha \Delta^{\bar{\omega}}_M\}}\| + O_p\left(\frac{1}{\delta}\right)\end{aligned}$$

where the first rate follows from Lemma A.15 and the second from step 2.
Step 5: To show: $M \sum_{j=1}^M \left(\frac{\hat{X}^C_j \hat{\Lambda}}{N}\right) \left(\frac{\hat{X}^C_j \hat{\Lambda}}{N}\right)^\top \left(\hat{X}^C_{j,i} - \frac{\hat{X}^C_j \hat{\Lambda}}{N}\hat{\Lambda}_i\right)^2$
$= \mathrm{M}\sum_{j=1}^M \left(H^{-1} F_j \mathbb{1}_{\{|F_j| \leq \alpha \Delta^{\bar{\omega}}_M\}}\right)^\top \left(H^{-1} F_j \mathbb{1}_{\{|F_j| \leq \alpha \Delta^{\bar{\omega}}_M\}}\right) \left(e^2_{j,i} \mathbb{1}_{\{|e_{j,i}| \leq \alpha \Delta^{\bar{\omega}}_M\}}\right) + o_p(1)$
Step 1 and 3 yield

$$\begin{aligned}&M \sum_{j=1}^M \left(\frac{\hat{X}^C_j \hat{\Lambda}}{N}\right) \left(\frac{\hat{X}^C_j \hat{\Lambda}}{N}\right)^\top \left(\hat{X}^C_{j,i} - \frac{\hat{X}^C_j \hat{\Lambda}}{N}\hat{\Lambda}_i\right)^2 \\ =&M \sum_{j=1}^M \left(\sum_{i=1}^N \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta^{\bar{\omega}}_M\}} \frac{\hat{\Lambda}_i \hat{\Lambda}_i^\top}{N} H^{-1} F_j\right)^\top \left(\sum_{i=1}^N \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta^{\bar{\omega}}_M\}} \frac{\hat{\Lambda}_i \hat{\Lambda}_i^\top}{N} H^{-1} F_j\right) \left(e^2_{j,i} \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta^{\bar{\omega}}_M\}}\right) \\ &+ o_p(1)\end{aligned}$$

We need to show

$$\sum_{i=1}^N \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta^{\bar{\omega}}_M\}} \frac{\hat{\Lambda}_i \hat{\Lambda}_i^\top}{N} H^{-1} F_j - H^{-1} F_j \mathbb{1}_{\{\{|F_j| \leq \alpha \Delta^{\bar{\omega}}_M\}} = o_p\left(\frac{1}{\sqrt{\delta}}\right).$$

By Mancini (2009) the threshold estimator correctly identifies the jumps for sufficiently large M. By Assumption 1.3 a jump in $X_{j,i}$ is equivalent to a jump in $\Lambda_i^\top F_j$ or/and a jump in $e_{j,i}$. Hence, it is sufficient to show that

$$\sum_{i=1}^N \mathbb{1}_{\{F_j^D\Lambda_i=0, e_i^D=0, |F_j^D|\neq 0\}} \frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N} + \sum_{i=1}^N \mathbb{1}_{\{e_i^D\neq 0\}} \frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N} - I_K \sum_{i=1}^N \mathbb{1}_{\{e_{j,i}^D\neq 0, |F_j^D|=0\}} = o_p(1)$$

Note that

$$\begin{aligned}\mathbb{P}\left(e_{j,i}^D \neq 0\right) &= \mathbb{E}\left[\mathbb{1}_{\{e_{j,i}^D\neq 0\}}\right] = \mathbb{E}\left[\int_{t_j}^{t_{j+1}}\int_{\mathbb{R}_{-0}} d\mu_{e_i}(ds,dx)\right] \\ &= \mathbb{E}\left[\int_{t_j}^{t_{j+1}}\int_{\mathbb{R}_{-0}} d\nu_{e_i}(ds,dx)\right] \leq C\int_{t_j}^{t_{j+1}} ds = O\left(\frac{1}{M}\right).\end{aligned}$$

It follows that $\sum_{i=1}^N \mathbb{1}_{\{e_i^D\neq 0\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N} = o_p(1)$ as

$$\mathbb{E}\left[\sum_{i=1}^N \mathbb{1}_{\{e_i^D\neq 0\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}\right] = \sum_{i=1}^N \mathbb{P}\left(e_i^D\neq 0\right)\frac{\hat{\Lambda}_i\hat{\Lambda}_i}{N} = O_p\left(\frac{1}{M}\right)$$

and

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=1}^N \mathbb{1}_{\{e_i^D\neq 0\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}\right)^2\right] =&\mathbb{E}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \hat{\Lambda}_i\hat{\Lambda}_i^\top\hat{\Lambda}_k\hat{\Lambda}_k^\top \mathbb{1}_{\{e_i^D\neq 0\}}\mathbb{1}_{\{e_k^D\neq 0\}}\right] \\ \leq& \left(\mathbb{E}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \|\hat{\Lambda}_i\hat{\Lambda}_i^\top\hat{\Lambda}_k\hat{\Lambda}_k^\top\|^2\right]\right)^{1/2}\cdot \\ &\left(\mathbb{E}\left[\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N \mathbb{1}^2_{\{e_i^D\neq 0\}}\mathbb{1}^2_{\{e_k^D\neq 0\}}\right]\right)^{1/2} \\ \leq& C\left(\mathbb{E}\left[\sum_{t_j}^{t_{j+1}}\frac{1}{N^2}\sum_{i=1}^N\sum_{k=1}^N G_{i,k}dt\right]\right)^{1/2} \\ \leq& \frac{C}{\sqrt{NM}}\end{aligned}$$

By the same logic it follows that $\sum_{i=1}^N \mathbb{1}_{\{e_{j,i}^D\neq 0, |F_j^D|=0\}} = o_p(1)$. Last but not least

$$\begin{aligned}\|\sum_{i=1}^N \mathbb{1}_{\{F_j^D\Lambda_i=0, e_i^D=0, |F_j^D|\neq 0\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}\| &\leq \|\sum_{i=1}^N \mathbb{1}_{\{|F_j^D|\neq 0\}}\frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}\| \\ &\leq \mathbb{1}_{\{|F_j^D|\neq 0\}}\|\sum_{i=1}^N \frac{\hat{\Lambda}_i\hat{\Lambda}_i^\top}{N}\| \leq O_p\left(\frac{1}{\sqrt{M}}\right).\end{aligned}$$

On the other hand there are only finitely many j for which $e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar\omega}\}} \neq e_{j,i}\mathbb{1}_{\{|e_{j,i}|\leq\alpha\Delta_M^{\bar\omega}\}}$ and the difference is $O_p\left(\frac{1}{\sqrt{M}}\right)$, which does not matter asymptotically for calculating the multipower variation.

Step 6: To show: $\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(\frac{\hat X_j^D\hat\Lambda}{N}\right)\left(\frac{\hat X_j^D\hat\Lambda}{N}\right)^\top\left(\sum_{h\in I(j)}\left(\hat X_{h,i}^C-\frac{\hat X_h^C\hat\Lambda}{N}\hat\Lambda_i\right)^2\right)$
$=\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(H^{-1}F_j\mathbb{1}_{\{|F_j|>\alpha\Delta_M^{\bar\omega}\}}\right)^\top\left(H^{-1}F_j\mathbb{1}_{\{|F_j|>\alpha\Delta_M^{\bar\omega}\}}\right)\left(\sum_{h\in I(j)}\left(e_{h,i}^2\mathbb{1}_{\{|e_{h,i}|\leq\alpha\Delta_M^{\bar\omega}\}}\right)\right)+o_p(1)$
We start by plugging in our results from Steps 2 and 3:

$$\begin{aligned}&\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(\frac{\hat X_j^D\hat\Lambda}{N}\right)\left(\frac{\hat X_j^D\hat\Lambda}{N}\right)^\top\left(\sum_{h\in I(j)}\left(\hat X_{h,i}^C-\frac{\hat X_h^C\hat\Lambda}{N}\hat\Lambda_i\right)^2\right)\\ =&\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(\sum_{i=1}^N\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar\omega}\}}\frac{\hat\Lambda_i\hat\Lambda_i^\top}{N}H^{-1}F_j\right)^\top\left(\sum_{i=1}^N\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar\omega}\}}\frac{\hat\Lambda_i\hat\Lambda_i^\top}{N}H^{-1}F_j\right)\\ &\cdot\left(\sum_{h\in I(j)}\left(e_{h,i}^2\mathbb{1}_{\{|X_{h,i}|\leq\alpha\Delta_M^{\bar\omega}\}}\right)\right)+o_p(1).\end{aligned}$$

We need to show that $\sum_{i=1}^N\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar\omega}\}}\frac{\hat\Lambda_i\hat\Lambda_i^\top}{N}H^{-1}F_j = H^{-1}F_j\mathbb{1}_{\{|F_j|>\alpha\Delta_M^{\bar\omega}\}}+o_p\left(\frac{1}{\sqrt{\delta}}\right)$. This follows from

$$\begin{aligned}&\sum_{i=1}^N\left(\mathbb{1}_{\{|F_j^D\Lambda_i|>0\}}\frac{\hat\Lambda_i\hat\Lambda_i^\top}{N}-I_K\mathbb{1}_{\{|F_j^D|\neq0\}}\right)-\sum_{i=1}^N\mathbb{1}_{\{|F_j^D\Lambda_i|>0,|F_j^D|>0,e_{j,i}^D=0\}}I_K+\sum_{i=1}^N\mathbb{1}_{\{e_{j,i}^D\neq0\}}\frac{\hat\Lambda_i\hat\Lambda_i^\top}{N}\\ =&o_p(1)\end{aligned}$$

which can be shown by the same logic as in step 5.

Step 7: To show: $\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(\hat X_{j,i}^D-\frac{\hat X_j^D\hat\Lambda}{N}\hat\Lambda_i\right)^2\left(\sum_{h\in I(j)}\left(\frac{\hat X_h^C\hat\Lambda}{N}\right)\left(\frac{\hat X_h^C\hat\Lambda}{N}\right)^\top\right)$
$=\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(e_{j,i}^2\mathbb{1}_{\{|e_{j,i}|>\alpha\Delta_M^{\bar\omega}\}}\right)\left(\sum_{h\in I(j)}\left(H^{-1}F_h\mathbb{1}_{\{|F_h|\leq\alpha\Delta_M^{\bar\omega}\}}\right)^\top\left(H^{-1}F_h\mathbb{1}_{\{|F_h|\leq\alpha\Delta_M^{\bar\omega}\}}\right)\right)+o_p(1)$
In light of the previous steps we only need to show how to deal with the first term. By step 4 we have

$$\begin{aligned}&\frac{M}{2k}\sum_{j=k+1}^{M-k}\left(\hat X_{j,i}^D-\frac{\hat X_j^D\hat\Lambda}{N}\hat\Lambda_i\right)^2\left(\sum_{h\in I(j)}\left(\frac{\hat X_h^C\hat\Lambda}{N}\right)\left(\frac{\hat X_h^C\hat\Lambda}{N}\right)^\top\right)\\ =&\frac{M}{2k}\sum_{j\in J}\left(e_{j,i}\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar\omega}\}}+O_p\left(\frac{1}{\delta}\right)+O_P\left(\frac{1}{\sqrt{\delta}}\right)\|F_j\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar\omega}\}}\|\right)^2\\ &\cdot\left(\sum_{h\in I(j)}\left(H^{-1}F_h\mathbb{1}_{\{|F_h|\leq\alpha\Delta_M^{\bar\omega}\}}\right)^\top\left(H^{-1}F_h\mathbb{1}_{\{|F_h|\leq\alpha\Delta_M^{\bar\omega}\}}\right)\right)+o_p(1)\end{aligned}$$

where J denotes the set of jumps of the process $X_i(t)$. Note that J contains only finitely many elements. The difference between $e_{j,i}\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar\omega}\}}$ and $e_{j,i}\mathbb{1}_{\{|e_{j,i}|>\alpha\Delta_M^{\bar\omega}\}}$ is of order $O_p\left(\frac{1}{\sqrt{M}}\right)$ as there might be increments j where there is a jump in the factors but not in the residuals. As we consider only finitely many increments j the result follows. □

Proof of Theorem 1.7:

Proof. Under cross-sectional independence of the error terms the asymptotic variance equals

$$\Theta_F = \underset{N,M\to\infty}{plim} H^\top \frac{\sum_{i=1}^N \Lambda_i[e_i,e_i]\Lambda_i^\top}{N} H$$

By Lemmas A.15 and A.22 we know that $\sum_{j=1}^M \hat{e}_{j,i}\hat{e}_{j,k} = [e_i, e_k] + o_p(1)$ and $\hat{\Lambda}_i = H^\top\Lambda_i + O_p\left(\frac{1}{\sqrt{\delta}}\right)$ and the result follows immediately. □

A.8 Separating Continuous and Jump Factors

Lemma A.23. ***Convergence rates for truncated covariations***
Under Assumptions 1.1 and 1.3 and for some $\alpha > 0$ and $\bar\omega \in \left(0, \frac{1}{2}\right)$ it follows that

$$\frac{1}{N}\sum_{i=1}^N \|\sum_{j=1}^M F_j e_{j,i}\mathbb{1}_{\{|X_{j,i}|\le\alpha\Delta_M^{\bar\omega}\}}\| = O_p\left(\frac{1}{\sqrt{M}}\right) + O_p\left(\frac{1}{N}\right)$$

$$\frac{1}{N}\sum_{i=1}^N \|\sum_{j=1}^M F_j e_{j,i}\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar\omega}\}}\| = O_p\left(\frac{1}{\sqrt{M}}\right)$$

$$\frac{1}{N}\sum_{i=1}^N \|\sum_{j=1}^M \left(e_{j,i}e_{j,k}\mathbb{1}_{\{|X_{j,i}|\le\alpha\Delta_M^{\bar\omega}\}}\mathbb{1}_{\{|X_{j,k}|\le\alpha\Delta_M^{\bar\omega}\}} - [e_i^C, e_k^C]\right)\| = O_p\left(\frac{1}{\sqrt{M}}\right) + O_p\left(\frac{1}{N}\right)$$

$$\frac{1}{N}\sum_{i=1}^N \|\sum_{j=1}^M \left(e_{j,i}e_{j,k}\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar\omega}\}}\mathbb{1}_{\{|X_{j,k}|>\alpha\Delta_M^{\bar\omega}\}} - [e_i^D, e_k^D]\right)\| = O_p\left(\frac{1}{\sqrt{M}}\right).$$

Proof. I will only prove the first statement as the other three statements can be shown analogously. By Theorem A.6

$$\sum_{j=1}^M F_j e_{j,i}\mathbb{1}_{\{\|F_j\|\le\alpha\Delta_M^{\bar\omega},\ e_{j,i}\le\alpha\Delta_M^{\bar\omega}\}} = O_p\left(\frac{1}{\sqrt{M}}\right).$$

However, as F and e_i are not observed our truncation is based on X. Hence we need to characterize

$$\sum_{j=1}^M F_j e_{j,i}\left(\mathbb{1}_{\{\|F_j\|\le\alpha\Delta_M^{\bar\omega},|\ e_{j,i}|\le\alpha\Delta_M^{\bar\omega}\}} - \mathbb{1}_{\{|X_{j,i}|\le\alpha\Delta_M^{\bar\omega}\}}\right).$$

If there is a jump in X, there has to be also a jump in e_i or F. By Assumption 1.3 if there is a jump in e_i or $\Lambda_i^\top F$, there has to be a jump in X. However, it is possible that two factors F_k and F_l jump at the same time but their weighted average $\Lambda_i^\top F$ is equal to zero. Hence, we could not identify these jumps by observing only X_i. This can only happen for a finite number of indices i as $\lim_{N\to\infty} \frac{\Lambda^\top \Lambda}{N} = \Sigma_\Lambda$ has full rank. Hence

$$\frac{1}{N}\sum_{i=1}^{N} \| \sum_{j=1}^{M} F_j e_{j,i} \left(\mathbb{1}_{\{\|F_j\|\leq \alpha\Delta_M^{\bar{\omega}},\ e_{j,i}\leq \alpha\Delta_M^{\bar{\omega}}\}} - \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \right) \| = O_p\left(\frac{1}{N}\right).$$

In the reverse case where we want to consider only the jump part, $|X_{j,i}| > \alpha\Delta_M^{\bar{\omega}}$ implies that either $\Lambda_i^\top F_j$ or $e_{j,i}$ has jumped. If we wrongly classify an increment $e_{j,i}$ as a jump although the jump happened in $\Lambda_i^\top F_j$, it has an asymptotically vanishing effect as we have only a finite number of jumps in total and the increment of a continuous process goes to zero with the rate $O_p\left(\frac{1}{\sqrt{M}}\right)$. □

Proof of Theorem 1.2:

Proof. I only prove the statement for the continuous part. The proof for the discontinuous part is completely analogous.

Step 1: Decomposition of the loading estimator:

First we start with the decomposition in Lemma A.9 that we get from substituting the definition of X into $\frac{1}{N}\hat{X}^{C\top}\hat{X}^C\hat{\Lambda}^C {V_{MN}^C}^{-1} = \hat{\Lambda}^C$. We choose H^C to set $\frac{1}{N}\Lambda^C {F^C}^\top F^C {\Lambda^C}^\top \hat{\Lambda}^C = \Lambda^C H V_{MN}^C$.

$$\begin{aligned} V_{MN}^C\left(\hat{\Lambda}_i^C - {H^C}^\top \Lambda_i^C\right) =& \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C e_{j,k} e_{j,i} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \\ &+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C {\Lambda_k^C}^\top F_j^C e_{j,i} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \\ &+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C e_{j,k} {F_j^C}^\top \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda_i^C + R^C \end{aligned}$$

with

$$\begin{aligned}
R^C =& + \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \Lambda_k^D e_{j,k} F_j^{D^\top} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda_i^C \\
&+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{D^\top} F_j^D e_{j,i} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \\
&+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{D^\top} F_j^D F_j^{D^\top} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda_i^D \\
&+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{C^\top} F_j^C F_j^{D^\top} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda_i^D \\
&+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{D^\top} F_j^C F_j^{D^\top} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda_i^C \\
&+ \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{C^\top} F_j^C F_j^{C^\top} \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}, |X_{j,k}|\leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda_i^C \\
&- \frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N} \hat{\Lambda}_k^C \Lambda_k^{C^\top} F_j^C F_j^{C^\top} \Lambda_i^C \\
=& o_p(1)
\end{aligned}$$

The convergence rate of R^C would be straightforward if the truncations were in terms of F and e_i instead of X. However using the same argument as in Lemma A.23, we can conclude that under Assumption 1.3 at most for a finite number of indices i it holds that $F_j \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}} - F_j \mathbb{1}_{\{\|F_j\|\leq \alpha\Delta_M^{\bar{\omega}}\}} = O_p\left(\frac{1}{\sqrt{\delta}}\right)$ for M sufficiently large and otherwise the difference is equal to 0. Likewise if there is no jump in F $e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}} = e_{j,i}\mathbb{1}_{\{|e_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}}$ except for a finite number of indices. Hence, we have a similar decomposition for $\left(\hat{\Lambda}_i^C - H^{C^\top}\Lambda_i^C\right)$ as in Lemma A.9 using only truncated observations.

Step 2: $\hat{\Lambda}_i^C - H^{C^\top}\Lambda_i^C = O_p\left(\frac{1}{\sqrt{\delta}}\right)$:

We need to show Lemmas A.10 and A.11 for the truncated observations. Note that Proposition A.1 does not hold any more because the truncated residuals are not necessarily local martingales any more. For this reason we obtain a lower convergence rate of $O_p\left(\frac{1}{\sqrt{\delta}}\right)$ instead of $O_p\left(\frac{1}{\delta}\right)$. The statement follows from a repeated use of Lemma A.23.

Step 3: Convergence of $\hat{F}_T^C - H^{C^{-1}}F_T^C$:

We try to extend Theorem 1.4 to the truncated variables. By abuse of notation I denote by $\Lambda^\top F\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}$ the matrix with elements $\Lambda_i^\top F_j \mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}}$ and similarly $e\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}$ is the matrix with elements $e_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}$.

$$
\begin{aligned}
\hat{F}^C - F^C H^{C^{-1}\top} =& \frac{1}{N}\hat{X}^C\hat{\Lambda}^C - F^C H^{C^{-1}\top} \\
=& \frac{1}{N}\left(F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{C^\top} + F^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{D^\top} + e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\right)\hat{\Lambda}^C - F^C H^{C^{-1}\top} \\
=& \frac{1}{N}F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{C^\top}\hat{\Lambda}^C - F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}H^{C^{-1}\top} + F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}H^{C^{-1}\top} \\
&+ \frac{1}{N}F^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{D^\top}\hat{\Lambda}^C + \frac{1}{N}e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\hat{\Lambda}^C - F^C H^{C^{-1}\top} \\
=& \frac{1}{N}F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\Lambda^{C^\top} - H^{C^{-1}\top}\hat{\Lambda}^{C\top}\right)\hat{\Lambda}^C + \left(F^C\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}} - F^C\right)H^{C^{-1}\top} \\
&+ F^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\frac{1}{N}\Lambda^{D^\top}\Lambda^C H^C\right) + F^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\frac{1}{N}\Lambda^{D^\top}\left(\hat{\Lambda}^C - \Lambda^C H^C\right) \\
&+ \frac{1}{N}e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\hat{\Lambda}^C - \Lambda^C H^C\right) + \frac{1}{N}e\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^C H^C.
\end{aligned}
$$

Using the result $\hat{\Lambda}_i^C - H^{C^\top}\Lambda_i^C = O_p\left(\frac{1}{\sqrt{\delta}}\right)$ and a similar reasoning as in Lemma A.23, we conclude that

$$
\hat{F}_T^C - H^{C^{-1}}F_T^C = o_p(1) + \left(\frac{1}{N}\Lambda^{D^\top}\Lambda^C H^C\right)^\top F_T^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}} + \frac{1}{N}H^{C^\top}\Lambda^{C^\top}e_T^\top\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}
$$

The term $F_T^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\left(\frac{1}{N}\Lambda^{D^\top}\Lambda^C H^C\right)$ goes to zero only if F^D has no drift term or Λ^D is orthogonal to Λ^C. Note that in general F^D can be written as a pure jump martingale and a finite variation part. Even when F^D does not jump its value does not equal zero because of the finite variation part. Hence in the limit $F_T^D\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ estimates the drift term of F^D. A similar argument applies to $\frac{1}{N}e_T\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}\Lambda^C H^C$. By definition e_i are local martingales. If the residuals also have a jump component, then this component can be written as a pure jump process minus its compensator, which is a predictable finite variation process. The truncation estimates the continuous part of e_i which is the continuous martingale plus the compensator process of the jump martingale. Hence, in the limit $e_i\mathbb{1}_{\{|X|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ is not martingale any more. In particular the weighted average of the compensator drift process does not vanish. In conclusion, if the jump factor process has a predictable finite variation part or more than finitely many residual terms have a jump component, there will be a predictable finite variation process as bias for the continuous factor estimator.

Step 4: Convergence of quadratic covariation:
The quadratic covariation estimator of the estimator $\hat{F}^C$ with another arbitrary process Y

is

$$\sum_{j=1}^{M} \hat{F^C}_j Y_j = H^{C^{-1}} \sum_{j=1}^{M} F_j^C Y_j + o_p(1) + \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^{C^\top} e_{j,i} Y_j \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}}$$
$$+ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^C \Lambda_i^{D^\top} F_j^D Y_j \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}}.$$

The first term converges to the desired quantity. Hence, we need to show that the other two terms go to zero.

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^{C^\top} e_{j,i} Y_j \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} = \frac{1}{N} \sum_{i=1}^{N} H^{C^\top} \Lambda_i^{C^\top} [e_i^C, Y]_T$$
$$+ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^{C^\top} e_{j,i} Y_j \left(\mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} - \mathbb{1}_{\{|e_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} \right)$$
$$+ \frac{1}{N} \sum_{i=1}^{N} H^{C^\top} \Lambda_i^{C^\top} \left(\sum_{j=1}^{M} e_{j,i} Y_j \mathbb{1}_{\{|e_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} - [e_i^C, Y]_T \right)$$

The last two term are $o_p(1)$ by a similar argument as in Lemma A.23. Applying the Cauchy Schwartz inequality and Assumption 1.1 to the first term yields

$$\| \frac{1}{N} \sum_{i=1}^{N} H^{C^\top} \Lambda_i^{C^\top} [e_i^C, Y]_T \|^2 \leq \| \frac{1}{N^2} H^{C^\top} \Lambda^{C^\top} [e^C, e^C]_T \Lambda^C H^C \| \cdot \| [Y, Y]_T \| = O_p\left(\frac{1}{N}\right)$$

Thus Assumption 1.1 implies that $\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^{C^\top} [e_i^C, Y]_T = O_p\left(\frac{1}{\sqrt{N}}\right)$. The last result follows from that fact that the quadratic covariation of a predictable finite variation process with a semimartingale is zero and $F_j^D \mathbb{1}_{\{\|F_j^D\| \leq \alpha \Delta_M^{\bar{\omega}}\}}$ converges to a predictable finite variation term:

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^C \Lambda_i^{D^\top} F_j^D Y_j \mathbb{1}_{\{|X_{j,i}| \leq \alpha \Delta_M^{\bar{\omega}}\}} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} H^{C^\top} \Lambda_i^C \Lambda_i^{D^\top} F_j^D Y_j \mathbb{1}_{\{\|F_j^D\| \leq \alpha \Delta_M^{\bar{\omega}}\}} + o_p(1)$$
$$= o_p(1)$$

□

A.9 Estimation of the Number of Factors

Lemma A.24. ***Weyl's eigenvalue inequality***
For any $M \times N$ matrices Q_i we have

$$\lambda_{i_1 + \ldots + i_K - (K-1)} \left(\sum_{k=1}^{K} Q_k \right) \leq \lambda_{i_1}(Q_1) + \ldots + \lambda_{i_K}(Q_K)$$

where $1 \leq i_1, ..., i_K \leq \min(N, M))$*,* $1 \leq i_1 + ... + i_K - (K-1) \leq \min(N, M)$ *and* $\lambda_i(Q)$ *denotes the* i*th largest singular value of matrix* Q*, which is another name for the square root of the* i*th largest eigenvalue of matrix* $QQ^\top$*.*

Proof. See Theorem 3.3.16 in Horn and Johnson (1991). □

Lemma A.25. ***Bound on non-systematic eigenvalues***
Assume Assumption 1.1 holds and $O\left(\frac{N}{M}\right) \leq O(1)$*. Then*

$$\lambda_k(X^\top X) \leq O_p(1) \qquad \text{for } k \geq K+1.$$

Proof. Note that the singular values of a symmetric matrix are equal to the eigenvalues of this matrix. By Weyl's inequality for singular values in Lemma A.24 we obtain

$$\lambda_k(X) \leq \lambda_k(F\Lambda^\top) + \lambda_1(e).$$

As $\lambda_k(F\Lambda^\top) = 0$ for $k \geq K+1$, we conclude

$$\lambda_k(X^\top X) \leq \lambda_1(e^\top e) \qquad \text{for } k \geq K+1$$

Now we need to show that $\lambda_k(e^\top e) \leq O_p(1)$ $\forall k \in [1, N]$. We start with a decomposition

$$\begin{aligned}\lambda_k(e^\top e) &= \lambda_k(e^\top e - [e,e] + [e,e]) \\ &\leq \lambda_1(e^\top e - [e,e]) + \lambda_k([e,e]).\end{aligned}$$

By Assumption 1.1 $[e,e]$ has bounded eigenvalues, which implies $\lambda_k([e,e]) = O_p(1)$.
Denote by V the eigenvector of the largest eigenvalue of $(e^\top e - [e,e])$.

$$\begin{aligned}\lambda_1\left(e^\top e - [e,e]\right) &= V^\top\left(e^\top e - [e,e]\right)V \\ &= \sum_{i=1}^N \sum_{l=1}^N V_i\left(e_i^\top e_l - [e_i, e_l]\right)V_l \\ &\leq \left(\sum_{i=1}^N\left(\sum_{l=1}^N (e_i^\top e_l - [e_i,e_l])V_l\right)^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^N V_i^2\right)^{\frac{1}{2}} \\ &\leq \left(\sum_{i=1}^N\left(\sum_{l=1}^N (e_i^\top e_l - [e_i,e_l])V_l\right)^2\right)^{\frac{1}{2}}\end{aligned}$$

as V is an orthonormal vector. Apply Proposition A.1 with $Y = e_i$ and $\bar{Z} = \sum_{l=1}^N e_l V_l$. Note that $[\bar{Z}] = V^\top[e,e]V$ is bounded. Hence

$$\sum_{l=1}^N \left(e_i^\top e_l - [e_i, e_l]\right)V_l = O_p\left(\frac{1}{\sqrt{M}}\right).$$

Therefore

$$\lambda_1\left(e^\top e-[e,e]\right)=\left(\sum_{i=1}^{N}O_p\left(\frac{1}{M}\right)\right)^{\frac{1}{2}}\leq O_p\left(\frac{\sqrt{N}}{\sqrt{M}}\right)\leq O_p(1).$$

□

Lemma A.26. ***Bound on systematic eigenvalues***
Assume Assumption 1.1 holds and $O\left(\frac{N}{M}\right)\leq O(1)$. *Then*

$$\lambda_k(X^\top X)=O_p(N)\qquad \text{for } k=1,...,K$$

Proof. By Weyl's inequality for singular values in Lemma A.24:

$$\lambda_k(F\Lambda^\top)\leq\lambda_k(X)+\lambda_1(-e)$$

By Lemma A.25 the last term is $\lambda_1(-e)=-\lambda_N(e)=O_p(1)$. Therefore

$$\lambda_k(X)\geq\lambda_k(F\Lambda^\top)+O_p(1)$$

which implies $\lambda_k(X^\top X)\geq O_p(N)$ as $\left(F^\top F\frac{\Lambda^\top\Lambda}{N}\right)$ has bounded eigenvalues for $k=1,...,K$. On the other hand

$$\lambda_k(X)\leq\lambda_k(F\Lambda^\top)+\lambda_1(e)$$

and $\lambda_1(e)=O_p(1)$ implies $\lambda_k(X^\top X)\leq O_p(N)$ for $k=1,...,K$. □

Lemma A.27. ***Bounds on truncated eigenvalues***
Assume Assumptions 1.1 and 1.3 hold and $O\left(\frac{N}{M}\right)\leq O(1)$. *Set the threshold identifier for jumps as* $\alpha\Delta_M^{\bar{\omega}}$ *for some* $\alpha>0$ *and* $\bar{\omega}\in\left(0,\frac{1}{2}\right)$ *and define* $\hat{X}^C_{j,i}=X_{j,i}\mathbb{1}_{\{|X_{j,i}|\leq\alpha\Delta_M^{\bar{\omega}}\}}$ *and* $\hat{X}^D_{j,i}=X_{j,i}\mathbb{1}_{\{|X_{j,i}|>\alpha\Delta_M^{\bar{\omega}}\}}$. *Then*

$$\begin{aligned}
\lambda_k\left(\hat{X}^{C\top}\hat{X}^C\right)&=O_p(N) && k=1,...,K_C\\
\lambda_k\left(\hat{X}^{C\top}\hat{X}^C\right)&\leq O_p(1) && k=K_C+1,...,N\\
\lambda_k\left(\hat{X}^{D\top}\hat{X}^D\right)&=O_p(N) && k=1,...,K_D\\
\lambda_k\left(\hat{X}^{D\top}\hat{X}^D\right)&\leq O_p(1) && k=K_D+1,...,N
\end{aligned}$$

where K^C *is the number of factors that contain a continuous part and* K^D *is the number of factors that have a jump component.*

Proof. By abuse of notation the vector $\mathbb{1}_{\{|e|\leq \alpha\Delta_M^{\bar{\omega}}\}}e$ has the elements $\mathbb{1}_{\{|e_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}}e_{j,i}$. e^C is the continuous martingale part of e and e^D denotes the jump martingale part.
Step 1: To show: $\lambda_k\left(\left(\mathbb{1}_{\{|e|\leq \alpha\Delta_M^{\bar{\omega}}\}}e\right)^\top\left(\mathbb{1}_{\{|e|\leq \alpha\Delta_M^{\bar{\omega}}\}}e\right)\right)\leq O_p(1)$ for $k=1,...,N$.
By Lemma A.24 it holds that

$$\lambda_k(\mathbb{1}_{\{|e|\leq \alpha\Delta_M^{\bar{\omega}}\}}e)\leq \lambda_1(\mathbb{1}_{\{|e|\leq \alpha\Delta_M^{\bar{\omega}}\}}e-e^C)+\lambda_k(e^C)$$

Lemma A.25 applied to e^C implies $\lambda_k(e^C)\leq O_p(1)$. The difference between the continuous martingale part of e and the truncation estimator $\mathbb{1}_{\{|e|\leq \alpha\Delta_M^{\bar{\omega}}\}}e-e^C$ equals a drift term from the jump martingale part plus a vector with finitely many elements that are of a small order:

$$\mathbb{1}_{\{|e_i|\leq \alpha\Delta_M^{\bar{\omega}}\}}e_i-e_i^C=b_{e_i}+d_{e_i}$$

where b_{e_i} is a vector that contains the finite variation part of the jump martingales which is classified as continuous and d_{e_i} is a vector that contains the negative continuous part $-e_{j,i}^C$ for the increments j that are correctly classified as jumps and hence are set to zero in $\mathbb{1}_{\{|e_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}}e_{j,i}$. Using the results of Mancini (2009) we have $\mathbb{1}_{\{e_{j,i}^D=0\}}=\mathbb{1}_{\{|e_{j,i}|\leq\alpha\Delta^{\bar{\omega}}\}}$ almost surely for sufficiently large M and hence we can identify all the increments that contain jumps. Note, that by Assumption 1.3 we have only finitely many jumps for each time interval and therefore d_{e_i}has only finitely many elements not equal to zero. By Lemma A.24 we have

$$\lambda_1(\mathbb{1}_{\{|e|\leq \alpha\Delta_M^{\bar{\omega}}\}}e-e^C)\leq\lambda_1(b_e)+\lambda_1(d_e)$$

It is well-known that the spectral norm of a symmetric $N\times N$ matrix A is bounded by N times its largest element: $\|A\|_2\leq N\max_{i,k}|A_{i,k}|$. Hence

$$\lambda_1(b_e^\top b_e)\leq N\cdot\max_{k,i}|b_{e_i}^\top b_{e_k}|\leq N\cdot O_p\left(\frac{1}{M}\right)\leq O_p\left(\frac{N}{M}\right)\leq O_p(1)$$

where we have use the fact that the increments of a finite variation term are of order $O_p\left(\frac{1}{M}\right)$. Similarly

$$\lambda_1\left(d_e^\top d_e\right)\leq N\cdot\max_{k,i}|d_{e_i}^\top d_{e_k}|\leq N\cdot O_p\left(\frac{1}{M}\right)\leq O_p\left(\frac{N}{M}\right)\leq O_p(1)$$

as d_{e_i} has only finitely many elements that are not zero and those are of order $O_p\left(\frac{1}{\sqrt{M}}\right)$.

Step 2: To show: $\lambda_k\left(\left(\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}e\right)^\top\left(\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}e\right)\right)\leq O_p(1)$ for $k=1,...,N$.
Here we need to show that the result of step 1 still holds, when we replace $\mathbb{1}_{\{|e_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}}$ with $\mathbb{1}_{\{|X_{j,i}|\leq \alpha\Delta_M^{\bar{\omega}}\}}$. It is sufficient to show that

$$\lambda_1\left(e\mathbb{1}_{\{|e|\leq\alpha\Delta^{\bar{\omega}}\}}-e\mathbb{1}_{\{|X|\leq\alpha\Delta^{\bar{\omega}}\}}\right):=\lambda_1(h)=O_p(1)$$

As by Assumption 1.3 only finitely many elements of h are non-zero and those are of order $O_p\left(\frac{1}{\sqrt{M}}\right)$, it follows that

$$\lambda_1(h) \leq N \max_{k,i} |h_i^\top h_k| \leq O_p\left(\frac{N}{M}\right) \leq O_p(1).$$

Step 3: To show: $\lambda_k(\hat{X}^{C\top}\hat{X}^C) \leq O_p(1)$ for $k \geq K_C + 1$.
By definition the estimated continuous movements are

$$\hat{X}^C = F^C \mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^C + F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{\text{pure jump}\top} + e\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}$$

where $F^{\text{pure jump}}$ denotes the pure jump factors that do not have a continuous component and $\Lambda^{\text{pure jump}}$ are the corresponding loadings. By Weyl's inequality for singular values in Lemma A.24 we have

$$\lambda_1\left(\hat{X}^C\right) \leq \lambda_1\left(F^C \mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^C\right) + \lambda_1\left(F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{\text{pure jump}\top}\right) + \lambda_1\left(e\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right)$$

For $k \geq K+1$ the first term vanishes $\lambda_1\left(F^C \mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^C\right) = 0$ and by step 2 the last term is $\lambda_1\left(e\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right) = O_p(1)$. The second term can be bounded by

$$\begin{aligned}\lambda_1\left(F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{\text{pure jump}\top}\right)^2 \leq & \|\Lambda^{\text{pure jump}\top}\Lambda^{\text{pure jump}}\|_2^2 \cdot \\ & \|\left(F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right)^\top F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\|_2^2\end{aligned}$$

The first factor is $\|\Lambda^{\text{pure jump}\top}\Lambda^{\text{pure jump}}\|_2^2 = O(N)$, while the truncated quadratic covariation in the second factor only contains the drift terms of the factors denoted by b_{F^D} which are of order $O_p\left(\frac{1}{M}\right)$:

$$\left\|\left(F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right)^\top F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right\|_2^2 \leq \|b_{F^D}{}^\top b_{F^D}\|_2^2 \leq O_p\left(\frac{1}{M}\right)$$

Step 4: To show: $\lambda_k\left(\left(\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}e\right)^\top\left(\mathbb{1}_{\{|X|>\alpha\Delta_M^{\bar{\omega}}\}}e\right)\right) \leq O_p(1)$ for $k = 1, ..., N$.
We decompose the truncated error terms into two components.

$$\lambda_k(\mathbb{1}_{\{|e|>\alpha\Delta_M^{\bar{\omega}}\}}e) > \lambda_1(\mathbb{1}_{\{|e|>\alpha\Delta_M^{\bar{\omega}}\}}e - e^D) + \lambda_k(e^D).$$

By Proposition A.1 the second term is $O_p(1)$. For the first term we can apply a similar logic as in step 1. Then we use the same arguments as in step 2.
Step 5: To show: $\lambda_k\left(\hat{X}^{C\top}\hat{X}^C\right) = O_p(N)$ for $k = 1, ..., K^C$.
By Lemma A.24 the first K^C singular values satisfy the inequality

$$\begin{aligned}&\lambda_k\left(F^C \mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{C\top}\right) \\ \leq & \lambda_k\left(\hat{X}^C\right) + \lambda_1\left(-F^{\text{pure jump}}\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\Lambda^{\text{pure jump}\top}\right) + \lambda_1\left(-e\mathbb{1}_{\{|X|\leq \alpha\Delta_M^{\bar{\omega}}\}}\right).\end{aligned}$$

Hence by the previous steps

$$\lambda_k\left(\hat{X}^C\right) \geq \lambda_k\left(F^C \mathbb{1}_{\{|X| \leq \alpha\Delta_M^{\bar{\omega}}\}} \Lambda^{C^\top}\right) + O_p(1).$$

By Assumption 1.1 for $k = 1, ..., K_C$

$$\lambda_k^2\left(F^C \Lambda^{C^\top}\right) = \lambda_k\left(F^{C^\top} F^C \frac{\Lambda^{C^\top}\Lambda^C}{N}\right) N = O_p(N).$$

On the other hand

$$\lambda_k\left(F^C \mathbb{1}_{\{|X| \leq \alpha\Delta^{\bar{\omega}}\}} \Lambda^{C^\top} - F^C \Lambda^{C^\top}\right)^2 \leq O_p\left(\frac{N}{M}\right) \leq O_p(1)$$

where we have used the fact that the difference between a continuous factor and the truncation estimator applied to the continuous part is just a finite number of terms of order $O_p\left(\frac{1}{\sqrt{M}}\right)$. Hence

$$\lambda_k^2\left(F^C \mathbb{1}_{\{|X| \leq \alpha\Delta^{\bar{\omega}}\}} \Lambda^{C^\top}\right) = O_p(N)$$

Similarly we get the reverse inequality for $\hat{X}^C$:

$$\lambda_k\left(\hat{X}^C\right) \leq \lambda_k\left(F^C \mathbb{1}_{\{|X| \leq \alpha\Delta^{\bar{\omega}}\}} \Lambda^{C^\top}\right) + \lambda_1\left(F^{\text{pure jump}} \mathbb{1}_{\{|X| \leq \alpha\Delta^{\bar{\omega}}\}} \Lambda^{\text{pure jump}^\top}\right) + \lambda_1\left(e \mathbb{1}_{\{|X| \leq \alpha\Delta^{\bar{\omega}}\}}\right)$$

which yields

$$O_p(N) \leq \lambda_k\left(\hat{X}^{C\top}\hat{X}^C\right) \leq O_p(N)$$

Step 6: To show: $\lambda_k\left(\hat{X}^{D\top}\hat{X}^D\right) = O_p(N)$ for $k = 1, ..., K^D$.
Analogous to step 5. □

Proof of Theorem 1.9:

Proof. I only prove the result for $\hat{K}(\gamma)$. The results for $\hat{K}^C(\gamma)$ and $\hat{K}^D(\gamma)$ follow exactly the same logic.
Step 1: ER_k for $k = K$
By Lemmas A.25 and A.26 the eigenvalue ratio statistic for $k = K$ is asymptotically

$$ER_k = \frac{\lambda_K + g}{\lambda_{K+1} + g} = \frac{\frac{O_p(N)}{g} + 1}{\frac{\lambda_{K+1}}{g} + 1} = \frac{\frac{O_p(N)}{g} + 1}{o_p(1) + 1} = O_p\left(\frac{N}{g}\right) \to \infty$$

Step 2: ER_k for $k \geq K+1$

$$ER_k = \frac{\lambda_k + g}{\lambda_{k+1} + g} = \frac{\frac{\lambda_k}{g} + 1}{\frac{\lambda_{k+1}}{g} + 1} = \frac{o_p(1) + 1}{o_p(1) + 1} = 1 + o_p(1).$$

Step 3: To show: $\hat{K}(\gamma) \xrightarrow{p} K$
As ER_k goes in probability to 1 for $k \geq K+1$ and grows without bounds for $k = K$, the probability for $ER_k > 1$ goes to zero for $k \geq K+1$ and to 1 for $k = K$.
Remark: Although it is not needed for this proof, note that for $k = 1, ..., K-1$

$$ER_k = \frac{\lambda_k + g}{\lambda_{k+1} + g} = \frac{O_p(N) + g}{O_p(N) + g} = \frac{O_p(1) + \frac{g}{N}}{O_p(1) + \frac{g}{N}} = O_p(1).$$

□

Proof of Proposition 1.1:

Proof. Apply Theorem A.7 to $\frac{1}{\sqrt{M}} X_{j,i} = \frac{1}{\sqrt{M}} F_j \Lambda_i^\top + \frac{1}{\sqrt{M}} e_{j,i}$. Note that $\frac{1}{\sqrt{M}} e$ can be written as $\frac{1}{\sqrt{M}} e = A\epsilon$ with $\epsilon_{j,i}$ being *i.i.d.* $(0,1)$ random variables with finite fourth moments. □

A.10 Identifying the Factors

Proof of Theorem 1.11:

Proof. Define

$$B = \begin{pmatrix} F^\top F & F^\top G \\ G^\top F & G^\top G \end{pmatrix} \qquad \hat{B} = \begin{pmatrix} \hat{F}^\top \hat{F} & \hat{F}^\top G \\ G^\top \hat{F} & G^\top G \end{pmatrix} \qquad B^* = \begin{pmatrix} H^{-1} F^\top F H^{-1^\top} & H^{-1} F^\top G \\ G^\top F H^{-1^\top} & G^\top G \end{pmatrix}.$$

As the trace is a linear function it follows that $\sqrt{M}\left(trace(B) - trace(\hat{B})\right) \xrightarrow{p} 0$ if $\sqrt{M}(B - \hat{B}) \xrightarrow{p} 0$. By assumption H is full rank and the trace of B is equal to the trace of B^*. Thus it is sufficient to show that $\sqrt{M}(\hat{B} - B^*) \xrightarrow{p} 0$. This follows from

(i) $\sqrt{M}\left((\hat{F}^\top \hat{F})^{-1} - (H^{-1} F^\top F H^{-1^\top})^{-1}\right) \xrightarrow{p} 0$

(ii) $\sqrt{M}\left(\hat{F}^\top G - H^{-1} F^\top G\right) \xrightarrow{p} 0.$

We start with (i). As

$$(\hat{F}^\top \hat{F})^{-1} - (H^{-1} F^\top F H^{-1^\top})^{-1} = (\hat{F}^\top \hat{F})^{-1} \left(H^{-1} F^\top F H^{-1^\top} - \hat{F}^\top \hat{F}\right) \left(H^{-1} F^\top F H^{-1^\top}\right)^{-1}$$

it is sufficient to show

$$\sqrt{M}\left(H^{-1}F^\top F H^{-1^\top} - \hat{F}^\top\hat{F}\right) = \sqrt{M}H^{-1}F^\top(FH^{-1^\top} - \hat{F}) + \sqrt{M}(H^{-1}F^\top - \hat{F}^\top)\hat{F} \xrightarrow{p} 0$$

It is shown in the proof of Theorem 1.4 that

$$\hat{F} - FH^{-1^\top} = \frac{1}{N}F(\Lambda - \hat{\Lambda}H^{-1})^\top\hat{\Lambda} + \frac{1}{N}e(\hat{\Lambda} - \Lambda H) + \frac{1}{N}e\Lambda H.$$

Hence the first term equals

$$-H^{-1}F^\top(\hat{F} - FH^{-1^\top}) = \frac{1}{N}H^{-1}F^\top F(\Lambda - \hat{\Lambda}H^{-1})^\top\hat{\Lambda} + \frac{1}{N}H^{-1}F^\top e(\hat{\Lambda} - \Lambda H) + \frac{1}{N}H^{-1}F^\top e\Lambda H$$

Lemmas A.10 and A.16 applied to the first summand yield $\frac{1}{N}H^{-1}F^\top F(\Lambda - \hat{\Lambda}H^{-1})^\top\hat{\Lambda} = O_p\left(\frac{1}{\delta}\right)$. Lemmas A.1 and A.10 provide the rate for the second summand as $\frac{1}{N}H^{-1}F^\top e(\hat{\Lambda} - \Lambda H) = O_p\left(\frac{1}{\delta}\right)$. Lemma A.1 bounds the third summand: $\frac{1}{N}H^{-1}F^\top e\Lambda H = O_p\left(\frac{1}{\sqrt{NM}}\right)$.

For the second term note that

$$\left(H^{-1}F^\top - \hat{F}^\top\right)\hat{F} = \left(H^{-1}F^\top - \hat{F}^\top\right)\left(FH^{-1^\top} - \hat{F}\right) + \left(H^{-1}F^\top - \hat{F}^\top\right)FH^{-1^\top}$$

Based on Lemmas A.10 and A.16 it is easy to show that $\left(H^{-1}F^\top - \hat{F}^\top\right)\left(FH^{-1^\top} - \hat{F}\right) = O_p\left(\frac{1}{\delta}\right)$.

Term (ii) requires the additional assumptions on G:

$$\left(\hat{F}^\top - H^{-1}F^\top\right)G = \left(\frac{1}{N}\hat{\Lambda}^\top\left(\Lambda - \hat{\Lambda}H^{-1}\right)F^\top G + \frac{1}{N}\left(\hat{\Lambda} - \Lambda H\right)^\top e^\top G + \frac{1}{N}H^\top\Lambda^\top e^\top G.\right.$$

By Lemma A.16 it follows that $\left(\frac{1}{N}\hat{\Lambda}^\top\left(\Lambda - \hat{\Lambda}H^{-1}\right)\right)F^\top G = O_p\left(\frac{1}{\delta}\right)$. Now let's first assume that G is independent of e. Then Proposition A.1 applies and $\frac{1}{N}H^\top\Lambda e^\top G = O_p\left(\frac{1}{\sqrt{NM}}\right)$. Otherwise assume that $G = \frac{1}{N}\sum_{i=1}^N X_i w_i^\top = F\frac{1}{N}\sum_{i=1}^N \Lambda_i w_i^\top + \frac{1}{N}\sum_{i=1}^N e_i w_i^\top$. Proposition A.1 applies to

$$\frac{1}{N}H^\top\Lambda e^\top F\left(\frac{1}{N}\sum_{i=1}^N \Lambda_i w_i^\top\right) = O_p\left(\frac{1}{\sqrt{NM}}\right)$$

and

$$\frac{1}{N}\sum_{i=1}^N\left(\frac{1}{N}H^\top\Lambda^\top\left(e^\top e_i - [e, e_i]\right)\right)w_i^\top = O_p\left(\frac{1}{\sqrt{NM}}\right)$$

separately. As by Assumption 1.2

$$\sum_{i=1}^{N} \frac{1}{N^2} H^\top \Lambda^\top [e, e_i] w_i^\top = \frac{1}{N^2} \left(\sum_{i=1}^{N} \sum_{k=1}^{N} H^\top \Lambda_k [e_k, e_i] w_i^\top \right) = O_p\left(\frac{1}{N}\right)$$

the statement in (ii) follows. The distribution result is a consequence of the delta method for the function

$$f\left(\begin{pmatrix} [F,F] \\ [F,G] \\ [G,F] \\ [G,G] \end{pmatrix}\right) = trace\left([F,F]^{-1}[F,G][G,G]^{-1}[G,F]\right)$$

which has the partial derivates

$$\begin{aligned}
\frac{\partial f}{\partial [F,F]} &= -\left([F,F]^{-1}[F,G][G,G]^{-1}[G,F][F,F]^{-1}\right)^\top \\
\frac{\partial f}{\partial [F,G]} &= [F,F]^{-1}[F,G][G,G]^{-1} \\
\frac{\partial f}{\partial [G,F]} &= [G,G]^{-1}[G,F][F,F]^{-1} \\
\frac{\partial f}{\partial [G,G]} &= -\left([G,G]^{-1}[G,F][F,F]^{-1}[F,G][G,G]^{-1}\right)^\top
\end{aligned}$$

Hence

$$\sqrt{M}\left(\hat{\bar{\rho}} - \bar{\rho}\right) = \xi^\top \sqrt{M}\left(vec\left(\begin{pmatrix} [F,F] & [F,G] \\ [G,F] & [G,G] \end{pmatrix} - B\right)\right) + \sqrt{M} \cdot trace\left(B^* - \hat{B}\right)$$

The last term is $O_p\left(\frac{\sqrt{M}}{\delta}\right)$ which goes to zero by assumption. □

Proof of Theorem 1.12:

Proof. The theorem is a consequence of Theorem 1.11 and Section 6.1.3 in Aït-Sahalia and Jacod (2014). □

A.11 Microstructure Noise

Lemma A.28. ***Limits of extreme eigenvalues***
Let Z be a $M \times N$ double array of independent and identically distributed random variables

with zero mean and unit variance. Let $S = \frac{1}{M} Z^\top Z$. *Then if* $\mathbb{E}[|Z_{11}|^4] < \infty$, *as* $M \to \infty$, $N \to \infty$, $\frac{N}{M} \to c \in (0,1)$, *we have*

$$\lim \lambda_{\min}(S) = \left(1 - \sqrt{c}\right)^2 \qquad a.s.$$
$$\lim \lambda_{\max}(S) = \left(1 + \sqrt{c}\right)^2 \qquad a.s.$$

where $\lambda_i(S)$ *denotes the* i*th eigenvalue of* S.

Proof. See Bai and Yin (1993) □

Proof of Theorem 1.10:

Proof. **Step 1:** To show: $\lambda_1\left(\frac{(e+\epsilon)^\top(e+\epsilon)}{N}\right) - \lambda_1\left(\frac{e^\top e}{N}\right) \leq \lambda_1\left(\frac{\epsilon^\top \epsilon}{N}\right) + \lambda_1\left(\frac{e^\top \epsilon}{N} + \frac{\epsilon^\top e}{N}\right)$
This is an immediate consequence of Weyl's eigenvalue inequality Lemma A.24 applied to the matrix

$$\frac{(e+\epsilon)^\top(e+\epsilon)}{N} = \frac{e^\top e}{N} + \frac{\epsilon^\top \epsilon}{N} + \frac{e^\top \epsilon}{N} + \frac{\epsilon^\top e}{N}.$$

Step 2: To show: $\lambda_1\left(\frac{e^\top \epsilon}{N} + \frac{\epsilon^\top e}{N}\right) = O_p\left(\frac{1}{N}\right)$
Let V be the eigenvector for the largest eigenvalue of $\frac{e^\top \epsilon}{N} + \frac{\epsilon^\top e}{N}$. Then

$$\begin{aligned}\lambda_1\left(\frac{e^\top \epsilon}{N} + \frac{\epsilon^\top e}{N}\right) &= V^\top \frac{e^\top \epsilon}{N} V + V^\top \frac{\epsilon^\top e}{N} V \\ &= 2\frac{1}{N}\sum_{j=1}^{M}\sum_{i=1}^{N}\sum_{k=1}^{N} V_i \epsilon_{j,i} e_{j,i} V_k.\end{aligned}$$

Define $\bar{\epsilon}_j = \sum_{i=1}^N V_i \epsilon_{j,i}$ and $\bar{e}_j = \sum_{k=1}^N V_k e_{j,k}$. As can be easily checked $\bar{\epsilon}_j \bar{e}_j$ form a martingale difference sequence and hence we can apply Burkholder's inequality in Lemma A.30:

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{j=1}^{M} \bar{\epsilon}_j \bar{e}_j\right)^2\right] &\leq C\sum_{j=1}^{M} \mathbb{E}\left[\bar{\epsilon}_j^2 \bar{e}_j^2\right] \leq C\sum_{j=1}^{M} \mathbb{E}\left[\bar{\epsilon}_j^2\right]\mathbb{E}\left[\bar{e}_j^2\right] \leq \frac{C}{M}\sum_{j=1}^{M}\mathbb{E}\left[\bar{\epsilon}_j^2\right] \\ &\leq \frac{C}{M}\mathbb{E}\left[\left(\sum_{i=1}^{N} V_i \epsilon_{j,i}\right)^2\right] \leq \frac{C}{M}\sum_{i=1}^{N} V_i^2 \mathbb{E}\left[\epsilon_{j,i}^2\right] \leq C.\end{aligned}$$

We have used the Burkholder inequality to conclude $\mathbb{E}\left[\bar{e}_j^2\right] \leq C V^\top \mathbb{E}[\Delta_j \langle e, e\rangle] V \leq \frac{C}{M}$. This shows that $V^\top \frac{e^\top \epsilon}{N} V = O_p\left(\frac{1}{N}\right)$.

Step 3: To show: $\lambda_1\left(\frac{\epsilon^\top\epsilon}{N}\right) \leq \frac{1}{c}\left(1+\sqrt{c}\right)^2 \lambda_1(B^\top B)\sigma_\epsilon^2 + o_p(1)$
Here we define B as

$$B = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

and note that $\epsilon = B\tilde{\epsilon}$ (up to the boundaries which do not matter asymptotically). Now we can split the spectrum into two components:

$$\lambda_1\left(\frac{\epsilon^\top\epsilon}{N}\right) = \lambda_1\left(\frac{\tilde{\epsilon}^\top B^\top B\tilde{\epsilon}}{N}\right) \leq \lambda_1\left(\frac{\tilde{\epsilon}^\top\tilde{\epsilon}}{N}\right)\lambda_1\left(B^\top B\right).$$

By Lemma A.28 it follows that

$$\lambda_1\left(\frac{\tilde{\epsilon}^\top\tilde{\epsilon}}{N}\right) = \frac{1}{c}\left((1+\sqrt{c})^2\sigma_\epsilon^2\right) + o_p(1).$$

Step 4: To show: $\sigma_\epsilon^2 \leq \frac{c}{\left(1-\sqrt{c}\right)^2}\frac{\lambda_s\left(\frac{Y^\top Y}{N}\right)}{\lambda_{s+K}(B^\top B)} + o_p(1)$
Weyl's inequality for singular values Lemma A.24 implies

$$\lambda_{s+K}(e+\epsilon) \leq \lambda_{K+1}(F\Lambda^\top) + \lambda_s(Y) \leq \lambda_s(Y)$$

as $\lambda_{K+1}(F\Lambda^\top) = 0$. Lemma A.6 in Ahn and Horenstein (2013) says that if A and B are $N \times N$ positive semidefinite matrices, then $\lambda_i(A) \leq \lambda_i(A+B)$ for $i = 1, ..., N$. Combining this lemma with step 2 of this proof, we get

$$\lambda_{s+K}\left(\frac{\epsilon^\top\epsilon}{N}\right) \leq \lambda_s\left(\frac{Y^\top Y}{N}\right)$$

Lemma A.4 in Ahn and Horenstein (2013) yields

$$\lambda_N(\tilde{\epsilon}^\top\tilde{\epsilon})\lambda_{s+K}(B^\top B) \leq \lambda_{s+K}(\epsilon^\top\epsilon)$$

Combining this with lemma A.28 gives us

$$\frac{1}{c}\left((1-\sqrt{c})^2\sigma_\epsilon^2\right)\lambda_{s+K}(B^\top B) + o_p(1) \leq \lambda_s\left(\frac{Y^\top Y}{N}\right)$$

Solving for σ_ϵ^2 yields the statement.

Step 5: To show: $\lambda_s(B^\top B) = 2\left(1+\cos\left(\frac{s+1}{N+1}\pi\right)\right)$
$B^\top B$ is a symmetric tridiagonal Toeplitz matrix with 2 on the diagonal and -1 on the off-diagonal. Its eigenvalues are well-known and equal $2 - 2\cos\left(\frac{N-s}{N+1}\pi\right) = 2\left(1+\cos\left(\frac{s+1}{N+1}\pi\right)\right)$.

Step 6: Combining the previous steps.

$$\lambda_1\left(\frac{(e+\epsilon)^\top(e+\epsilon)}{N}\right)-\lambda_1\left(\frac{e^\top e}{N}\right)\leq\left(\frac{1+\sqrt{c}}{1-\sqrt{c}}\right)^2\frac{2\left(1+\cos\left(\frac{2}{N+1}\pi\right)\right)}{2\left(1+\cos\left(\frac{s+1+K}{N}\pi\right)\right)}\lambda_s\left(\frac{Y^\top Y}{N}\right)+o_p(1)$$
$$\leq\left(\frac{1+\sqrt{c}}{1-\sqrt{c}}\right)^2\frac{2}{1+cos\left(\frac{s+K+1}{N}\pi\right)}\lambda_s\left(\frac{Y^\top Y}{N}\right)+o_p(1)$$

for all $s\in[K+1,N_K]$. Here we have used the continuity of the cosinus function. □

A.12 Collection of Limit Theorems

Theorem A.1. ***Localization procedure***
Assume X is a d-dimensional Itô semimartingale on $(\Omega,\mathfrak{F},(\mathfrak{F}_t)_{t\geq0},\mathbb{P})$ defined as

$$X_t=X_0+\int_0^t b_s ds+\int_0^t\sigma_s dW_s+\int_0^t\int_E \mathbb{1}_{\{\|\tilde{\delta}\|\leq1\}}\delta(s,x)(\mu-\nu)(ds,dx)$$
$$+\int_0^t\int_E \mathbb{1}_{\{\|\tilde{\delta}\|>1\}}\delta(s,x)\mu(ds,dx)$$

where W is a d-dimensional Brownian motion and μ is a Poisson random measure on $\mathbb{R}_+\times E$ with $(E,\mathbb{E})$ an auxiliary measurable space on the space $(\Omega,\mathfrak{F},(\mathfrak{F}_t)_{t\geq0},\mathbb{P})$ and the predictable compensator (or intensity measure) of μ is $\nu(ds,dx)=ds\times v(dx)$.

The volatility σ_t is also a d-dimensional Itô semimartingale of the form

$$\sigma_t=\sigma_0+\int_0^t\tilde{b}_s ds+\int_0^t\tilde{\sigma}_s dW_s+\int_0^t\tilde{\sigma}'_s dW'_s+\int_0^t\int_E \mathbb{1}_{\{\|\tilde{\delta}\|\leq1\}}\tilde{\delta}(s,x)(\mu-\nu)(ds,dx)$$
$$+\int_0^t\int_E \mathbb{1}_{\{\|\tilde{\delta}\|>1\}}\tilde{\delta}(s,x)\mu(ds,dx)$$

where W' is another Wiener process independent of (W,μ). Denote the predictable quadratic covariation process of the martingale part by $\int_0^t a_s ds$ and the compensator of $\int_0^t\int_E \mathbb{1}_{\{\|\tilde{\delta}\|>1\}}\tilde{\delta}(s,x)\mu(ds,dx)$ by $\int_0^t\tilde{a}_s ds$.

Assume local boundedness denoted by Assumption H holds for X:

1. *The process b is locally bounded and cádlág.*
2. *The process σ is càdlàg.*
3. *There is a localizing sequence τ_n of stopping times and, for each n, a deterministic nonnegative function Γ_n on E satisfying $\int\Gamma_n(z)^2 v(dz)<\infty$ and such that $\|\delta(\omega,t,z)\|\wedge 1\leq\Gamma_n(z)$ for all (ω,t,z) with $t\leq\tau_n(\omega)$.*

The volatility process also satisfies a local boundedness condition denoted by Assumption K:

1. *The processes* $\tilde{b}$, a *and* $\tilde{a}$ *are locally bounded and progressively measurable*
2. *The process* $\tilde{\sigma}$ *is càdlàg or càglàd and adapted*

We introduce a global boundedness condition for X *denoted by Assumption SH: Assumption H holds and there are a constant* C *and a nonnegative function* Γ *on* E *such that*

$$\|b_t(\omega)\| \leq C \qquad \|\sigma_t(\omega)\| \leq C \qquad \|X_t(\omega)\| \leq C \qquad \|\delta(\omega,t,z)\| \leq \Gamma(z)$$
$$\Gamma(z) \leq C \qquad \int \Gamma(z)^2 v(dz) \leq C.$$

Similarly a global boundedness condition on σ *is imposed and denoted by Assumption SK: We have Assumption K and there are a constant and a nonnegative function* Γ *on* E*, such that Assumption SH holds and also*

$$\|\tilde{b}_t(\omega)\| \leq C \qquad \|\tilde{\sigma}_t(\omega)\| \leq C \qquad \|a_t(\omega)\| \leq C \qquad \|\tilde{a}_t(\omega)\| \leq C.$$

The processes $U^n(X)$ *and* $U(X)$ *are subject to the following conditions, where* X *and* X' *are any two semimartingales that satisfy the same assumptions and* S *is any* $(\mathfrak{F}_t)$*-stopping time:*
$X_t = X'_t$ *a.s.* $\forall t < S \Rightarrow$

- $t < S \Rightarrow U^n(X)_t = U^n(X')_t$ *a.s.*
- *the* $\mathfrak{F}$*-conditional laws of* $(U(X)_t)_{t<S}$ *and* $(U(X')_t)_{t<S}$ *are a.s. equal.*

The properties of interest for us are either one of the following properties:

- *The processes* $U^n(X)$ *converge in probability to* $U(X)$
- *The variables* $U^n(X)_t$ *converge in probability to* $U(X)_t$
- *The processes* $U^n(X)$ *converge stably in law to* $U(X)$
- *The variables* $U^n(X)_t$ *converge stably in law to* $U(X)_t$.

If the properties of interest hold for Assumption SH, then they also hold for Assumption H. Likewise, if the properties of interest hold for Assumption SK, they also hold for Assumption K.

Proof. See Lemma 4.4.9 in Jacod and Protter (2012). □

Theorem A.2. ***Central limit theorem for quadratic variation***
Let X be an Itô semimartingale satisfying Definition A.1. Then the $d \times d$-dimensional processes $\bar{Z}^n$ defined as

$$\bar{Z}^n_t = \frac{1}{\sqrt{\Delta}} \left([X,X]^n_t - [X,X]_{\Delta[t/\Delta]} \right)$$

converges stably in law to a process $\bar{Z} = \left(\bar{Z}^{ij}\right)_{1 \leq i,j \leq d}$ defined on a very good filtered extension $(\tilde{\Omega}, \tilde{\mathfrak{F}}, (\tilde{\mathfrak{F}}_t)_{t\geq 0}, \tilde{\mathbb{P}})$ of $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$ and which, conditionally on $\mathfrak{F}$, is centered with independent increments and finite second moments given by

$$\mathbb{E}\left[\bar{Z}^{ij}_t \bar{Z}^{kl}_t | \mathfrak{F}\right] = \frac{1}{2} \sum_{s \leq t} \Bigg(\Delta X^i_s \Delta X^k_s (c^{jl}_{s-} + c^{jl}_s) + \Delta X^i_s \Delta X^l_s (c^{jk}_{s-} + c^{jk}_s) + \Delta X^j_s \Delta X^k_s (c^{il}_{s-} + c^{il}_s) + \Delta X^j_s \Delta X^l_s (c^{ik}_{s-} + c^{ik}_s) \Bigg) + \int_0^t \left(c^{ik}_s c^{jl}_s + c^{il}_s c^{jk}_s \right) ds$$

with $c_t = \sigma_t^\top \sigma_t$. This process $\bar{Z}$ is $\mathfrak{F}$-conditionally Gaussian, if the process X and σ have no common jumps.

Moreover, the same is true of the process $\frac{1}{\sqrt{\Delta}}([X,X]^n - [X,X])$, when X is continuous, and otherwise for each t we have the following stable convergence of variables

$$\frac{1}{\sqrt{\Delta}} \left([X,X]^n_t - [X,X]_t \right) \overset{L-s}{\to} \bar{Z}_t.$$

Proof. See Jacod and Protter (2013) Theorem 5.4.2. □

Theorem A.3. ***Consistent Estimation of Covariance in Theorem A.2***
We want to estimate

$$D_t = \sum_{s \leq t} |\Delta X|^2 (\sigma_{s-} + \sigma_s)$$

Let X be an Itô semimartingale satisfying Definition A.1. In addition for some $0 \leq r < 1$ it satisfies the stronger assumption that there is a localizing sequence τ_n of stopping times and for each n a deterministic nonnegative function Γ_n on E satisfying $\int \Gamma_n(z)\lambda(dz) < \infty$ and such that $\|\delta(\omega,t,z)\|^r \wedge 1 \leq \Gamma_n(z)$ for all (ω,t,z) with $t \leq \tau_n(\omega)$.

Assume that $\frac{1}{2(2-r)} \leq \bar{\omega} < \frac{1}{2}$ and let u_M be proportional to $\frac{1}{M^{\bar{\omega}}}$. Choose a sequence k_n of integers with the following property:

$$k \to \infty, \qquad \frac{k}{M} \to 0$$

We set

$$\hat{\sigma}(\bar{\omega})_j = \frac{M}{k} \sum_{m=0}^{k-1} (\Delta_{j+m} X)^2 \mathbb{1}_{\{|\Delta_{j+m} X| \leq u_M\}}$$

Define $\hat{D} = \sum_{j=k+1}^{[t\cdot M]-k} |\Delta_j X|^2 \mathbb{1}_{\{|\Delta_j X| > u_M\}} \cdot (\hat{\sigma}_{j-k} + \hat{\sigma}_{j+1})$ *Then*

$$\hat{D} \xrightarrow{p} D$$

Proof. See Theorem A.7 in Aït-Sahalia and Jacod (2014). □

Lemma A.29. ***Martingale central limit theorem***
Assume $Z_n(t)$ *is a sequence of local square integrable martingales and* Z *is a Gaussian martingale with quadratic characteristic* $\langle Z, Z \rangle$. *Assume that for any* $t \in (t, T]$

1. $\int_0^t \int_{|z|>\epsilon} z^2 \nu^n(ds, dx) \xrightarrow{p} 0 \ \forall \epsilon \in (0, 1]$

2. $[Z_n, Z_n]_t \xrightarrow{p} [Z, Z]_t$

Then $Z_n \xrightarrow{D} Z$ *for* $t \in (0, T]$.

Proof. See Lipster and Shiryayev (1980) □

Theorem A.4. ***Martingale central limit theorem with stable convergence***
Assume $X^n = \{(X_t^n, \mathfrak{F}_t^n; 0 \leq t \leq 1\}$ *are cádlág semimartingales with* $X_0^n = 0$ *and histories* $\mathfrak{F}^n = \{\mathfrak{F}_t^n; 0 \leq t \leq 1\}$.

$$\begin{aligned} X_t^n =& X_0^n + \int_0^t b_s^{X^n} ds + \int_0^t \sigma_s^{X^n} dW_s + \int_0^t \int_E \mathbb{1}_{\{\|x\| \leq 1\}} (\mu^{X^n} - \nu^{X^n})(ds, dx) \\ &+ \int_0^t \int_E \mathbb{1}_{\{\|x\| > 1\}} \mu^{X^n}(ds, dx) \end{aligned}$$

We require the nesting condition of the $\mathfrak{F}^n$: *There exists a sequence* $t_n \downarrow 0$ *such that*

1. $\mathfrak{F}_{t_n}^n \subseteq \mathfrak{F}_{t_{n+1}}^{n+1}$
2. $\bigvee_n \mathfrak{F}_{t_n}^n = \bigvee_n \mathfrak{F}_1^n$

Define $C = \{$ *g: continuous real functions, zero in a neighborhood of zero, with limits at* ∞ $\}$ *Suppose*

1. D *is dense in* $[0, 1]$ *and* $1 \in D$.

2. X *is a quasi-left continuous semimartingale.*

3. *a)* $\forall t \in D \ \sup_{s \leq t} |b_s^{X^n} - b_s^X| \xrightarrow{p} 0$.

   b) $\forall t \in D \ \langle X^{nc} \rangle_t + \int_0^t \int_{|x|<1} x^2 d\nu^{X^n} - \sum_{s \leq t} |\Delta b_s^{X^n}|^2 \xrightarrow{p} \langle X^c \rangle_t + \int_0^t \int_{|x|<1} x^2 \nu^X(ds, dx)$.

   c) $\forall t \in D \ \forall g \in C \ \int_0^t \int_{\mathbb{R}} g(x) \nu^{X^n}(ds, dx) \xrightarrow{p} \int_0^t \int_{\mathbb{R}} g(x) \nu^X(ds, dx)$.

Then

$$X^n \stackrel{L-s}{\to} X$$

in the sense of stable weak convergence in the Skorohod topology.

Proof. See Theorem 1 in Feigin (1984). □

Lemma A.30. ***Burkholder's inequality for discrete martingales***
Consider a discrete time martingale $\{S_j, \mathfrak{F}_j, 1 \leq j \leq M\}$. *Define* $X_1 = S_1$ *and* $X_j = S_j - S_{j-1}$ *for* $2 \leq j \leq M$. *Then, for* $1 < p < \infty$, *there exist constants* C_1 *and* C_2 *depending only on* p *such that*

$$C_1 \mathbb{E} \left[\sum_{j=1}^{M} X_i^2 \right]^{p/2} \leq \mathbb{E}|S_M|^p \leq C_2 \mathbb{E} \left[\sum_{j=1}^{M} X_j^2 \right]^{p/2}.$$

Proof. See Theorem 2.10 in Hall and Heyde (1980). □

Lemma A.31. ***Burkholder-Davis-Gundy inequality***
For each real $p \geq 1$ *there is a constant* C *such that for any local martingale* M *starting at* $M_0 = 0$ *and any two stopping times* $S \leq T$, *we have*

$$E\left[\sup_{t \in \mathbb{R}^+ : S \leq t \leq T} |M_t - M_S|^p | \mathfrak{F}_S \right] \leq CE\left[([M,M]_T - [M,M]_S)^{p/2} | \mathfrak{F}_S \right].$$

Proof. See Section 2.1.5 in Jacod and Protter (2012). □

Lemma A.32. ***Hölder's inequality applied to drift term***
Consider the finite variation part of the Itô semimartingale defined in Definition A.1. We have

$$\sup_{0 \leq u \leq s} \| \int_T^{T+u} b_r dr \|^2 \leq s \int_T^{T+s} \|b_u\|^2 du.$$

Proof. See Section 2.1.5 in Jacod and Protter (2012). □

Lemma A.33. ***Burkholder-Davis-Gundy inequality for continuous martingales***
Consider the continuous martingale part of the Itô semimartingale defined in Definition A.1. There exists a constant C *such that*

$$E\left[\sup_{0 \leq u \leq s} \| \int_T^{T+u} \sigma_r dW_r \|^2 | \mathfrak{F}_T \right] \leq CE\left[\int_T^{T+s} \|\sigma_u\|^2 du | \mathfrak{F}_T \right]$$

Proof. See Section 2.1.5 in Jacod and Protter (2012). □

Lemma A.34. ***Burkholder-Davis-Gundy inequality for purely discontinuous martingales***
Suppose that $\int_0^t \int \|\delta(s,z)\|^2 v(dz)ds < \infty$ *for all* t, *i.e. the process* $Y = \delta \star (\mu - \nu)$ *is a locally square integrable martingale. There exists a constant* C *such that for all finite stopping times* T *and* $s > 0$ *we have*

$$E\left[\sup_{0\le u\le s} \|Y_{T+u} - Y_T\|^2 | F_T\right] \le CE\left[\int_T^{T+s}\int \|\delta(u,z)\|^2 v(dz)du|\mathfrak{F}_T\right].$$

Proof. See Section 2.1.5 in Jacod and Protter (2012). □

Theorem A.5. ***Detecting Jumps***
Assume X *is an Itô-semimartingale as in Definition A.1 and in addition has only finite jump activity, i.e. on each finite time interval there are almost surely only finitely many bounded jumps. Denote* $\Delta_M = \frac{T}{M}$ *and take a sequence* v_M *such that*

$$v_M = \alpha \Delta_M^{\bar{\omega}} \qquad \textit{for some } \bar{\omega} \in \left(0, \frac{1}{2}\right) \textit{ and a constant } \alpha > 0.$$

Our estimator classifies an increment as containing a jump if

$$\Delta_j X > v_M.$$

Denote by $I_M(1) < ... < I_M(\hat{R})$ *the indices* j *in* $1, ..., M$ *such that* $\Delta_j X > v_M$. *Set* $\hat{T}_{jump}(q) = I_M(q) \cdot \Delta_M$ *for* $q = 1, ..., \hat{R}$. *Let* $R = \sup(q : T_{jump}(q) \le T)$ *be the number of jumps of* X *within* $[0, T]$. *Then we have*

$$\mathbb{P}\left(\hat{R} = R, T_{jump}(q) \in (\hat{T}_{jump}(q) - \Delta_M, \hat{T}_{jump}(q)] \quad \forall q \in \{1, ..., R\}\right) \to 1$$

Proof. See Theorem 10.26 in Aït-Sahalia and Jacod (2014). □

Theorem A.6. ***Estimation of continuous and discontinuous quadratic covariation***
Assume X *is an Itô-semimartingale as in Definition A.1 and in addition has only finite jump activity, i.e. on each finite time interval there are almost surely only finitely many bounded jumps. Denote* $\Delta_M = \frac{T}{M}$ *and take some* $\bar{\omega} \in \left(0, \frac{1}{2}\right)$ *and a constant* $\alpha > 0$. *Define the continuous component of* X *by* X^C *and the discontinuous part by* X^D. *Then*

$$\sum_{j=1}^M X_j^2 \mathbb{1}_{\{|X_j| \le \alpha \Delta_M^{\bar{\omega}}\}} = [X^C, X^C] + O_p\left(\frac{1}{\sqrt{M}}\right)$$
$$\sum_{j=1}^M X_j^2 \mathbb{1}_{\{|X_j| > \alpha \Delta_M^{\bar{\omega}}\}} = [X^D, X^D] + O_p\left(\frac{1}{\sqrt{M}}\right).$$

Proof. See Theorem A.16 in Aït-Sahalia and Jacod (2014). Actually they make a much stronger statement and characterize the limiting distribution of the truncation estimators. □

Theorem A.7. ***Onatski estimator for the number of factors***
Assume a factor model holds with

$$X = F\Lambda^\top + e$$

where X is a $M \times N$ matrix of N cross-sectional units observed over M time periods. Λ is a $N \times K$ matrix of loadings and the factor matrix F is a $M \times K$ matrix. The idiosyncratic component e is a $M \times N$ matrix and can be decomposed as

$$e = A\epsilon B$$

with a $M \times M$ matrix A, a $N \times N$ matrix B and a $M \times N$ matrix ϵ.

Define the eigenvalue distribution function of a symmetric $N \times N$ matrix S as

$$\mathcal{F}^S(x) = 1 - \frac{1}{N}\#\{i \le N : \lambda_i(S) > x\}$$

where $\lambda_1(S) \ge ... \ge \lambda_N(S)$ are the ordered eigenvalues of S. For a generic probability distribution having bounded support and cdf $\mathcal{F}(x)$, let $u(\mathcal{F})$ be the upper bound of the support, i.e. $u(\mathcal{F}) = \min\{x : (x) = 1\}$. The following assumptions hold:

1. *For any constant $C > 0$ and $\delta > 0$ there exist positive integers N_0 and M_0 such that for any $N > N_0$ and $M > M_0$ the probability that the smallest eigenvalue of $\frac{\Lambda^\top\Lambda}{N}\frac{F^\top F}{M}$ is below C is smaller than δ.*
2. *For any positive integers N and M, the decomposition $e = A\epsilon B$ holds where*
   a) *$\epsilon_{t,i}$, $1 \le i \le N$, $1 \le t \le M$ are i.i.d. and satisfy moment conditions $\mathbb{E}[\epsilon_{t,i}] = 0$, $\mathbb{E}[\epsilon_{t,i}^2] = 1$ and $\mathbb{E}[\epsilon_{t,i}^4] < \infty$.*
   b) *$\mathcal{F}^{AA^\top}$ and $\mathcal{F}^{BB^\top}$ weakly converge to probability distribution functions $\mathcal{F}_A$ and $\mathcal{F}_B$ respectively as N and M go to infinity.*
   c) *Distributions $\mathcal{F}_A$ and $\mathcal{F}_B$ have bounded support, $u(\mathcal{F}^{AA^\top}) \to u(\mathcal{F}_A) > 0$ and $u(\mathcal{F}^{BB^\top}) \to u(\mathcal{F}_B) > 0$ almost surely as N and M go to infinity. $\liminf_{\delta\to 0} \delta^{-1}\int_{u(\mathcal{F}_A)-\delta}^{u(\mathcal{F}_A)} d\mathcal{F}_A(\lambda) = k_A > 0$ and $\liminf_{\delta\to 0} \delta^{-1}\int_{u(\mathcal{F}_B)-\delta}^{u(\mathcal{F}_B)} d\mathcal{F}_B(\lambda) = k_B > 0$.*
3. *Let $M(N)$ be a sequence of positive integers such that $\frac{N}{M(N)} \to c > 0$ as $N \to \infty$.*
4. *Let ϵ either have Gaussian entries or either A or B are a diagonal matrix*

Then as $N \to \infty$, we have

1. *For any sequence of positive integers $r(N)$ such that $\frac{r(N)}{N} \to 0$ as $N \to \infty$ and $r(N) > K$ for large enough N the $r(N)$th eigennvalue of $\frac{X^\top X}{NM}$ converges almost surely to $u(\mathcal{F}^{c,A,B})$ where $\mathcal{F}^{c,A,B}$ is the distribution function defined in Onatski (2010).*

2. *The K-th eigenvalue of $\frac{X^\top X}{NM}$ tends to infinity in probability.*

3. *Let $\{K^N_{max}, N \in \mathbb{N}\}$ be a slowly increasing sequence of real numbers such that $K^N_{max}/N \to 0$ as $N \to \infty$. Define*

$$\hat{K}^\delta = \max\{i \leq K^N_{max} : \lambda_i - \lambda_{i+1} \geq \delta\}$$

For any fixed $\delta > 0$ $\hat{K}(\delta) \to K$ in probability as $N \to \infty$.

Proof. See Onatski (2010). □

Appendix B

Appendix to Chapter 2

B.1 Empirical Appendix

B.1.1 Equity Data

I collect the price data from the TAQ database for the time period 2003 to 2012. I construct the log-prices for 5 minutes sampling, which gives us on average 250 days per year with 77 daily increments. Overnight returns are removed so that there is no concern of price changes due to dividend distributions or stock splits. For each year I take the intersection of stocks traded each day with the stocks that have been in the S&P500 index at any point during 1993-2012. This gives us a cross-section N of around 500 to 600 firms for each year. I apply standard data cleaning procedures:

- Delete all entries with a time stamp outside 9:30am-4pm
- Delete entries with a transaction price equal to zero
- Retain entries originating from a single exchange
- Delete entries with corrected trades and abnormal sale condition.
- Aggregate data with identical time stamp using volume-weighted average prices

In each year I eliminate stocks from our data set if any of the following conditions is true:

- All first 10 5-min observations are missing in any of the day of this year
- There are in total more than 50 missing values before the first trade of each day for this year
- There are in total more than 500 missing values in the year

Table 2.2 in the main text shows the number of observations after the data cleaning.

Missing observations are replaced by interpolated values. For each day if the first n observations are missing, I interpolate the first values with the $(n+1)$th observation. Otherwise I take the previous observation. As my estimators are based on increments, the interpolated values will result in increments of zeros, which do not contribute to the quadratic covariation.

Daily returns and industry classifications (SIC codes) for the above stocks are from CRSP. I rely on Kenneth R. French's website for daily returns on the Fama-French-Carhart four-factor portfolios. I define three different industry factors as equally weighted portfolios of assets with the following SIC codes

1. Oil and gas: 1200; 1221; 1311; 1381; 1382; 1389; 2870; 2911; 3533; 4922; 4923
2. Banking and finance: 6020; 6021; 6029; 6035; 6036; 6099; 6111; 6141; 6159; 6162; 6189; 6199; 6282; 6311; 6331; 6351; 6798
3. Energy: 4911; 4931; 4991

B.1.2 Factor Analysis

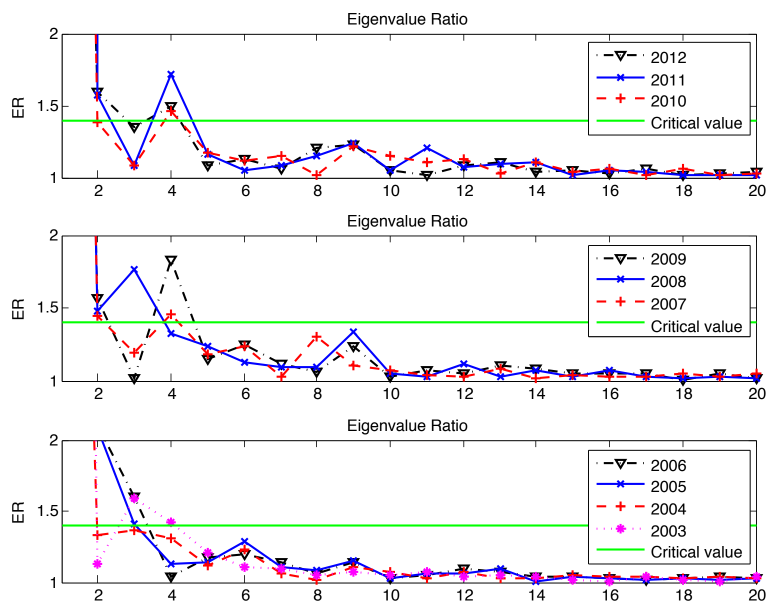


Figure B.1: Number of continuous factors using unperturbed eigenvalue ratios

B.1.3 Jump Factors

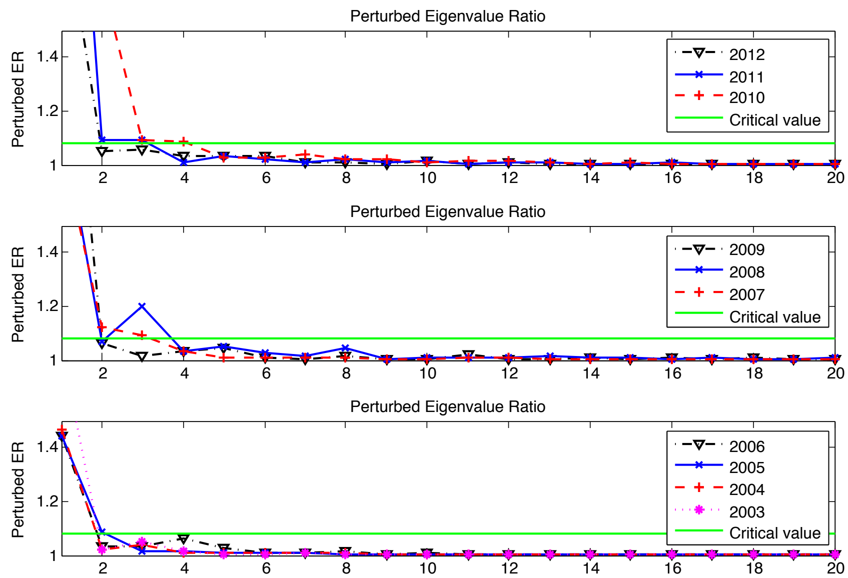


Figure B.2: Number of jump factors with truncation level $a = 3$.

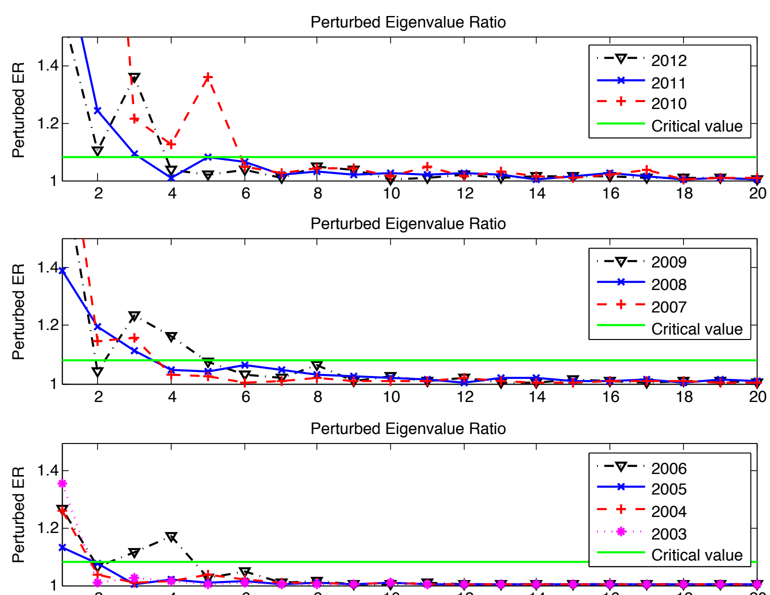


Figure B.3: Number of jump factors with truncation level $a = 4$.

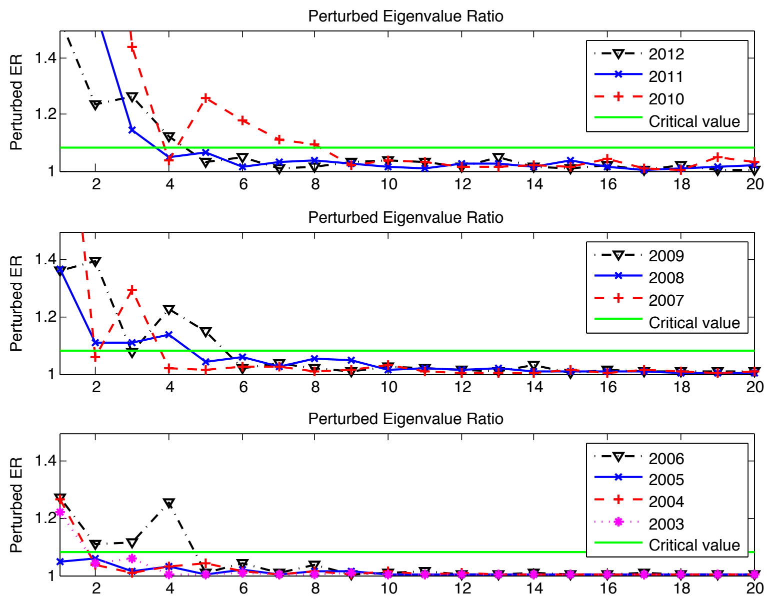


Figure B.4: Number of jump factors with truncation level $a = 4.5$.

1	2	3	4	5	6	7	8	9	10	11	12
Generalized correlations of monthly with yearly jump factors (a=3)											
0.98	0.96	0.99	0.98	0.99	1.00	1.00	1.00	1.00	0.99	1.00	1.00
0.62	0.68	0.87	0.74	0.88	0.76	0.95	0.95	0.96	0.87	0.95	0.80
0.22	0.49	0.41	0.45	0.39	0.58	0.60	0.93	0.58	0.81	0.73	0.42
0.08	0.18	0.20	0.16	0.18	0.14	0.11	0.76	0.42	0.73	0.18	0.11
Generalized correlations of monthly with yearly jump loadings (a=3)											
0.85	0.82	0.90	0.85	0.89	0.96	0.94	0.97	0.97	0.90	0.96	0.94
0.29	0.32	0.42	0.38	0.48	0.43	0.66	0.77	0.56	0.52	0.64	0.44
0.11	0.22	0.16	0.26	0.17	0.30	0.22	0.71	0.30	0.42	0.33	0.19
0.03	0.08	0.09	0.06	0.07	0.05	0.05	0.40	0.19	0.32	0.08	0.04
Generalized correlations of monthly with yearly jump factors (a=4)											
0.73	0.75	0.80	0.77	0.90	1.00	0.99	0.88	1.00	0.89	1.00	0.97
0.35	0.20	0.63	0.44	0.82	0.89	0.93	0.73	0.97	0.71	1.00	0.80
0.06	0.11	0.56	0.21	0.37	0.21	0.76	0.42	0.50	0.47	0.98	0.45
0.02	0.01	0.28	0.03	0.03	0.08	0.32	0.11	0.14	0.08	0.76	0.30
Generalized correlations of monthly with yearly jump loadings (a=4)											
0.35	0.29	0.31	0.32	0.42	0.95	0.60	0.24	0.96	0.30	0.95	0.53
0.10	0.06	0.23	0.12	0.19	0.25	0.29	0.17	0.41	0.11	0.89	0.15
0.02	0.03	0.15	0.05	0.06	0.03	0.17	0.08	0.06	0.08	0.69	0.09
0.01	0.00	0.07	0.01	0.00	0.02	0.05	0.01	0.02	0.01	0.11	0.05
Generalized correlations of monthly with yearly jump factors (a=4.5)											
0.67	0.72	0.69	0.66	0.91	1.00	0.97	0.72	0.99	0.53	1.00	0.95
0.31	0.36	0.63	0.31	0.66	0.64	0.73	0.69	0.90	0.32	1.00	0.66
0.28	0.30	0.32	0.11	0.45	0.26	0.51	0.29	0.25	0.14	0.85	0.44
0.05	0.05	0.20	0.04	0.18	0.04	0.13	0.21	0.02	0.03	0.04	0.13
Generalized correlations of monthly with yearly jump loadings (a=4.5)											
0.22	0.19	0.20	0.18	0.31	0.93	0.40	0.11	0.31	0.09	0.96	0.32
0.09	0.11	0.15	0.08	0.12	0.10	0.11	0.09	0.12	0.05	0.94	0.09
0.08	0.08	0.06	0.03	0.08	0.04	0.06	0.04	0.04	0.02	0.77	0.07
0.01	0.01	0.04	0.01	0.03	0.01	0.01	0.03	0.00	0.01	0.01	0.02

Table B.1: Persistence of jump factors in 2011. Generalized correlation of monthly jump factors and loadings with yearly jump factors and loadings. The yearly number of factors is $K = 4$.

B.1.4 Comparison with Daily Data and Total Factors

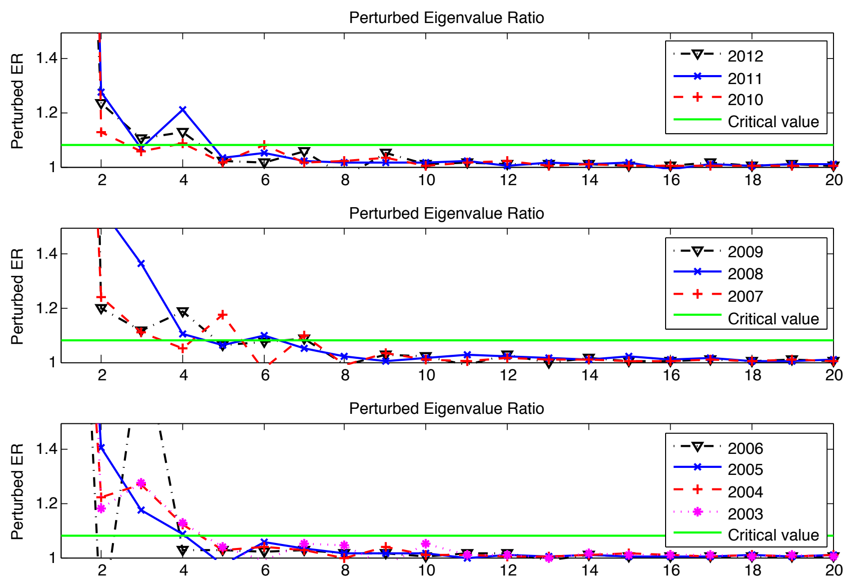


Figure B.5: Number of daily factors

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Generalized correlations between continuous and total factors									
1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.9999	0.9999	1.0000	1.0000	0.9999	1.0000	0.9999	0.9997	1.0000	0.9999
0.9997	0.9996	0.9998	0.9998	0.9997	0.9999	0.9999	0.9993	0.9999	0.9999
0.9979	0.9982	0.9983	0.9748	0.9995	0.9989	0.9998	0.9855	0.9997	0.9997
Generalized correlations between continuous and total loadings									
0.9992	0.9982	0.9998	0.9997	0.9988	0.9994	0.9991	0.9993	0.9997	0.9992
0.9977	0.9982	0.9995	0.9994	0.9976	0.9988	0.9991	0.9720	0.9993	0.9991
0.9967	0.9974	0.9972	0.9982	0.9965	0.9988	0.9988	0.9720	0.9993	0.9991
0.9967	0.9974	0.9972	0.9708	0.9965	0.9950	0.9988	0.9698	0.9987	0.9988
Generalized correlations between continuous and daily factors									
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.99	0.97	0.99	0.99	0.99	1.00	0.99	0.98	0.99	0.99
0.98	0.94	0.94	0.97	0.95	0.98	0.98	0.98	0.98	0.97
0.55	0.65	0.83	0.17	0.47	0.76	0.98	0.96	0.93	0.93
Generalized correlations between continuous and daily loadings									
0.99	0.96	0.98	0.98	0.99	0.98	0.99	0.99	0.96	0.98
0.82	0.86	0.83	0.89	0.76	0.95	0.91	0.89	0.96	0.92
0.82	0.86	0.83	0.86	0.56	0.83	0.91	0.86	0.83	0.86
0.51	0.48	0.73	0.13	0.41	0.61	0.91	0.86	0.83	0.82

Table B.2: Generalized correlations between continuous factors and loadings based on continuous data and on total HF and daily data for $K = 4$ factors and for each year. I use the loadings estimated from the different data sets to construct continuous factors and estimate the distance between the different sets of continuous factors.

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Generalized correlations between continuous and jump factors (a=3)									
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.99	0.99	1.00	1.00	0.98	0.98	0.99	0.96	0.98	0.99
0.98	0.97	0.98	0.64	0.98	0.77	0.97	0.56	0.93	0.87
0.82	0.58	0.87	0.29	0.18	0.38	0.93	0.39	0.67	0.38
Generalized correlations between continuous and jump loadings (a=3)									
0.94	0.98	0.97	0.95	0.96	0.93	0.97	0.95	0.92	0.96
0.94	0.90	0.86	0.72	0.50	0.32	0.83	0.32	0.77	0.79
0.84	0.90	0.84	0.34	0.30	0.31	0.80	0.32	0.63	0.44
0.68	0.31	0.84	0.14	0.30	0.31	0.80	0.26	0.48	0.15
Generalized correlations between continuous and jump factors (a=4)									
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.97	0.66	0.85	0.81	0.92	0.95	0.88	0.89	0.89	0.88
0.88	0.17	0.29	0.58	0.65	0.65	0.45	0.41	0.60	0.69
0.10	0.11	0.21	0.21	0.08	0.07	0.22	0.30	0.25	0.54
Generalized correlations between continuous and jump loadings (a=4)									
0.86	0.73	0.82	0.83	0.88	0.79	0.78	0.76	0.87	0.81
0.31	0.13	0.34	0.20	0.50	0.26	0.28	0.33	0.31	0.32
0.25	0.08	0.09	0.17	0.08	0.26	0.14	0.24	0.19	0.32
0.25	0.08	0.09	0.17	0.08	0.03	0.08	0.10	0.09	0.22
Generalized correlations between continuous and jump factors (a=4.5)									
1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.84	0.66	0.79	0.81	0.92	0.94	0.83	0.87	0.79	0.81
0.55	0.23	0.37	0.52	0.49	0.60	0.47	0.59	0.52	0.66
0.04	0.10	0.27	0.16	0.03	0.03	0.13	0.25	0.18	0.43
Generalized correlations between continuous and jump loadings (a=4.5)									
0.73	0.60	0.75	0.79	0.82	0.58	0.54	0.54	0.75	0.72
0.27	0.14	0.25	0.19	0.38	0.45	0.25	0.43	0.21	0.26
0.08	0.14	0.12	0.14	0.08	0.32	0.15	0.24	0.10	0.22
0.08	0.04	0.12	0.14	0.02	0.01	0.04	0.11	0.10	0.22

Table B.3: Generalized correlations between continuous factors and loadings based on continuous data and on jump data for $K = 4$ factors and for each year. I use the loadings estimated from the different data sets to construct continuous factors and estimate the distance between the different sets of continuous factors.

B.1.5 Implied Volatility Data

I use daily prices for standard call and put options from OptionMetrics for the same firms and time periods as for the high-frequency data. OptionMetrics provides implied volatilities for 30 days at the money options using a linearly interpolated volatility surface. I average the implied call and put volatilities for each asset and each day. Then I apply the following data cleaning procedure in order to identify outliers. For each year I remove a stock if

- for days 1-15 any of the volatilities is greater than 200% of the average volatility of the first 31 days
- for the last 15 days any of the volatilities is greater than 200% of the average volatility of the last 31 days
- for all the other days any of the volatilities is greater than 200% of the average of a 31 days moving window centered at that day.

The observations after the data cleaning are reported in Table 2.13 in the main text.

B.1.6 Componentwise Leverage Effect

The following plots depict the sorted correlations between total, systematic and idiosyncratic log-prices with total, systematic and idiosyncratic implied volatility. I use 4 asset factors and 1 volatility factor.

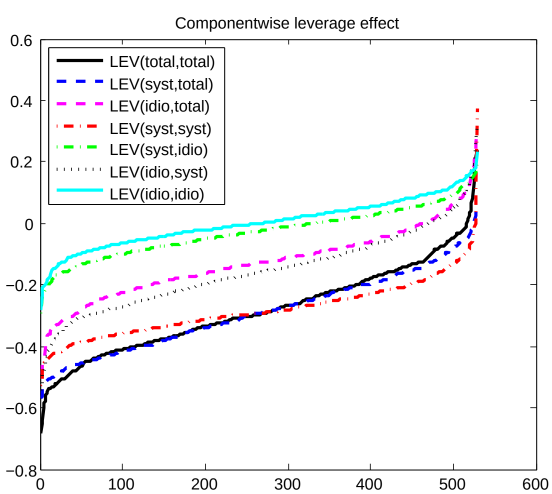


Figure B.6: Componentwise leverage effect in 2012

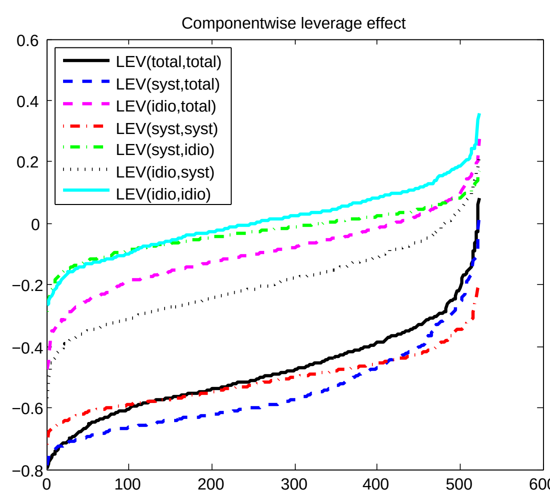


Figure B.7: Componentwise leverage effect in 2011

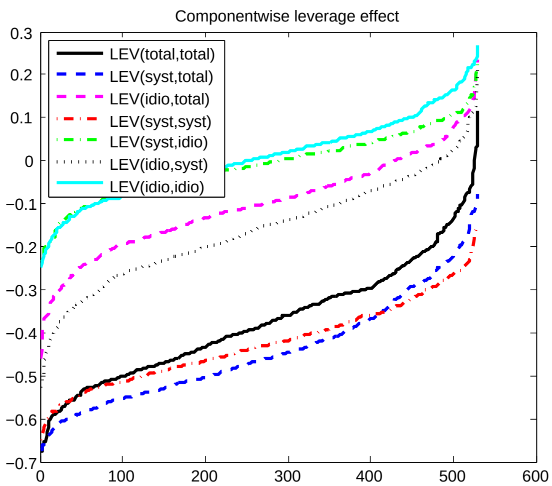


Figure B.8: Componentwise leverage effect in 2010

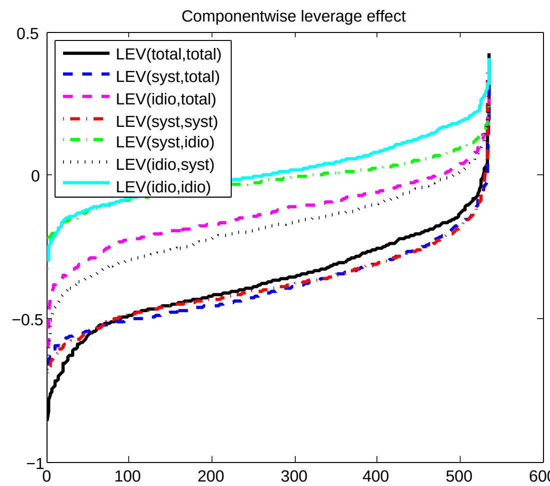


Figure B.9: Componentwise leverage effect in 2009

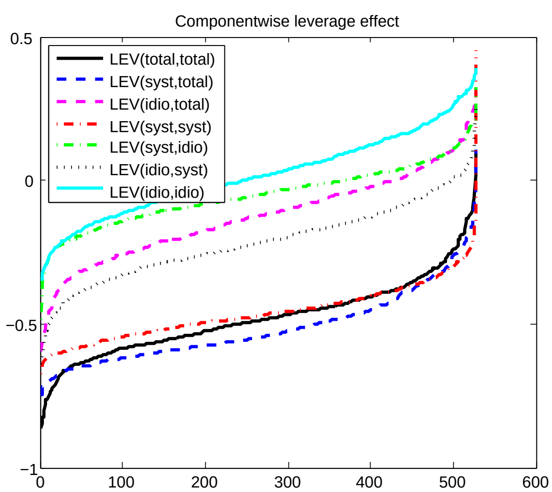


Figure B.10: Componentwise leverage effect in 2008

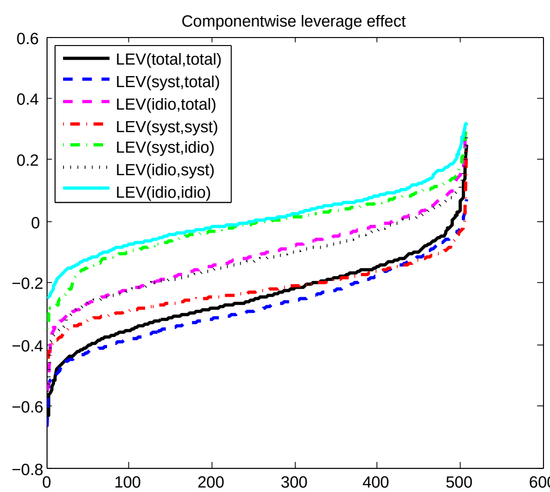


Figure B.11: Componentwise leverage effect in 2007

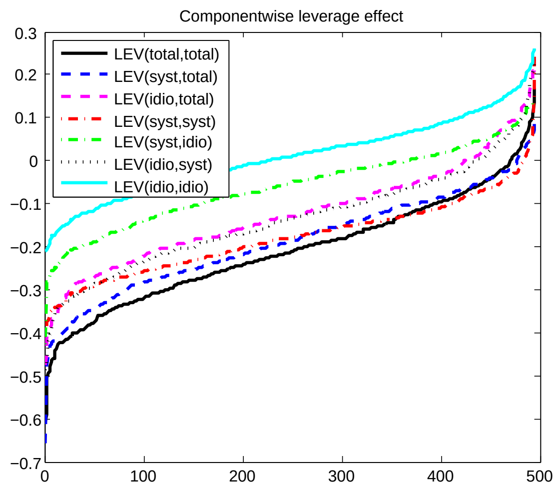


Figure B.12: Componentwise leverage effect in 2006

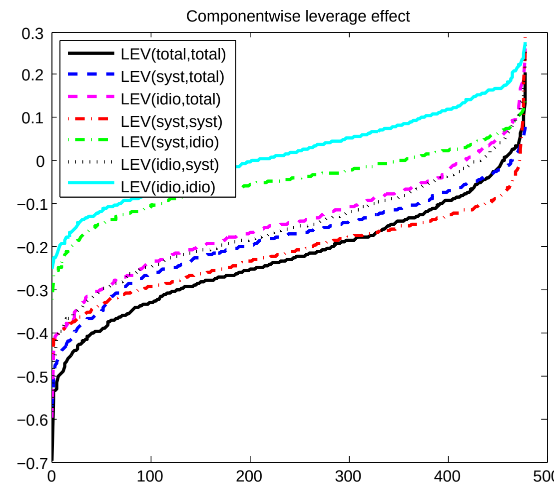


Figure B.13: Componentwise leverage effect in 2005

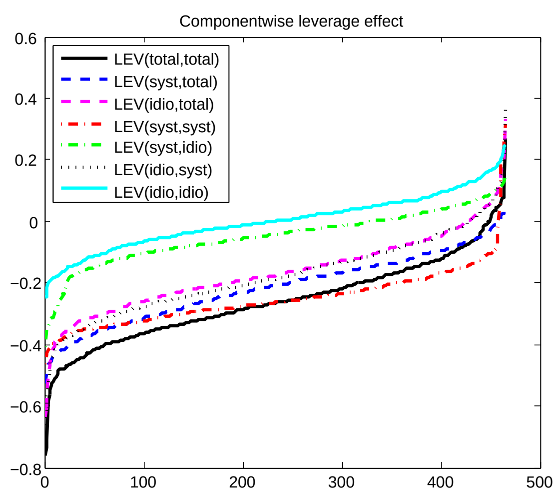


Figure B.14: Componentwise leverage effect in 2004

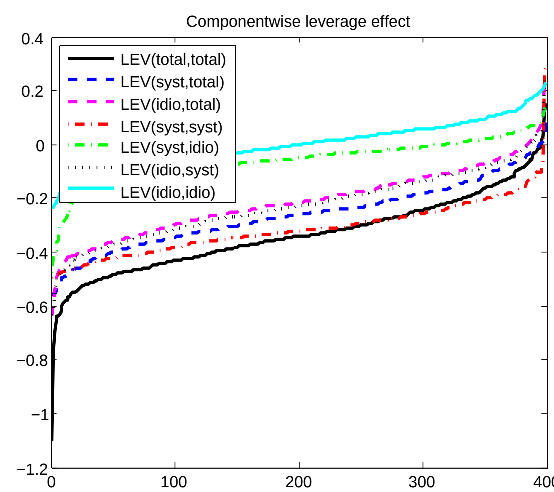


Figure B.15: Componentwise leverage effect in 2003

B.1.7 Comparison of Componentwise Leverage Effect based on Implied Volatilities and High-Frequency Volatilities

The following plots compare the componentwise leverage effect based on implied volatilities and high-frequency volatilities. I use the four continuous factors for separating the return into a systematic and idiosyncratic component.

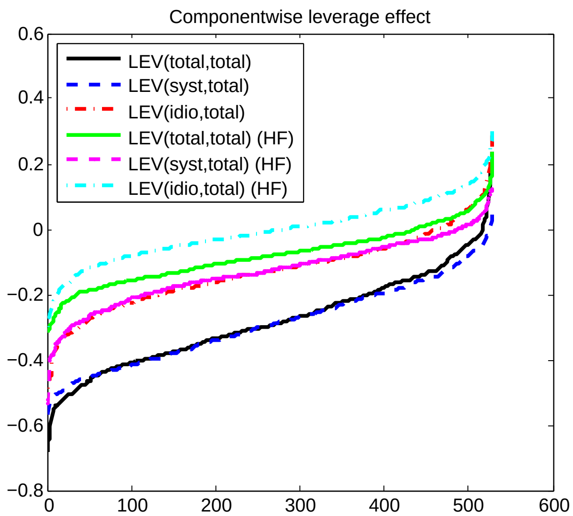


Figure B.16: Componentwise leverage effect in 2012 based on implied and high-frequency volatilities.

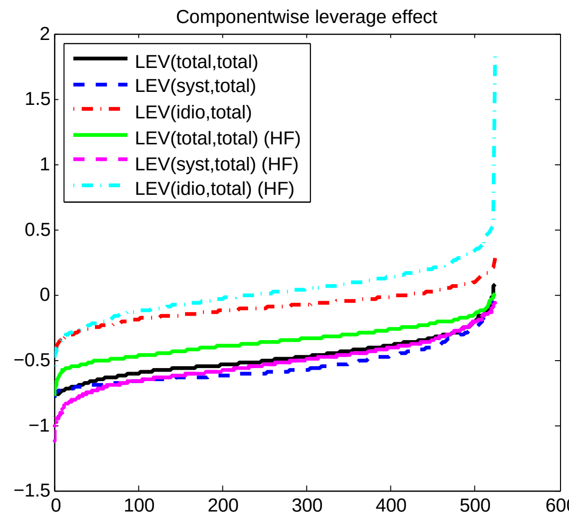


Figure B.17: Componentwise leverage effect in 2011 based on implied and high-frequency volatilities.

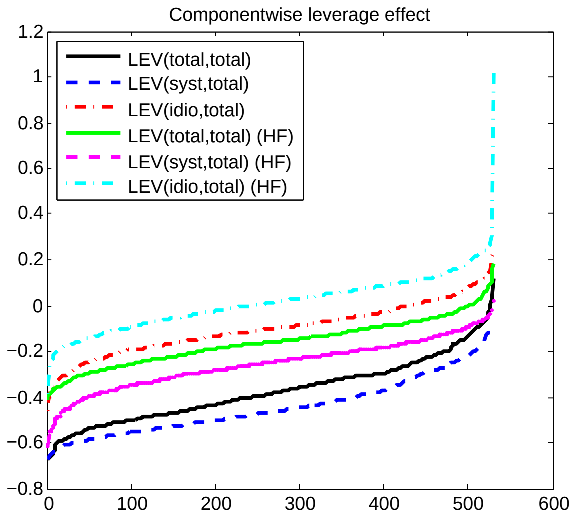


Figure B.18: Componentwise leverage effect in 2010 based on implied and high-frequency volatilities.

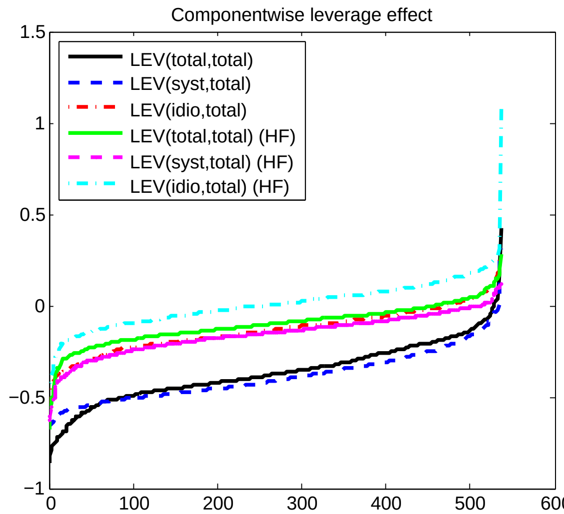


Figure B.19: Componentwise leverage effect in 2009 based on implied and high-frequency volatilities.

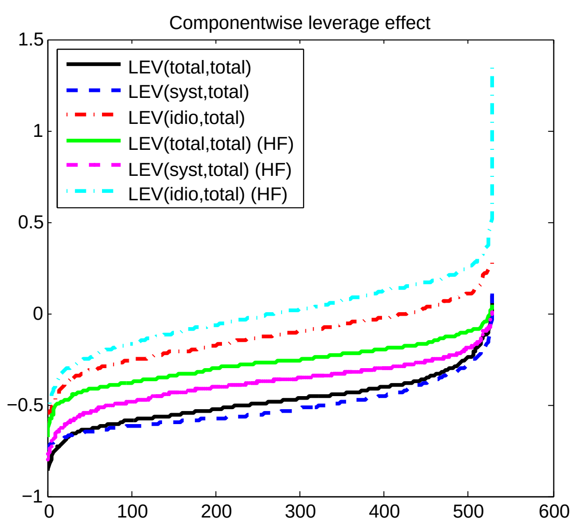


Figure B.20: Componentwise leverage effect in 2008 based on implied and high-frequency volatilities.

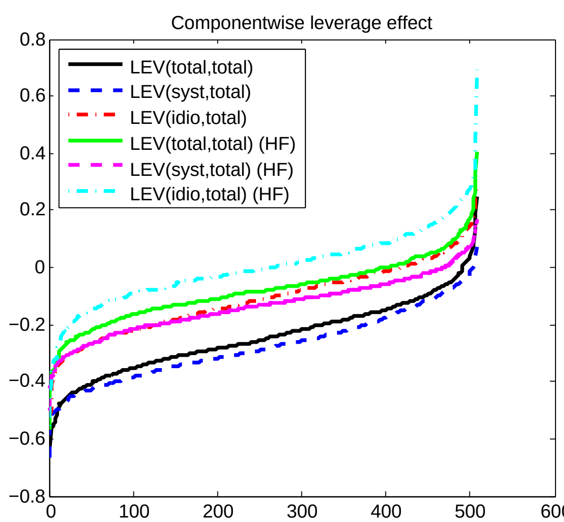


Figure B.21: Componentwise leverage effect in 2007 based on implied and high-frequency volatilities.

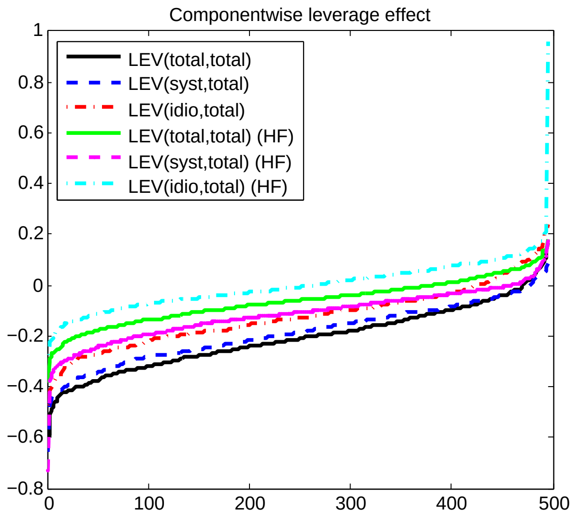


Figure B.22: Componentwise leverage effect in 2006 based on implied and high-frequency volatilities.

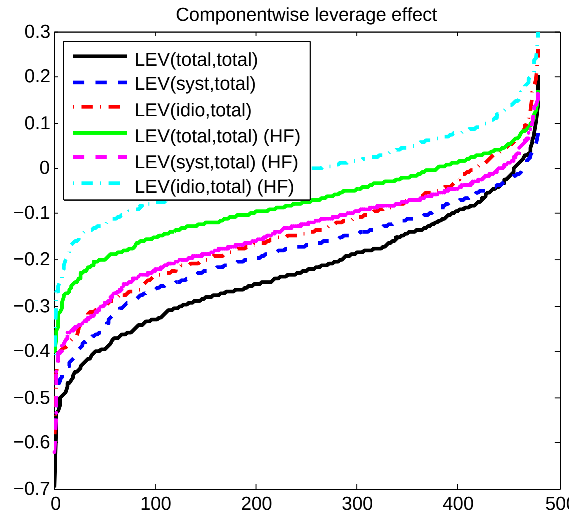


Figure B.23: Componentwise leverage effect in 2005 based on implied and high-frequency volatilities.

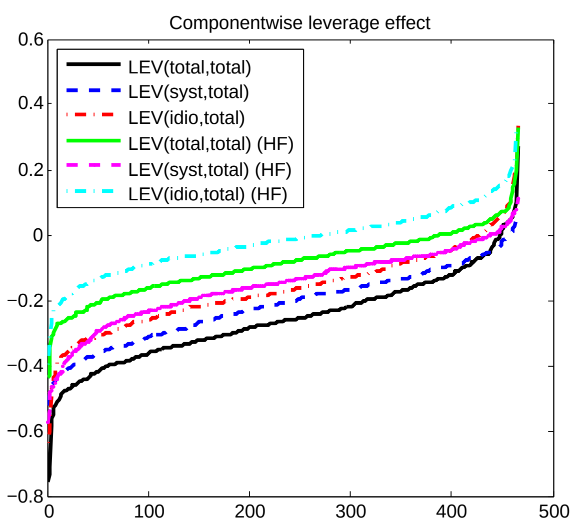


Figure B.24: Componentwise leverage effect in 2004 based on implied and high-frequency volatilities.

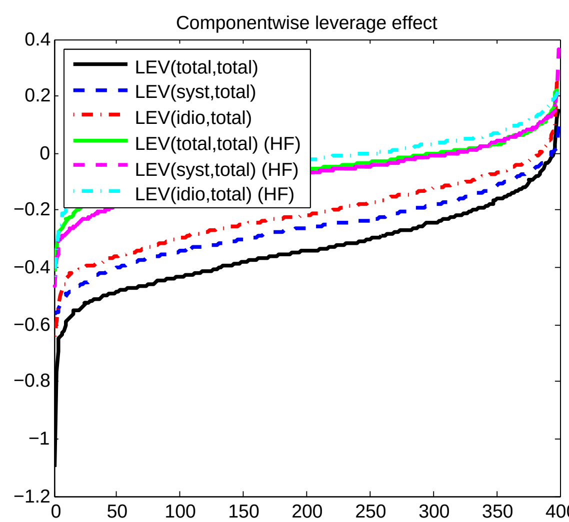


Figure B.25: Componentwise leverage effect in 2004 based on implied and high-frequency volatilities.

B.1.8 Decomposition of the Leverage Effect

The following figures show the decomposition of the leverage effect into a systematic and idiosyncratic part based on implied volatilities and high-frequency volatilities. I use 4 continuous asset factors.

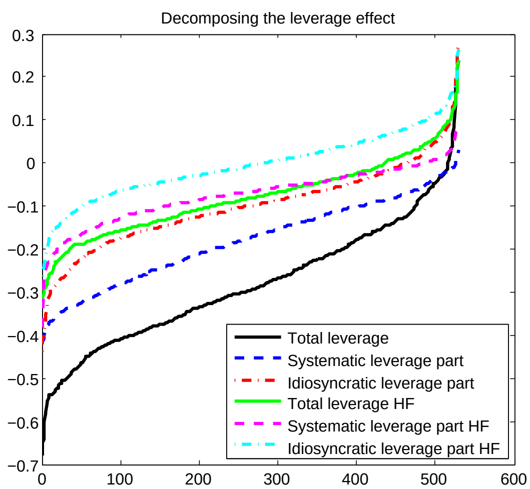


Figure B.26: Decomposition of LEV in 2012

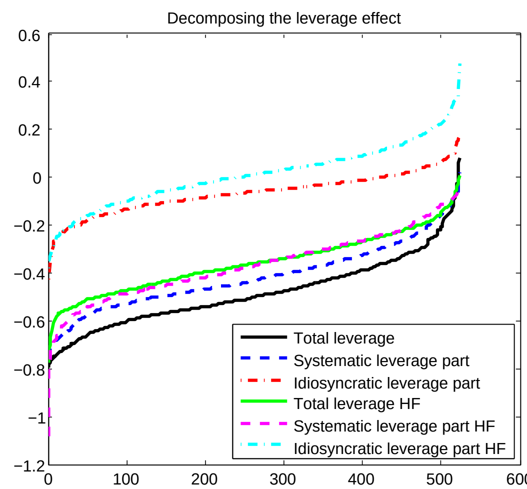


Figure B.27: Decomposition of LEV in 2011

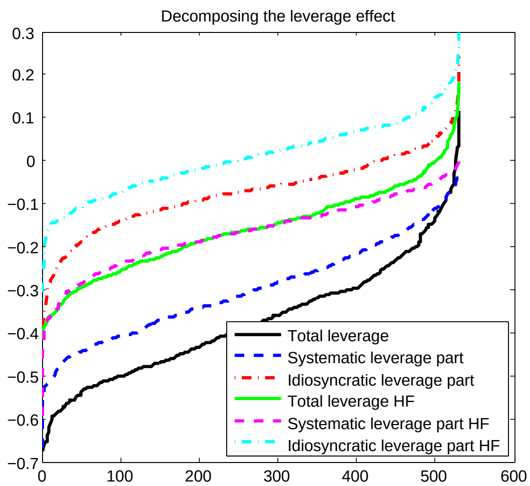


Figure B.28: Decomposition of LEV in 2010

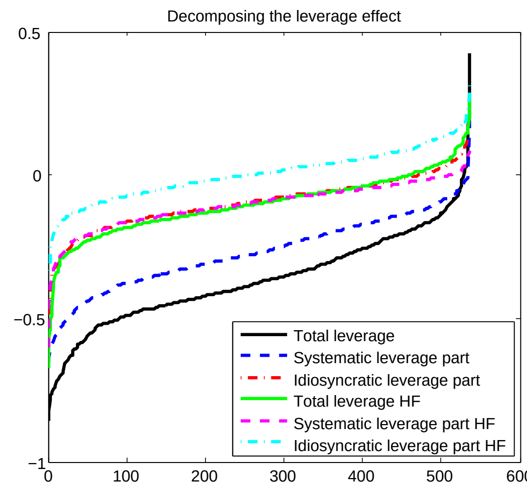


Figure B.29: Decomposition of LEV in 2009

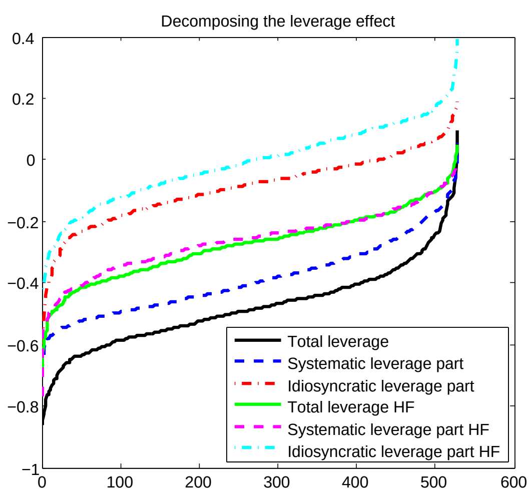


Figure B.30: Decomposition of LEV in 2008

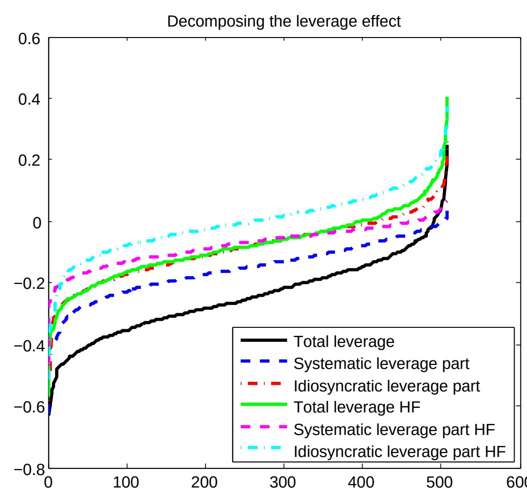


Figure B.31: Decomposition of LEV in 2007

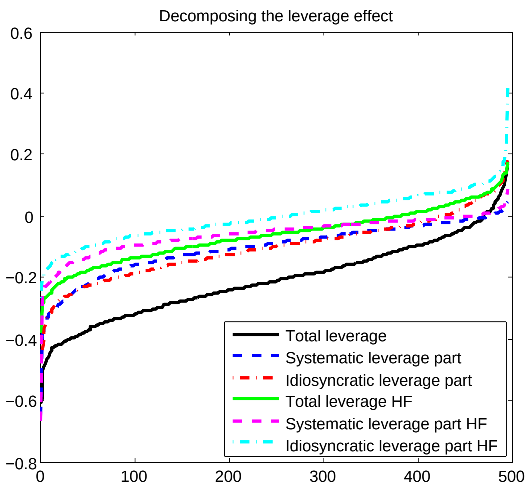


Figure B.32: Decomposition of LEV in 2006

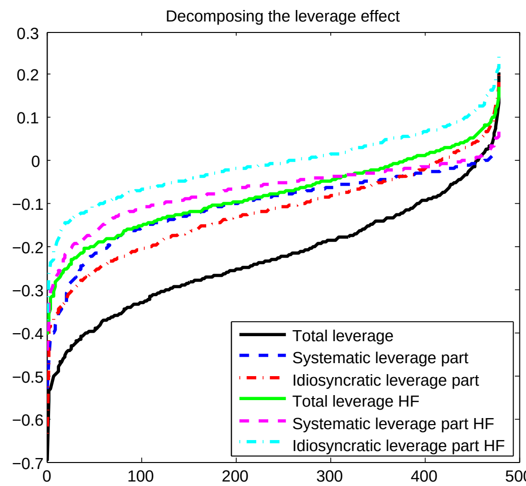


Figure B.33: Decomposition of LEV in 2005

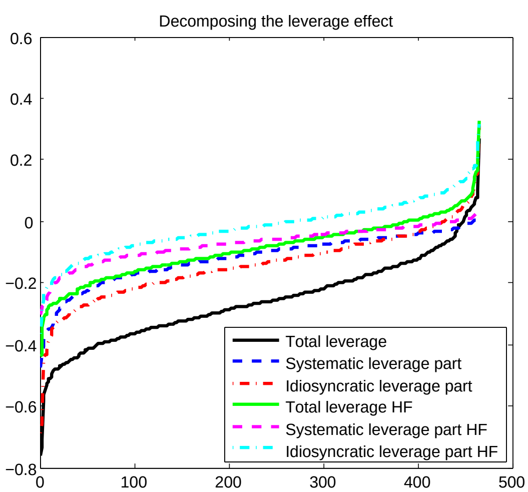


Figure B.34: Decomposition of LEV in 2004

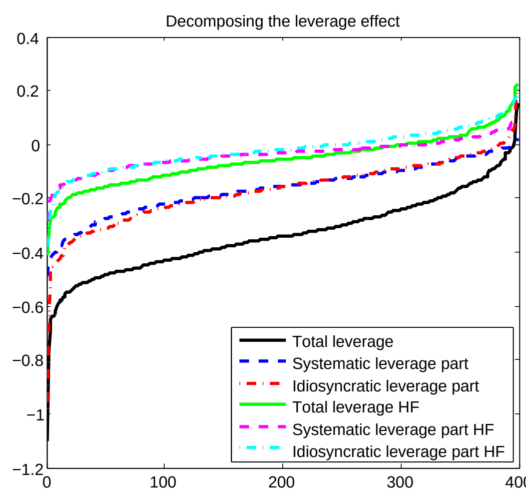


Figure B.35: Decomposition of LEV in 2003

B.1.9 Componentwise Leverage Effect Using Daily Return Data

The following figures compare the componentwise leverage based on implied volatilities calculated either with the daily accumulated continuous log price increments or with daily CRSP returns. I use 4 continuous factors for separating the systematic from the idiosyncratic part.

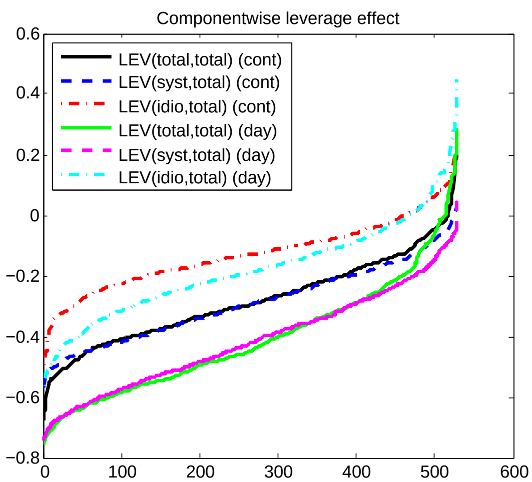


Figure B.36: Componentwise leverage effect in 2012 with daily continuous log price increments and daily returns.

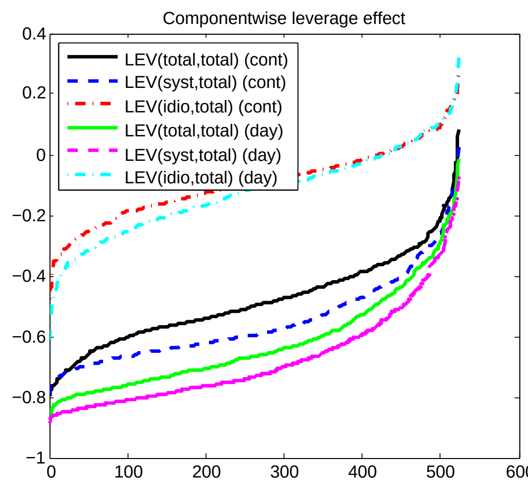


Figure B.37: Componentwise leverage effect in 2011 with daily continuous log price increments and daily returns.

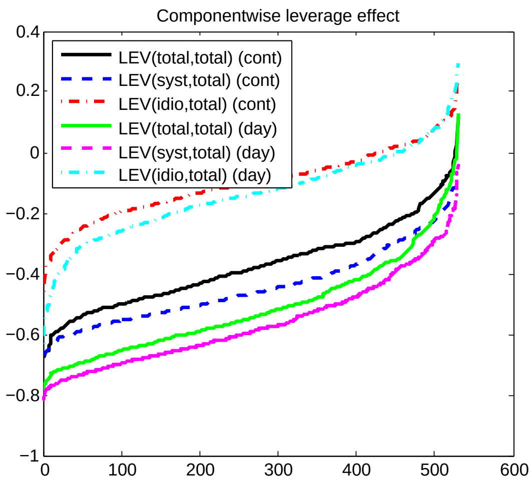


Figure B.38: Componentwise leverage effect in 2010 with daily continuous log price increments and daily returns.

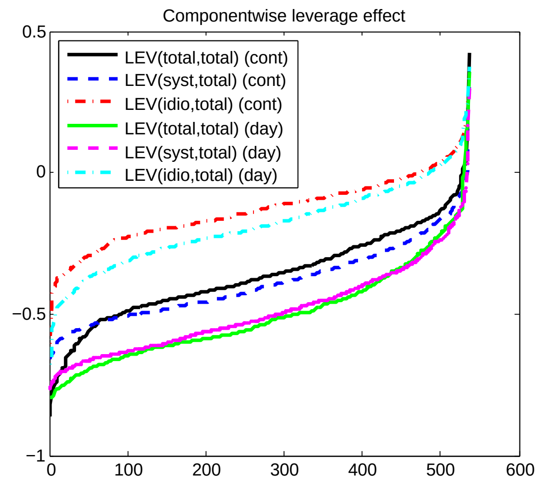


Figure B.39: Componentwise leverage effect in 2009 with daily continuous log price increments and daily returns.

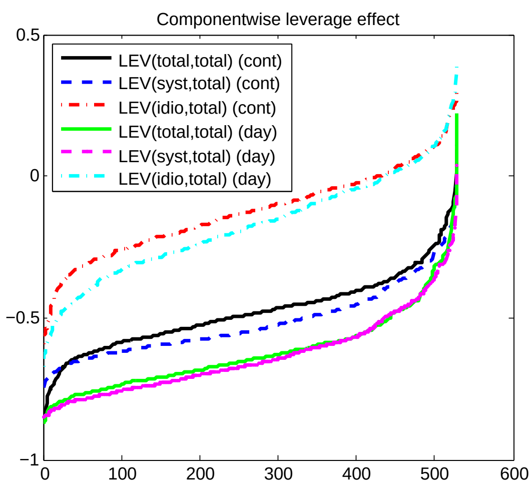


Figure B.40: Componentwise leverage effect in 2008 with daily continuous log price increments and daily returns.

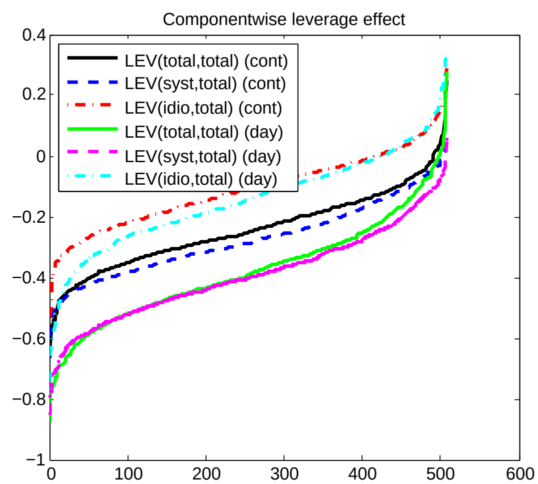


Figure B.41: Componentwise leverage effect in 2007 with daily continuous log price increments and daily returns.

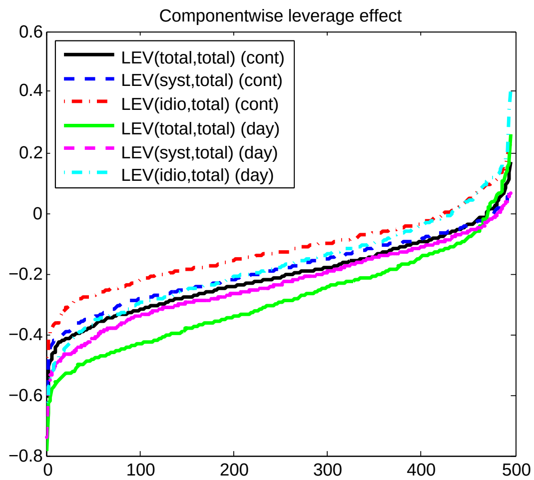


Figure B.42: Componentwise leverage effect in 2006 with daily continuous log price increments and daily returns.

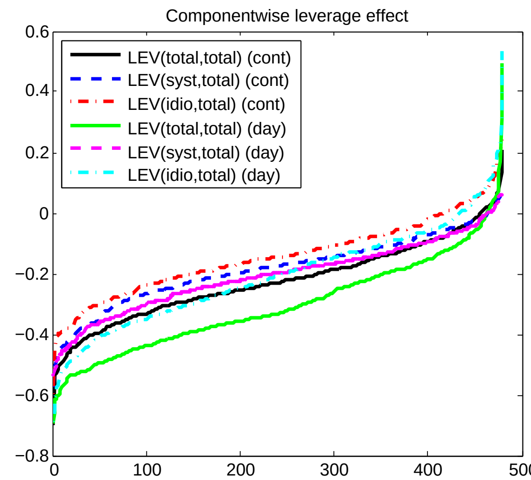


Figure B.43: Componentwise leverage effect in 2006 with daily continuous log price increments and daily returns.

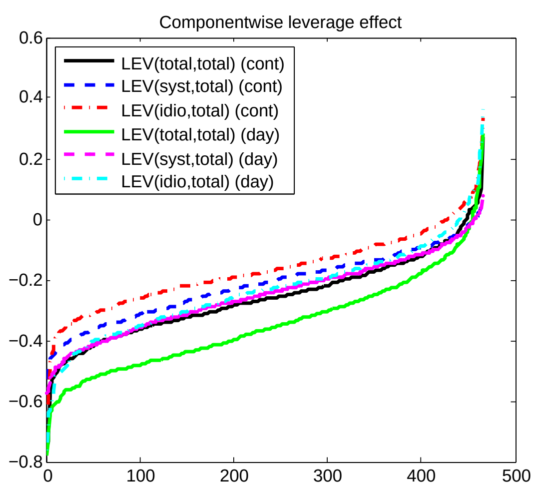


Figure B.44: Componentwise leverage effect in 2004 with daily continuous log price increments and daily returns.

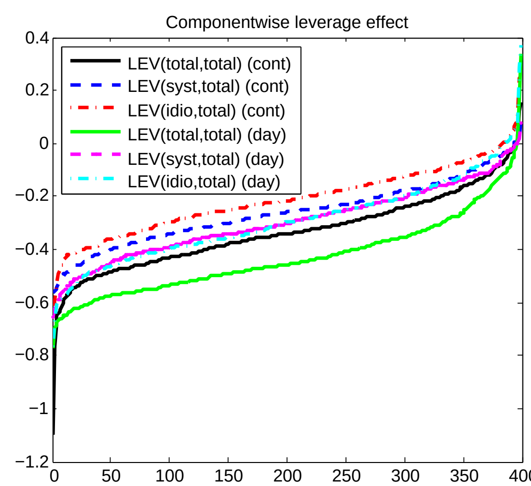


Figure B.45: Componentwise leverage effect in 2003 with daily continuous log price increments and daily returns.

B.1.10 Componentwise Leverage Effect with Systematic Risk Based on Fama-French-Carhart Factors

The following figures show the componentwise leverage effect based on implied volatilities and daily CRSP returns. The systematic part is either calculated with the 4 continuous factors or with the 4 Fama-French-Carhart factors (market, size, value and momentum).

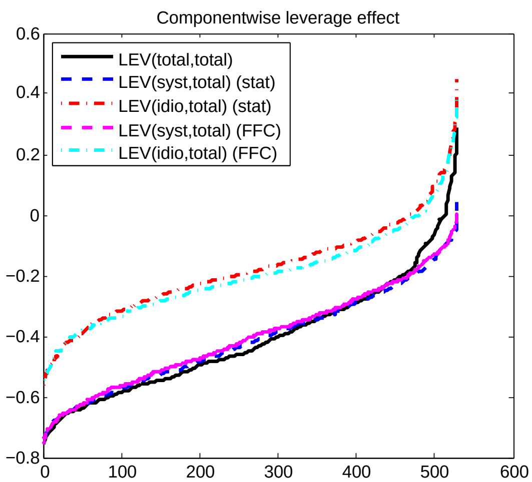


Figure B.46: Componentwise leverage effect in 2012 with 4 continuous or 4 Fama-French-Carhart factors.

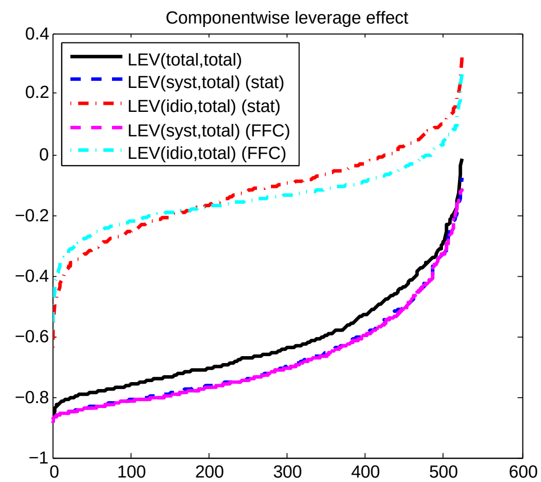


Figure B.47: Componentwise leverage effect in 2011 with 4 continuous or 4 Fama-French-Carhart factors.

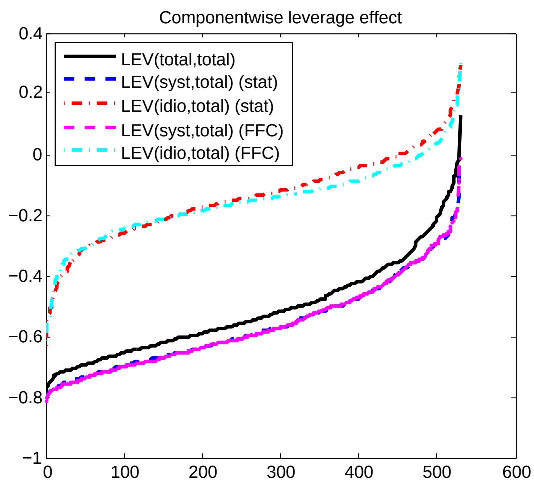


Figure B.48: Componentwise leverage effect in 2010 with 4 continuous or 4 Fama-French-Carhart factors.

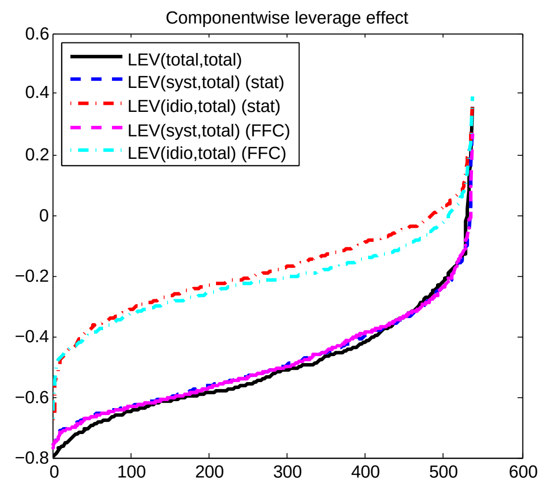


Figure B.49: Componentwise leverage effect in 2009 with 4 continuous or 4 Fama-French-Carhart factors.

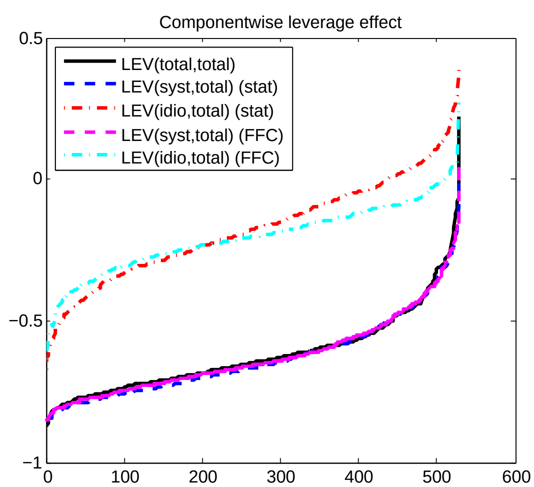


Figure B.50: Componentwise leverage effect in 2008 with 4 continuous or 4 Fama-French-Carhart factors.

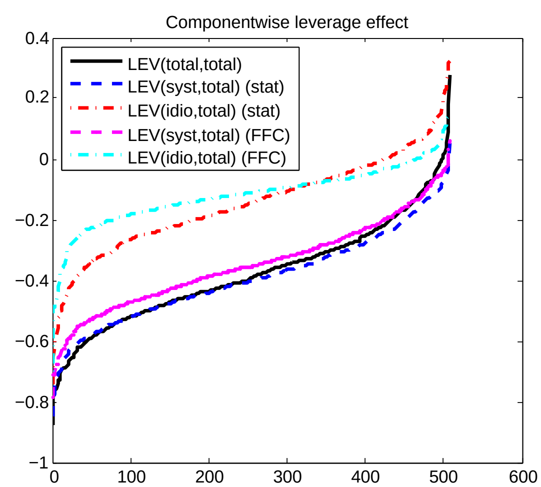


Figure B.51: Componentwise leverage effect in 2007 with 4 continuous or 4 Fama-French-Carhart factors.

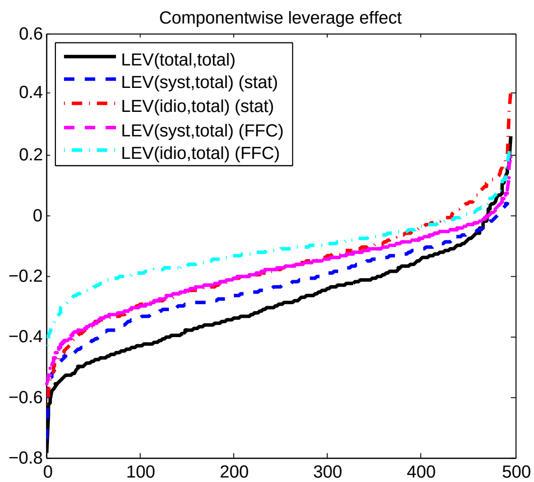


Figure B.52: Componentwise leverage effect in 2006 with 4 continuous or 4 Fama-French-Carhart factors.

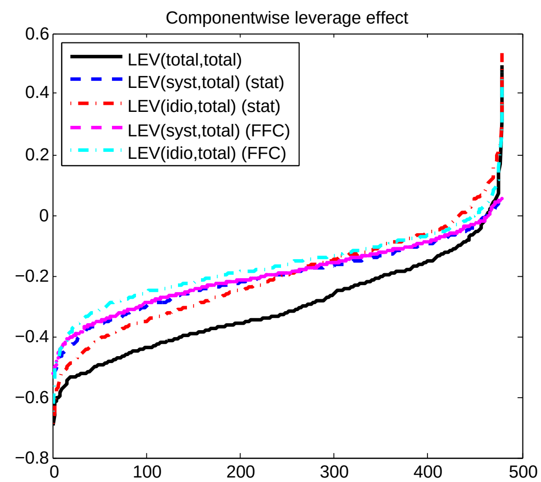


Figure B.53: Componentwise leverage effect in 2005 with 4 continuous or 4 Fama-French-Carhart factors.

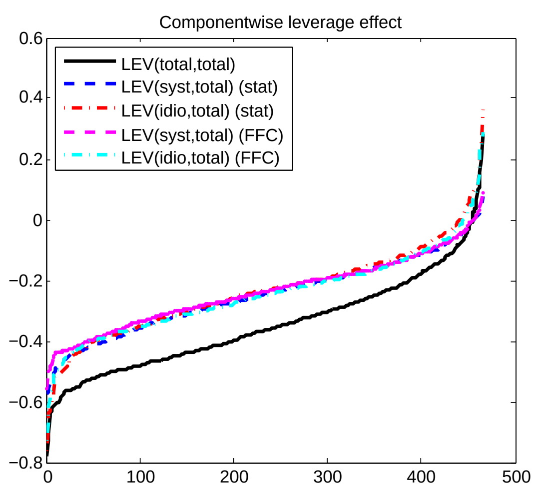


Figure B.54: Componentwise leverage effect in 2004 with 4 continuous or 4 Fama-French-Carhart factors.

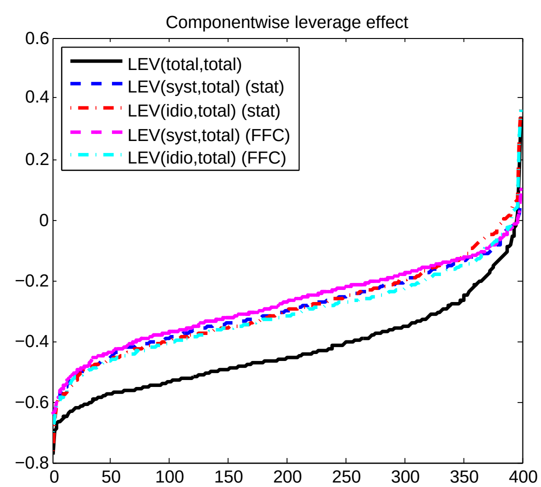


Figure B.55: Componentwise leverage effect in 2003 with 4 continuous or 4 Fama-French-Carhart factors.

B.2 Theoretical Appendix

Theorem B.1. ***Estimation of the leverage effect with high-frequency data***
Assume Y is a 1-dimensional Itô-semimartingale as in Definition A.1 and in addition has only finite jump activity and its volatility process σ_Y^2 is continuous. We want to estimate the leverage effect defined as

$$LEV = \frac{[\sigma_Y^2, Y]_T^C}{\sqrt{[Y,Y]_T^C}\sqrt{[\sigma_Y^2, \sigma_Y^2]_T^C}}.$$

Denote the M increments of Y as $Y_j = Y_{t_{j+1}} - Y_{t_j}$. A consistent estimator of the leverage effect is

$$\widehat{LEV} = \frac{\widehat{[\sigma_Y^2, Y]}_T^C}{\sqrt{\widehat{[Y,Y]}_T^C}\sqrt{\widehat{[\sigma_Y^2, \sigma_Y^2]}_T^C}}.$$

with

$$\begin{aligned}
\hat{Y}_j^C &= Y_j \mathbb{1}_{\{|Y_j| \leq \alpha \Delta_M^{\bar{\omega}}\}} \\
\hat{\sigma}_l^2 &= \frac{1}{k\Delta_M} \sum_{j=1}^{k} \hat{Y}_{l+j}^{C\;2} \\
\widehat{[\sigma_Y^2, Y]}_T^C &= \frac{2}{k} \sum_{l=0}^{M-2k} \left(\hat{\sigma}^2_{(l+k)\Delta_M} - \hat{\sigma}^2_{l\Delta_M}\right)\left(\hat{Y}^C_{(l+k)\Delta_M} - \hat{Y}^C_{l\Delta_M}\right) \\
\widehat{[Y,Y]}_T^C &= \sum_{j=1}^{M} \hat{Y}_j^{C2} \\
\widehat{[\sigma_Y^2, \sigma_Y^2]}_T &= \frac{3}{2k} \sum_{l=1}^{M-2k} \left(\hat{\sigma}^2_{(l+k)\Delta_M} - \hat{\sigma}^2_{l\Delta_M}\right)^2 - \sum_{j=0}^{M-2k} \frac{6}{k^2}\left(1 - \frac{2}{k}\right)\hat{\sigma}^4_{l\Delta_M}.
\end{aligned}$$

Let $\Delta_M = \frac{T}{M}$, $k \sim \Delta_M^{-1/2}$, $\alpha > 0$ and $\bar{\omega} \in \left(0, \frac{1}{2}\right)$. Then for any fixed T and as $M \to \infty$

$$\widehat{LEV} \xrightarrow{p} LEV$$

Proof. See Theorem 8.14 in Aït-Sahalia and Jacod (2014) and Theorem 3 in Kalnina and Xiu (2014). □

Appendix C

Appendix to Chapter 4

C.1 Dynamic Structural Models and Bank Capital Structure

Our analysis builds on the capital structure model of Leland (1994) and Leland and Toft (1996), as extended by Chen and Kou (2009) to include jumps. These models were developed for general firms and not specifically for financial firms, so here we discuss their application to banks. Recent work applying the Leland (1994) framework and its extensions to financial firms includes Auh and Sundaresan (2014), Diamond and He (2014), Harding, Liang, and Ross (2013), He and Xiong (2012), and Sundaresan and Wang (2014a). Albul et al. (2010) use the Leland (1994) model in their analysis of contingent convertible debt, as does the Bank of England in its analysis (Murphy, Walsh, and Willison (2012)). The models of Hilscher and Raviv (2011), Koziol and Lawrenz (2012), and Pennacchi (2010) may be viewed as extensions of Merton (1974) rather than Leland (1994) in the sense that they treat default exogenously. In the remainder of this section, we discuss special features of the banking context and how they are addressed within our framework.

Deposit insurance. Deposit insurance differentiates one of the most important sources of bank funding from funding available to nonfinancial firms. In our model, we can interpret P_1 as the face value of deposits when deposit insurance is fairly priced, as follows. Suppose for simplicity that $c_1 = r$ in (4.7). When it issues P_1 in debt, the bank collects only $B(V; V_b)$ in cash. We have interpreted the difference as the compensation for default risk demanded by bond holders, given that they may not receive full repayment. But we could just as well assume that the bank issues debt at par, collecting P_1 in cash, and then pays $P_1 - B(V; V_b)$ for deposit insurance that guarantees full repayment to depositors. The net effect for the bank (and our analysis) is the same; see also Section 4.3.4. We can approximate the demand feature of bank deposits by taking their average maturity $1/m$ to be small.

In an earlier version of this paper (Chen et al. (2012)), we modeled deposits as a separate category of debt, and we included an explicit insurance premium proportional to the par value of deposits issued. This allows us to capture a possible subsidy through mispriced deposit insurance. Here, we take the simpler approach introduced in Section 4.2.2 and include any such subsidy as a funding benefit.

Liquidity value of deposits. Deposits are not simply a source of funding for a bank — they are also an important service to its customers. Customers value the safety and ready availability of bank deposits and are willing to pay (or accept a lower interest rate) for this convenience. As discussed in Section 4.2.2, we model this effect as a funding benefit. Similar approaches are used in DeAngelo and Stulz (2013) and Sundaresan and Wang (2014a).

Capital structure tradeoffs. In the Leland (1994) model, a firm's optimal debt level is driven by the tradeoff between the tax benefit of debt and bankruptcy costs. Tax deductibility takes on added importance with the inclusion of CoCos, which have tax-deductible coupons in some jurisdictions but not others. The difference appears to be an important factor in explaining the widespread issuance of CoCos by banks in Europe and Japan but not the United States. Our model captures this distinction through different values of κ_i, $i = 1, 2$, for straight debt and CoCos.

As explained above, the mechanism we use for the tax benefit also captures subsidized deposit insurance and the liquidity value of deposits, leading to a richer set of features that makes Leland's tradeoff more interesting for banks, not less. In a very different setting, Allen, Carletti, and Marquez (2013) develop a model of bank capital structure in which equity reduces bankruptcy costs but is more costly because of market segmentation: equity investors have outside options but depositors do not. This effect is once again reflected in our setting through κ_1, which reflects the benefits of deposit financing.

Default endogeneity. We view this element of the Leland (1994) model as the most important feature for our analysis: we cannot hope to study the incentive effects of CoCos without first modeling the optimal decision of shareholders to keep the firm operating. Endogenous default captures the shareholders' option to close the firm or keep it running. In applying the model to a large bank, we interpret this decision as reflecting the bank's continued access to equity financing, given its debt load and its investment opportunities. The most important aspect of endogenous default is that it determines the value of equity at all asset levels, not just at the default boundary. The bank's anticipated ability to operate following a decline in assets determines the value of its equity at higher asset levels.

Regulatory capital requirements. Banks face regulatory capital requirements that influence their capital structure in ways that do not affect nonfinancial firms. These constraints would be important in analyzing optimal capital structure, an issue we do not address. Our analysis compares changes in the levels of different debt levels, but these debt levels could be the result of capital requirements.

Asset dynamics. Geometric Brownian motion is a rough approximation to asset value for any firm. But it is a more plausible approximation for the assets of a bank than a nonfinancial firm. The main deficiency of the model in studying bank risk is the absence of jumps, and this deficiency is resolved by the Chen-Kou (2009) model.

Stationary capital structure. The framework of continuous debt rollover we use, extending Leland and Toft (1996), is particularly well suited for a financial firm rather than a nonfinancial firm. It leads to qualitatively different results than a model with a single debt maturity, whether finite or infinite: the costs of debt rollover allow shareholders to benefit from reducing default risk and thus mitigate debt overhang. The stationarity of the model — the debt outstanding remains constant — provides tractability. Diamond and He (2014, p.736, p.747, p.750) argue that this type of model, with a constant refinancing rate, is particularly appropriate for financial firms. A drawback of the stationary capital structure is that it leads to shrinking leverage as the firm's assets grow. In the setting of Theorem 5.3, increasing the level of debt at a higher asset level would only add to the possibility of debt-induced collapse.

Secured debt. Many types of firms issue secured debt; for financial firms, it often takes the form of repurchase agreements. Auh and Sundaresan (2014) and Sundaresan and Wang (2014a) develop extensions of Leland (1994) to combine secured and unsecured debt, with particular focus on banks. Interestingly, they show that a bank will optimally set its level of secured debt in such a way that the endogenous default boundary is unaffected. This suggests that for purposes of determining the default boundary, one may omit secured debt. Holding fixed the default boundary and the total amount of senior debt, CoCo valuation is unaffected by the composition of the types of senior debt.

C.2 Proofs for Section 4.2

C.2.1 Optimal Default Barrier Without CoCos

Chen and Kou (2009) have shown that for a firm with only straight debt P_1, the optimal default barrier is $V_b^{\mathsf{PC}} = P_1\epsilon_1$, with with

$$\epsilon_1 = \frac{\frac{c_1+m}{r+m}\gamma_{1,r+m}\gamma_{2,r+m} - \frac{\kappa_1 c_1}{r}\gamma_{1,r}\gamma_{2,r}}{(1-\alpha)(\gamma_{1,r}+1)(\gamma_{2,r}+1) + \alpha(\gamma_{1,r+m}+1)(\gamma_{2,r+m}+1)}\frac{\eta+1}{\eta}. \tag{C.1}$$

where $-\gamma_{1,\rho} > -\eta > -\gamma_{2,\rho}$ are the two negative roots of the equation

$$G(x) = \left(r - \delta - \frac{1}{2}\sigma^2 - \lambda(\frac{\eta}{\eta+1} - 1)\right) + \frac{1}{2}\sigma^2 x^2 + \lambda\left(\frac{\eta}{\eta+x} - 1\right) = \rho.$$

A similar argument shows that the constant ϵ_2 that we need for V_b^{NC} is given by

$$\epsilon_2 = \frac{\frac{c_2+m}{r+m}\gamma_{1,r+m}\gamma_{2,r+m} - \frac{\kappa_2 c_2}{r}\gamma_{1,r}\gamma_{2,r}}{(1-\alpha)(\gamma_{1,r}+1)(\gamma_{2,r}+1) + \alpha(\gamma_{1,r+m}+1)(\gamma_{2,r+m}+1)}\frac{\eta+1}{\eta}. \tag{C.2}$$

C.2.2 Proof of Theorem 5.3

We will use the following lemma:

Lemma C.1. *If $V_b^{PC} \leq V_c \leq V_b^{NC}$ and $E^{BC}(V; V_b^{PC}) \geq 0$ for all $V \in (V_c, V_b^{NC})$, then $E^{BC}(V; V_b^{PC}) \geq E^{NC}(V; V_b^{NC}) \geq 0$ for all V.*

Proof of Lemma C.1. For $V \leq V_c$, we have $E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) = E^{\mathsf{PC}}(V; V_b^{\mathsf{PC}})$, by definition, and $E^{\mathsf{PC}}(V; V_b^{\mathsf{PC}}) \geq 0$ because V_b^{PC} is the optimal default barrier for the post-conversion firm and thus preserves limited liability for the post-conversion firm. Combining this with the hypothesis in the lemma yields $E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) \geq 0$ for all $V \leq V_b^{\mathsf{NC}}$. But for $V \leq V_b^{\mathsf{NC}}$, $E^{\mathsf{NC}}(V; V_b^{\mathsf{NC}}) = 0$, so the conclusion of the lemma holds for all $V \leq V_b^{\mathsf{NC}}$.

Now consider $V > V_b^{\mathsf{NC}}$. The value of equity before conversion is the difference between firm value and debt value and is given explicitly by

$$\begin{aligned} E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) &= V - (1-\alpha)\mathbb{E}^{\mathbb{Q}}[V_{\tau_b} e^{-r\tau_b}] \\ &\quad + \frac{P_1 \kappa_1 c_1}{r} \mathbb{E}^{\mathbb{Q}}[1 - e^{-r\tau_b}] + \frac{P_2 \kappa_2 c_2}{r} \mathbb{E}^{\mathbb{Q}}[1 - e^{-r\tau_c}] \\ &\quad - P_1 \left(\frac{c_1 + m}{r + m}\right) \mathbb{E}^{\mathbb{Q}}[1 - e^{-(r+m)\tau_b}] - P_2 \left(\frac{c_2 + m}{r + m}\right) \mathbb{E}^{\mathbb{Q}}[1 - e^{-(r+m)\tau_c}] \\ &\quad - \alpha \mathbb{E}^{\mathbb{Q}}[V_{\tau_b} e^{-(r+m)\tau_b}] - \frac{\Delta P_2}{1 + \Delta P_2} \mathbb{E}^{\mathbb{Q}}[e^{-(r+m)\tau_c} E^{\mathsf{PC}}(V_{\tau_c}; V_b^{\mathsf{PC}})]. \end{aligned} \tag{C.3}$$

Similarly, if we let τ_b^{NC} denote the first time V is at or below V_b^{NC}, we have

$$\begin{aligned} E^{\mathsf{NC}}(V; V_b^{NC}) &= V - (1-\alpha)\mathbb{E}^{\mathbb{Q}}[V_{\tau_b^{\mathsf{NC}}} e^{-r\tau_b^{\mathsf{NC}}}] \\ &\quad + \frac{(\kappa_1 c_1 P_1 + \kappa_2 c_2 P_2)}{r} \mathbb{E}^{\mathbb{Q}}[1 - e^{-r\tau_b^{\mathsf{NC}}}] \\ &\quad - \left\{ P_1 \left(\frac{c_1 + m}{r + m}\right) + P_2 \left(\frac{c_2 + m}{r + m}\right) \right\} \mathbb{E}^{\mathbb{Q}}[1 - e^{-(r+m)\tau_b^{\mathsf{NC}}}] \\ &\quad - \alpha \mathbb{E}^{\mathbb{Q}}[V_{\tau_b^{\mathsf{NC}}} e^{-(r+m)\tau_b^{\mathsf{NC}}}]. \end{aligned} \tag{C.4}$$

Note the fact that that $e^{-r\tau_b} \leq e^{-r\tau_b^{\mathsf{NC}}}$ and $V_{\tau_b} \leq V_{\tau_b^{\mathsf{NC}}}$. We then have

$$\mathbb{E}[(V_{\tau_b^{\mathsf{NC}}} e^{-r\tau_b^{\mathsf{NC}}} - V_{\tau_b} e^{-r\tau_b})(1 - e^{-m\tau_b^{\mathsf{NC}}})] \geq 0.$$

Using this inequality and taking the difference between (C.3) and (C.4), we get

$$\begin{aligned} &E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) - E^{\mathsf{NC}}(V; V_b^{\mathsf{NC}}) \\ &\geq E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) - E^{\mathsf{NC}}(V; V_b^{\mathsf{NC}}) - (1-\alpha)\mathbb{E}[(V_{\tau_b^{\mathsf{NC}}} e^{-r\tau_b^{\mathsf{NC}}} - V_{\tau_b} e^{-r\tau_b})(1 - e^{-m\tau_b^{\mathsf{NC}}})] \\ &= -(1-\alpha)\mathbb{E}^{\mathbb{Q}}[V_{\tau_b} e^{-r\tau_b - m\tau_b^{\mathsf{NC}}}] + \mathbb{E}^{\mathbb{Q}}[V_{\tau_b^{\mathsf{NC}}} e^{-(r+m)\tau_b^{\mathsf{NC}}}] \\ &\quad + \frac{P_1 \kappa_1 c_1}{r} \mathbb{E}^{\mathbb{Q}}[e^{-r\tau_b^{\mathsf{NC}}} - e^{-r\tau_b}] + \frac{P_2 \kappa_2 c_2}{r} \mathbb{E}^{\mathbb{Q}}[e^{-r\tau_b^{\mathsf{NC}}} - e^{-r\tau_c}] \\ &\quad - P_1 \left(\frac{c_1 + m}{r + m}\right) \mathbb{E}^{\mathbb{Q}}[e^{-(r+m)\tau_b^{\mathsf{NC}}} - e^{-(r+m)\tau_b}] - P_2 \left(\frac{c_2 + m}{r + m}\right) \mathbb{E}^{\mathbb{Q}}[e^{-(r+m)\tau_b^{\mathsf{NC}}} - e^{-(r+m)\tau_c}] \\ &\quad - \alpha \mathbb{E}^{\mathbb{Q}}[V_{\tau_b} e^{-(r+m)\tau_b}] - \frac{\Delta P_2}{1 + \Delta P_2} \mathbb{E}^{\mathbb{Q}}[e^{-(r+m)\tau_c} E^{\mathsf{PC}}(V_{\tau_c}; V_b^{\mathsf{PC}})]. \end{aligned} \tag{C.5}$$

On the other hand, we know that $E^{\mathsf{BC}}(V;V_b^{\mathsf{PC}}) \geq 0$ for all $V \leq V_b^{\mathsf{NC}}$, so it follows that

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-(r+m)\tau_b^{\mathsf{NC}}} E^{BC}(V_{\tau_b^{\mathsf{NC}}}; V_b^{\mathsf{PC}})\right] \geq 0.$$

Recall that the expectations here and in (C.3) and (C.4) are conditional expectations given the current time t and value $V_t = V$, though we have suppressed the conditioning to simplify the notation. To make the conditioning explicit, let $\mathbb{E}^{\mathbb{Q}}[\cdot|V_{\tau_b^{\mathsf{NC}}}]$ denote the expectation conditioned on $V = V_{\tau_b^{\mathsf{NC}}}$. Substituting (C.3) for E^{BC}, evaluated at $V = V_{\tau_b^{\mathsf{NC}}}$ and $t = \tau_b^{\mathsf{NC}}$ yields

$$\begin{aligned}
0 \leq \mathbb{E}^{\mathbb{Q}}\Bigg[& e^{-(r+m)\tau_b^{\mathsf{NC}}} V_{\tau_b^{NC}} - e^{-(r+m)\tau_b^{\mathsf{NC}}}(1-\alpha)\mathbb{E}^{\mathbb{Q}}\left[V_{\tau_b} e^{-r(\tau_b-\tau_b^{\mathsf{NC}})}|V_{\tau_b^{NC}}\right] \\
& + e^{-(r+m)\tau_b^{\mathsf{NC}}} \frac{P_1\kappa_1 c_1}{r}\mathbb{E}^{\mathbb{Q}}\left[1 - e^{-r(\tau_b-\tau_b^{\mathsf{NC}})}|V_{\tau_b^{\mathsf{NC}}}\right] + e^{-(r+m)\tau_b^{\mathsf{NC}}} \frac{P_2\kappa_2 c_2}{r}\mathbb{E}^{\mathbb{Q}}\left[1 - e^{-r(\tau_c-\tau_b^{\mathsf{NC}})}|V_{\tau_b^{\mathsf{NC}}}\right] \\
& - e^{-(r+m)\tau_b^{\mathsf{NC}}} P_1\left(\frac{c_1+m}{r+m}\right)\mathbb{E}^{\mathbb{Q}}\left[1 - e^{-(r+m)(\tau_b-\tau_b^{\mathsf{NC}})}|V_{\tau_b^{\mathsf{NC}}}\right] \\
& - e^{-(r+m)\tau_b^{\mathsf{NC}}} P_2\left(\frac{c_2+m}{r+m}\right)\mathbb{E}^{\mathbb{Q}}\left[1 - e^{-(r+m)(\tau_c-\tau_b^{\mathsf{NC}})}|V_{\tau_b^{\mathsf{NC}}}\right] \\
& - e^{-(r+m)\tau_b^{\mathsf{NC}}} \alpha\mathbb{E}^{\mathbb{Q}}\left[V_{\tau_b} e^{-(r+m)(\tau_b-\tau_b^{\mathsf{NC}})}|V_{\tau_b^{\mathsf{NC}}}\right] \\
& - e^{-(r+m)\tau_b^{\mathsf{NC}}} \frac{\Delta P_2}{1+\Delta P_2}\mathbb{E}^{\mathbb{Q}}\left[e^{-(r+m)(\tau_c-\tau_b^{\mathsf{NC}})} E^{\mathsf{PC}}(V_{\tau_c}; V_b^{\mathsf{PC}})|V_{\tau_b^{\mathsf{NC}}}\right]\Bigg].
\end{aligned} \tag{C.6}$$

The right side of (C.6) simplifies to

$$\begin{aligned}
& -(1-\alpha)\mathbb{E}^{\mathbb{Q}}[V_{\tau_b} e^{-r\tau_b - m\tau_b^{\mathsf{NC}}}] + \mathbb{E}^{\mathbb{Q}}[V_{\tau_b^{\mathsf{NC}}} e^{-(r+m)\tau_b^{\mathsf{NC}}}] \\
& + \frac{P_1\kappa_1 c_1}{r}\mathbb{E}^{\mathbb{Q}}[e^{-m\tau_b^{\mathsf{NC}}}(e^{-r\tau_b^{\mathsf{NC}}} - e^{-r\tau_b})] + \frac{P_2\kappa_2 c_2}{r}\mathbb{E}^{\mathbb{Q}}[e^{-m\tau_b^{\mathsf{NC}}}(e^{-r\tau_b^{\mathsf{NC}}} - e^{-r\tau_c})] \\
& - P_1\left(\frac{c_1+m}{r+m}\right)\mathbb{E}^{\mathbb{Q}}[e^{-(r+m)\tau_b^{\mathsf{NC}}} - e^{-(r+m)\tau_b}] - P_2\left(\frac{c_2+m}{r+m}\right)\mathbb{E}^{\mathbb{Q}}[e^{-(r+m)\tau_b^{\mathsf{NC}}} - e^{-(r+m)\tau_c}] \\
& - \alpha\mathbb{E}^{\mathbb{Q}}[V_{\tau_b} e^{-(r+m)\tau_b}] - \frac{\Delta P_2}{1+\Delta P_2}\mathbb{E}^{\mathbb{Q}}[e^{-(r+m)\tau_c} E^{\mathsf{PC}}(V_{\tau_c}; V_b^{\mathsf{PC}})],
\end{aligned}$$

which is less than or equal to the right side of (C.5) because $e^{-m\tau_b^{\mathsf{NC}}} \leq 1$. We have thus shown that $E^{\mathsf{BC}}(V;V_b^{\mathsf{PC}}) \geq E^{\mathsf{NC}}(V;V_b^{\mathsf{NC}}) \geq 0$ for $V > V_b^{\mathsf{NC}}$ and thus for all V. □

We now turn to the proof of the theorem. The post-conversion (PC) firm and the no-conversion (NC) firm have only straight debt, but the NC firm has more debt, so $V_b^{\mathsf{PC}} \leq V_b^{\mathsf{NC}}$, and the inequality is strict if $P_2 > 0$. (If $P_2 = 0$, the result holds trivially.) We distinguish three cases based on the position of the conversion trigger relative to these default barriers.

Case 1: $V_b^{\mathsf{PC}} \le V_c \le V_b^{\mathsf{NC}}$. For all barrier choices V_b with $V_b \le V_c$, conversion precedes default,[1] and the only choice among those that satisfies the commitment condition is V_b^{PC}. For all feasible barrier choices $V_b \ge V_c$, default precedes conversion, so $E^{\mathsf{BC}}(V; V_b) = E^{\mathsf{NC}}(V; V_b)$, and the optimal choice among such barriers is V_b^{NC}. Thus, these are the only two candidates for the optimal barrier level. If V_b^{PC} is consistent with limited liability for the BC firm, then Lemma C.1 implies that $E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) \ge E^{\mathsf{NC}}(V; V_b^{\mathsf{NC}}) \equiv E^{\mathsf{BC}}(V; V_b^{\mathsf{NC}})$, for all V. Thus, $V_b^* = V_b^{\mathsf{PC}}$ if V_b^{PC} is feasible, and otherwise $V_b^* = V_b^{\mathsf{NC}}$.

Case 2: $V_c < V_b^{\mathsf{PC}}$. For $V_b < V_b^{\mathsf{PC}}$, it follows from Chen and Kou (2009) that the equity valuation $V \mapsto E^{\mathsf{PC}}(V; V_b)$ violates limited liability, so no $V_b \le V_c$ is feasible in this case. For all $V_b > V_c$, we have $E^{\mathsf{BC}}(V; V_b) \equiv E^{\mathsf{NC}}(V; V_b)$, so the optimal choice is V_b^{NC}.

Case 3: $V_b^{\mathsf{NC}} < V_c$. If default precedes conversion, equity value is given by $E^{\mathsf{NC}}(V; V_b)$. For each V, this is a decreasing function of V_b for $V_b \ge V_b^{\mathsf{NC}}$; thus, no $V_b > V_c$ can be optimal. Among barriers $V_b \le V_c$ for which conversion precedes default, only V_b^{PC} satisfies the commitment condition. Thus, we need to compare the default barriers V_b^{PC} and V_c, with default preceding conversion in the latter case. The argument in Lemma C.1 now applies directly, replacing τ_b^{NC} with τ_c, and shows that $E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) \ge E^{\mathsf{NC}}(V; V_c)$, for all $V > V_c$. The inequality is strict at $V = V_c$, so the optimal barrier is V_b^{PC}. □

C.2.3 Proof of Theorem 4.2

Set $\bar{P}_1 = V_c/\epsilon_1$. If $P_2 = 0$ (so that the firm has only straight debt), then the optimal default barrier is $V_b^* = P_1\epsilon_1$. Thus, $V_b^* \le V_c$ if $P_1 \le \bar{P}_1$, and $V_b^* > V_c$ if $P_1 > \bar{P}_1$, which confirms that $\bar{P}_1$ is indeed the critical debt level in the absence of CoCos.

Now suppose $P_1 < \bar{P}_1$ and $P_2 > 0$. From Theorem 5.3, we know that debt-induced collapse occurs when setting the default barrier at V_b^{PC} is infeasible because it violates limited liability; that is, when $E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) < 0$ for some $V > V_c$. For any $V > V_c$, we have

$$E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) = E^{\mathsf{PC}}(V; V_b^{\mathsf{PC}}) + P_2 A - P_2 B - \frac{\Delta P_2}{1 + \Delta P_2} M, \tag{C.7}$$

where

$$A = \kappa_2 \frac{c_2}{r} \mathbb{E}^{\mathbb{Q}}\left[1 - e^{-r\tau_c}\right], \qquad B = \left(\frac{c_2 + m}{m + r}\right) \mathbb{E}^{\mathbb{Q}}\left[1 - e^{-(r+m)\tau_c}\right]$$

and

$$M = \mathbb{E}^{\mathbb{Q}}\left[e^{-(r+m)\tau_c} E^{\mathsf{PC}}(V_{\tau_c}; V_b^{\mathsf{PC}})\right] \ge 0.$$

Here, A gives the normalized value of the funding benefits from CoCos, and B is the normalized value of the coupons and principal for the CoCos. Each of these (and M) is a function of the current asset level V, though we suppress this dependence in the notation.

[1]Recall our convention that when we write $V_b \le V_c$, the order of events at $V_b = V_c$ is taken to be consistent with $V_b < V_c$, and when we write $V_b \ge V_c$ the opposite order of events is assumed.

Suppose $A < B$. This means that the funding benefit received is less than the value of the payments made on the debt, as we would expect in practice. In this case, the right side of (C.7) is decreasing continuously and without bound as P_2 increases. We may therefore define $\bar{P}_2^V$ to be the smallest P_2 at which (C.7) equals zero and then set

$$\bar{P}_2 = \inf\{\bar{P}_2^V : V > V_c\}.$$

If $P_2 > \bar{P}_2$, then $P_2 > \bar{P}_2^V$ for some $V > V_c$, and then $E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) < 0$ for some $V > V_c$, so limited liability fails, V_b^{PC} is infeasible, and we have debt-induced collapse. If $P_2 \leq \bar{P}_2$, then $P_2 \leq \bar{P}_2^V$ for all $V > V_c$ and $E^{\mathsf{BC}}(V; V_b^{\mathsf{PC}}) \geq 0$, so V_b^{PC} is feasible and then optimal.

For the alternative case $A \leq B$, it is not hard to see that $E^{\mathsf{PC}}(V; V_b^{\mathsf{PC}}) \geq M$, so (C.7) remains positive at all $P_2 \geq 0$, and the result holds with $\bar{P}_2 = \infty$. □

C.3 The Extended Model

The extended model used in the numerical illustrations of Sections 4.4-4.7 extends (4.1) and (4.2) to allow two types of jumps — firm-specific jumps and market-wide jumps, with respective arrival rates λ_f and λ_m, and mean jump sizes η_f and η_m. In both cases, the jump sizes are exponentially distributed and decrease asset value. The extended model also allows more layers of debt: deposits (with or without insurance), ordinary debt, subordinated debt, and CoCos, in decreasing seniority. The multiple layers can be valued using the approach in Section 4.2.3. All parts of the capital structure can be valued in terms of transforms of τ_b and τ_c, and these transforms can be expressed in terms of roots of an equation; see Chen et al. (2012) for details. Table C.1 shows parameter values for the numerical examples. The subscripts $1a$, $1b$, and $1c$ distinguish the three layers of non-convertible debt.

Parameter		Value
initial asset value	V_0	100
debt principal	(P_{1a}, P_{1b}, P_{1c})	$(40, 30, 15)$
risk free rate	r	6%
volatility	σ	8%
payout rate	δ	1%
funding benefit	κ	35%
firm specific jump intensity	λ_f	.2
market jump intensity	λ_m	.05
firm specific jump exponent	η_f	4
market jump exponent	η_m	3
coupon rates	$(c_{1a}, c_{1b}, c_{1c}, c_2)$	$(r, r+3\%, r+3\%, r)$
deposits insurance premium rate	φ	1%
contingent capital principal	P_2	1 or 5
maturity profile exponent – base case	$(m_{1a}, m_{1b}, m_{1c}, m_2)$	$(1, 1/4, 1/4, 1/4)$
maturity profile exponent – long maturity	$(m_{1a}, m_{1b}, m_{1c}, m_2)$	$(1, 1/16, 1/16, 1/16)$ or $(1, 1/25, 1/25, 1/25)$
conversion trigger	V_c	75 (in most cases)
conversion loss (if applied)		20% of value of shares

Table C.1: Parameters for extended model. Asset returns have a total volatility (combining jumps and diffusion) of 20.6% and overall drift rate of 3.3%. In the base case, the number of shares Δ issued at conversion is set such that if conversion happens at exactly V_c, the market value of shares delivered is the same as the face value of the converted debt.

Appendix D

Appendix to Chapter 5

D.1 Alternative Contract Formulation

In this section we present the details for modeling FVC2s. The number of shares granted at conversion is determined based on the stock price $S(V_{\tau_C})$, i.e. the contingent convertible bondholders receive $n' = \frac{\ell P_C}{S(V_{\tau_C})}$ shares. We do not require Assumption 5.2 to be satisfied.

Assume that the value of equity for $V_{\tau_C} = V_C$ is sufficient to pay the conversion value. Then, in a model without jumps and $\ell = 1$ the contingent convertible bond with face value 1 has the same features as a riskless bond with face value 1. If we include jumps the contingent convertible bondholders face the additional risks that conversion and bankruptcy happen simultaneously or that the value of the equity after conversion is not sufficient to pay the promised conversion value. Hence, jumps introduce two additional sources of risk.

D.1.1 Evaluation of FVC2s

The conversion value for FVC2s requires us to distinguish several cases. If $\tau_C < \tau$, i.e. the downward movement of V_{τ_C} is not sufficient to trigger bankruptcy, the contingent convertible bondholders receive a payment. If on the one hand the value of the equity is sufficiently large, they get a number of stocks such that the value of the total payment equals ℓP_C. If on the other hand the value of the equity is insufficient to make the promised payment to the contingent convertible bond holders, they take possession of the whole equity and the old shareholders are completely diluted out. We assume that the face value of all contingent convertible debt is P_C and thus a bondholder with a bond with face value 1 gets a fraction $1/P_C$ of the value of the equity $EQ(V_{\tau_C})$ after conversion in this case. We start by proving proposition 5.7.

Proof of Proposition 5.7:

Proof. We only need to do the calculations for the conversion value:

$$
\begin{aligned}
CONV =& \int_0^\infty p_C \Psi(t) d_C(V, V_B, V_C, t) dt \\
=& \ell p_C \mathbb{E}\left[e^{-r\tau_C} \int_{\tau_C}^\infty e^{-mt} dt \mathbb{1}_{\{\ell P_C \leq EQ(V_{\tau_C})\}} \mathbb{1}_{\{\tau_C < \tau\}} \right] \\
&+ \frac{p_C}{P_C} \mathbb{E}\left[e^{-r\tau_C} EQ(V_{\tau_C}) \int_{\tau_C}^\infty e^{-mt} dt \mathbb{1}_{\{\tau_C < \tau\}} \mathbb{1}_{\{\ell P_C > EQ(V_{\tau_C})\}} \right] \\
=& \ell \frac{p_C}{m} \mathbb{E}\left[e^{-(r+m)\tau_C} \mathbb{1}_{\{\ell P_C \leq EQ(V_{\tau_C})\}} \mathbb{1}_{\{\tau_C < \tau\}} \right] \\
&+ \frac{p_C}{m P_C} \mathbb{E}\left[e^{-(r+m)\tau_C} EQ(V_{\tau_C}) \mathbb{1}_{\{\tau_C < \tau\}} \mathbb{1}_{\{\ell P_C > EQ(V_{\tau_C})\}} \mathbb{1}_{\{\tau_C < \infty\}} \right]
\end{aligned}
$$

□

Equipped with the results of Section 5.4, we can derive the price of FVC2s.

Theorem D.1. *The price of the FVC2s equals*

$$
\begin{aligned}
CB =& \frac{C_C + m P_C}{r+m} \left(1 - \frac{\beta_{4,r+m}}{\eta_2} \frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}} \left(\frac{V_C}{V} \right)^{\beta_{3,r+m}} - \frac{\beta_{3,r+m}}{\eta_2} \frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}} \left(\frac{V_C}{V} \right)^{\beta_{4,r+m}} \right) \\
&+ \ell P_C \cdot G\left(\log\left(\frac{V_C}{V_0} \right), \log\left(\frac{V_C}{\max(T^{-1}(\ell P_C), V_B)} \right), m+r \right) \mathbb{1}_{\{V_C > T^{-1}(\ell P_C)\}} \\
&+ \sum \alpha_i V_0^{\theta_i} \Bigg(J\left(\log\left(\frac{V_C}{V_0} \right), \theta_i, \log\left(\frac{V_C}{V_B} \right), m+r \right) \\
&- J\left(\log\left(\frac{V_C}{V_0} \right), \theta_i, \log\left(\frac{V_C}{T^{-1}(\ell P_C)} \right), m+r \right) \mathbb{1}_{\{V_C > T^{-1}(\ell P_C)\}} \Bigg) \mathbb{1}_{\{V_B < T^{-1}(\ell P_C)\}}
\end{aligned}
$$

Proof. We only need to prove the formula for the conversion value.

$$\begin{aligned}
CONV =& \ell P_C \mathbb{E}\left[e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\tau\}}\mathbb{1}_{\{\ell P_C \le EQ(V_{\tau_C})\}}\right] \\
&+ \mathbb{E}\left[EQ(V_{\tau_C})e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\tau\}}\mathbb{1}_{\{\ell P_C > EQ(V_{\tau_C})\}}\mathbb{1}_{\{\tau_C<\infty\}}\right] \\
=& \ell P_C \mathbb{E}\left[e^{-(m+r)\tau_C}\mathbb{1}_{\{V_{\tau_C}>\max(T^{-1}(\ell P_C),V_B)\}}\right] \\
&+ \mathbb{E}\left[EQ(V_{\tau_C})e^{-(m+r)\tau_C}\mathbb{1}_{\{V_B<V_{\tau_C}\le T^{-1}(\ell P_C)\}}\mathbb{1}_{\{\tau_C<\infty\}}\right]\mathbb{1}_{\{V_B<T^{-1}(\ell P_C)\}} \\
=& \ell P_C \mathbb{E}\left[e^{-(m+r)\tau_C}\mathbb{1}_{\left\{-(X(\tau_C)-x_C)<-\log\left(\frac{\max(T^{-1}(\ell P_C),V_B)}{V_C}\right)\right\}}\right] \\
&+ \sum \alpha_i V_0^{\theta_i}\mathbb{E}\left[e^{-(m+r)\tau_C+\theta_i X(\tau_C)}\mathbb{1}_{\{V_B<V_{\tau_C}\le T^{-1}(\ell P_C)\}}\mathbb{1}_{\{\tau_C<\infty\}}\right]\mathbb{1}_{\{V_B<T^{-1}(\ell P_C)\}} \\
=& \ell P_C \mathbb{E}\left[e^{-(m+r)\tau_C}\mathbb{1}_{\left\{-(X(\tau_C)-x_C)<-\log\left(\frac{\max(T^{-1}(\ell P_C),V_B)}{V_C}\right)\right\}}\right] \\
&+ \sum \alpha_i V_0^{\theta_i} E\Bigg[e^{-(m+r)\tau_C+\theta_i X(\tau_C)}\left(\mathbb{1}_{\{-(X_{\tau_C}-x_C)<\log(V_C/V_B)\}} - \mathbb{1}_{\{-(X_{\tau_C}-x_C)<\log(V_C/T^{-1}(\ell P_C))\}}\right) \\
&\mathbb{1}_{\{\tau_C<\infty\}}\Bigg]\mathbb{1}_{\{V_B<T^{-1}(\ell P_C)\}}
\end{aligned}$$

□

D.1.2 Costs of Dilution for FVC2s

For contingent convertible bonds with a flexible number of shares at conversion, the dilution costs are much more complicated. Here, the number of shares depends on the stock price value $S(\tau_C)$ at the time of conversion. In more detail, the number of shares of the old shareholders are

$$n = \frac{EQ(V_0) - DC(V_0)}{S(0)}$$

while the number of new shares of the bondholders equal

$$n' = \frac{\ell P_C}{S(\tau_C)}$$

if the equity value $EQ(V_{\tau_C})$ is sufficiently high. Otherwise, contingent convertible bondholders own the whole remaining equity. Note, that n' is a random variable at time $t = 0$. Therefore the dilution costs DC at time $t = 0$ are

$$\begin{aligned}
DC(V_0) =& \mathbb{E}\left[EQ(V_{\tau_C})e^{-(r+m)\tau_C}\frac{\ell P_C}{(EQ(V_0)-DC(V_0))S(\tau_C)/S(0)+\ell P_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{EQ(V_{\tau_C})\ge \ell P_C\}}\right] \\
&+ \mathbb{E}\left[EQ(V_{\tau_C})e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{EQ(V_{\tau_C})<\ell P_C\}}\right].
\end{aligned}$$

Under certain assumptions on the stock price process this equation boils down to $CONV(V_0)$.

D.1.3 Modeling the Stock Price Process for FVC2s

If FVC2s are included in the capital structure, there is the possibility that old shareholders are completely diluted out before bankruptcy. Hence, the stock price of the "old" shares can be zero, although the company has not defaulted. However, there will be the "new" shares of the former contingent convertible bondholders with a positive value. Hence, we need to distinguish between "old" and "new" shares:

Definition D.1. *The endogenous stock price for the old shares is defined as*

$$S_{old}(t) = S_{old}(V_t) = \begin{cases} \frac{EQ(V_t)-DC(V_t)}{n} & \textit{if } t > \tau_C \\ \frac{EQ(V_t)}{n+n'} & \textit{if } \tau_C \leq t < \tau \textit{ and } EQ(\tau_C) \geq \ell P_C \\ 0 & \textit{if } \tau_C \geq t \textit{ and } EQ(\tau_C) < \ell P_C \textit{ or if } \tau \leq t \end{cases}$$

and the price for the new shares is

$$S_{new}(t) = S_{new}(V_t) = \begin{cases} S_{old}(t) & \textit{if } t < \tau_C \\ S_{old}(t) & \textit{if } \tau_C \leq t < \tau \textit{ and } EQ(\tau_C) \geq \ell P_C \\ EQ(V_t) & \textit{if } \tau_C \leq t < \tau \textit{ and } EQ(\tau_C) < \ell P_C \textit{ (we have normalized } n' = 1\textit{)} \end{cases}$$

where n is the number of "old" shares

$$n = \frac{EQ(V_0) - DC(V_0)}{S_{old}(0)}$$

and n' is the number of "new" shares issued at conversion

$$n' = \frac{\ell P_C}{S_{old}(\tau_C)}.$$

In the case of FVC2s the conversion value and dilution costs must also coincide:

Proposition D.1. *The conversion value for FVC2s equals the dilution costs:*

$$CONV(V_t) = DC(V_t)$$

Proof. Recall that n and n' are set as

$$n = \frac{EQ(V_0) - DC(V_0)}{S_{old}(0)}$$

and

$$n' = \frac{\ell P_C}{S_{old}(\tau_C)}.$$

At the time of conversion the stock price $S_{old}(\tau_C)$ equals the equity value divided by the number of old and new shares:

$$\begin{aligned} S_{old}(\tau_C)\mathbb{1}_{\{EQ(V_{\tau_C})\geq \ell P_C\}} &= \frac{EQ(V_{\tau_C})}{n+n'}\mathbb{1}_{\{EQ(V_{\tau_C})\geq \ell P_C\}} \\ &= \frac{EQ(V_{\tau_C})}{n+\frac{\ell P_C}{S_{old}(\tau_C)}}\mathbb{1}_{\{EQ(V_{\tau_C})\geq \ell P_C\}} \end{aligned}$$

which is equivalent to

$$\begin{aligned} (S_{old}(\tau_C)n+\ell P_C)\mathbb{1}_{\{EQ(V_{\tau_C}\geq \ell P_C\}} &= EQ(V_{\tau_C})\mathbb{1}_{\{EQ(V_{\tau_C}\geq \ell P_C\}} \qquad \Leftrightarrow \\ S_{old}(\tau_C)\mathbb{1}_{\{EQ(V_{\tau_C}\geq \ell P_C\}} &= \frac{EQ(V_{\tau_C})-\ell P_C}{n}\mathbb{1}_{\{EQ(V_{\tau_C}\geq \ell P_C\}} \qquad \Leftrightarrow \\ \frac{S_{old}(\tau_C)}{S_{old}(0)}\mathbb{1}_{\{EQ(V_{\tau_C}\geq \ell P_C\}} &= \frac{EQ(V_{\tau_C})-\ell P_C}{EQ(V_0)-DC(V_0)}\mathbb{1}_{\{EQ(V_{\tau_C}\geq \ell P_C\}}. \end{aligned}$$

Hence, the dilution costs simplify to

$$\begin{aligned} &DC(V_0) \\ =&\mathbb{E}\left[EQ(V_{\tau_C})e^{-(r+m)\tau_C}\frac{\ell P_C}{(EQ(V_0)-DC(V_0))S(\tau_C)/S(0)+\ell P_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{EQ(V_{\tau_C}\geq \ell P_C\}}\mathbb{1}_{\{\tau_C<\tau\}}\right] \\ &+\mathbb{E}\left[EQ(V_{\tau_C})e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C<\infty\}}\mathbb{1}_{\{EQ(V_{\tau_C})<\ell P_C\}}\mathbb{1}_{\{\tau_C<\tau\}}\right] \\ =&\ell P_C\mathbb{E}\left[e^{-(r+m)\tau_C}\mathbb{1}_{\{\ell P_C\leq EQ(V_{\tau_C})\}}\mathbb{1}_{\{\tau_C<\tau\}}\right] \\ &+\mathbb{E}\left[e^{-(r+m)\tau_C}EQ(V_{\tau_C})\mathbb{1}_{\{\tau_C<\tau\}}\mathbb{1}_{\{\ell P_C>EQ(V_{\tau_C})\}}\mathbb{1}_{\{\tau_C<\infty\}}\right] \\ =&CONV(V_t). \end{aligned}$$

□

D.2 Comparing Contract Specifications

The two contracts FVC and FSC differ in the specification of the number of shares n' granted to the contingent convertible shareholders in the event of conversion. We have seen that FVC1s are actually a particular version of FSCs. Here, we want to analyze whether we can make any statement about n' in the both cases.

The old shareholders own a number of shares n that is fixed at time $t=0$ and that is equal to the value of equity to them divided by the price of the stock at time $t=0$:

$$n=\frac{EQ(V_0)-DC(V_0)}{S(0)}$$

Assume that in the case of FSCs the new shareholders (i.e. the holders of contingent convertible bonds) receive a fixed number of shares n' that satisfies a posteriori the following condition:

$$n'_{FSC} = \frac{P_C \ell}{S_0} = \frac{n \ell P_C}{EQ(V_0) - DC(V_0)}.$$

In the case of FVC1s the corresponding number is

$$n'_{FVC1} = \frac{\ell P_C}{S(V_C)} = \frac{n \ell P_C}{EQ(V_C) - \ell P_C}.$$

Note that the last equality follows from the assumption that

$$\ell P_C = n'_{FVC1} S(V_C) = \frac{n'_{FVC1}}{n + n'_{FVC1}} EQ(V_C).$$

How do the two numbers n'_{FSC} and n'_{FVC1} relate to each other? Denote the contract parameters for the two specifications by ℓ_{FSC} respectively ℓ_{FVC1}. Assume that $\ell_{FVC1} \geq \ell_{FSC}$. By the definition of n we can conclude

$$\begin{aligned} n'_{FVC1} &= \frac{n \ell_{FVC1} P_C}{EQ(V_C) - \ell P_C} \\ &= \frac{EQ(V_0) - DC(V_0)}{EQ(V_C) - \ell_{FVC1} P_C} \frac{\ell_{FSC} P_C}{S_0} \frac{\ell_{FVC1}}{\ell_{FSC}} \\ &= \frac{EQ(V_0) - DC(V_0)}{EQ(V_C) - \ell_{FVC1} P_C} n'_{FSC} \frac{\ell_{FVC1}}{\ell_{FSC}}. \end{aligned}$$

Hence

$$\frac{n'_{FVC1}}{n'_{FSC}} = \underbrace{\frac{EQ(V_0) - DC(V_0)}{EQ(V_C) - \ell_{FVC1} P_C}}_{>1} \frac{\ell_{FVC1}}{\ell_{FSC}}.$$

The inequality $\frac{EQ(V_0) - DC(V_0)}{EQ(V_C) - \ell_{FVC1} P_C} > 1$ holds in our model because we will show later that $EQ(.)$ is a strictly increasing function and that $DC(V_0) < \ell_{FSC} P_C$.

Lemma D.1. *If $\ell_{FVC1} \geq \ell_{FSC}$ then $n'_{FVC1} > n'_{FSC}$.*

It is important to note, that the numbers n'_{FVC1} and n'_{FSC} are both constants independently of future realizations of V_t.

D.3 Pure Diffusion Case

In the pure diffusion case the value of the firm's assets which follows a geometric Brownian motion is given by

$$dV_t = V_t((r-\delta)dt + \sigma dW_t).$$

Thus the process X is simply

$$X_t = \left(r - \delta - \frac{1}{2}\sigma^2\right) t + \sigma W_t.$$

and its Laplace exponent is given by

$$\psi(z) = \frac{1}{2}\sigma^2 z^2 + \left(r - \delta - \frac{1}{2}\sigma^2\right) z.$$

The default time τ is defined as $\tau = \tau_x = \inf(t \geq 0 : X(t) \leq x)$ with $x = \log(V_B/V)$ and $\tau_C = \tau_{x_C} = \inf(t \geq 0 : X(t) \leq x_C)$ with $x_C = \log(V_C/V)$. The Laplace exponent of τ is calculated for example in Duffie (2001):

Lemma D.2. *The Laplace exponent of τ_x for the pure diffusion process X equals*

$$E\left[e^{-\lambda \tau_x}\right] = e^{\beta_\lambda x}$$

where $\beta_\lambda = \frac{\gamma + \sqrt{\gamma^2 + 2\sigma^2\lambda}}{\sigma^2}$ and $\gamma = r - \delta - \frac{1}{2}\sigma^2$. Note that $-\beta_\lambda$ is the negative root of the equation $\psi(z) = \lambda$.

The evaluation of the straight debt is presented in Leland (1994b):

Proposition D.2. *The value of the debt equals*

$$D = D(V, V_B) = \frac{C_D + mP_D}{r+m}\left(1 - \left(\frac{V_B}{V}\right)^{\beta_{r+m}}\right) + (1-\alpha)V_B\left(\frac{V_B}{V}\right)^{\beta_{r+m}}$$

while the total value of the firm is

$$G_{debt}(V, V_B) = V + \frac{\bar{c}C_D}{r}\left(1 - \left(\frac{V_B}{V}\right)^{\beta_r}\right) - \alpha V_B\left(\frac{V_B}{V}\right)^{\beta_r}.$$

The law of the first passage time equals:

$$\mathbb{P}(\tau \leq t) = \Phi(h_1) + \exp\left(\frac{2\mu x}{\sigma^2}\right)\Phi(h_2),$$

where $h_1 = \frac{x - \gamma t}{\sigma\sqrt{t}}$, $h_2 = \frac{x + \gamma t}{\sigma\sqrt{t}}$ and $x = \log(\frac{V_B}{V_0})$. Finally, the optimal barrier level is:

$$V_B^* = \frac{\frac{C_D + mP_D}{r+m}\beta_{r+m} - \frac{\bar{c}C_D}{r}\beta_r}{1 + \alpha\beta_r + (1-\alpha)\beta_{r+m}}.$$

D.3.1 FVCs in a Pure Diffusion Model

The distinction between FVC1s and FVC2s was due to the possibility of jumps. In a pure diffusion model both contracts coincide. Here we require that Assumption 5.2 is satisfied, i.e. we consider only contracts where the value of the equity after conversion is sufficient to make the promised payment.

Proposition D.3. *In a pure diffusion model the two contracts FVC1 and FVC2 are identical. The price of the CCBs is given by*

$$CB(V,V_C) = \left(\frac{c_C P_C + mP_C}{m+r}\right) + P_C\left(\frac{(m+r)\ell - c_C - m}{m+r}\right)\left(\frac{V_C}{V}\right)^{\beta_{m+r}}.$$

The price of FVCs is completely independent of any features of the straight debt.

Proof. First note that $V_{\tau_C} = V_C$ and hence $EQ(V_{\tau_C}) = EQ(V_C)$.

$$\begin{aligned} CB(V,V_C) &= \frac{cP_C + mP_C}{m+r} E\left[1 - e^{-(m+r)\tau_C}\right] + \ell P_C E\left[e^{-(m+r)\tau_C}\right] \\ &= \frac{cP_C + mP_C}{m+r} + P_C\left(\frac{(m+r)\ell - c - m}{m+r}\right) E\left[e^{-(m+r)\tau_C}\right]. \end{aligned}$$

□

Lemma D.3. *The limit for $m \to 0$ corresponds to the case where only consol bonds are issued. The price of FVCs simplifies to*

$$CB(V,V_C) = \frac{c_C P_C}{r} + P_C\left(\frac{r\ell - c_C}{r}\right)\left(\frac{V_C}{V}\right)^{\beta_r}.$$

Remark D.1. *Albul, Jaffee and Tchistyi's (2010) model is the special case for $m \to 0$.*

D.3.2 FSCs in a Pure Diffusion Model

Proposition D.4. *The price of FSCs in a pure diffusion model equals*

$$CB(V,V_B,V_C) = \frac{c_C P_C + mP_C}{m+r}\left(1 - \left(\frac{V_C}{V}\right)^{\beta_{m+r}}\right) + \frac{n'}{n+n'} EQ_{debt}(V_C)\left(\frac{V_C}{V}\right)^{\beta_{m+r}}$$

where

$$\begin{aligned} EQ_{debt}(V_C) =& V_C + \frac{\bar{c}C_D}{r}\left(1 - \left(\frac{V_B}{V_C}\right)^{\beta_r}\right) - \alpha V_B\left(\frac{V_B}{V_C}\right)^{\beta_r} \\ &- \frac{C_D + mP_D}{r+m}\left(1 - \left(\frac{V_B}{V_C}\right)^{\beta_{r+m}}\right) + (1-\alpha)V_B\left(\frac{V_B}{V_C}\right)^{\beta_{r+m}}. \end{aligned}$$

Proof. First note that $V_{\tau_C} = V_C$ and hence $EQ(V_{\tau_C}) = EQ(V_C)$.

$$CB(V, V_B, V_C) = \frac{c_C P_C + m P_C}{m + r} \left(1 - E\left[e^{-(m+r)\tau_C}\right]\right) + \frac{n'}{n + n'} EQ_{debt}(V_C) E\left[e^{-(m+r)\tau_C}\right].$$

□

In contrast to FVCs the price of FSCs explicitly depends on V_B and thus on the features of the straight debt.

Lemma D.4. *For $m \to 0$ we obtain the special case of consol bonds. The pricing formula simplifies to*

$$CB(V, V_B, V_C) = \frac{c_C P_C}{r} \left(1 - \left(\frac{V_C}{V}\right)^{\beta_r}\right) + \frac{n'}{n + n'} EQ_{debt}(V_C) \left(\frac{V_C}{V}\right)^{\beta_r}$$

where

$$EQ_{debt}(V, V_B, V_C) = V + \frac{\bar{c} C_D}{r} + \frac{\bar{c} C_D}{r} - \frac{\bar{c} C_D + \bar{c} C_C + r\ell P_C - c_C P_C}{r} \left(\frac{V_C}{V}\right)^{\beta_r}.$$

D.4 Proofs

D.4.1 Proofs for Section 5.2

Proof of Proposition 5.2:

Proof. By definition, the total value of the firm equals

$$\begin{aligned}
G_{debt}(V, V_B) &= V + TB_D(V, V_B) - BC(V, V_B) \\
&= V + \bar{c} C_D \mathbb{E}\left[\int_0^\tau e^{-rt} dt\right] - \alpha \mathbb{E}\left[V(\tau) e^{-r\tau} \mathbb{1}_{\{\tau < \infty\}}\right] \\
&= V + \frac{\bar{c} C_D}{r} \mathbb{E}\left[1 - e^{-r\tau}\right] - \alpha \mathbb{E}\left[V(\tau) e^{-r\tau} \mathbb{1}_{\{\tau < \infty\}}\right].
\end{aligned}$$

□

D.4.2 Proofs for Section 5.3

Proof of Proposition 5.3:

Proof. First note that

$$\Psi(s) = \int_s^\infty \varphi(y) dy = \int_s^\infty m e^{-my} dy = e^{-sm}$$

Hence, it follows that $P_C = p_C \int_0^\infty \Psi(s)ds = p_C \int_0^\infty e^{-ms}ds = \frac{p_C}{m}$. Therefore, we get

$$
\begin{aligned}
CB(V, V_B, V_C) = &\int_0^\infty p\Psi(t) d_C(V, V_B, V_C, t)dt \\
=& p_C c_C \mathbb{E}\left[\int_0^{\tau_C} e^{-rt} \int_t^\infty \Psi(s)dsdt\right] + p_C \mathbb{E}\left[\int_0^{\tau_C} e^{-rs}\Psi(t)dt\right] \\
&+ \int_0^\infty p_C \cdot \Psi(t) \cdot conv(V, V_B, V_C, t)dt \\
=& p_C c_C \mathbb{E}\left[\int_0^{\tau_C} e^{-rt}\frac{1}{m}e^{-mt}dt\right] + p_C \mathbb{E}\left[\int_0^{\tau_C} e^{-rt}e^{-mt}dt\right] + CONV(V, V_B, V_C) \\
=& \left(\frac{p_C c_C}{m} + p_C\right)\mathbb{E}\left[\frac{1}{-(r+m)}\left(e^{-(r+m)\tau_C} - 1\right)\right] + CONV(V, V_B, V_C) \\
=& \frac{c_C P_C + mP_C}{m+r}\mathbb{E}\left[1 - e^{-(m+r)\tau_C}\right] + CONV(V, V_B, V_C).
\end{aligned}
$$

□

Proof of Proposition 5.4:

Proof. We only need to calculate the total conversion value:

$$
\begin{aligned}
CONV =& \int_0^\infty p_C \Psi(t) d_C(V, V_B, V_C, t)dt \\
=& \frac{n'}{P_C}\int_0^\infty p_C e^{-mt}\mathbb{E}\left[S(\tau_C)e^{-r\tau_C}\mathbb{1}_{\{\tau_C \le t\}}\mathbb{1}_{\{V_{\tau_C} > V_B\}}\right]dt \\
=& \frac{n'}{P_C}p_C \mathbb{E}\left[S(\tau_C)e^{-r\tau_C}\int_{\tau_C}^\infty e^{-ms}ds \mathbb{1}_{\{V_{\tau_C} > V_B\}}\mathbb{1}_{\{\tau_C < \infty\}}\right] \\
=& \frac{n'}{P_C}p_C \mathbb{E}\left[S(\tau_C)e^{-r\tau_C}\frac{1}{m}e^{-m\tau_C}\mathbb{1}_{\{V_{\tau_C} > V_B\}}\mathbb{1}_{\{\tau_C < \infty\}}\right] \\
=& \frac{n'}{P_C}\frac{p_C}{m}\mathbb{E}\left[S(\tau_C)e^{-(m+r)\tau_C}\mathbb{1}_{\{\tau_C < \infty\}}\mathbb{1}_{\{V_{\tau_C} > V_B\}}\right].
\end{aligned}
$$

□

Proof of Proposition 5.5:

Proof. Equation 5.7 has to be satified for V_0, which implies

$$DC(V_0) = \frac{n'S_0}{EQ(V_0) - DC(V_0) + n'S_0}\mathbb{E}\left[EQ(V_{\tau_C})e^{-(r+m)\tau_C}\mathbb{1}_{\{\tau_C < \infty\}}\mathbb{1}_{\{\tau_C < \tau\}}\right]$$

Solving for $DC(V_0)$ yields

$$DC(V_0) = \frac{EQ(V_0) + n'S_0}{2} \pm \sqrt{\left(\frac{EQ(V_0) + n'S_0}{2}\right)^2 - \mathbb{E}\left[EQ(V_{\tau_C})e^{-(r+m)\tau_C}\mathbb{1}_{\{\tau_C < \infty\}}\mathbb{1}_{\{\tau_C < \tau\}}\right]n'S_0}.$$

As the price of FSCs at time zero is a function of $DC(0)$, multiplicity of the market value of the dilution costs results in multiple equilibrium prices. □

Proof of Proposition 5.6:

Proof. Recall that under Assumption 5.2 n and n' are set as

$$n = \frac{EQ(V_0) - DC(V_0)}{S_0} \qquad \text{and} \qquad n' = \frac{\ell P_C}{S(V_C)}.$$

At the time of conversion the stock price $S(V_C)$ equals the equity value divided by the number of old and new shares:

$$S(V_C) = \frac{EQ(V_C)}{n+n'} = \frac{EQ(V_C)}{n + \frac{\ell P_C}{S(V_C)}}$$

which is equivalent to

$$\begin{aligned} S(V_C)n + \ell P_C &= EQ(V_C) && \Leftrightarrow \\ S(V_C) &= \frac{EQ(V_C) - \ell P_C}{n} && \Leftrightarrow \\ \frac{S(V_C)}{S_0} &= \frac{EQ(V_C) - \ell P_C}{EQ(V_0) - DC(V_0)}. \end{aligned}$$

Hence, the dilution costs simplify to

$$\begin{aligned} DC(V_0) =& \frac{\ell P_C}{(EQ(V_0) - DC(V_0))S(V_C)/S(0) + \ell P_C} \mathbb{E}\left[EQ(V_{\tau_C}) e^{-(r+m)\tau_C} \mathbb{1}_{\{\tau_C<\infty\}} \mathbb{1}_{\{\tau_C<\tau\}}\right] \\ =& \ell P_C \mathbb{E}\left[\frac{EQ(V_{\tau_C})}{EQ(V_C)} e^{-(r+m)\tau_C} \mathbb{1}_{\{\tau_C<\infty\}} \mathbb{1}_{\{\tau_C<\tau\}}\right]. \end{aligned} \tag{D.1}$$

By Lemma 5.1 it follows

$$CONV(V_0) = DC(V_0).$$

Note, that the value of the equity after conversion is independent of any features of the contingent convertible debt. In particular, the dilution costs do not appear on the RHS of equation D.1. □

D.4.3 Proofs for Section 5.4

Proof of Proposition 5.8:

Proof.

$$
\begin{aligned}
&\mathbb{E}\left[e^{-\rho\tau+\theta X_\tau}\mathbb{1}_{\{\tau<\infty,-(X_\tau-x)<y\}}\right]\\
=&\mathbb{E}\left[e^{\rho\tau+\theta X_\tau}\mathbb{1}_{\{\tau<\infty,X_\tau=x\}}\right]+e^{\theta x}\mathbb{E}\left[e^{-\rho\tau+\theta(X_\tau-x)}\mathbb{1}_{\{\tau<\infty,0<-(X_\tau-x)<y\}}\right]\\
=&e^{\theta x}\mathbb{E}\left[e^{-\rho\tau}\mathbb{1}_{\{\tau<\infty,X_\tau=x\}}\right]+e^{\theta x}\mathbb{E}\left[e^{\rho\tau}\mathbb{1}_{\{\tau<\infty,0<-(X_\tau-x)<y\}}\right]\frac{\int_{-y}^{0}e^{\theta Y}\eta_2 e^{\eta_2 Y}dY}{\int_{-y}^{0}\eta_2 e^{\eta_2 Y}dY}\\
=&e^{\theta x}\mathbb{E}\left[e^{-\rho\tau}\mathbb{1}_{\{\tau<\infty,X_\tau=x\}}\right]+e^{\theta x}\mathbb{E}\left[e^{-\rho\tau}\mathbb{1}_{\{\tau<\infty,0<-(X_\tau-x)<y\}}\right]\frac{\eta_2}{\theta+\eta_2}\frac{\left(1-e^{-(\theta+\eta_2)y}\right)}{\left(1-e^{-\eta_2 y}\right)}
\end{aligned}
$$

Note that

$$
\mathbb{E}\left[e^{-\rho\tau}\mathbb{1}_{\{\tau<\infty,0<-(X_\tau-x)<y\}}\right]=\mathbb{E}\left[e^{-\rho\tau}\right]-E\left[e^{-\rho\tau}\mathbb{1}_{\{X_\tau=x\}}\right]-\mathbb{E}\left[e^{-\rho\tau}\mathbb{1}_{\{-(X_\tau-x)>y\}}\right]
$$

Using the following results from Kou and Wang (2003) we can finish the proof:

$$
E\left[e^{-\rho\tau}\mathbb{1}_{\{X_\tau=x\}}\right]=\frac{\eta_2-\beta_{3,\rho}}{\beta_{4,\rho}-\beta_{3,\rho}}e^{x\beta_{3,\rho}}+\frac{\beta_{4,\rho}-\eta_2}{\beta_{4,\rho}-\beta_{3,\rho}}e^{x\beta_{4,\rho}} \tag{D.2}
$$

$$
E\left[e^{-\rho\tau}\mathbb{1}_{\{-(X_\tau-x)\geq y\}}\right]=e^{-\eta_2 y}\frac{\eta_2-\beta_{3,\rho}}{\eta_2}\frac{\beta_{4,\rho}-\eta_2}{\beta_{4,\rho}-\beta_{3,\rho}}\left(e^{x\beta_{3,\rho}}-e^{x\beta_{4,\rho}}\right)\qquad y>0 \tag{D.3}
$$

□

Lemma D.5. *The total value of the equity at conversion $EQ(V_{\tau_C})$ satisfies*

$$
EQ(V_{\tau_C})=EQ_{debt}(V_{\tau_C})=\sum_i \alpha_i V_{\tau_C}^{\theta_i}=\sum_i V_0^{\theta_i}\alpha_i e^{X(\tau_C)\theta_i}
$$

with

$$\alpha_1 = 1 \qquad \theta_1 = 1$$
$$\alpha_2 = -\frac{\bar{c}C_D}{r}\frac{\beta_{4,r}}{\eta_2}\frac{\eta_2 - \beta_{3,r}}{\beta_{4,r} - \beta_{3,r}}(V_B)^{\beta_{3,r}} \qquad \theta_2 = -\beta_{3,r}$$
$$\alpha_3 = -\frac{\bar{c}C_D}{r}\frac{\beta_{3,r}}{\eta_2}\frac{\beta_{4,r} - \eta_2}{\beta_{4,r} - \beta_{3,r}}(V_B)^{\beta_{4,r}} \qquad \theta_3 = -\beta_{4,r}$$
$$\alpha_4 = -\alpha V_B\frac{\beta_{4,r} + 1}{\eta_2 + 1}\frac{\eta_2 - \beta_{3,r}}{\beta_{4,r} - \beta_{3,r}}(V_B)^{\beta_{3,r}} \qquad \theta_4 = -\beta_{3,r}$$
$$\alpha_5 = -\alpha V_B\frac{\beta_{3,r} + 1}{\eta_2 + 1}\frac{\beta_{4,r} - \eta_2}{\beta_{4,r} - \beta_{3,r}}(V_B)^{\beta_{4,r}} \qquad \theta_5 = -\beta_{4,r}$$
$$\alpha_6 = \frac{C_D + mP}{r+m}\frac{\beta_{4,r+m}}{\eta_2}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}(V_B)^{\beta_{3,r+m}} \qquad \theta_6 = -\beta_{3,r+m}$$
$$\alpha_7 = \frac{C_D + mP}{r+m}\frac{\beta_{3,r+m}}{\eta_2}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}(V_B)^{\beta_{4,r+m}} \qquad \theta_7 = -\beta_{4,r+m}$$
$$\alpha_8 = -(1-\alpha)V_B\frac{\beta_{4,r+m} + 1}{\eta_2 + 1}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}(V_B)^{\beta_{3,r+m}} \qquad \theta_8 = -\beta_{3,r+m}$$
$$\alpha_9 = -(1-\alpha)V_B\frac{\beta_{3,r+m} + 1}{\eta_2 + 1}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}(V_B)^{\beta_{4,r+m}} \qquad \theta_9 = -\beta_{4,r+m}$$
$$\alpha_{10} = \frac{\bar{c}C_D}{r} - \frac{C_D + mP}{r+m} \qquad \theta_{10} = 0.$$

Proof. The total equity value is the difference between the total value of the firm and the value of actual debt payments:

$$EQ(V_t) = G(V_t) - D(V_t) - CB(V_t) + CONV(V_t)$$

Hence, we conclude that

$$
\begin{aligned}
&EQ(V) = \\
&V + \frac{\bar{c}C_D}{r}\left(1 - \frac{\beta_{4,r}}{\eta_2}\frac{\eta_2 - \beta_{3,r}}{\beta_{4,r} - \beta_{3,r}}\left(\frac{V_B}{V}\right)^{\beta_{3,r}} - \frac{\beta_{3,r}}{\eta_2}\frac{\beta_{4,r} - \eta_2}{\beta_{4,r} - \beta_{3,r}}\left(\frac{V_B}{V}\right)^{\beta_{4,r}}\right) \\
&+ \frac{\bar{c}C_C}{r}\left(1 - \frac{\beta_{4,r}}{\eta_2}\frac{\eta_2 - \beta_{3,r}}{\beta_{4,r} - \beta_{3,r}}\left(\frac{V_C}{V}\right)^{\beta_{3,r}} - \frac{\beta_{3,r}}{\eta_2}\frac{\beta_{4,r} - \eta_2}{\beta_{4,r} - \beta_{3,r}}\left(\frac{V_C}{V}\right)^{\beta_{4,r}}\right) \\
&- \alpha V_B\left(\frac{\beta_{4,r} + 1}{\eta_2 + 1}\frac{\eta_2 - \beta_{3,r}}{\beta_{4,r} - \beta_{3,r}}\left(\frac{V_B}{V}\right)^{\beta_{3,r}} + \frac{\beta_{3,r} + 1}{\eta_2 + 1}\frac{\beta_{4,r} - \eta_2}{\beta_{4,r} - \beta_{3,r}}\left(\frac{V_B}{V}\right)^{\beta_{4,r}}\right) \\
&- \frac{C_D + mP}{r + m}\left(1 - \frac{\beta_{4,r+m}}{\eta_2}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_B}{V}\right)^{\beta_{3,r+m}} - \frac{\beta_{3,r+m}}{\eta_2}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_B}{V}\right)^{\beta_{4,r+m}}\right) \\
&- (1 - \alpha)V_B\left(\frac{\beta_{4,r+m} + 1}{\eta_2 + 1}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_B}{V}\right)^{\beta_{3,r+m}} + \frac{\beta_{3,r+m} + 1}{\eta_2 + 1}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_B}{V}\right)^{\beta_{4,r+m}}\right) \\
&- \frac{C_C + mP_C}{r + m}\left(1 - \frac{\beta_{4,r+m}}{\eta_2}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_C}{V}\right)^{\beta_{3,r+m}} - \frac{\beta_{3,r+m}}{\eta_2}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}\left(\frac{V_C}{V}\right)^{\beta_{4,r+m}}\right)
\end{aligned}
$$

This has a structure of the form

$$EQ(V_t) = \sum_i \alpha_i V_t^{\theta_i}$$

where V_t is the only time dependent variable. Note that at the time of conversion τ_C the value of the tax benefits of the contingent convertible bonds and the value of $CB - CONV$ are zero. Hence, the corresponding terms in $EQ(V_{\tau_C})$ disappear. □

D.4.4 Proofs for Section 5.5

Proof of Theorem 5.3:

Proof. We will first show when V_B^* is optimal. Assume the following five conditions are all satisfied:

1. Assume that first

$$EQ_{old}(V, V_B^*, V_C) \geq 0 \qquad \text{for all } V \geq V_C$$

   i.e. the default barrier V_B^* satisfies the limited liability constraint.

2. Second, the equity of the old shareholders for $V_B > V_C$ is given by

$$EQ_{old}(V, V_B, V_C, P_D, P_C, C_D, C_C) = EQ_{debt}(V, V_B, P_D + P_C, C_D + C_C),$$

   i.e. the equity value of the older shareholders for $V_B > V_C$ is the same as for a firm that issues only straight debt in the amount $P_D + P_C$ with coupon $C_D + C_C$.

3. Third, if the shareholders want to default before conversion the optimal default barrier is V_B^{**} if $V_B^{**} > V_C$ and V_C otherwise.

4. Fourth, it should hold that

$$EQ_{old}(V, V_B^*, V_C) \geq EQ_{old}(V, V_B^{**}, V_B^{**}) \qquad \text{for } V_B^{**} > V_C \text{ and for all } V \geq V_C$$

i.e. the older shareholders do not want to default before conversion for $V_B^{**} > V_C$.

5. Fifth,

$$EQ_{old}(V, V_B^*, V_C) \geq EQ_{old}(V, V_C, V_C) \qquad \text{for all } V > V_C,$$

i.e. the old shareholders prefer the default barrier V_B^* to the default barrier V_C.

The fourth and the fifth condition imply that the shareholders will never choose a default barrier higher than or equal to V_C, because this would result in a lower equity value. By the commitment condition, the optimal default barrier after conversion is V_B^*. This is a valid solution to our optimization problem, if and only if the limited liability constraint is satisfied. The first condition ensures that the limited liability constraint is satisfied for $V \geq V_C$. After conversion V_B^* trivially satisfies the limited liability constraint.

The first and fourth condition are stated as assumptions in our theorem. Now we need to show that the second, third and fifth condition are always satisfied. We start with the second statement. Recall that if the conversion barrier is smaller than the default barrier, we can treat this case as if the default and conversion barrier are the same. Hence, for all $V_B = y \geq V_C$

$$\begin{aligned}
&EQ_{old}(V, y, V_C, P_D, C_D, P_C, C_C) \\
=&EQ_{debt}(V, y, P_D, C_D) + TB_C(V, y) - CCB(V, y, y, P_C, C_C) \\
=&V + \frac{\bar{c}C_D}{r}\mathbb{E}\left[1 - e^{-r\tau}\right] + \alpha E\left[V(\tau)e^{-r\tau}\mathbb{1}_{\{\tau<\infty\}}\right] \\
&-\left(\frac{cP_D + mP_D}{m+r}\mathbb{E}\left[1 - e^{-(m+r)\tau}\right] + (1-\alpha)\mathbb{E}\left[V(\tau)e^{-(m+r)\tau}\mathbb{1}_{\{\tau<\infty\}}\right]\right) \\
&+\frac{\bar{c}C_C}{r}E\left[1 - e^{-r\tau}\right] - \left(\left(\frac{cP_C + mP_C}{m+r}\right)\mathbb{E}\left[1 - e^{-(m+r)\tau}\right] + 0\right) \\
=&EQ_{debt}(V, y, P_D + P_C, C_D + C_C)
\end{aligned}$$

This means that the equity value of the old shareholders is the same as in the case with only straight debt, but with a higher face and coupon value. Chen and Kou (2009) have shown the $EQ_{debt}(V, V_B)$ is strictly decreasing in V_B for $V \geq V_B$ and $V_B \geq V_B^{**}$. Hence,

$$EQ_{debt}(V, V_B^{**}) \geq EQ_{debt}(V, y) \qquad \text{for all } y \geq V_B^{**}, \text{ for all } V \geq V_B^{**}.$$

We have already shown, that in the case of only straight debt V_B^{**} is the optimal default barrier for an amount of debt $P_D + P_C$ and a coupon value of $C_D + C_C$. However, if $V_B^{**} < V_C$, the commitment problem rules out V_B^{**} as a solution. The old shareholders would maximize their equity value by choosing $V_B = V_C$. This proves the third statement. In order to show the fifth statement, we have to reformulate it:

$$\begin{aligned} & EQ_{old}(V, V_B^*, V_C) - EQ_{old}(V, V_C, V_C) \\ =& EQ_{debt}(V, V_B^*) + TB_C(V, V_C) - CCB(V, V_B^*, V_C) - EQ_{debt}(V, V_C) - TB_C(V, V_C) \\ & + CCB(V, V_C, V_C) \\ =& EQ_{debt}(V, V_B^*) - EQ_{debt}(V, V_C) - CONV(V, V_B^*, V_C) \end{aligned}$$

The equality holds as the conversion value for $V_B = V_C$ is equal to zero: $CONV(V, V_C, V_C) = 0$. Assume first, that at conversion the old shareholders are completely diluted out, i.e. all the equity is given to the new shareholders. In this case, the event of conversion is like the default event for the old shareholders and they are indifferent between V_B^* and V_C:

$$EQ_{old}(V, V_B^*, V_C) = EQ_{old}(V, V_C, V_C)$$

which is equivalent to

$$EQ_{debt}(V, V_B^*) - EQ_{debt}(V, V_C) = CONV(V, V_B^*, V_C)$$

i.e. the conversion value equals exactly the gain in the equity value due to a lower default barrier. Obviously, complete dilution gives the highest possible conversion value and hence establishes an upper bound on $CONV(V, V_B^*, V_C)$. Therefore, for an arbitrary amount of shares granted at conversion the following inequality has to hold:

$$EQ_{debt}(V, V_B^*) - EQ_{debt}(V, V_C) \geq CONV(V, V_B^*, V_C)$$

which is equivalent to

$$EQ_{old}(V, V_B^*, V_C) \geq EQ_{old}(V, V_C, V_C)$$

and thus proves the statement.

What happens, if the first condition (limited liability for V_B^*) is violated? The value of the equity will be zero before conversion and hence, default will be triggered for a value of the firm's assets that is larger than V_C. Anticipating this, the old shareholders will choose a default barrier $V_B > V_C$ such that the value of their equity is maximized. As we have shown before, this problem is equivalent to maximizing

$$\begin{aligned} & \max_{V_B \geq V_C} EQ_{debt}(V, V_B, P_D + P_C, C_D, C_C) \\ & \text{s.t. } EQ_{debt}(V', V_B, P_D + P_C, C_D + C_C) > 0 \ \forall \ V' > V_B \end{aligned}$$

A firm with only straight debt, that has face value $P_D + P_C$ and coupons $C_D + C_C$, would ideally choose V_B^{**}. However, if $V_B^{**} < V_C$, the commitment problem does not allow it to take V_B^{**}. As EQ_{debt} is strictly decreasing in the default barrier, the firm would choose the smallest possible default barrier such that $V_B > V_C$, which is obviously V_C. The limited liability constraint is trivially satisfied as EQ_{debt} is strictly increasing in the firm's value. A similar reasoning applies to the case where the fourth condition is violated. □

Proof of Proposition 5.15:

Proof. From Theorem 5.3 we already know, that the optimal default barrier is either V_B^* or $\max\{V_B^{**}, V_C\}$. In the case, where $V_B = V_B^*$, the old shareholders' equity can be at most $EQ_{old}(V, V_B^*, V_C^*)$, as V_C^* is chosen such that is maximizes their equity value. In the other case the old shareholders will choose a conversion and default barrier, such that default happens before conversion, i.e. $V_B \geq V_C$. As we have seen in the proof of Theorem 5.3 in this case $EQ_{old}(V, V_B, V_C, P_D, C_D, P_C, C_C) = EQ_{debt}(V, V_B, P_D + P_C, C_D + C_C)$. The optimal default barrier is then V_B^{**}. In order to ensure, that default happens before conversion the conversion barrier must be smaller than V_B^{**}, i.e $V_C \leq V_B^{**}$. As V_C is a choice variable, the old shareholders can always ensure that this condition holds by setting $V_C = V_B^{**}$. Hence, the old shareholders can get at most $EQ_{debt}(V, V_B^{**}, V_B^{**})$ in this case. Comparing the two maximal values yields the optimal choice. □

Proof of Proposition 5.16:

Proof. For $V \geq V_C \geq V_B$ the total value of the firm is a strictly decreasing function in the conversion barrier:

$$\frac{\partial G(V, V_B, V_C)}{\partial V_C} < 0$$

This is due to the fact that

$$G(V, V_B, V_C) = V + TB_D(V, V_B) + TB_C(V, V_C) - BC(V, V_B)$$

and hence the total value of the firm is only affected by V_C through the tax benefits. It is easy to verify that

$$\frac{\partial TB_C(V, V_C)}{\partial V_C} = \frac{\partial \frac{\bar{c}C_C}{r} E\left[1 - e^{-r\tau_C}\right]}{\partial V_C} < 0$$

for any Markov process V. Therefore, the firm would always choose the lowest possible conversion barrier for a given V_B in the first stage, which is $V_C = V_B$. However, by choosing $V_C = V_B$ the default barrier does not stay fixed, but can change as well. As we have seen in the proof of Theorem 5.3 in this case $EQ_{old}(V, V_B, V_C, P_D, C_D, P_C, C_C) = EQ_{debt}(V, V_B, P_D + P_C, C_D + C_C)$. The optimal default barrier is then V_B^{**} and the resulting total value of the firm is $G(V, V_B^{**}, V_B^{**})$. If the firm chooses the lowest possible value of $\bar{V}_C$, such that V_B^* satisfies the limited liability constraint, then the total value of the firm equals $G(V, V_B^*, \bar{V}_C)$. Comparing the optimal total values of the firm for the two cases yields the optimal solution. □

D.4.5 Proofs for Section 5.6

Proof of Lemma 5.9:

Proof.

$$\begin{aligned}
B(C_D) &\geq D(V, V_B) && \Leftrightarrow \\
\frac{C_D + mP_D}{r+m} &\geq \frac{C_D + mP_D}{r+m}\left(1 - \tilde{A} - \tilde{B}\right) + (1-\alpha)V_B\left(\tilde{C} + \tilde{D}\right) && \Leftrightarrow \\
\frac{C_D + mP_D}{r+m} &\geq (1-\alpha)V_B \frac{\tilde{C} + \tilde{D}}{\tilde{A} + \tilde{B}}
\end{aligned}$$

with

$$\begin{aligned}
\tilde{A} &= \frac{\beta_{4,r+m}}{\eta_2} \frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}} \\
\tilde{B} &= \frac{\beta_{3,r+m}}{\eta_2} \frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}} \\
\tilde{C} &= \frac{\beta_{4,r+m} + 1}{\eta_2 + 1} \frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}} \\
\tilde{D} &= \frac{\beta_{3,r+m} + 1}{\eta_2 + 1} \frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}
\end{aligned}$$

All we need to show is that

$$\frac{\tilde{C} + \tilde{D}}{\tilde{A} + \tilde{B}} < 1$$

$$\begin{aligned}
\frac{\tilde{C} + \tilde{D}}{\tilde{A} + \tilde{B}} &= \frac{\eta_2}{\eta_2 + 1} \frac{(\beta_{4,r+m} + 1)(\eta_2 - \beta_{3,r+m}) + (\beta_{3,r+m} + 1)(\beta_{4,r+m} - \eta_2)}{\beta_{4,r+m}(\eta_2 - \beta_{3,r+m}) + \beta_{3,r+m}(\beta_{4,r+m} - \eta_2)} \\
&= \frac{\eta_2}{\eta_2 + 1}\left(1 + \frac{\eta_2 - \beta_{3,r+m} + \beta_{3,r+m} - \eta_2}{\beta_{4,r+m}(\eta_2 - \beta_{3,r+m}) + \beta_{3,r+m}(\beta_{4,r+m} - \eta_2)}\right) \\
&= \frac{\eta_2}{\eta_2 + 1} \\
&< 1
\end{aligned}$$

□

Proof of Lemma 5.10:

Proof. Define $x = V_B/V$. Obviously, it holds

$$\frac{\partial D(V)}{\partial V} = \frac{\partial D(V)}{\partial x} \frac{\partial x}{\partial V} = \frac{\partial D(V)}{\partial x} \frac{-V_B}{V^2}.$$

Hence, we need to show $\frac{\partial D(V)}{\partial x} < 0$.

We will use Lemma B.1 from the Appendix of Chen and Kou (2009) which states the following: Consider the function $g(x) = Ax^{\alpha_1} + Bx^{\beta_1} - Cx^{\alpha_2} - Dx^{\beta_2}$, $0 \le x \le 1$. In the case of $0 \le \alpha_1 \le \alpha_2 \le \beta_1 \le \beta_2$, $A + B \ge C + D$ and $A > C$, then $g(x) \ge 0$ for all $0 \le x \le 1$.

In our case the debt as a function of x is given by

$$\begin{aligned} D(x) = \frac{C_D + mP_D}{r+m}\Bigg(1 - \frac{\beta_{4,r+m}}{\eta_2}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}x^{\beta_{3,r+m}} - \frac{\beta_{3,r+m}}{\eta_2}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}x^{\beta_{4,r+m}}\Bigg) \\ + (1-\alpha)V_B\Bigg(\frac{\beta_{4,r+m}+1}{\eta_2+1}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}x^{\beta_{3,r+m}} + \frac{\beta_{3,r+m}+1}{\eta_2+1}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}x^{\beta_{4,r+m}}\Bigg) \end{aligned}$$

Therefore,

$$\begin{aligned} D'(x) = &- \frac{C_D + mP_D}{r+m}\Bigg(\frac{\beta_{3,r+m}\beta_{4,r+m}}{\eta_2}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}x^{\beta_{3,r+m}-1} \\ &+ \frac{\beta_{4,r+m}\beta_{3,r+m}}{\eta_2}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}x^{\beta_{4,r+m}-1}\Bigg) \\ &+ (1-\alpha)V_B\Bigg(\frac{\beta_{3,r+m}(\beta_{4,r+m}+1)}{\eta_2+1}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}}x^{\beta_{3,r+m}-1} \\ &+ \frac{\beta_{4,r+m}(\beta_{3,r+m}+1)}{\eta_2+1}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}}x^{\beta_{4,r+m}-1}\Bigg) \\ = &- \left(Ax^{\alpha_1} + Bx^{\beta_1} - Cx^{\alpha_2} - Dx^{\beta_2}\right) \end{aligned}$$

with

$$\begin{aligned} A &= \frac{C_D + mP_D}{r+m}\frac{\beta_{3,r+m}\beta_{4,r+m}}{\eta_2}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}} \\ B &= \frac{C_D + mP_D}{r+m}\frac{\beta_{4,r+m}\beta_{3,r+m}}{\eta_2}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}} \\ C &= (1-\alpha)V_B\frac{\beta_{3,r+m}(\beta_{4,r+m}+1)}{\eta_2+1}\frac{\eta_2 - \beta_{3,r+m}}{\beta_{4,r+m} - \beta_{3,r+m}} \\ D &= (1-\alpha)V_B\frac{\beta_{4,r+m}(\beta_{3,r+m}+1)}{\eta_2+1}\frac{\beta_{4,r+m} - \eta_2}{\beta_{4,r+m} - \beta_{3,r+m}} \end{aligned}$$

and

$$\begin{aligned} \alpha_1 = \alpha_2 &= \beta_{3,r+m} - 1 \\ \beta_1 = \beta_2 &= \beta_{4,r+m} - 1 \end{aligned}$$

From Kou and Wang (2003) we know that $\beta_{4,r+m} > \eta_2 > \beta_{3,r+m} > 0$. Hence $0 \le \alpha_1 \le \alpha_2 \le \beta_1 \le \beta_2$. Next,

$$A + B = \frac{C_D + mP_D}{r+m}\left(\frac{\beta_{3,r+m}\beta_{4,r+m}}{\eta_2}\right)$$

and

$$C+D=(1-\alpha)V_B\left(\frac{\beta_{3,r+m}\beta_{4,r+m}}{\eta_2+1}+\frac{\beta_{3,r+m}(\eta_2-\beta_{3,r+m})+\beta_{4,r+m}(\beta_{4,r+m}-\eta_2)}{(\eta_2+1)(\beta_{4,r+m}-\beta_{3,r+m})}\right)$$

We need to show that $A+B\geq C+D$. Note, that

$$\begin{aligned}
&\frac{A+B}{C+D}\geq 1 && \Leftrightarrow\\
&\frac{\frac{C_D+mP_D}{r+m}}{(1-\alpha)V_B}\geq\frac{\frac{\beta_{3,r+m}\beta_{4,r+m}}{\eta_2+1}+\frac{\beta_{3,r+m}(\eta_2-\beta_{3,r+m})+\beta_{4,r+m}(\beta_{4,r+m}-\eta_2)}{(\eta_2+1)(\beta_{4,r+m}-\beta_{3,r+m})}}{\frac{\beta_{3,r+m}\beta_{4,r+m}}{\eta_2}} && \Leftrightarrow\\
&\frac{\frac{C_D+mP_D}{r+m}}{(1-\alpha)V_B}\geq\frac{\eta_2}{\eta_2+1}+\frac{\eta_2}{\eta_2+1}\left(\frac{\eta_2-\beta_{3,r+m}}{\beta_{4,r+m}(\beta_{4,r+m}-\beta_{3,r+m})}+\frac{\beta_{4,r+m}-\eta_2}{\beta_{3,r+m}(\beta_{4,r+m}-\beta_{3,r+m})}\right) && \Leftrightarrow\\
&\frac{\frac{C_D+mP_D}{r+m}}{(1-\alpha)V_B}\geq\frac{\eta_2}{\eta_2+1}\left(1+\frac{\beta_{3,r+m}(\eta_2-\beta_{3,r+m})+\beta_{4,r+m}(\beta_{4,r+m}-\eta_2)}{\beta_{3,r+m}\beta_{4,r+m}(\beta_{4,r+m}-\beta_{3,r+m})}\right)
\end{aligned}$$

We know that $0<\beta_{3,r+m}<\eta_2<\beta_{4,r+m}$. Hence, it holds that

$$\frac{\beta_{3,r+m}(\eta_2-\beta_{3,r+m})+\beta_{4,r+m}(\beta_{4,r+m}-\eta_2)}{\beta_{3,r+m}\beta_{4,r+m}(\beta_{4,r+m}-\beta_{3,r+m})}<\frac{1}{\beta_{3,r+m}}$$

By assumption we have

$$\frac{\frac{C_D+mP_D}{r+m}}{(1-\alpha)V_B}\geq\frac{\eta_2}{\eta_2+1}\frac{\beta_{3,r+m}+1}{\beta_{3,r+m}}=\frac{\eta_2}{\eta_2+1}\left(1+\frac{1}{\beta_{3,r+m}}\right)$$

Thus, we conclude

$$\begin{aligned}
\frac{\frac{C_D+mP_D}{r+m}}{(1-\alpha)V_B}&\geq\frac{\eta_2}{\eta_2+1}\left(1+\frac{1}{\beta_{3,r+m}}\right)\\
&\geq\frac{\eta_2}{\eta_2+1}\left(1+\frac{\beta_{3,r+m}(\eta_2-\beta_{3,r+m})+\beta_{4,r+m}(\beta_{4,r+m}-\eta_2)}{\beta_{3,r+m}\beta_{4,r+m}(\beta_{4,r+m}-\beta_{3,r+m})}\right).
\end{aligned}$$

Last but not least, we need to show $A>C$.

$$A-C=\frac{C_D+mP_D}{r+m}\frac{\beta_{3,r+m}\beta_{4,r+m}}{\eta_2}\frac{\eta_2-\beta_{3,r+m}}{\beta_{4,r+m}-\beta_{3,r+m}}-(1-\alpha)V_B\frac{\beta_{3,r+m}(\beta_{4,r+m}+1)}{\eta_2+1}\frac{\eta_2-\beta_{3,r+m}}{\beta_{4,r+m}-\beta_{3,r+m}}$$

By assumption $\frac{C_D+mP_D}{r+m}>(1-\alpha)V_B$ as $\frac{\eta_2}{\eta_2+1}\frac{\beta_{3,r+m}+1}{\beta_{3,r+m}}>1$. Hence, it is sufficient to show the following:

$$\begin{aligned}
&\frac{\beta_{3,r+m}\beta_{4,r+m}}{\eta_2}\frac{\eta_2-\beta_{3,r+m}}{\beta_{4,r+m}-\beta_{3,r+m}}-\frac{\beta_{3,r+m}(\beta_{4,r+m}+1)}{\eta_2+1}\frac{\eta_2-\beta_{3,r+m}}{\beta_{4,r+m}-\beta_{3,r+m}}\\
=&\frac{\beta_{3,r+m}\beta_{4,r+m}(\eta_2+1)(\eta_2-\beta_{3,r+m})-\eta_2\beta_{3,r+m}(\beta_{4,r+m}+1)(\eta_2-\beta_{3,r+m})}{\eta_2(\eta_2+1)(\beta_{4,r+m}-\beta_{3,r+m})}\\
=&\frac{\beta_{3,r+m}(\eta_2-\beta_{3,r+m})(\beta_{4,r+m}-\eta_2)}{\eta_2(\eta_2+1)(\beta_{4,r+m}-\beta_{3,r+m})}\\
>&0
\end{aligned}$$

The last step follows from the fact that $0<\beta_{3,r+m}<\eta_2<\beta_{4,r+m}$.

Therefore, all the conditions for Lemma B.1 from Chen and Kou (2009) are satisfied and thus $D'(x)<0$. □

Proof of Lemma 5.11:

Proof. Plugging in the expression for V_B^* yields

$$\frac{C_D+mP_D}{r+m}\geq\frac{\eta_2}{\eta_2+1}\frac{\beta_{3,r+m}+1}{\beta_{3,r+m}}(1-\alpha)V_B \quad\Leftrightarrow$$

$$\frac{C_D+mP_D}{r+m}\geq\frac{\beta_{3,r+m}+1}{\beta_{3,r+m}}(1-\alpha)\frac{\frac{C_D+mP_D}{r+m}\beta_{3,r+m}\beta_{4,r+m}-\frac{\bar{c}C_D}{r}\beta_{3,r}\beta_{4,r}}{\alpha(\beta_{3,r}+1)(\beta_{4,r}+1)+(1-\alpha)(\beta_{3,r+m}+1)(\beta_{4,r+m}+1)}$$

This expression is equivalen to

$$\frac{C_D+mP_D}{r+m}\underbrace{\left(\frac{\beta_{3,r+m}}{\beta_{3,r+m}+1}\alpha(\beta_{3,r}+1)(\beta_{4,r}+1)+(1-\alpha)\beta_{3,r+m}\right)}_{>0}\geq-\frac{\bar{c}C_D}{r}\underbrace{(1-\alpha)\beta_{3,r}\beta_{4,r}}_{>0}$$

As long as $C_D>0$, i.e. the firm has to make positive coupon payments, the above expression will always hold. However, Assumption 5.5 implies that $C_D>0$. □

D.4.6 Proofs for Section 5.8

Proof of Proposition 5.20:

Proof. By lowering the firm's value process to V_{manip} the contingent convertible bondholder enforce conversion and will receive equity with the market value $\min\{\ell P_C, EQ(V_{manip})\}$. The total equity $EQ(V)=EQ_{debt}(V)$ is a continuous and strictly monotonic function for $V\in(V_B,V_C)$. Hence for any given ℓ there exists a V_{manip} such that $\ell P_C>EQ(V_{manip})$. Hence, by manipulating the market the contingent convertible bondholders can always completely dilute out the old shareholders and take control over the firm. We know that before conversion $EQ_{old}(V_t)>0$. This is equivalent to $EQ_{debt}(V_t)+TB_C(V_t)>CCB(V_t)$. If the tax benefits are sufficiently low, the equity value after manipulation $EQ(V_t) = EQ_{debt}(V_t)$, which is then owned only by the contingent convertible bondholders, is larger than the bondholder's value without manipulation $CCB(V_t)$. In this case, manipulation is profitable. □

Proof of Proposition 5.21:

Proof. At time 0, the contingent convertible bonds sell at par:

$$CCB(V_0, V_C) = P_C$$

Hence, the inequality is satisfied

$$\underbrace{TB_C(V_0)}_{\geq 0} + P_C \underbrace{\frac{EQ_{debt}(V_0)}{EQ_{debt}(V_C)}}_{\geq 1} \geq P_C$$

As the conversion value is less than the face value, i.e. $CONV(V_t, V_B, V_C) \leq P_C$ for $V_0 \geq V_t \geq V_C$, it holds that

$$CCB(V_t, V_B, V_C) \leq P_C \qquad \text{for } V_0 \geq V_t \geq V_C.$$

Therefore

$$TB_C(V_t) + P_C \frac{EQ_{debt}(V_t)}{EQ_{debt}(V_C)} \geq CCB(V_t, V_B, V_C).$$

□

D.4.7 Proofs for Section 5.9

Proof of Theorem 5.5:

Proof.

$$\begin{aligned} \mathbb{E}\left[e^{-\tau\rho}\right] &= \mathbb{E}\left[e^{-\tau_C\rho}e^{-(\tau-\tau_C)\rho}\mathbb{1}_{\{\tau>\tau_C\}} + e^{-\tau_C\rho}\mathbb{1}_{\{\tau=\tau_C\}}\right] \\ &= \mathbb{E}\left[e^{-\tau_C\rho}\mathbb{E}\left[e^{-(\tau-\tau_C)\rho}\mathbb{1}_{\{\tau>\tau_C\}}|X_{\tau_C}, \tau_C\right] + e^{-\tau_C\rho}\mathbb{1}_{\{\tau=\tau_C\}}\right] \end{aligned} \tag{D.4}$$

We will first consider the conditional expectation:

$$\mathbb{E}\left[e^{-(\tau-\tau_C)\rho}\mathbb{1}_{\{\tau>\tau_C\}}|X_{\tau_C}, \tau_C\right] = \mathbb{E}\left[e^{-\bar{\tau}\rho}\right]$$

where $\bar{X}_t$ is defined as

$$\bar{X}_t = (r - \delta_2)t + \sigma W_t^* + \sum_{i=1}^{N_t} Y_i$$

and $\bar{\tau} = \inf\big(t \in [0,\infty) : \bar{X}_t \leq \log(V_B/V(\tau_C))\big)$. The above equality is true on account of the Markov property of X_t. Hence,

$$\mathbb{E}\left[e^{-(\tau-\tau_C)\rho}|X_{\tau_C},\tau_C\right] = \bar{c}_1\left(\frac{V_B}{V(\tau_C)}\right)^{\bar{\beta}_{3,\rho}} + \bar{c}_2\left(\frac{V_B}{V(\tau_C)}\right)^{\bar{\beta}_{4,\rho}}$$
$$= \bar{c}_1\left(\frac{V_B}{V_0}\right)^{\bar{\beta}_{3,\rho}} e^{-X(\tau_C)\bar{\beta}_{3,\rho}} + \bar{c}_2\left(\frac{V_B}{V_0}\right)^{\bar{\beta}_{4,\rho}} e^{-X(\tau_C)\bar{\beta}_{4,\rho}}$$

where

$$\bar{c}_1 = \frac{\eta_2 - \bar{\beta}_{3,\rho}}{\eta_2}\frac{\bar{\beta}_{4,\rho}}{\bar{\beta}_{4,\rho}-\bar{\beta}_{3,\rho}}$$
$$\bar{c}_2 = \frac{\bar{\beta}_{4,\rho} - \eta_2}{\eta_2}\frac{\bar{\beta}_{3,\rho}}{\bar{\beta}_{4,\rho}-\bar{\beta}_{3,\rho}}$$

and $-\bar{\beta}_{3,\rho} > -\bar{\beta}_{4,\rho}$ are the two negative roots of the equation

$$\bar{\psi}(\beta) = \rho$$

with $\bar{\psi}$ being the Lévy exponent of $\bar{X}_t$. The first expectation in D.4 equals

$$\mathbb{E}\left[e^{-\tau_C\rho}\left(\bar{c}_1\left(\frac{V_B}{V_0}\right)^{\bar{\beta}_{3,\rho}} e^{-X(\tau_C)\bar{\beta}_{3,\rho}} + \bar{c}_2\left(\frac{V_B}{V_0}\right)^{\bar{\beta}_{4,\rho}} e^{-X(\tau_C)\bar{\beta}_{4,\rho}}\right)\mathbb{1}_{\{\tau>\tau_C\}}\right]$$
$$=\mathbb{E}\left[e^{-\tau_C\rho-\bar{\beta}_{3,\rho}X(\tau_C)}\mathbb{1}_{\{x_C-X(\tau_C)<x_C-x_B\}}\right]\bar{c}_1\left(\frac{V_B}{V_0}\right)^{\bar{\beta}_{3,\rho}}$$
$$+\,\mathbb{E}\left[e^{-\tau_C\rho-\bar{\beta}_{4,\rho}X(\tau_C)}\mathbb{1}_{\{x_C-X(\tau_C)<x_C-x_B\}}\right]\bar{c}_2\left(\frac{V_B}{V_0}\right)^{\bar{\beta}_{4,\rho}}$$

where $x_C = \log(V_C/V_0)$ and $x_B = \log(V_B/V_0)$. The condition $\mathbb{1}_{\{x_C-X(\tau_C)<x_C-x_B\}}$ ensures that the downward jumps are not large enough to trigger conversion and bankruptcy. Now we can apply Proposition 5.8 to the two expectations. The second expectation in equation D.4 equals

$$\mathbb{E}\left[e^{-\tau_C\rho}\mathbb{1}_{\{\tau=\tau_C\}}\right] = \mathbb{E}\left[e^{\tau_C\rho}\mathbb{1}_{\{x_C-X(\tau_C)>x_C-x_B\}}\right]$$
$$= e^{-\eta_2(x_C-x_B)}\frac{\eta_2-\beta_{3,\rho}}{\eta_2}\frac{\beta_{4,\rho}-\eta_2}{\beta_{4,\rho}-\beta_{3,\rho}}\left(e^{x_C\beta_{3,\rho}} - e^{x_C\beta_{4,\rho}}\right)$$
$$= \left(\frac{V_B}{V_C}\right)^{\eta_2}\frac{\eta_2-\beta_{3,\rho}}{\eta_2}\frac{\beta_{4,\rho}-\eta_2}{\beta_{4,\rho}-\beta_{3,\rho}}\left(\left(\frac{V_C}{V_0}\right)^{\beta_{3,\rho}} - \left(\frac{V_C}{V_0}\right)^{\beta_{4,\rho}}\right)$$

where we have applied equation D.3 in the second line. □

Proof of Theorem 5.6:

Proof.

$$
\begin{aligned}
&\mathbb{E}\left[e^{-\tau\rho+\theta X_\tau}\mathbb{1}_{\{\tau<\infty\}}\right] \\
=&\mathbb{E}\left[e^{-\rho\tau_C+\theta X_{\tau_C}}\mathbb{E}\left[e^{-\rho(\tau-\tau_C)+\theta(X_\tau-X_{\tau_C})}|X(\tau_C),\tau_C\right]\mathbb{1}_{\{\tau>\tau_C\}}+e^{-\rho\tau_C+\theta X(\tau_C)}\mathbb{1}_{\{\tau=\tau_C\}}\right] \\
=&\mathbb{E}\left[e^{-\rho\tau_C+\theta X_{\tau_C}}\left(\bar{d}_1\left(\frac{V_B}{V(\tau_C)}\right)^{\theta+\bar{\beta}_{3,\rho}}+\bar{d}_2\left(\frac{V_B}{V(\tau_C)}\right)^{\theta+\bar{\beta}_{4,\rho}}\right)\mathbb{1}_{\{\tau<\infty,\tau>\tau_C\}}+e^{-\rho\tau_C+\theta X(\tau_C)}\mathbb{1}_{\{\tau<\infty,\tau=\tau_C\}}\right] \\
=&\mathbb{E}\left[\left(e^{-\rho\tau_C-\bar{\beta}_{3,\rho}X(\tau_C)}\bar{d}_1\left(\frac{V_B}{V_0}\right)^{-\theta-\bar{\beta}_{3,\rho}}+e^{-\rho\tau_C-\bar{\beta}_{4,\rho}X(\tau_C)}\bar{d}_2\left(\frac{V_B}{V_0}\right)^{-\theta-\bar{\beta}_{4,\rho}}\right)\mathbb{1}_{\{\tau<\infty,\tau>\tau_C\}}\right] \\
&+\mathbb{E}\left[e^{-\rho\tau_C+\theta X(\tau_C)}\mathbb{1}_{\{\tau<\infty,\tau=\tau_C\}}\right]
\end{aligned}
$$

The first expectation can be calculated using Proposition 5.8. The second expectation equals

$$
\begin{aligned}
\mathbb{E}\left[e^{-\rho\tau_C+\theta X(\tau_C)}\mathbb{1}_{\{\tau_C<\infty,\tau=\tau_C\}}\right]&=\mathbb{E}\left[e^{-\rho\tau_C+\theta X(\tau_C)}\mathbb{1}_{\{\tau_C<\infty,-(X(\tau_C)-x_C)\geq x_C-x_B\}}\right] \\
&=\mathbb{E}\left[e^{-\rho\tau_C+\theta X(\tau_C)}\mathbb{1}_{\{\tau_C<\infty\}}\right]-\mathbb{E}\left[e^{-\rho\tau_C+\theta X(\tau_C)}\mathbb{1}_{\{-(X(\tau_C)-x_C)<x_C-x_B\}}\right]
\end{aligned}
$$

which can also be calculated using Proposition 5.8. □

D.4.8 Proofs for Section 5.10

Proof of Proposition 5.24:

Proof. If $V_B^*(c^*) \geq V_C$, then $V_B^*(c^*)$ is feasible. Hence, the firm will choose a combination of P_D and P_C which leads to $c^*(P_D+P_C) = P_D c_D + P_C c_C$. However, because of the commitment problem, $V_B = V_B^*(c^*) < V_C$ is not a feasible solution. By the strict concavity of $G_{debt}(\tilde{P})$ we conclude, that the optimal default barrier will be $V_B = V_C$. Hence, the firm will choose $\{P_D, P_C\}$ such that $V_B^{**} = V_C$. As long as $V_B = V_C$ the total value of the firm is a strictly increasing function in P_D and P_C. The firm solves the following problem:

$$\max_{P_D,P_C} V + TB_D(V_C) + TB_C(V_C) - BC(V_C)$$

Therefore, the firm choses the highest values for P_D and P_C such that $V_B^{**}(P_D, P_C) = V_C$. As $c_D < c_C$ the marginal increase in tax benefits for contingent convertible debt is higher than for straight debt. Hence, the optimal debt choice $\{P_D, P_C\}$ is the highest amount of P_C such that $V_B^{**}(P_D, P_C) = V_C$ under the constraint that Assumption 5.3 is not satisfied. □

Proof of Lemma 5.15:

Proof. The total value of the firm for debt P_D and no CCBs is $G_{debt}(P_D)$. By assumption

$$
\begin{aligned}
G_{debt}(P_D^*) &= V + TB_D(P_D^*) - BC(P_D^*) \\
&= G_{debt}(\rho_i) + TB_C(\phi_i^C)
\end{aligned}
$$

which is equivalent to

$$BC(P_D^*) - BC(\rho_i) = TB_D(P_D^*) - TB_D(\rho_i) - TB_C(\phi_i^C).$$

As $P_D^* \geq \rho_i$, and the default barrier is strictly increasing in the amount of debt, it holds that $BC(P_D^*) > BC(\rho_i)$, which yields

$$TB_D(P_D^*) > TB_D(\rho_i) + TB_C(\phi_i^C).$$

□

Proof of Lemma 5.16:

Proof. The total leverage with the CCB regulation scheme is

$$TL_1 = \frac{\rho_i + \phi_i^C}{G_D(\rho_i) + TB_C(\phi_i^C)}.$$

The total leverage for the regulation without contingent convertible bonds equals

$$TL_2 = \frac{\rho_i}{G_D(\rho_i)}.$$

Obviously, we have $\frac{\rho_i}{G_D(\rho_i)} < 1$. By Assumption 5.7 it holds $\frac{\phi_i^C}{TB_C(\phi_i^C)} > 1$. Hence, we conclude $\frac{\phi_i^C}{TB_C(\phi_i^C)} > \frac{\rho_i}{G_D(\rho_i)}$. This is equivalent to $TL_1 > TL_2$ as the following chain of equivalent statements shows.

$$\begin{aligned}
\frac{\rho_i + \phi_i^C}{G_D(\rho_i) + TB_C(\phi_i^C)} &> \frac{\rho_i}{G_D(\rho_i)} && \Leftrightarrow \\
\frac{\rho_i}{G_D(\rho_i)} \frac{G_D(\rho_i)}{G_D(\rho_i) + TB_C(\phi_i^C)} + \frac{\phi_i^C}{G_D(\rho_i) + TB_C(\phi_i^C)} &> \frac{\rho_i}{G_D(\rho_i)} && \Leftrightarrow \\
\frac{\rho_i}{G_D(\rho_i)} \left(\frac{G_D(\rho_i)}{G_D(\rho_i) + TB_C(\phi_i^C)} - 1 \right) + \frac{\phi_i^C}{G_D(\rho_i) + TB_C(\phi_i^C)} &> 0 && \Leftrightarrow \\
\frac{\phi_i^C}{G_D(\rho_i) + TB_C(\phi_i^C)} \frac{G_D(\rho_i) + TB_C(\phi_i^C)}{TB_C(\phi_i^C)} &> \frac{\rho_i}{G_D(\rho_i)} && \Leftrightarrow \\
\frac{\phi_i^C}{TB_C(\phi_i^C)} &> \frac{\rho_i}{G_D(\rho_i)}.
\end{aligned}$$

□