

Understanding Financial Market Volatility

Anne Opschoor



Understanding Financial Market Volatility

ISBN: 978 90 361 0379 4

© Anne Opschoor, 2013

All rights reserved. Save exceptions stated by the law, no part of this publication may be reproduced, stored in a retrieval system of any nature, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, included a complete or partial transcription, without the prior written permission of the author, application for which should be addressed to the author.

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 574 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Understanding Financial Market Volatility

Een beter begrip van volatiliteit op financiële markten

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 20 februari 2014 om 13:30 uur

door

Anne Opschoor
geboren te Krimpen aan den IJssel



Promotiecommissie

Promotor: Prof.dr. D.J.C. van Dijk

Overige leden: Prof.dr. H.P. Boswijk
Prof.dr. R. Paap
Dr. J. Rombouts

Copromotor: Dr. M. van der Wel

Acknowledgements

Although you can read my name on the cover of this booklet, it is a serious misunderstanding to think that this thesis is fruit from my own field. Several people deserve my most sincere thanks for their direct contribution to this thesis and for their support during the last four years. To summarize these years, I imagine the evolution of a sinusoidal function through time with a time-varying amplitude, combined with a positive drift. That is, I experienced ups and downs of various size, but in the end I can say that I am positive about pursuing a PhD.

First and foremost, this is due to my supervisor Dick van Dijk, who let me think about becoming a PhD student. Dick, many times I felt you were the right person at the right time in the right place. I appreciated the moments I entered your office and had a quick and efficient conversation about our methodology, results or how to continue the research project. In addition, I would like to thank Michel van der Wel for being my co-promotor. Michel, you supported me a lot with writing a paper, as well as giving me the freedom to think individually about solutions of problems in the research process. Besides, I thank both Dick and Michel for their enthusiasm and cooperation, and for giving me the opportunity to visit conferences at London, Berlin, Stockholm, Stavanger, Oxford, Vancouver and Gothenburg.

I also thank Lennart Hoogerheide and Herman van Dijk for being co-authors of the paper that forms the basis of Chapter 2 of this thesis. Lennart, thanks for all (hilarious) chats and your support during the course Financial Case Studies and during the time I wrote my Master's thesis. Both courses gave me a good first insight in doing research. Thanks in addition for our cooperation and I look forward to meet you again in Amsterdam! Herman, many thanks for taking place in the larger committee.

Further, I am indebted to Jeroen Rombouts, Peter Boswijk and Richard Paap for taking part in my doctoral committee, for reading my thesis thoroughly and for their helpful comments. Richard, I also thank you for our nice conversations about classical music when I passed by your office. I would like to thank Siem Jan Koopman for taking place in the larger committee.

Beyond doing research, I also enjoyed the cooperation with Philip Hans Franses and Dick van Dijk when we worked on the book about time series econometrics. Philip Hans, I appreciated our meetings and your great enthusiasm.

The Econometric Institute and the Tinbergen Institute have provided a very nice work environment. I thank all staff members for their friendly support. In particular, I would like to thank Judith and Caroline for all nice chats, which were a pleasant distraction from my daily work. Further, I enjoyed the funny moments with Marjon at the 11th floor. Ursula, thanks for the Friday afternoon chats, I will miss them! Further, it was always nice to walk around at the 11th floor, which led to conversations with Christiaan Heij, Alex Koning, Erik Kole or Bart Diris. Emöke, I enjoyed assisting you in teaching the course Basic Mathematics.

I was lucky to have a couple of sociable colleagues around me which added value to my daily working life as a PhD student. H9-18 was one of my favorite rooms to enter around 4 or 5 p.m., with the desk lamp turned on instead of the ugly fluorescent lamp. Sjoerd, thanks for all good moments and (sometimes hilarious) conversations about various topics. You have a special kind of humor which made me laugh many times. Jorn, thanks for all chats we had and the advices you gave. Rianne, I met you as a first year student, being my tutor. It was nice to meet you again as a PhD student and to share a room with you for more than a year. I had a great time with you and I apologize for being a noisy roommate, however the good part is that this is also a result of the fact that I appreciated your presence very much. Jeanine, thanks for being my next roommate, it was nice to share an office with you and to experience your enthusiasm. I will not forget the very strong coffee you (also) made! In addition, I would like to thank Twan for dropping by at my office and for the social drinks at the Smitse. Martin, I appreciated the coffee breaks which made me laugh if you talked about your updated ideas about your career, work, house or gardening. I also enjoyed the moments dropping by at the office of Aart

and the coffee breaks with Mehtap. Zara, you were a cheerful neighbor at the 8th floor during the last year. I appreciated our smoking sessions! Dana, thanks for the interesting chats during social drinks, I like your enthusiasm (also about classical music). Nalan, it was always nice to enter your office for a short talk or to discuss things about research. For the sake of space, I would finally thank my colleagues Barbara, Bart, Bruno, Eran, Evsen, Francine, Karolina, Niels, Oke, Sanne, Saeedeh, Saskia, Tim, Tom, Thijs, Victor, Willem and Wim for all conversations we had.

Furthermore, I am grateful to my paranympths, Suzanne and Dennis. Suzanne, you are a sociable and enthusiastic woman. Many thanks for all your support and the coffee breaks which closed each Friday at the university in a good way. Dennis, we had a lot of nice moments during the last year, both when we discussed our research, or just when we had a chat.

I am not only blessed with great colleagues, but also with a great family and friends. Bas, thanks for taking me out of my working area and just for being a friend. Francis, I always enjoyed our talks after you played with the RPHO. Jan, Chris and Caroline, Gerard, Hilda and Bert, you all deserve gratitude for your love and interest. Mom and Dad, thanks for being the parents all one could wish for, for always supporting my decisions in life, for always being available whenever I need a listening ear, for your love. Finally, I should be indebted to the Lord, who always gave me the strength and a good health such that I was able to do my work.

Anne Opschoor

Rotterdam, November 2013

Contents

1	Introduction and outline	1
1.1	Introduction	1
1.2	Outline	4
2	MitISEM: An adaptive sampling method for posterior and predictive simulation	7
2.1	Introduction	7
2.2	Related literature	9
2.3	MitISEM	11
2.3.1	Application: Bayesian analysis of the DCC-GARCH model . . .	20
2.4	Sequential MitISEM	24
2.4.1	Tempered MitISEM	29
2.5	Conclusion	34
2.A	Appendix	34
2.A.1	Derivation of the IS-weighted EM algorithm for mixtures of Student- t distributions	34
2.A.2	Two alternative adaptive simulation methods	39
3	On the effects of private information on volatility	41
3.1	Introduction	41
3.2	Related literature	45
3.3	Data and summary statistics	47
3.3.1	U.S. Treasury bond futures data	48
3.3.2	Private information variables	49

3.3.3	Public information variables	51
3.3.4	Public and private information effects: A first impression	55
3.4	Modeling the response of Treasury futures to public and private information	55
3.4.1	Public and private information matter for returns	56
3.4.2	Public and private information also matter for volatility	60
3.5	Robustness checks and extensions	68
3.5.1	Controlling for liquidity effects	69
3.5.2	An additional proxy for private information	71
3.5.3	Conditioning on the state of the economy	73
3.5.4	Conditioning on the dispersion of beliefs	77
3.6	Conclusion	80
4	Predicting covariance matrices with financial conditions indexes	81
4.1	Introduction	81
4.2	The modeling framework	84
4.2.1	The Spline-GARCH model for volatility with DCC for correlation	85
4.2.2	The Factor-Spline-GARCH model	88
4.2.3	Estimation	90
4.3	Data	92
4.4	In-sample results	97
4.5	Estimating portfolio Value-at-Risk	105
4.6	The role of the VIX in the FCI	110
4.7	Conclusion	113
4.A	Appendix	113
4.A.1	VaR backtests	113
4.A.2	In-sample volatilities of all assets	116
5	Improving density forecasts and Value-at-Risk estimates by combining den-	119
	sities	
5.1	Introduction	119
5.2	Combining density forecasts	122
5.3	Models and distributions	125

5.3.1	Univariate volatility models	125
5.3.2	Conditional distributions	128
5.4	Application	130
5.4.1	Data and implementation details	130
5.4.2	Evaluation	132
5.4.3	Results	135
5.5	Conclusion	149
5.A	Appendix	150
5.A.1	Optimizing weights	150
5.A.2	Pooling results of the log score function	152
	Nederlandse samenvatting (Summary in Dutch)	155
	Bibliography	159

Chapter 1

Introduction and outline

“... what distinguishes financial economics is the central role that uncertainty plays in both financial theory and its empirical implementation... Indeed in the absence of uncertainty, the problems of financial economics reduce to exercises in basic microeconomics” (p.3).

Campbell, Lo and MacKinlay, 1997

1.1 Introduction

Volatility has been one of the most active and successful areas of research in time series econometrics and economic forecasting in recent decades. Loosely speaking, volatility is defined as the average magnitude of fluctuations observed in some phenomenon over time. Within the area of economics, this definition narrows to the variability of an unpredictable random component of a time series variable. Typical examples in finance are returns on assets, such as individual stocks or a stock index like the S&P 500 index. As indicated by the quote from Campbell *et al.* (1997), (financial market) volatility is central to financial economics. Since it is the most common measure of the risk involved in investments in traded securities, it plays a crucial role in portfolio management, risk management, and pricing of derivative securities including options and futures contracts. Volatility is therefore closely tracked by private investors, institutional investors like pension funds,

central bankers and policy makers. For example, the so-called Basel accords contain regulations where banks are required to hold a certain amount of capital to cover the risks involved in their consumer loans, mortgages and other assets. An estimate of the volatility of these assets is a crucial input for determining these capital requirements. In addition, the financial crisis in 2007-2008 has proven that the impact of financial market volatility is not only limited to the financial industry. It shows that volatility may be costly for the economy as a whole. For example, extreme stock market volatility may negatively influence aggregate investments behavior, in particular as companies often require equity as a source of external financing.

This thesis contributes to the volatility literature by investigating several relevant aspects of volatility. First, we focus on the *parameter estimation* of multivariate volatility models, which is problematic if the number of considered assets increases. Second, we consider the question what exactly *causes* financial market volatility? In this context, we relate volatility with various types of information. In addition, we pay attention to *modeling* volatility, by adapting volatility models such that they allow for including possible exogenous variables. Finally, we turn to *forecasting* techniques of volatility, with the focus on the combination of density forecasts.

The literature on modeling volatility dynamics dates back to Engle (1982) and Bollerslev (1986), who develop the (G)ARCH volatility model. This model takes into account two important aspects of asset return volatilities: they are time-varying and persistent. This so-called phenomenon of “volatility clustering” - periods of high volatility and low volatility tend to alternate - was originally put forward by Mandelbrot (1963). As pointed out by Andersen *et al.* (2006), this feature is observed across assets, asset classes, time periods, and countries. The GARCH model is still popular after 20 years since it provides a simple intuitive way to incorporate this volatility clustering. When applied to daily returns, it relates today’s conditional volatility linearly to the squared return of yesterday. Hence a large shock in yesterday’s return boosts the volatility today and vice versa. The GARCH model (and many extensions which are proposed in the literature) is easy to estimate and can be seen as the benchmark volatility model in the field of financial econometrics.

Given this popularity, it is not surprising that the idea of univariate volatility clustering is extended to the multivariate case (i.e. to describe volatilities and correlations). Indeed, various multivariate extensions to model also comovements are proposed in the literature. These range from modeling the full covariance matrix in one step, like the BEKK model (Engle and Kroner, 1995), to modeling the correlations and volatilities separately in order to keep the model parsimonious, such as the DCC model of Engle (2002). The last model circumvents the *curse of dimensionality*; which refers to the problem that the number of parameters to be estimated grows rapidly if the number of assets becomes large. In general, estimating parameters of Multivariate GARCH models remains a difficult task, especially using a frequentist approach. Chapter 2 of this thesis addresses this issue therefore from a Bayesian viewpoint. We contribute to the literature by proposing a new algorithm that improves the Bayesian estimation of the parameters of both univariate mixture GARCH models as well as multivariate GARCH models.

A next important theme beyond the modeling aspect of financial market volatility is to understand its determinants. Put differently, what exactly *causes* volatility? Despite the widely recognized importance of volatility, we do not yet fully understand this. In general, volatility changes due to the arrival of new information. But what type of information? Chapter 3 differentiates in this context between various types of information, such as public information (e.g. macro announcements of inflation and employment or interest rate decisions of the central bank) and private information. The literature (e.g. Andersen, Bollerslev, Diebold and Vega, 2003; Andersen *et al.*, 2007) shows that the former type indeed affects volatility. The second type of information refers to news that is distributed asymmetrically among market participants, and may be revealed by the trading process and changes in the asset price itself, since informed investors may buy or sell the asset based on their privately held knowledge. By documenting a relation between financial market volatility and private information, we contribute to the literature by connecting the volatility modeling literature with the literature on financial market microstructure, which looks in detail at how assets are traded and the associated costs, informational effects and the behavior of asset prices from trade to trade.

A closely related subject to various types of information that may affect financial market volatility are financial conditions of a certain country. The recent 2007-2008 crisis

clearly pointed out that financial conditions are an important driver of the economy at large. This crisis boosted the interest in these conditions, also because they affect monetary policy and macroeconomic variables, as shown by the literature (Hatzius *et al.*, 2010). Financial conditions, defined in this literature as the current state of financial variables that influence economic behavior and the future state of the economy, typically include equity returns and volatility, the shape of the yield curve and measures of credit risk. In Chapter 4, we relate these conditions to the volatility and comovement of asset returns.

The final major theme considered in this thesis is *forecasting* volatility. As noted earlier in the example earlier in this introduction, a forecast of volatility is crucial to determine the capital requirements of financial companies. Due to the relatively high persistence of volatility compared to returns, it is quite predictable. But which model should a decision-maker use to forecast volatility? Academic literature as well as expertise of companies in the financial industry provide a huge set of volatility models. In addition, the decision-maker faces *model uncertainty*, as he does not know which model is correct. Moreover, each model is in a certain way an incomplete description of reality. Chapter 5 answers this question in the context of combining density forecasts of different recently developed univariate volatility models, with the focus on the left tail of the predictive density. Although the idea of combining *point* forecasts is well established in the literature (see Bates and Granger, 1969), combining *density* forecasts has attracted interest only recently and hence the literature about this topic is scarce. In addition, Chapter 5 contributes to the literature by focusing on the left tail, motivated by the fact that using the whole density may result in suboptimal forecasts if one is only interested in a particular region of the density.

1.2 Outline

The chapters of this thesis are self-contained and can thus be read independently. All chapters deal with financial market volatility. The first and the last chapter focus on estimation and forecasting techniques in the context of multivariate and univariate volatility models respectively. The second and third chapter relate volatility with possible economic/financial determinants.

Chapter 2 is based on Hoogerheide *et al.* (2012) and introduces an adaptive sampling method for efficient posterior and predictive simulation. The method makes use of sequences of importance weighted Expectation Maximization steps in order to efficiently construct a mixture of Student- t densities that approximates accurately the target distribution. This is typically a posterior distribution, of which we only require a kernel in the sense that the Kullback-Leibler divergence between target and mixture is minimized. We label this approach Mixture of t by Importance Sampling weighted Expectation Maximization (MitISEM). The constructed mixture is used as a candidate density for quick and reliable application of either Importance Sampling (IS) or the Metropolis - Hastings (MH) method. We also introduce an extension of the basic MitISEM approach by proposing a method for applying MitISEM in a sequential manner, so that the candidate distribution for posterior simulation is cleverly updated when new data become available. Our results indicate that the proposed methods can substantially reduce the computational burden in Multivariate GARCH models (i.e. the DCC model) or mixture GARCH models.

In Chapter 3, we study the impact of private information on volatility. We develop a comprehensive framework to investigate this link while controlling for the effects of both public information (such as macroeconomic news releases) and private information on prices and the effect of public information on volatility. Using high-frequency 30-year U.S. Treasury bond futures data, we find that private information, measured by order flow, is statistically and economically significant for explaining volatility. This result is robust to controlling for liquidity, using market sidedness as alternative proxy for private information, and using data at a daily frequency.

The relationship between financial conditions and asset market volatility and correlation is put forward in Chapter 4. We propose extensions of (factor-)GARCH models for volatility and DCC models for correlation that allow for including indexes that measure financial conditions. In our empirical application we consider daily stock returns of US deposit banks during the period 1994-2011. We proxy financial conditions by the Bloomberg Financial Conditions Index (FCI) which comprises the money, bond, and equity markets. We find that worse financial conditions are associated with both higher volatility and higher average correlations between stock returns. Especially during crises the additional impact of the FCI indicator is considerable, with an increase in correlations

by 0.15. Moreover, including the FCI in volatility and correlation modeling improves Value-at-Risk forecasts, particularly at short horizons.

Chapter 5 investigates the added value of combining density forecasts for asset return prediction in a specific region of interest. We develop a new technique that takes into account model uncertainty by assigning weights to individual predictive densities using a scoring rule based on the censored likelihood. We apply this approach in the context of recently developed univariate volatility models (including HEAVY and Realized GARCH models), using daily returns from the S&P 500, DJIA, FTSE and Nikkei stock market indexes from 2000 until 2013. The results show that combined density forecasts based on the censored likelihood scoring rule significantly outperform pooling based on equal weights or the log scoring rule and individual density forecasts. In addition, using our technique improves VaR estimates at short horizons.

Chapter 2

MitISEM: An adaptive sampling method for posterior and predictive simulation

Based on Hoogerheide et al. (2012).

2.1 Introduction

Since a few decades there is considerable interest in Bayesian analysis using computer generated pseudo random draws from the posterior and predictive distribution. Markov Chain Monte Carlo (MCMC) techniques are useful for this purpose and a popular MCMC technique is the Metropolis-Hastings algorithm, developed by Metropolis *et al.* (1953) and generalized by Hastings (1970). Several updates of this sampler are proposed in the literature, especially the idea of adapting the proposal distribution given sampled draws.

Monte Carlo procedures based on Importance Sampling (IS), see Hammersley and Handscomb (1964), are an alternative. This idea has been introduced in Bayesian inference by Kloek and Van Dijk (1978) and is further developed by Van Dijk and Kloek (1980, 1984) and, in particular, by Geweke (1989). Cappé *et al.* (2008) discuss that there exists renewed interest in Importance Sampling. This is due to its relatively simple properties which allow for the development of parallel implementation. The increased popularity of

Importance Sampling goes jointly with the development of multiple core machines and computer clusters.

In this chapter we specify a class of adaptive sampling methods for efficient and reliable posterior and predictive simulation. The proposed methods are robust in the sense that they can handle target distributions that exhibit non-elliptical shapes such as multimodality and skewness. These methods are especially useful for posteriors where the convergence of alternative simulation methods is slow or even doubtful, such as high serial correlation in Gibbs sequences that may be caused by large numbers of latent variables or non-elliptical shapes. Importance Sampling and Gibbs sampling are not necessarily substitutes: given that diagnostic checks can never fully guarantee that results have converged to the true values (that is, that convergence has been reached and that no errors have been made in the derivations and code), the use of both simulation methods that have completely different theory and implementation can be a useful validity check. Further, an appropriate candidate distribution can be used to draw initial values for multiple Gibbs sequences, whereas a sample of Gibbs draws can be used to obtain initial values for the mean and covariance matrix in the process of constructing an approximating candidate distribution. Our proposed methods make use of the novel *Mixture of t by Importance Sampling weighted Expectation Maximization* (MitISEM) approach. This approach uses sequences of importance weighted steps in an Expectation Maximization algorithm in order to relatively quickly construct a mixture of Student- t densities, which is used as an efficient and reliable candidate density for Importance Sampling (IS) or the Metropolis-Hastings (MH) method. Next to assessing possibly non-elliptical posterior distributions, MitISEM is particularly useful for accurately estimating marginal and predictive likelihoods via IS.

We introduce also an extension beyond specifying the basic approach of MitISEM. More specifically, we propose a method for applying MitISEM in a *sequential* manner, so that the candidate distribution for posterior simulation is cleverly updated when new data become available. Our results show that the computational effort reduces enormously, while the quality of the approximation remains almost unchanged, as compared with an ‘ad hoc’ procedure in which the construction of the MitISEM candidate is performed ‘from scratch’ at every moment in time. This sequential approach can be combined with a

tempering approach, which facilitates the simulation from densities with multiple modes that are far apart. The proposed tempering method moves sequentially from a tempered target density kernel, the target density kernel to the power of a positive number that is smaller than 1, towards the real target density kernel. The tempered target distribution is more diffuse and hence the probability of detecting far-away modes is higher. The idea of tempering was introduced by Geyer (1991), see also Hukushima and Nemoto (1996).

The outline of this chapter is as follows. We first review the literature in section 2.2. In section 2.3 we introduce the MitISEM method, and we show applications in a multivariate GARCH model with 17 parameters. Section 2.4 introduces the Sequential MitISEM method, and includes a subsection on Tempered MitISEM. Section 2.5 concludes. The appendix provides the derivation of the IS-weighted EM method, and discusses the alternative simulation methods of Roberts and Rosenthal (2009) and Giordani and Kohn (2010).

2.2 Related literature

Several approaches of adaptive sampling using mixtures exist in the literature. Keith *et al.* (2008) developed adaptive independence samplers by minimizing the Kullback-Leibler (KL) divergence in order to provide the best candidate density, which consists of a mixture of Gaussian densities. The minimization of the KL-divergence is done by applying the EM algorithm of Dempster *et al.* (1977) and the number of mixture components is selected through information criteria like AIC (Akaike, 1974), BIC (Schwarz, 1978) or DIC (Gelman *et al.*, 2003). Our basic approach is a ‘bottom up’ procedure that starts with one Student-*t* distribution (instead of a Gaussian distribution) and Student-*t* components are added iteratively until a certain stop criterion is met. We emphasize that the IS-weighted version of the EM algorithm is applied in order to use all candidate draws without requiring the Metropolis-Hastings algorithm to transform the candidate draws into a set of posterior draws. Cappé *et al.* (2008) and Cornuet *et al.* (2012) also use IS-weights in the EM algorithm with a mixture of Student-*t* densities as candidate density. Cappé *et al.* (2008) developed the M-PMC (Mixture Population Monte Carlo) algorithm, which is an adaptive algorithm that iteratively updates both the weights and component parameters

of a mixture importance sampling density. An important difference between Cappé *et al.* (2008) (and also Cornuet *et al.*, 2012) and the present chapter is the choice of the number of mixture components and the starting values of the candidate mixture's Student- t components' means and covariances in the EM optimization procedure. Regarding the first issue, in earlier papers the number of mixture components is chosen a priori, where we let the algorithm choose the required number of components. Second, we choose the starting values based on the draws that correspond to the highest IS-weights for the previous mixture of Student- t candidate in the algorithm, where Cappé *et al.* (2008) do not provide a strategy for choosing starting values. Although the EM procedure is guaranteed to converge to a *local* optimum, the choice of the starting values may still be crucial, given that the KL divergence between target and candidate (as a function of the candidate mixture's means, covariances, degrees of freedom and component weights) is a highly non-elliptical, multimodal function. Moreover, we provide extensions (sequential, tempered) that facilitate simulation for specific applications and for particular statistical and econometric models.

A different strand of literature is the use of adaptive MCMC algorithms where the parameters of the candidate density are automatically tuned during the sampling procedure. Learning about candidate density parameter values leading to more efficient sampling while maintaining the ergodicity property for asymptotic convergence is, of course, important. Roberts and Rosenthal (2009) consider an adaptive random walk Metropolis sampler and Giordani and Kohn (2010) use a mixture of densities in their adaptive independent MH sampler. We differ from these authors by using a two-stage approach. Using the Kullback-Leibler distance function, we fit during a first stage of *preliminary adaptation* a flexible candidate to the target with the IS-weighted EM algorithm. In the second stage we insert the obtained candidate in a 'standard', non-adaptive IS or MH algorithm. So, in terms of Giordani and Kohn (2010), we do not perform *strict adaptation*; our second phase of non-adaptive IS or MH ensures that the simulation output converges to the correct distribution. In section 2.3.1, we compare the efficiency of our approach with those from Roberts and Rosenthal (2009) and Giordani and Kohn (2010) in the context of a DCC-GARCH model with 17 parameters. The results indicate that our approach compares favorably with these alternative adaptive MCMC schemes, but we emphasize

that a systematic study of the relevant merits of alternative sampling schemes for a variety of target density shapes is a topic of great interest, which is however beyond the scope of the present study.

A final remark considering the literature regards the Adaptive Mixture of t (AdMit) approach of Hoogerheide *et al.* (2007). Whereas the idea behind AdMit and MitISEM is the same, i.e. iteratively constructing an approximation of a target distribution by a mixture of Student- t distributions, there are two substantial differences. First, AdMit aims at minimizing the variance of the IS estimator directly, whereas MitISEM aims at this goal indirectly by minimizing the Kullback-Leibler divergence. As a result, AdMit optimizes the mixture component weights using a non-linear optimization procedure that requires considerable computational effort. Second, in the AdMit method, means and covariance matrices of the candidate components are chosen heuristically and are never updated when additional components are added to the mixture, whereas in MitISEM all mixture parameters are optimized jointly by means of the relatively quick EM algorithm. This implies a large reduction of the computing time in the approximation procedure, and is expected to lead to a better candidate in most applications. One relative advantage of the AdMit approach is the step in which the importance weight function is maximized with respect to the parameter vector, which may lead to finding relevant areas of the parameter space that were ‘missed’ by all draws from the previous candidate. We intend to investigate the use of such an AdMit step within MitISEM in further research.

2.3 MitISEM

If one uses Importance Sampling or the Metropolis-Hastings algorithm to conduct posterior analysis, a key issue is to find a candidate density which approximates the target distribution. This can be quite difficult if the target density is not elliptical. This chapter proposes to specify the candidate distribution as a mixture of Student- t distributions. As discussed by Hoogerheide *et al.* (2007), the usage of mixtures of Student- t distributions has several advantages. First, they can provide an accurate approximation to a wide variety of target densities. For example, they can exhibit substantial skewness or irregularly curved contours such as multimodality. Zeevi and Meir (1997) show that under

certain conditions any density function may be approximated to arbitrary accuracy by a convex combination of ‘basis’ densities; the mixture of Student- t densities falls within their framework. Second, simulation from the Student- t distribution and evaluation of the Student- t density are performed easily and efficiently. Third, Student- t distributions have fatter tails than normal distributions, which reduces the risk that the tails of the candidate density are thinner than those of the target distribution. Fourth, a mixture of t approximation to a target distribution can be constructed in a quick, automatic, reliable manner by our novel procedure.

We will use the notation $f(\theta)$ for the target density kernel of θ , the k -dimensional vector of interest. $f(\theta)$ is typically a posterior density kernel, but it can also be a density kernel of observable variables or a density kernel of both parameters and observable variables. $g(\theta)$ is the candidate density, a mixture of H Student- t densities:

$$g(\theta) = g(\theta|\zeta) = \sum_{h=1}^H \eta_h t_k(\theta|\mu_h, \Sigma_h, \nu_h), \quad (2.1)$$

where ζ is the set of modes μ_h , scale matrices Σ_h , degrees of freedom ν_h , and mixing probabilities η_h ($h = 1, \dots, H$) of the k -dimensional Student- t components with density:

$$t_k(\theta|\mu_h, \Sigma_h, \nu_h) = \frac{\Gamma\left(\frac{\nu_h+k}{2}\right)}{\Gamma\left(\frac{\nu_h}{2}\right) (\pi\nu_h)^{k/2} |\Sigma_h|^{-1/2}} \times \left(1 + \frac{(\theta - \mu_h)' \Sigma_h^{-1} (\theta - \mu_h)}{\nu_h}\right)^{-(k+\nu_h)/2}. \quad (2.2)$$

Here Σ_h is positive definite, $\eta_h \geq 0$ and $\sum_{h=1}^H \eta_h = 1$. We further restrict ν_h such that $\nu_h \geq 1$.

First, assume that the number of components H is given. In the sequel of this section we will propose a ‘bottom up’ procedure that starts with one Student- t distribution and which iteratively adds Student- t components until a certain stop criterion is met. The aim is to choose the candidate mixture density $g(\theta)$ in such a way that it provides a good approximation of the target density $\tilde{f}(\theta)$ of which $f(\theta)$ is a kernel. Typically, $\tilde{f}(\theta)$ will represent the posterior density, whereas $f(\theta)$ represents the unnormalized posterior density kernel which can be evaluated explicitly. We do this by choosing ζ such that it mini-

mizes the Kullback-Leibler divergence (or Cross-entropy distance) (Kullback and Leibler, 1951), which is defined as

$$\mathcal{D}_1(\tilde{f} \rightarrow g) = \int \tilde{f}(\theta) \log \frac{\tilde{f}(\theta)}{g(\theta|\zeta)} d\theta. \quad (2.3)$$

This is obviously equivalent with minimizing

$$\mathcal{D}_1(f \rightarrow g) = \int f(\theta) \log \frac{f(\theta)}{g(\theta|\zeta)} d\theta. \quad (2.4)$$

as long as the same kernel f of the target density \tilde{f} is used throughout the minimization. Since

$$\begin{aligned} \mathcal{D}_1(f \rightarrow g) &= \int f(\theta) \log \frac{f(\theta)}{g(\theta|\zeta)} d\theta = \\ &= \int f(\theta) \log f(\theta) d\theta - \int f(\theta) \log g(\theta|\zeta) d\theta, \end{aligned} \quad (2.5)$$

where only the second term on the right-hand side of (2.5) depends on ζ , this amounts to maximizing

$$\int f(\theta) \log g(\theta|\zeta) d\theta = E_{\theta \sim f(\theta)}[\log g(\theta|\zeta)] = \quad (2.6)$$

$$\int g_0(\theta) \frac{f(\theta)}{g_0(\theta)} \log g(\theta|\zeta) d\theta = E_{\theta \sim g_0(\theta)} \left[\frac{f(\theta)}{g_0(\theta)} \log g(\theta|\zeta) \right], \quad (2.7)$$

where $g_0(\theta)$ is a given candidate density that has been obtained in a previous step. For $H = 1$ the density $g_0(\theta)$ is an initial candidate distribution, such as a Student- t distribution around the posterior mode with scale matrix equal to minus the inverse Hessian of the log-posterior at the mode, or an adapted version thereof. For $H \geq 2$, g_0 is a mixture of $H - 1$ Student- t components, that has been obtained in the previous step of the ‘bottom up’ construction procedure.

We use an Expectation-Maximization (EM) algorithm for minimizing the stochastic counterpart of (2.7) in order to find

$$\zeta^* = \arg \max_{\zeta} \frac{1}{N} \sum_{i=1}^N W^i \log g(\theta^i|\zeta) \quad \text{with} \quad W^i = \frac{f(\theta^i)}{g_0(\theta^i)},$$

where θ^i ($i = 1, 2, \dots, N$) are independent draws from g_0 . Note that both the θ^i and W^i are given during the optimization; θ^i and W^i ($i = 1, 2, \dots, N$) do not depend on ζ . We emphasize that the importance weighted version of the EM algorithm is applied, rather than minimizing the stochastic counterpart of (2.6) by a ‘regular’ EM algorithm, in order to use all candidate draws without requiring the Metropolis-Hastings algorithm to transform the candidate draws into a set of posterior draws. This has three advantages. First, we do not require a burn-in sample. Second, the use of all candidate draws θ^i ($i = 1, 2, \dots, N$) helps to prevent numerical problems with estimating candidate covariance matrices; also draws with relatively small, but positive importance weights are helpful for this purpose. Third, the use of all candidate draws may lead to a better approximation.

The EM algorithm (Dempster *et al.*, 1977)) is based on the idea that a complex model for some observable ‘data’ θ with parameters ζ can be formulated in a simpler form with latent data $\tilde{\theta}$ in addition to θ and ζ . If the latent data $\tilde{\theta}$ were observed, the computation of the Maximum Likelihood estimator of θ would be relatively straightforward. Each iteration L of the EM algorithm consists of two (iterative) steps, the Expectation and Maximization step. The first (Expectation) step takes the expectation of the log-likelihood function with respect to the latent data $\tilde{\theta}$ (given the parameter values $\zeta^{(L-1)}$ from the previous iteration). The second (Maximization) step maximizes this expected log-likelihood with respect to the parameters.

In our situation we maximize the *weighted* log-density

$$\frac{1}{N} \sum_{i=1}^N W^i \log g(\theta^i | \zeta),$$

where $g(\cdot | \zeta)$ is the mixture of Student- t densities (2.1). The mixture of Student- t densities (2.1) for θ^i is equivalent with the specification

$$\theta^i \sim N(\mu_h, w_h^i \Sigma_h) \quad \text{if} \quad z_h^i = 1,$$

where z^i is a latent H -dimensional vector indicating from which Student- t component the observation θ^i stems: if θ^i stems from component h , then $z_h^i = 1$, $z_j^i = 0$ for $j \neq h$; $\Pr[z^i = e_h] = \eta_h$ with e_h the h -th column of the identity matrix; w_h^i has the Inverse-

Gamma distribution $IG(\nu_h/2, \nu_h/2)$. For a more extensive explanation of the continuous scale mixing representation of Student- t distributions we refer to Rubin (1983) and to Lange *et al.* (1989) who consider the more general situation with unknown degrees of freedom. For mixtures of Student- t distributions we refer to Peel and McLachlan (2000).

Here we have latent ‘data’ $\tilde{\theta}^i$ ($i = 1, \dots, N$)

$$\tilde{\theta}^i = \{z_h^i, w_h^i | h = 1, \dots, H\}$$

and the so-called data-augmented density is given by

$$\begin{aligned} \log p(\theta^i, w^i, z^i | \zeta) &= \log p(\theta^i | w^i, z^i, \zeta) + \log p(w^i | \zeta) + \log p(z^i | \zeta) \\ &= \sum_{h=1}^H z_h^i \log \left[\text{pdf}_{N(\mu_h, w_h^i \Sigma_h)}(\theta^i) \right] \\ &\quad + \sum_{h=1}^H \log \text{pdf}_{IG(\nu_h/2, \nu_h/2)}(w_h^i) + \sum_{h=1}^H z_h^i \log(\eta_h) \\ &= \sum_{h=1}^H z_h^i \left\{ -\frac{k}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_h| - \frac{k}{2} \log(w_h^i) \right. \\ &\quad \left. - \frac{1}{2} \frac{(\theta^i - \mu_h)'(\Sigma_h)^{-1}(\theta^i - \mu_h)}{w_h^i} \right\} + \sum_{h=1}^H \left\{ \frac{\nu_h}{2} \log\left(\frac{\nu_h}{2}\right) \right. \\ &\quad \left. - \left(\frac{\nu_h}{2} - 1\right) \log(w_h^i) - \frac{\nu_h}{2} \frac{1}{w_h^i} - \log\left(\Gamma\left(\frac{\nu_h}{2}\right)\right) \right\} \\ &\quad + \sum_{h=1}^H z_h^i \log(\eta_h) \end{aligned} \tag{2.8}$$

where w^i and z^i are *a priori* independent. The expressions of the latent variables w^i and z^i that appear in terms which also involve the parameters ζ to be optimized are z_h^i , $\frac{z_h^i}{w_h^i}$, $\log w_h^i$, and $\frac{1}{w_h^i}$. The conditional expectations given θ^i and $\zeta = \zeta^{(L-1)}$, the optimal

parameters in the previous EM iteration, are as follows:

$$\tilde{z}_h^i \equiv E[z_h^i | \theta^i, \zeta = \zeta^{(L-1)}] = \frac{t(\theta^i | \mu_h, \Sigma_h, \nu_h) \eta_h}{\sum_{j=1}^H t(\theta^i | \mu_j, \Sigma_j, \nu_j) \eta_j}, \quad (2.9)$$

$$\widetilde{z/w}_h^i \equiv E\left[\frac{z_h^i}{w_h^i} \middle| \theta^i, \zeta = \zeta^{(L-1)}\right] = \tilde{z}_h^i \frac{k + \nu_h}{\rho_h^i + \nu_h}, \quad (2.10)$$

$$\begin{aligned} \xi_h^i &\equiv E[\log w_h^i | \theta^i, \zeta = \zeta^{(L-1)}] = \\ &= \left[\log\left(\frac{\rho_h^i + \nu_h}{2}\right) - \psi\left(\frac{k + \nu_h}{2}\right) \right] \tilde{z}_h^i \\ &\quad + \left[\log\left(\frac{\nu_h}{2}\right) - \psi\left(\frac{\nu_h}{2}\right) \right] (1 - \tilde{z}_h^i), \end{aligned} \quad (2.11)$$

$$\delta_h^i \equiv E\left[\frac{1}{w_h^i} \middle| \theta^i, \zeta = \zeta^{(L-1)}\right] = \frac{k + \nu_h}{\rho_h^i + \nu_h} \tilde{z}_h^i + (1 - \tilde{z}_h^i), \quad (2.12)$$

with $\rho_h^i = (\theta^i - \mu_h)' \Sigma_h^{-1} (\theta^i - \mu_h)$, $\psi(\cdot)$ the digamma function (the derivative of the logarithm of the gamma function $\log \Gamma(\cdot)$), and all parameters $\mu_h, \Sigma_h, \nu_h, \eta_h$ elements of $\zeta^{(L-1)}$. For the derivations of these expectations we refer to the appendix.

Define $\log \tilde{p}(\theta^i, w^i, z^i | \zeta)$ as the result of substituting the expectations (2.9)-(2.12) into $\log p(\theta^i, w^i, z^i | \zeta)$ in (2.8). The Maximization step amounts to computing the ζ that maximizes

$$\zeta^{(L)} = \arg \max_{\zeta} \frac{1}{N} \sum_{i=1}^N W^i \log \tilde{p}(\theta^i, w^i, z^i | \zeta).$$

Using the analogy with Maximum Likelihood estimation for the Seemingly Unrelated Regression model with Gaussian errors (for the k elements of θ^i) and the same ‘regressor’ (a constant term) in each equation, in which case the Ordinary Least Squares (OLS) estimator provides the Maximum Likelihood Estimator, and Maximum Likelihood estimation for the multinomial distribution, it is easily derived that $\zeta^{(L)}$ consists of:

$$\mu_h^{(L)} = \left[\sum_{i=1}^N W_i \widetilde{z/w}_h^i \right]^{-1} \left[\sum_{i=1}^N W_i \widetilde{z/w}_h^i \theta^i \right], \quad (2.13)$$

$$\hat{\Sigma}_h^{(L)} = \frac{\sum_{i=1}^N W_i \widetilde{z/w}_h^i (\theta^i - \mu_h^{(L)}) (\theta^i - \mu_h^{(L)})'}{\sum_{i=1}^N W_i \tilde{z}_h^i}, \quad (2.14)$$

$$\eta_h^{(L)} = \frac{\sum_{i=1}^N W_i \tilde{z}_h^i}{\sum_{i=1}^N W_i}. \quad (2.15)$$

Further, $\nu_h^{(L)}$ is solved from the first order condition of ν_h :

$$-\psi(\nu_h/2) + \log(\nu_h/2) + 1 - \frac{\sum_{i=1}^N W_i \xi_h^i}{\sum_{i=1}^N W_i} - \frac{\sum_{i=1}^N W_i \delta_h^i}{\sum_{i=1}^N W_i} = 0. \quad (2.16)$$

Cappé *et al.* (2008) only update the expectations and covariance structures of the Student- t distributions and not the degrees of freedom, because there is no closed-form solution for the latter. We propose to optimize also the degrees of freedom parameter ν_h during the EM procedure for three reasons. First, the larger flexibility may lead to a better approximation of the target distribution. Second, solving ν_h from (2.16) requires only a one-dimensional root finder, which requires little computation time. Moreover, $1 - \frac{\sum_{i=1}^N W_i \xi_h^i}{\sum_{i=1}^N W_i} - \frac{\sum_{i=1}^N W_i \delta_h^i}{\sum_{i=1}^N W_i}$ is constant with respect to ν_h , so that it only has to be evaluated once in the process of solving the equation. Third, the resulting values of ν_h ($h = 1, \dots, H$) may provide information on the shape of the target distribution (e.g. whether the kurtosis is small, moderate or large).

We now discuss two remaining issues: (1) how to choose the number of components H ; (2) how to specify the initial values in the EM algorithm. In order to deal with both issues, we use a ‘bottom up’ procedure that starts with one Student- t distribution and which iteratively adds Student- t components until a certain stop criterion is met:

Algorithm 1. The *basic MitISEM* approach for obtaining an approximation to a target density:

- (0) **Initialization:** Simulate draws $\theta^1, \dots, \theta^N$ from the naive proposal density g_{naive} where g_{naive} denotes a Student- t distribution with mode and scale matrix equal to the target distribution’s mode and minus the inverse Hessian of the log-target density kernel evaluated at the mode.
- (1) **Adaptation:** Estimate the target distribution’s mean and covariance matrix using IS with the draws $\theta^1, \dots, \theta^N$ from g_{naive} . Use these estimates as the mode and scale matrix of Student- t distribution $g_{adaptive}$. Draw a sample $\theta^1, \dots, \theta^N$ from this adaptive Student- t distribution $g_0 = g_{adaptive}$, and compute the IS weights for this sample.

- (2) Apply the **IS-weighted EM algorithm** given the latest IS weights and the drawn sample of step 1. The output consists of the new candidate density g with optimized ζ , the set of $\mu_h, \Sigma_h, \nu_h, \eta_h$ ($h = 1, \dots, H$). Draw a new sample $\theta^1, \dots, \theta^N$ from this proposal density and compute corresponding IS weights.
- (3) **Iterate on the number of mixture components:** Given the current mixture of H components with corresponding μ_h, Σ_h, ν_h and η_h ($h = 1, \dots, H$), take $x\%$ of the sample $\theta^1, \dots, \theta^N$ that correspond to the highest IS weights. Construct with these draws and IS weights a new mode μ_{H+1} and scale matrix Σ_{H+1} which are starting values for the additional component in the mixture candidate density. The reason behind this choice is that the new component is meant to cover a region of the parameter space in which the previous candidate mixture had relatively too little probability mass. Starting values for η_{H+1} and ν_{H+1} are at each iteration set at 0.10 and 1, respectively. Obvious starting values for μ_h, Σ_h and ν_h ($h = 1, \dots, H$) are the optimal values in the mixture of H components, while η_h is 0.90 times the previously optimal value. Given the latest IS weights and the drawn sample from the current mixture of H components, apply the IS-weighted EM algorithm to optimize *each* mixture component μ_h, Σ_h, ν_h and η_h ($h = 1, \dots, H + 1$). Draw a new sample from the mixture of $H + 1$ components and compute corresponding IS weights.
- (4) **Evaluate the IS weights** by computing the Coefficient of Variation (C.o.V.), i.e. the standard deviation of the IS weights divided by their mean. Stop the algorithm when this coefficient has converged. Otherwise return to step 3.

Step (1) can be seen as an intermediate step which quickly tries to improve the initial candidate distribution g_0 , before calling the IS-weighted EM algorithm. If during the EM algorithm, a scale matrix Σ_h of a Student- t component (with very small weight η_h) becomes (nearly) singular, then this h -th component is removed from the mixture. We emphasize that in the iteration on the number of mixture components, the EM algorithm is applied to optimize *all* components. This is a qualitative improvement compared to the AdMit approach of Hoogerheide *et al.* (2007), which fixes the Student- t densities once they are formed.

There are still two strategic issues to be discussed about the MitISEM algorithm. The first issue relates to the following question: what is an efficient simulation method? Is this a simulation method that, given a certain amount of computing time, provides an estimate of a quantity of interest with the highest possible precision? Or is this a simulation method that, given a certain required precision, needs the shortest computing time. The optimal number of Student- t components may depend on the available computing time or the required precision. The more computing time is available, or the higher the required precision, the more rewarding a large ‘investment’ in an accurate approximation may be. Moreover, in order to choose the optimal number of Student- t components, we need to know the quantity of interest. That is, for a particular quantity of interest and a particular desired precision (or available amount of computing time), one could attempt to compute an optimal allocation of computing time over the construction of the candidate and the subsequent use in IS or the MH algorithm. We intend to investigate this issue in future research. In the current chapter, we propose a heuristic procedure that continues adding Student- t components until the approximation’s quality ‘hardly’ improves. We define the latter as a relative change in the C.o.V. of the IS weights that is smaller than 10%.

We discuss examples in which the posterior distribution is itself approximated, which seems a reasonable choice when we are interested in quantities such as the posterior mean, median or covariance. For the specific application of multi-step-ahead forecasting of Value at Risk (VaR), it is arguably wise to approximate the optimal importance density of Geweke (1989), see Hoogerheide and Van Dijk (2010). In the latter case, one may monitor the Numerical Standard Error (NSE) of the estimated VaR, as an alternative to the C.o.V. of IS weights.

Second, although the EM procedure is guaranteed to converge to a *local* optimum, the choice of the starting values may still be crucial, given that the KL divergence between target and candidate (as a function of the candidate mixture’s means, covariances, degrees of freedom and component weights) is a highly non-elliptical, multimodal function. MitISEM uses $x\%$ of the sample $\theta_1, \dots, \theta_N$ that correspond to the highest IS weights, in order to compute starting values for the mode μ_{H+1} and scale matrix Σ_{H+1} of the additional component in the mixture candidate density. The optimal choice of $x\%$ depends on the particular target distribution and the current candidate mixture of H Student- t compo-

nents. Therefore, we apply the EM algorithm with three different starting values (based on 1%, 5% or 10% of the draws $\theta_1, \dots, \theta_N$), and continue the algorithm with the resulting mixture of $H + 1$ Student- t components that yields the lowest C.o.V. value of the IS weights among the three approaches.

The results in the present chapter suggest that the current implementation of MitISEM is successful at constructing approximations that are useful candidate distributions. It should be stressed that we do *not* require the globally optimal candidate distribution: it suffices to have a ‘good’ approximation that makes a trade-off between the computing time of constructing a candidate distribution and the efficiency during the subsequential simulation.

2.3.1 Application: Bayesian analysis of the DCC-GARCH model

In this subsection the MitISEM approach is applied to the popular Dynamic Conditional Correlation (DCC) GARCH model of Engle (2002). This multivariate GARCH model allows the conditional correlation between multiple time series to be time-varying, whereas it allows flexible GARCH specifications for the univariate processes. For the Bayesian estimation of this model, a regular Gibbs sampling approach is not feasible due to the recursive structure of GARCH models. One could apply a Griddy-Gibbs sampler (Ritter and Tanner, 1992), but this sampler is known to be relatively very slow. We use the MH sampler with a candidate density resulting from the MitISEM algorithm, and compare the performance of the MitISEM candidate density with two samplers from the literature: the adaptive Metropolis (AM) sampler of Roberts and Rosenthal (2009) and the adaptive independent Metropolis-Hastings (AIMH) sampler of Giordani and Kohn (2010).

In our example, the d -dimensional vector y_t ($t = 1, \dots, T$) consists of (demeaned) returns of asset prices and is supposed to follow the following conditional distribution:

$$y_t | \mathcal{I}_{t-1} \sim N(0, H_t) \quad (2.17)$$

with \mathcal{I}_{t-1} the information set at time $t - 1$ and H_t representing the time-varying conditional covariance matrix of the returns. Decomposing H_t into conditional variances and

correlations, H_t can be written as

$$H_t = D_t R_t D_t, \quad (2.18)$$

where D_t represents a $d \times d$ time-varying diagonal matrix containing the square root of the conditional variances $h_{i,t}$ ($i = 1, \dots, d$) of the asset returns $y_{i,t}$. The d conditional variances follow a GARCH process,

$$h_{i,t} = \omega_i + \alpha_i y_{i,t-1}^2 + \beta_i h_{i,t-1} \quad (i = 1, \dots, d) \quad (2.19)$$

with the usual restrictions $\omega_i \geq 0, \alpha_i \geq 0$ and $\beta_i \geq 0$ in order to ensure positive values of the conditional variance. To ensure covariance stationarity of y_t , one must impose $\alpha_i + \beta_i \leq 1$ ($i = 1, \dots, d$).

Regarding the correlations, Engle (2002) suggests a dynamic process

$$Q_t = (1 - A - B) \bar{Q} + A(z_{t-1} z'_{t-1}) + B Q_{t-1} \quad (2.20)$$

with scalars A and B satisfying $A \geq 0, B \geq 0$ and $A+B \leq 1$, and with \bar{Q} representing the unconditional correlation matrix of the standardized residuals $z_t = D_t^{-1} y_t$. This matrix \bar{Q} is estimated by the sample correlation matrix $\frac{1}{T} \sum_{t=1}^T \hat{z}_t \hat{z}'_t$. The stated conditions plus a positive definite initial matrix Q_0 guarantee a positive definite matrix Q_t . The matrix Q_t has to be rescaled however to produce a valid time-varying *correlation* matrix (with diagonal elements 1):

$$R_t = (I \circ Q_t)^{-1/2} Q_t (I \circ Q_t)^{-1/2}, \quad (2.21)$$

where ‘ \circ ’ denotes the Hadamard product. We take uniform priors on the parameter vector θ , which consists of 17 parameters (5 times a univariate GARCH process plus the Dynamic Correlation process). For α_i, β_i, A and B , we are restricted to $[0,1]$ plus the covariance stationarity restrictions. For the remaining five parameters ω_i we use a truncated uniform prior $[0, P]$ in order to have proper non-informative priors. When using truncated uniform priors, many draws may fall outside the feasible prior region for a naive Student- t candidate distribution. An advantage of the MitISEM algorithm is that it produces a

rather close approximation to the posterior (including the 0 level outside the ‘allowed range’) that will have almost all probability mass inside the feasible region.

We take returns from five indices: the MSCI World, the MSCI Emerging Markets, the Barclays Global Bond Index, the DJ AIG-Spot Commodity Index and the FTSE EPRA/NAREIT Global Real Estate Index. The first series is a stock market index of 1500 world stocks of 23 different developed countries, maintained by Morgan Stanley Capital International (MSCI inc.). The Emerging Markets Index is a market capitalization index that consists of indices in 26 emerging economies. The third series is often used to represent investment grade bonds being traded in the United States. The Dow Jones-AIG Commodity Index is a benchmark index for the commodities markets. It is composed of futures contracts on 19 physical commodities. The last series represents trends in real estate equities worldwide. From these indices we use daily observations on log return (100 times the change of the logarithm of the closing price) from January 3, 2000 to November 3, 2003, which corresponds with 1,000 observations.

Posterior means of the model parameters are estimated by using the independence chain MH algorithm with the candidate density produced by MitISEM. We compare the performance of MitISEM to the AM sampler of Roberts and Rosenthal (2009) and the AIMH sampler of Giordani and Kohn (2010). The first sampler is based on a mixture of two multivariate normal distributions, where both covariance matrices are multiplied by factors that aim at an optimal random walk proposal distribution and a high acceptance rate of the local sampler. The second sampler consists of a mixture of normal densities, which are estimated by the k -harmonic means clustering algorithm instead of the EM-algorithm that is used in MitISEM. See the appendix for a brief description of both samplers.

Table 2.1 shows posterior means estimated by the MH algorithms. For all methods, we simulate 100,000 draws after a burn-in sample of 5000 draws. Numerical standard errors (NSE) are obtained by using the integrated autocorrelation time (IACT),

$$IACT = 1 + 2 \sum_{\tau=1}^{\infty} \rho_{\theta}(\tau), \quad (2.22)$$

Table 2.1: Application of the basic MitISEM algorithm

This table provides results of comparing the basic MitISEM algorithm with the AIMH sampler of Giordani and Kohn (2010), and the AM sampler of Roberts and Rosenthal (2009) to the posterior in the DCC-GARCH model with 17 parameters, using daily data from five indices: the MSCI World, the MSCI Emerging Markets, the Barclays Global Bond Index, the DJ AIG-Spot Commodity Index and the FTSE EPRA/NAREIT Global Real Estate Index. The columns report the estimated posterior means, corresponding numerical standard errors (NSEs) times 100 and first order autocorrelations of draws $\rho_\theta(1)$. Results are based on 100,000 draws after a burn-in period of 5,000 draws. We report 100 times the NSE values which are obtained by equation (2.23). The sample goes from January 3, 2000 through November 3, 2003 (1,000 observations).

Panel A: Posterior results									
	MitISEM			AIMH (GK)			AM (RR)		
	mean	NSE	$\rho_\theta(1)$	mean	NSE	$\rho_\theta(1)$	mean	NSE	$\rho_\theta(1)$
ω_1	0.056	0.022	0.716	0.056	0.028	0.832	0.056	0.041	0.970
α_1	0.076	0.016	0.679	0.075	0.021	0.817	0.076	0.034	0.970
β_1	0.872	0.031	0.698	0.872	0.041	0.827	0.872	0.061	0.969
ω_2	0.123	0.065	0.723	0.124	0.088	0.853	0.124	0.119	0.971
α_2	0.090	0.027	0.700	0.090	0.033	0.828	0.090	0.048	0.968
β_2	0.802	0.076	0.716	0.801	0.105	0.852	0.801	0.136	0.970
ω_3	0.005	0.002	0.722	0.005	0.003	0.843	0.005	0.004	0.973
α_3	0.096	0.026	0.687	0.095	0.035	0.821	0.095	0.053	0.970
β_3	0.749	0.073	0.709	0.747	0.104	0.839	0.750	0.147	0.972
ω_4	0.353	0.120	0.705	0.355	0.162	0.830	0.355	0.237	0.969
α_4	0.130	0.046	0.710	0.132	0.061	0.837	0.131	0.083	0.968
β_4	0.409	0.174	0.705	0.405	0.235	0.835	0.407	0.338	0.968
ω_5	0.072	0.018	0.685	0.072	0.024	0.826	0.072	0.036	0.970
α_5	0.173	0.040	0.697	0.173	0.051	0.828	0.172	0.073	0.970
β_5	0.659	0.064	0.687	0.660	0.090	0.831	0.660	0.123	0.969
A	0.014	0.003	0.692	0.014	0.004	0.824	0.014	0.005	0.971
B	0.974	0.008	0.725	0.974	0.008	0.834	0.975	0.012	0.973

Panel B: Diagnostics			
	MitISEM	AIMH(GK)	AM(RR)
Acceptance rate	34%	19%	29%
Computational time (in seconds)			
Constructing candidate	2170		
MH-sampler	3400	4625	5590

where we truncate this sum of τ -th order autocorrelations $\rho_\theta(\tau)$ at $\tau_{max} = 50$. Hence the variance of the sample mean $\hat{\theta}_{mean}$ after N iterations of the MCMC algorithm is equal to:

$$\text{var}(\hat{\theta}_{mean}) = \sigma_\theta^2 / N \times IACT. \quad (2.23)$$

The main result from Table 2.1 is that MitISEM outperforms the competing algorithms in this DCC-GARCH application, since the NSE values of the estimated posterior means are smaller than the corresponding values of AIMH and AM. MitISEM combines a higher

acceptance rate with lower first order autocorrelations than the competing algorithms. MitISEM does require more computing time (on an AMD Athlon™ II X2 B24 processor) than AIMH, but if we give AIMH the same computing time (generating more draws), then its NSEs only drop approximately 10%, so that these are still clearly worse than those of MitISEM. If we compare the AIMH and AM algorithms, then the acceptance rate of AM is higher than the rate of AIMH, but the high serial correlation of the AM draws increases the IACT, causing higher NSE values.

Note that the relative quality of the AIMH and AM algorithms, as compared with MitISEM, may improve for parameter spaces with higher dimension. In such cases a comparison of AIMH and AM with the basic MitISEM approach would be particularly interesting. A systematic study of the relevant merits of alternative sampling schemes for a variety of target density shapes and dimensions is a topic of great interest, which is however beyond the scope of the present study. In any case, we expect that no algorithm will dominate in all applications. Moreover, given that diagnostic checks can never fully guarantee that simulation results have converged to the true values, the use of multiple simulation methods can be a quite useful validity check.

2.4 Sequential MitISEM

In this section, we propose a method for applying MitISEM in a sequential manner, so that the candidate distribution for posterior simulation is cleverly updated when new data become available. Our results show that the computational effort reduces enormously, while the quality of the approximation remains almost unchanged, as compared with an ‘ad hoc’ procedure in which the construction of the MitISEM candidate is performed ‘from scratch’ at every moment in time. In the next subsection we show how this sequential approach can be combined with a tempering approach, which facilitates the simulation from densities with multiple modes that are far apart. For sequential Monte Carlo methods, we refer to Liu and Chen (1998), Doucet *et al.* (2001), and Chopin (2002). The latter explicitly takes into account that a candidate proposal will not be updated until the sequential weights become very variable.

The previous section showed that, although the IS-weighted EM steps are relatively efficient, the construction of an appropriate candidate distribution may still require considerable computing time. This may seem a serious disadvantage if one requires multiple estimates over time, for example daily Bayesian forecasts. However, the idea behind the procedure in this section is that the posterior for the parameters given the data $y_{1:T+1} = \{y_1, \dots, y_T, y_{T+1}\}$ is typically not so different from the posterior for the parameters given the data $y_{1:T} = \{y_1, \dots, y_T\}$. That is, for large T , $p(\theta|y_1, \dots, y_{T+1}) \approx p(\theta|y_1, \dots, y_T)$ where

$$p(\theta|y_1, \dots, y_{T+1}) = \frac{p(y_{T+1}|\theta, y_1, \dots, y_T)p(\theta|y_1, \dots, y_T)}{p(y_{T+1}|y_1, \dots, y_T)}.$$

The variance of the incremental weight $p(y_{T+1}|\theta, y_1, \dots, y_T)$, with respect to $p(\theta|y_1, \dots, y_T)$, declines with T , as shown by Chopin (2002). Therefore, one can ‘recycle’ the same candidate distribution. At many moments, the candidate distribution can simply be reused. Further, if the candidate distribution needs to be updated, i.e. if its quality falls below a certain level, then we still do not require to start from scratch. It may suffice to perform an update using the IS-weighted EM algorithm, keeping the number H of Student- t components the same. Only if the resulting quality is still below a desired level, then we start the MitISEM procedure, adding components until convergence has been reached.

Suppose that at time $T + \tau$ ($\tau = 1, 2, \dots$) we want to analyze the posterior based on data $y_{1:T+\tau} = \{y_1, \dots, y_{T+\tau}\}$, and that time T was the last time when we had to update the candidate density. That is, the current candidate distribution has been estimated using the data $y_{1:T}$. Then at time $T + \tau$ we perform the following algorithm:

Algorithm 2. The *Sequential MitISEM* approach for obtaining a candidate density for the posterior density for data $y_{1:T+\tau}$ ($\tau = 1, 2, \dots$):

- (1) Compute *C.o.V. (no update)*, the C.o.V. value that is based on the posterior density kernel for data $y_{1:T+\tau}$ and the current candidate density.
- (2) Compare *C.o.V. (no update)* with *C.o.V. (T)*, the C.o.V. value of the last time when the candidate was updated. If the change is below a certain threshold (10%), stop. Otherwise go to step (3).

- (3) Run the IS-weighted EM algorithm with the current mixture of H Student- t densities as starting values. Sample from the new distribution (with the same number of components H) and compute IS weights and the corresponding C.o.V. value $C.o.V.$ (only EM update). Since the IS-weighted EM algorithm updates all mixture components, it can easily perform a useful shift of the candidate density.
- (4) Judge the value of $C.o.V.$ (only EM update). If the change of quality is below a certain threshold (10%), stop. Otherwise go to step (5).
- (5) Iterate on the number of components until the C.o.V. value has converged.

When a particular Student- t component gets a minimal weight, then the practical relevance is negligible. In such a case we delete the Student- t component from the mixture. So, the number of Student- t components is not monotonically increasing over time. In step (2) we compare $C.o.V.$ (no update) with $C.o.V.$ (T) rather than the C.o.V. for the posterior at time $y_{T+\tau-1}$, since in the latter case a series of small increases of the C.o.V. may eventually lead to a much worse candidate density, without the algorithm ever being ‘alarmed’ to update the candidate.

We apply the Sequential MitISEM algorithm to the univariate two-component Gaussian Mixture EGARCH model, which is given by:

$$\begin{aligned}
 y_t &= \mu + \sqrt{h_t} \varepsilon_t, \\
 \log(h_t) &= \omega + \gamma \frac{y_{t-1} - \mu}{\sqrt{h_{t-1}}} + \alpha \left(\frac{|y_{t-1} - \mu|}{\sqrt{h_{t-1}}} - \frac{E|y_{t-1} - \mu|}{\sqrt{h_{t-1}}} \right) + \beta \log(h_{t-1}) \\
 \varepsilon_t &\sim \begin{cases} N(0, \sigma^2) & \text{with probability } \rho \\ N(0, \sigma^2/\lambda) & \text{with probability } 1 - \rho \end{cases}, \tag{2.24}
 \end{aligned}$$

with h_t the conditional variance of y_t given the information set $\mathcal{I}_{t-1} = \{y_{t-1}, y_{t-2}, y_{t-3}, \dots\}$. See Nelson (1991) for the original (one-component) EGARCH model. In addition, $0 < \lambda < 1$, and $\sigma^2 \equiv 1/(\rho + (1 - \rho)/\lambda)$ so that $\text{var}(\varepsilon_t) = 1$; h_0 is treated as a known constant. We restrict $|\beta| \leq 1$ to ensure covariance stationarity of h_t and impose the prior restriction $0.5 < \lambda < 1$, so that it is ensured that the state with smaller variance has larger probability than the state with larger variance. Moreover, we truncate

μ , ω , α and γ such that these have proper (non-informative) priors. For the parameter vector $\theta = (\rho, \lambda, \mu, \omega, \gamma, \alpha, \beta)'$ of dimension $k = 7$ we have a uniform prior on $[0.5, 1] \times [0, 1] \times [-1, 1] \times [-1, 1] \times [-1, 1] \times [0, 1] \times [0, 1]$.

The returns y_t are taken from the S&P 500 index. From this index we use daily observations y_t ($t = 1, \dots, T$) on the log return (100 times the change of the logarithm of the closing price) from January 2, 1998 to March 6, 2003 (1350 observations).

We estimate the model on the first 1300 observations and recycle the obtained candidate density by adding iteratively one observation of the forecast sample to the existing sample. At each time $t = 1301, \dots, 1350$, we compute the predictive likelihood, see Gelfand and Dey (1994) and Eklund and Karlsson (2007), who provide an overview of several approaches including the fractional Bayes factor of O'Hagan (1995) and the intrinsic Bayes factor of Berger and Pericchi (1996). In principle, the marginal likelihood exists for this flat, proper prior. However, if we would perform a model comparison, then we could make the marginal likelihood for the mixture EGARCH model as low as we want, for example, by increasing the prior range for the parameter μ or ω . For the predictive likelihood this does not hold, since the effect of a lower (exact) prior density due to a wider prior range for μ or ω drops out of the ratio in (2.26) below. For other adaptive sampling methods for estimating marginal and predictive likelihoods we refer to Frühwirth-Schnatter and Wagner (2008) and Pitt *et al.* (2010).

The predictive likelihood, a useful quantity in Bayesian inference for model comparison, is computed as follows. By splitting the data $y = (y_1, \dots, y_T)$ into $y^* = (y_1, \dots, y_m)$ and $\tilde{y} = (y_{m+1}, \dots, y_T)$, the predictive likelihood of model M is given by:

$$p(\tilde{y}|y^*, M) = \int p(\tilde{y}|\theta, y^*, M)p(\theta|y^*, M)d\theta, \quad (2.25)$$

which is actually the marginal likelihood if we consider \tilde{y} as 'the data' and $p(\theta|y^*, M)$, the exact posterior density after observing y^* , as 'the prior'. Using Bayes' rule for this exact posterior density $p(\theta|y^*, M)$ and substituting into (2.25) yields

$$p(\tilde{y}|y^*, M) = \frac{\int p(y|\theta, M)p(\theta|M)d\theta}{\int p(y^*|\theta, M)p(\theta|M)d\theta}, \quad (2.26)$$

where $p(y|\theta, M)$ is the likelihood of the model M and $p(\theta|M)$ the prior density of the parameters θ in the model. Hence this predictive likelihood is simply the ratio of the marginal likelihood for all observations over the marginal likelihood for the first part of the data. In our example, the training sample y^* (for the marginal likelihood in the denominator of the predictive likelihood) consists of 500 observations, and remains fixed.

We compare the Sequential MitISEM approach with the ‘ad hoc MitISEM approach’, which runs the MitISEM algorithm from scratch at each time $t = 1301, \dots, 1350$. The comparison is twofold. First we compare the computing time that is required by both methods. Second the quality of the estimates of the predictive likelihood is compared. In order to fulfill the second comparison measure, we repeat the calculation of the predictive likelihoods 100 times and compute the NSE as the standard deviation over the repetitions.

Table 2.2 compares both methods in computational effort and provides more details about the results of the Sequential MitISEM algorithm. During the forecast sample, the constructed candidate density is adapted only one time (step (3)). In all other cases, it was not necessary in our strategy to adapt the candidate density.

Using the Sequential MitISEM algorithm implies a huge computational advantage, as it is more than 45 times faster than the ‘ad hoc MitISEM method’. The Sequential MitISEM algorithm is visualized in the top panel of Figure 2.1. The blue line represents $C.o.V.(T)$, the Coefficient of Variation that is used in step (2) for comparison, whereas the green line denotes $C.o.V.(no\ update)$. Finally the red line gives an impression of the quality of the ‘ad hoc MitISEM approach’: the average C.o.V. value of the ‘ad hoc MitISEM approach’ over the same period. When the dataset includes the 25th observation of the forecast sample, the new C.o.V. value is relatively too high. In this case the candidate density is updated which is shown by the upward shift of the dashed line, representing the new value of $C.o.V.(T)$ (and the new moment T of the latest update). The figure suggests that the quality of Sequential MitISEM is approximately the same as the ‘ad hoc MitISEM approach’, since the difference in C.o.V. values is quite small. (Note that the y-axis corresponds to merely the interval $[0.66, 0.84]$.)

An additional indication is given by the bottom panel of Figure 2.1 which shows the mean of 100 predictive likelihoods with 95% confidence bounds. Since the blue and red asterisks lie most of the time in both confidence intervals, we suggest again that the

Table 2.2: Application of Sequential MitISEM

This table reports results of an application of Sequential MitISEM and ‘ad hoc MitISEM’ (which simply runs the MitISEM algorithm from scratch on each sample $y_{1:t}$ ($t = 1301, \dots, 1350$)) to a Gaussian Mixture EGARCH model of (2.24), estimated using daily returns from the S&P 500 index. The number of times adapted denotes the case when the candidate is only updated, using IS-weighted EM, while the number of components is held constant. When the candidate is adapted and extended, the number of components increases. Reusing the candidate density implies that the same candidate density is held, hence no updating occurs. The computational time corresponds to the construction of 50 candidate densities over the period (1301 – 1350). The sample goes from January 2, 1998 to March 6, 2003 (1350 observations).

	Sequential MitISEM	Adhoc MitISEM
Sequential MitISEM steps		
# adapted	1	
# adapted and extended	0	
# reused	48	
Computational effort	117 s	5602 s

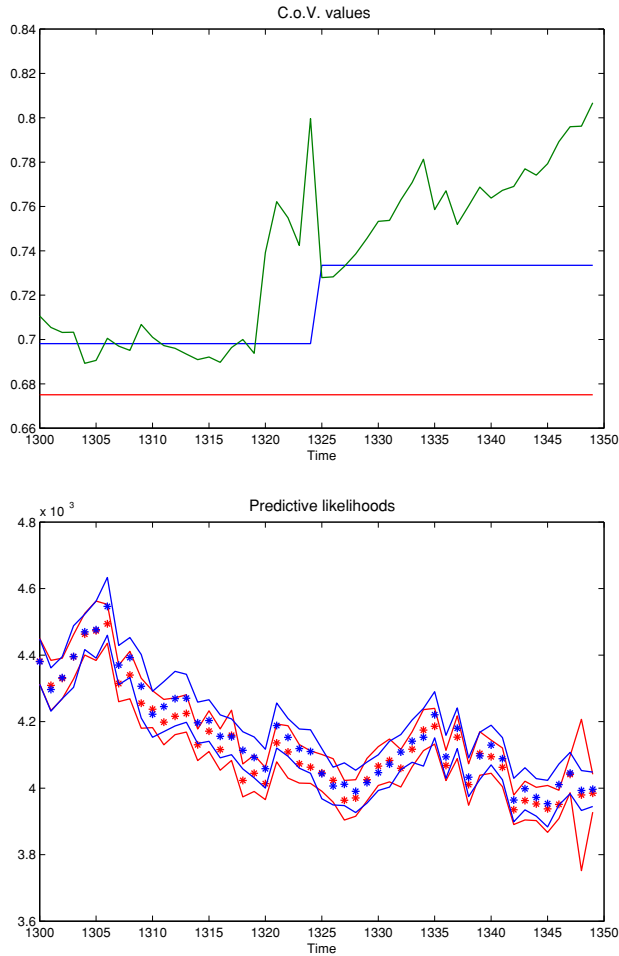
quality of the Sequential MitISEM algorithm is of the same order as the ‘ad hoc MitISEM approach’. We further note that the same procedure can be used if one makes use of a *moving window* instead of the *expanding window* of data that we use. To conclude this subsection, Sequential MitISEM is far more efficient compared to an ‘ad hoc approach’ as it produces approximately the same quality of candidate distributions for predictive likelihood estimation with considerably less computational effort.

2.4.1 Tempered MitISEM

Although the MitISEM approach can approximate multimodal target distributions, it may occur in extreme cases that the modes of a target distribution are so wide apart that one or more of the modes are ‘missed’. To decrease the probability that distant modes are ‘missed’, one can combine MitISEM with a tempering approach. The proposed tempering method moves sequentially from a tempered target density kernel, the target density kernel to the power of a positive number that is smaller than 1, towards the real target density kernel. The tempered target distribution is more diffuse, roughly stated ‘more uniform’, and hence the probability of detecting far-away modes is higher. The idea of tempering was introduced by Geyer (1991), see also Hukushima and Nemoto (1996). The tempering idea is also used in the Equi-Energy sampler, developed by Kou *et al.* (2006).

Figure 2.1: Sequential MitISEM: C.o.V values and predictive likelihoods

This figure illustrates the Sequential MitISEM approach for predictive likelihood estimation in a Gaussian Mixture EGARCH model of (2.24). The part of the figure represents $C.o.V. (T)$ values. In particular the blue line depicts the Coefficient of Variation that is used for comparison in step (2) of the Sequential MitISEM approach, whereas the green line denotes $C.o.V. (no\ update)$. Finally the red line gives an impression of the quality of the ‘ad hoc MitISEM approach’: the average C.o.V. value of the ‘ad hoc MitISEM approach’ over the same period. The lower part of the figure shows predictive likelihood estimates. The asterisks show at each time the mean of 100 predictive likelihoods; the red and blue lines correspond with 95% confidence bounds. The red asterisks and confidence bounds are based on the ‘ad hoc MitISEM approach’, where each day the MitISEM approach is applied from scratch. The blue asterisks and confidence bounds correspond with the Sequential MitISEM algorithm.



We apply the tempering approach in the following way as a Sequential MitISEM algorithm. Given a target density kernel $f(\theta)$, we temper this kernel by raising it to the power $(1/P_0)$ with $P_0 > 1$, i.e. $f(\theta)^{1/P_0}$. The MitISEM algorithm is applied to this tempered kernel $f(\theta)^{1/P_0}$. The resulting mixture of Student- t densities is used as input for the updated tempered target kernel, say $f(\theta)^{1/P_1}$, with $1 \leq P_1 < P_0$. This approach is repeated by decreasing P_n ($n = 0, 1, 2, \dots, \tilde{n}$) iteratively to $P_{\tilde{n}} = 1$, corresponding to the real target kernel. Many possible choices can be made on the number of iterations and the distance between the P_n . We follow Kou *et al.* (2006), and take equidistant steps of $\log(P_n)$. We label this approach the Tempered MitISEM procedure:

Algorithm 2*. *The Tempered MitISEM approach for obtaining an approximation to a multimodal target density with kernel $f(\theta)$:* Apply the Sequential MitISEM algorithm to $f(\theta)^{1/P_n}$ ($n = 0, 1, 2, \dots, \tilde{n}$) with P_n monotonically decreasing to $P_{\tilde{n}} = 1$.

To illustrate the Tempered MitISEM approach, we apply it to the same highly multimodal density that is used by Kou *et al.* (2006), a two-dimensional normal mixture for $\theta = (x_1, x_2)'$ with 20 modes that are relatively very far apart. Since most local modes are 15 standard deviations away from the nearest one, this mixture distribution is a good test for our approach. We compare three methods. First the Tempered MitISEM approach is used. In more detail, we choose $P_0 = 5$ and apply the MitISEM algorithm to the tempered target density. That is, we start with a ‘naive’ Student- t distribution around one of the modes, with scale matrix equal to minus the inverse Hessian of the log-density. We use this ‘naive’ Student- t distribution as a candidate in IS to obtain a first estimate of the mean and covariance matrix of the target distribution. We then continue with an ‘adaptive’ Student- t distribution with mode and scale matrix given by the first estimates of the target distribution’s mean and covariance matrix. After that, the usual steps 2-4 of Algorithm 1 in Section 2.3 are conducted. Given a candidate density, we move sequentially in five steps to $P_5 = 1$ with equally (log) spaced intervals. The second method applies the basic MitISEM algorithm to the real target density. Here no tempering approach is used. The final method is the aforementioned ‘adaptive’ candidate density, which is the Student- t

Table 2.3: Application of Tempered MitISEM

This table reports results of simulation from the two-dimensional normal mixture of Kou *et al.* (2006) using three different candidates: an (adaptive) Student- t density, and mixtures of Student- t densities from basic and Tempered MitISEM. The number of components of (Tempered) MitISEM and the corresponding C.o.V. of IS weights correspond with the last iteration of the MitISEM algorithm, as described in Algorithms 1 and 2*.

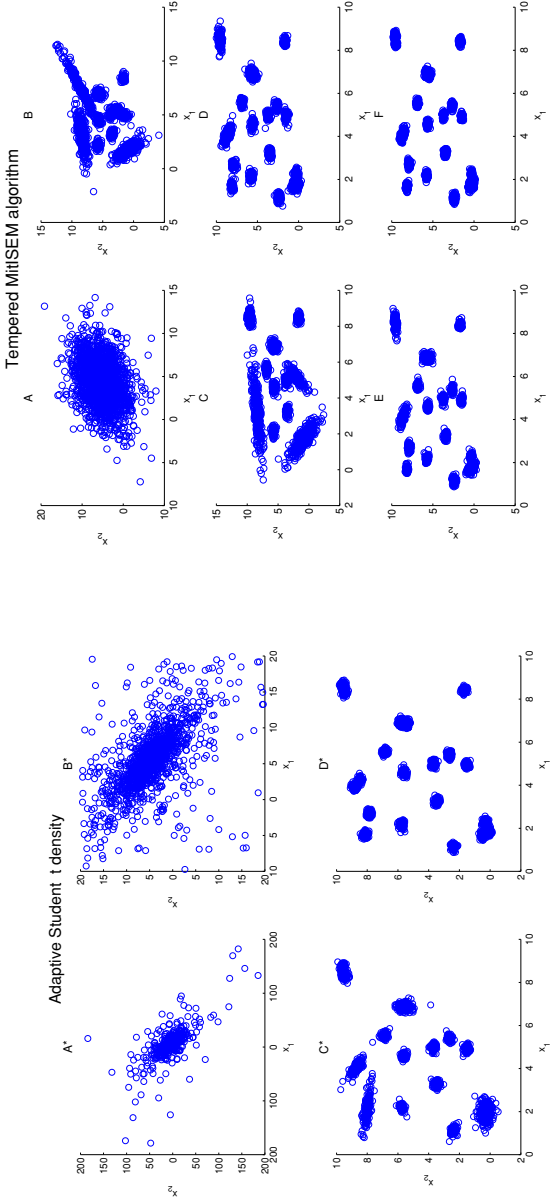
	Adaptive t	Basic MitISEM	Tempered MitISEM
# components in candidate mixture	1	14	16
C.o.V. of IS weights	21.57	0.78	0.43

distribution with adapted mode and scale matrix. That is, for the ‘adaptive’ candidate density we perform only step (0) and step (1) of the original MitISEM algorithm.

Figure 2.2 and Table 2.3 show simulation results from these three methods. All figures are based on 10,000 simulated draws. Panels (A^*) and (B^*) of Figure 2.2 show simulated draws from the adaptive candidate density, where panel (B^*) is similar to panel (A^*) but zoomed in on a closer interval. These panels plus the huge C.o.V. of IS weights in the table suggest that the ‘adaptive’ Student- t density produces poor results. In other words, one really needs advanced samplers to handle multimodal target kernels. Second, the basic MitISEM approach without tempering is a serious improvement, as the C.o.V. value decreases substantially from 22 to 0.78. The MitISEM algorithm is able to detect most of the modes, however by comparing panel (C^*) - simulated draws from the candidate density that is produced by MitISEM - to panel (D^*) of Figure 2.2, which represents simulated draws from the real target density, not all modes are covered. The mode around (8.41, 1.68) is missed by MitISEM. This reflects that if the mode lies too far away from the remaining modes, MitISEM may not be able to detect this important subdomain of the target density. Finally, the tempered MitISEM approach is shown in the right-hand panels of Figure 2.2. From panel (A) to (E), simulated candidate draws from the resulting candidate density of MitISEM applied to the target density $p(\theta)^{1/P}$ are shown, where P is equally log-spaced from 5 to 1. The importance of sequentially lowering the value of P_n lies in the fact that first the global area of interest is captured. Then a lower P_n in the subsequent panels shows an increasing precision of the local modes. In the end, the improvement of tempered MitISEM over basic MitISEM is clearly illustrated in panel (E), since all 20 modes are covered. The quality of the final candidate density is also

Figure 2.2: Simulation results of Tempered MitISEM

This figure shows an application of Tempered MitISEM to the bivariate multimodal distribution of Kou *et al.* (2006). The left part of the figure shows samples generated by the adaptive Student- t density (Panel A* and B*). These panels represent the same draws, but panel B* focuses on a smaller interval. Panel C* shows draws resulting from applying MitISEM to the original target density, while Panel D* depicts simulated draws from the real target distribution. The right part of the figure shows samples generated from each step of the Tempered MitISEM algorithm. Starting from panel A to E, simulated draws are shown from the candidate density that is produced by applying MitISEM to the target density $f(\theta)^{1/P}$, with P equally log-spaced from 5 to 1. Panel F shows draws simulated from the real target distribution.



confirmed by Table 2.3, as the *C.o.V.* value drops further from 0.77 to 0.43. We stress that the reported numbers of Student-*t* components are not chosen beforehand by the user; these are automatically found by the basic and tempered MitISEM methods.

2.5 Conclusion

We introduce a new class of adaptive sampling methods for efficient and reliable posterior and predictive simulation. Multiple examples show the possible relevance of the novel methods, as a substitute for worse candidate distributions in Importance Sampling or the Metropolis-Hastings algorithm, or as a substitute or complement (e.g., as a validity check for estimated posterior moments or marginal likelihoods) for Gibbs sampling.

In this chapter we deal with flexible approximations to posterior distributions of parameter vectors that have a small or moderate dimension (up to 17 parameters), and where each data observation has at most dimension 5. Regarding the application of MitISEM to situations with higher dimensional data, our new approach is obviously no substitute for theoretical a priori restrictions on model structures that may be required in case of higher dimensional data. For example, VAR or multivariate GARCH models for large numbers of time series may still require certain restrictive ‘shrinkage’ priors. The use of the MitISEM approach within a diagnostic procedure to assess the implications of different types of priors is an interesting topic for further research, but clearly beyond the scope of the present chapter.

2.A Appendix

2.A.1 Derivation of the IS-weighted EM algorithm for mixtures of Student-*t* distributions

This appendix provides the derivation of the IS-weighted EM algorithm that is considered in this chapter.

The candidate density $g(\theta)$ is a mixture of H Student- t densities ($h = 1, \dots, H$):

$$g(\theta) = g(\theta|\zeta) = \sum_{h=1}^H \eta_h t_k(\theta|\mu_h, \Sigma_h, \nu_h), \quad (2.A.1)$$

where ζ is the set of coefficients μ_h , scale matrices Σ_h , degrees of freedom ν_h , and mixing probabilities η_h of the k -dimensional Student- t components with density:

$$t_k(\theta|\mu_h, \Sigma_h, \nu_h) = \frac{\Gamma\left(\frac{\nu_h+k}{2}\right)}{\Gamma\left(\frac{\nu_h}{2}\right) (\pi\nu_h)^{k/2} |\Sigma_h|^{-1/2}} \times \left(1 + \frac{(\theta - \mu_h)' \Sigma_h^{-1} (\theta - \mu_h)}{\nu_h}\right)^{-(k+\nu_h)/2}. \quad (2.A.2)$$

Here Σ_h is positive definite, $\nu_h \geq 1$, $\eta_h \geq 0$ and $\sum_{h=1}^H \eta_h = 1$.

In our situation we maximize the *weighted* log-likelihood

$$\frac{1}{N} \sum_{i=1}^N W^i \log g(\theta^i|\zeta)$$

where $g(\cdot|\zeta)$ is the mixture of Student- t densities (2.A.1).

The mixture of Student- t densities (2.A.1) for θ^i is equivalent with the specification

$$\theta^i \sim N(\mu_h, w_h^i \Sigma_h) \quad \text{if} \quad z_h^i = 1,$$

where z^i is a latent H -dimensional vector, indicating from which Student- t component the observation θ^i stems: if θ^i stems from component h , then $z_h^i = 1$, $z_j^i = 0$ for $j \neq h$; $\Pr[z_h = 1] = \eta_h$; w_h^i has the Inverse-Gamma distribution $IG(\nu_h/2, \nu_h/2)$. For a more extensive explanation of this continuous scale mixing representation of (mixtures of) Student- t distributions we refer to Peel and McLachlan (2000). Here we have latent ‘data’ $\tilde{\theta}^i$ ($i = 1, \dots, N$)

$$\tilde{\theta}^i = \{z_h^i, w_h^i | h = 1, \dots, H\}$$

and the so-called data-augmented density is given by

$$\begin{aligned}
\log p(\theta^i, w^i, z^i | \zeta) &= \log p(\theta^i | w^i, z^i, \zeta) + \log p(w^i | \zeta) + \log p(z^i | \zeta) \\
&= \sum_{h=1}^H z_h^i \log \left[\text{pdf}_{N(\mu_h, w_h^i \Sigma_h)}(\theta^i) \right] \\
&\quad + \sum_{h=1}^H \log \text{pdf}_{IG(\nu_h/2, \nu_h/2)}(w_h^i) + \sum_{h=1}^H z_h^i \log(\eta_h) \\
&= \sum_{h=1}^H z_h^i \left\{ -\frac{k}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_h| - \frac{k}{2} \log(w_h^i) \right. \\
&\quad \left. - \frac{1}{2} \frac{(\theta^i - \mu_h)'(\Sigma_h)^{-1}(\theta^i - \mu_h)}{w_h^i} \right\} + \sum_{h=1}^H \left\{ \frac{\nu_h}{2} \log\left(\frac{\nu_h}{2}\right) \right. \\
&\quad \left. - \left(\frac{\nu_h}{2} - 1\right) \log(w_h^i) - \frac{\nu_h}{2} \frac{1}{w_h^i} - \log\left(\Gamma\left(\frac{\nu_h}{2}\right)\right) \right\} \\
&\quad + \sum_{h=1}^H z_h^i \log(\eta_h), \tag{2.A.3}
\end{aligned}$$

where w^i and z^i are *a priori* independent.

The expressions of the latent variables w^i and z^i that appear in terms which also involve the parameters ζ to be optimized are z_h^i , $\frac{z_h^i}{w_h^i}$, $\log w_h^i$, and $\frac{1}{w_h^i}$. Therefore, we derive the conditional expectations of z_h^i , $\frac{z_h^i}{w_h^i}$, $\log w_h^i$, and $\frac{1}{w_h^i}$ given θ^i and $\zeta = \zeta^{(L-1)}$, the optimal parameters in the previous EM iteration:

- (1) **Expectation of z_h^i** : in order to speed up the convergence of the (IS weighted) EM algorithm we compute the expectation

$$\tilde{z}_h^i \equiv E[z_h^i | \theta^i, \zeta = \zeta^{(L-1)}] = \Pr[z_h^i = 1 | \theta^i, \zeta = \zeta^{(L-1)}]$$

not given w_h^i ; that is, w_h^i is integrated out:

$$\begin{aligned}
p(\theta^i, z^i | \zeta) &= \prod_{h=1}^H [p(\theta^i | z_h^i = 1, \zeta) \Pr[z_h^i = 1 | \zeta]]^{z_h^i} \\
&= \prod_{h=1}^H [t(\theta^i | \mu_h, \Sigma_h, \nu_h) \eta_h]^{z_h^i},
\end{aligned}$$

which is a kernel of a probability function of a multinomial distribution for the set of z_h^i ($h = 1, \dots, H$) given θ^i and ζ , with probabilities $\Pr[z_h^i = 1 | \theta^i, \zeta = \zeta^{(L-1)}]$ equal to

$$\tilde{z}_h^i \equiv E[z_h^i | \theta^i, \zeta = \zeta^{(L-1)}] = \frac{t(\theta^i | \mu_h, \Sigma_h, \nu_h) \eta_h}{\sum_{j=1}^J t(\theta^i | \mu_h, \Sigma_j, \eta_j) \eta_j}. \quad (2.A.4)$$

(2) **Expectation of $\frac{z_h^i}{w_h^i}$:**

$$\begin{aligned} \widetilde{z/w_h^i}^i \equiv E \left[z_h^i \frac{1}{w_h^i} \middle| \theta^i, \zeta = \zeta^{(L-1)} \right] &= \Pr[z_h^i = 1 | \theta^i, \zeta = \zeta^{(L-1)}] \times \\ &E \left[\frac{1}{w_h^i} \middle| z_h^i = 1, \theta^i, \zeta = \zeta^{(L-1)} \right]. \end{aligned}$$

Given $z_h^i = 1$, i.e. given that θ^i stems from of Student- t component h , the situation reduces to the case of the EM algorithm for a Student- t distribution without mixtures (see Hu (2005) for an extensive explanation):

$$E \left[\frac{1}{w_h^i} \middle| z_h^i = 1, \theta^i, \zeta \right] = \frac{k + \nu_h}{\rho_h^i + \nu_h}.$$

with

$$\rho_h^i = (\theta^i - \mu_h)' \Sigma_h^{-1} (\theta^i - \mu_h).$$

Therefore we have

$$\widetilde{z/w_h^i}^i \equiv E \left[z_h^i \frac{1}{w_h^i} \middle| \theta^i, \zeta = \zeta^{(L-1)} \right] = \tilde{z}_h^i \frac{k + \nu_h}{\rho_h^i + \nu_h}. \quad (2.A.5)$$

(3) **Expectation of $\log w_h^i$:**

$$\begin{aligned} \xi_h^i &\equiv E [\log w_h^i | \theta^i, \zeta = \zeta^{(L-1)}] = \\ &= E [\log w_h^i | z_h = 1, \theta^i, \zeta = \zeta^{(L-1)}] \Pr[z_h^i = 1 | \theta^i, \zeta = \zeta^{(L-1)}] \\ &\quad + E [\log w_h^i | z_h^i = 0, \theta^i, \zeta = \zeta^{(L-1)}] \Pr[z_h^i = 0 | \theta^i, \zeta = \zeta^{(L-1)}] \\ &= \left[\log \left(\frac{\rho_h^i + \nu_h}{2} \right) - \psi \left(\frac{k + \nu_h}{2} \right) \right] \tilde{z}_h^i \\ &\quad + \left[\log \left(\frac{\nu_h}{2} \right) - \psi \left(\frac{\nu_h}{2} \right) \right] (1 - \tilde{z}_h^i), \end{aligned} \quad (2.A.6)$$

where $\psi(\cdot)$ is the digamma function (the derivative of the logarithm of the gamma function $\log \Gamma(\cdot)$), and where we again used that given $z_h = 1$ the situation reduces to the case of the EM algorithm for a Student- t distribution without mixtures (see Hu (2005) for an extensive explanation). For $z_h^i = 0$, the conditional distribution of w_h^i given θ^i, ζ is the distribution given only ζ (since the observation θ^i does not depend on w_h^i) which is Inverse-Gamma $IG(\nu_h/2, \nu_h/2)$:

$$E[\log w_h^i | z_h^i = 0, \theta^i, \zeta = \zeta^{(L-1)}] = \log\left(\frac{\nu_h}{2}\right) - \psi\left(\frac{\nu_h}{2}\right).$$

(4) **Expectation of $\frac{1}{w_h^i}$:**

$$\begin{aligned} \delta_h^i &\equiv E\left[\frac{1}{w_h^i} \middle| \theta^i, \zeta = \zeta^{(L-1)}\right] \\ &= E\left[\frac{1}{w_h^i} \middle| z_h^i = 1, \theta^i, \zeta = \zeta^{(L-1)}\right] \Pr[z_h^i = 1 | \theta^i, \zeta = \zeta^{(L-1)}] \\ &\quad + E\left[\frac{1}{w_h^i} \middle| z_h^i = 0, \theta^i, \zeta = \zeta^{(L-1)}\right] \Pr[z_h^i = 0 | \theta^i, \zeta = \zeta^{(L-1)}] \\ &= \frac{k + \nu_h}{\rho_h^i + \nu_h} \tilde{z}_h^i + (1 - \tilde{z}_h^i). \end{aligned} \tag{2.A.7}$$

where if $z_h^i = 0$, $1/w_h^i$ has the $\text{Gamma}(\nu_h/2, \nu_h/2)$ distribution with

$$E[1/w_h^i | z_h^i = 0, \theta^i, \zeta = \zeta^{(L-1)}] = 1.$$

Define $\log \tilde{p}(\theta^i, w^i, z^i | \zeta)$ as the result of substituting the expectations (2.A.4)-(2.A.7) into $\log p(\theta^i, w^i, z^i | \zeta)$ in (2.A.3). The Maximization step amounts to computing the ζ that maximizes

$$\zeta^{(L)} = \arg \max_{\zeta} \frac{1}{N} \sum_{i=1}^N W^i \log \tilde{p}(\theta^i, w^i, z^i | \zeta).$$

Using the analogy with Maximum Likelihood estimation for the Seemingly Unrelated Regression model with Gaussian errors (for the k elements of θ^i) and the same ‘regressor’ (a constant term) in each equation, in which case the Ordinary Least Squares (OLS) estimator provides the Maximum Likelihood Estimator, and with Maximum Likelihood

estimation for the multinomial distribution, it is easily derived that $\zeta^{(L)}$ consists of:

$$\mu_h^{(L)} = \left[\sum_{i=1}^N W^i \widetilde{z/w_h^i} \right]^{-1} \left[\sum_{i=1}^N W^i \widetilde{z/w_h^i} \theta^i \right], \quad (2.A.8)$$

$$\hat{\Sigma}_h^{(L)} = \frac{\sum_{i=1}^N W^i \widetilde{z/w_h^i} (\theta^i - \mu_h^{(L)}) (\theta^i - \mu_h^{(L)})'}{\sum_{i=1}^N W^i \widetilde{z_h^i}}, \quad (2.A.9)$$

$$\eta_h^{(L)} = \frac{\sum_{i=1}^N W^i \widetilde{z_h^i}}{\sum_{i=1}^N W^i}. \quad (2.A.10)$$

Further, $\nu_h^{(L)}$ is solved from the first order condition of ν_h :

$$-\psi(\nu_h/2) + \log(\nu_h/2) + 1 - \frac{\sum_{i=1}^N W^i \xi_h^i}{\sum_{i=1}^N W^i} - \frac{\sum_{i=1}^N W^i \delta_h^i}{\sum_{i=1}^N W^i} = 0 \quad (2.A.11)$$

using a procedure for one-dimensional root finding.

2.A.2 Two alternative adaptive simulation methods

Adaptive Metropolis (AM) of Roberts and Rosenthal (2009)

The Adaptive Metropolis (AM) sampler is proposed on page 3 of Roberts and Rosenthal (2009). It is a version of the AM algorithm of Haario *et al.* (2001). Defining V as the covariance matrix of the Laplace approximation to the posterior at the posterior mode, at iteration j the proposal distribution is given by:

$$q_j(\theta^c, \cdot) = N(\theta^c, (0.1)^2 V/k) \quad \text{if } j < 5k, \quad (2.A.12)$$

$$q_j(\theta^c, \cdot) = (1 - \beta) N(\theta^c, (2.38)^2 \Sigma_j/k) + \beta N(\theta^c, (0.1)^2 V/k) \quad \text{if } j \geq 5k,$$

with θ^c the current value of θ , k the dimension of θ , $\beta = 0.05$ and Σ_j the current empirical estimate of the covariance matrix of the target distribution based on the iterations thus far. The scalar 0.1 tries to achieve a high acceptance rate by moving the sampler locally. From previous literature, (see Roberts and Rosenthal, 2001) it is known that the proposal $N(\theta^c, (2.38)^2 \Sigma/k)$ is optimal in a particular large-dimensional context. In the original setting, Roberts and Rosenthal (2009) propose I_k instead of the covariance matrix V when $j < 5k$. Here we follow Giordani and Kohn (2010) and replace the identity matrix by V .

Adaptive Independence Metropolis-Hastings (AIMH) sampler of Giordani and Kohn (2010)

Giordani and Kohn (2010) propose a mixture with four terms as candidate density in their adaptive independent Metropolis-Hastings approach. Starting with the general form, the candidate density at iteration j is given by

$$q_j(\theta, \lambda_n) = \omega_1 g_0(\theta) + (1 - \omega_1) g_j(\theta; \lambda_j), \quad (2.A.13)$$

where λ_j denotes a parameter vector that evolves over time. The density $g_0(\theta)$ is constant and given by a mixture of the form $g_0(\theta) = 0.6 \phi_0(\theta) + 0.4 \tilde{\phi}_0(\theta)$ where $\phi_0(\theta)$ is a mixture of normals, initialized at iteration $j = 1$ by a Laplace expansion in which case it is a multivariate normal. The density $\tilde{\phi}_0(\theta)$ is a mixture of normals with similar parameters as $\phi_0(\theta)$, however the covariance matrices are multiplied by a factor k_1 . Omitting the parameter vector λ_j , the density $g_j(\theta)$ is given by:

$$g_j(\theta) = \omega_2^* g_j^*(\theta) + (1 - \omega_2^*) \tilde{g}_j^*(\theta), \quad (2.A.14)$$

where both $g_j^*(\theta)$ and $\tilde{g}_j^*(\theta)$ are mixtures of normals with again the same parameters, except that the covariance matrices of the last density are multiplied by a scalar k_2 . These parameters are estimated by the k -harmonic means clustering algorithm, see Giordani and Kohn (2009) for a discussion of this k -harmonic means clustering algorithm.

Chapter 3

On the effects of private information on volatility

Joint work with Michel van der Wel, Dick van Dijk and Nick Taylor.

3.1 Introduction

One of the most fundamental questions in financial economics is what drives asset prices and volatility. Both quantities are believed to change due to the arrival of new information. In this context it is useful to distinguish between public and private information. Public information concerns news that becomes available to all market participants at the same point in time, for example in the form of announcements of important macroeconomic variables. As all investors are equally informed at the same time, the arrival of public information typically causes an immediate change in asset prices and volatility. Private information usually refers to news that is distributed asymmetrically among market participants. The presence of such private information may be revealed by the trading process and changes in the asset price itself, as informed investors may buy or sell the asset based on their privately held knowledge. Other investors observe the trading process and make inferences on this private information, giving rise to further price adjustments.¹

¹This idea was first formalized by Kyle (1985) and Glosten and Milgrom (1985).

The effect of public information on prices and volatility and the effect of private information on prices have been established for many financial markets.² Andersen *et al.* (2007), for example, document the link between public information releases and prices and volatility in foreign exchange, Treasury and equity markets. Based on high-frequency intraday data they demonstrate that surprises in macro announcements (i.e. the difference between the actual release and the consensus market expectation) affect the conditional mean and volatility of exchange rates, Treasuries and stocks. Recent literature also documents a relation between private information and prices for equity (Hasbrouck, 1991; Albuquerque *et al.*, 2009; Albuquerque and Vega, 2009), foreign exchange (Evans and Lyons, 2008), and the Treasury market (Brandt and Kavajecz, 2004; Green, 2004; Pasquariello and Vega, 2007; Li *et al.*, 2009; Menkveld *et al.*, 2012).

The goal of this chapter is to examine whether private information also influences the volatility of assets. We investigate this issue empirically for the 30-year U.S. Treasury bond futures, using high-frequency (5 minutes) intraday data for the period 2004-2009. We conduct our analysis in a comprehensive modeling framework that allows to control for the effects of both public and private information on prices and the effects of public information on volatility. We proxy private information by order flow, which is defined as the difference between the volume in buyer-initiated transactions and the volume in seller-initiated transactions.

Our results demonstrate that this private information proxy indeed affects volatility. Specifically, a higher level of private information increases the uncertainty surrounding Treasury futures prices. This positive relation between private information and volatility is significant in both statistical and economic terms. For example, the average effect of a one standard deviation shock in order flow on volatility is around 2.7 basis points in fifteen minutes. This effect is about one-fourth of the economic significance of Nonfarm Payroll Employment, the most influential announcement in our analysis. The evident implication of our findings is that the level of private information should be taken into account when modeling volatility.

²This literature dates back to (at least) French and Roll (1986) and Cutler, Poterba and Summers (1989), and includes important contributions by Ederington and Lee (1993), Berry and Howe (1992), Fleming and Remolona (1999), Balduzzi *et al.* (2001), Boyd *et al.* (2005), Andersen, Bollerslev, Diebold and Vega (2003); Andersen *et al.* (2007), Faust *et al.* (2007), Bartolini *et al.* (2008), and Brenner *et al.* (2009).

We use a high-frequency data set consisting of transaction prices and volumes for the 30-year U.S. Treasury bond futures. The Treasury futures market is highly liquid and generates an average monthly trading volume of 5.2 million contracts (based on 2009 data). The 30-year U.S. Treasury bond futures is among the most actively traded long-term interest rate contracts in the world. Trading in Treasuries with a maturity of 30 years takes place almost exclusively in the futures market, hence we focus on this venue. Modeling returns and volatility of Treasuries has the advantage that public information is aptly captured by macroeconomic news releases and thus easily tracked and measured. The availability of analysts' forecasts for each macro announcement provides an estimate of the market consensus, which in turn can be used to construct the surprise component of each news release.

Two ideas need some further clarification. First, the use of order flow (also sometimes referred to as signed volume) to proxy private information. As mentioned earlier, order flow measures the difference between buyer-initiated and seller-initiated trade volume. Trades get initiated by investors wanting to take (or close) a position immediately. The initiators have to pay a premium to traders willing to facilitate the need for such immediacy. The idea is best illustrated in a limit order market setting, where trade initiators would submit orders meeting or crossing prices of previously submitted limit orders and the providers of immediacy are the traders that submitted these previous limit orders. The aforementioned literature on private information and prices finds that a reason for paying for such immediacy is trading due to (short-lived) informational advantages.³ In this literature, order flow is found to be significant in explaining returns. This significance is interpreted as order flow transmitting information relevant to prices to the market, in particular information that is held asymmetrically among market participants. A large order flow (in absolute sense) thus coincides with a period of high information asymmetry and thus a high level of private information.

A second idea worth clarifying further is how private information in the Treasury market can be thought of. The conventional interpretation of private information, as typically applied to individual stocks and corporate bonds, is advance knowledge of firm-related

³Based on this idea, Sarkar and Schwartz (2009) develop a measure of 'market sidedness' that provides insights into the motives for trade initiation. We will use this measure in our robustness section as alternative private information proxy.

news, concerning earnings announcements, new investment projects, or changes in management, among others. Obviously, this interpretation does not straightforwardly apply to Treasuries. Green (2004) provides several alternatives for the interpretation of private information in the Treasury market, such as information about endowments or the interpretation of macroeconomic news due to differential information processing skills. Despite the different interpretation, we can use similar variables to proxy private information in the Treasury market as commonly used for equities, such as order flow.

Apart from uncovering the effect of private information on volatility, our study also contributes to the methodology for analyzing information effects on asset prices and volatility. We develop a modeling framework that allows us to simultaneously assess the effects of public and private information on returns and volatility. Specifically, we propose a model specifying the dynamics of both returns and volatility in such a way that both equations can be estimated jointly. We split volatility in two (multiplicative) components, following the Spline-GARCH model of Engle and Rangel (2008). One component describes the effects of private and public information, while the other captures short-run GARCH-type behavior. This joint modeling approach extends and improves upon the specification of Andersen, Bollerslev, Diebold and Vega (2003); Andersen *et al.* (2007). In their set-up, equations for return and volatility effects are estimated separately. Thus, parameter uncertainty of the return equation is neglected when estimating the volatility equation and, consequently, the parameters of both equations are not estimated efficiently. In addition, the approach could suffer from negative fitted values of the time-varying conditional volatility. In our framework, both return and volatility equations are estimated jointly by means of (quasi) maximum likelihood, such that the parameters are estimated efficiently. Also, by construction, our model automatically avoids negative values of the volatility.

We provide several robustness checks and extensions of our main result that private information affects volatility. First, we use the bid-ask spread, the volume of trades and the illiquidity ratio of Amihud (2002) to control for liquidity effects. This does not change the main result that private information matters for volatility. Second, we strengthen our main result as we re-estimate the model on a daily basis and include the market sidedness of Sarkar and Schwartz (2009) as an additional measure of private information in

our framework. Both measures of private information are significantly related to volatility. Third, we document that the effect of private information is stronger in crises periods compared to normal times. Finally, we find that the effect of private information on volatility depends on the heterogeneity of analysts' expectations concerning upcoming public information releases. Order flow influences volatility to a greater extent in times characterized by a high level of dispersion of beliefs.

The remainder of this chapter is organized as follows. Section 3.2 discusses related literature. Section 3.3 describes the data and presents summary statistics. Section 3.4 develops the joint modeling framework for analyzing the effects of public and private information on returns and volatility and provides the main empirical results. Section 3.5 reports extensions of the main analysis, including the control for liquidity effects on volatility, considering market sidedness as an additional measure of private information at the daily level and allowing for the effects of public and private information to depend on whether the period is characterized by a crisis or on the level of disagreement among analysts concerning upcoming public information releases. Section 3.6 concludes.

3.2 Related literature

Only a few papers study issues related to private information and volatility. The paper closest to this chapter is Berger *et al.* (2009), who take a long term perspective and find volatility variation is directly related to information flow and the impact of that information flow on prices for the foreign exchange market. In their analysis they use order flow to measure information flow. We follow suit but label order flow as a measure for private information. We do so as in addition we consider public information to explicitly contrast the two and determine the relative contribution of each.⁴ Berger *et al.* (2009) consider daily volatility, measured by the realized variance being the sum of squared returns over all 5-minute intervals in the day. We differ by focusing on the raw higher-frequency data at the 5-minute level, and jointly modeling return and volatility to ensure no return ef-

⁴In footnote 18 Berger *et al.* (2009) point out that the largest residuals remaining after modeling information flow are due to important macro news announcements. This finding highlights the importance of incorporating macro news announcements in the model as we do. In fact, it may be the case that part of the effect Berger *et al.* (2009) attribute to information flow is actually due to macro news releases as literature shows order flow is more informative around such releases (Green, 2004).

fects are attributed to volatility (and vice versa). In addition, we consider an alternative measure of market sidedness at the daily level instead of only order flow.

Related contributions identify the effects of private information effects on volatility only through its effects on returns. Evans and Lyons (2008) use a variance decomposition of the conditional mean of foreign exchange returns, which implies that they consider a link between unconditional volatility and order flow. They show that the arrival of macro news can account for more than 30% of the daily variance of exchange rates, a number that includes the impact through order flow. Likewise, He *et al.* (2009) start with a structural model for price changes due to public information shocks and information asymmetry and obtain a model-implied variance of price changes in the U.S. Treasury market. In contrast to these two papers, in our work the term volatility implies a parameterization of the conditional volatility. Our framework is thus in essence a more general set-up of the aforementioned papers that identify the private information effects on volatility solely through the effects on the returns. We do not rely on such identification, and estimate the impact of public and private information on return and on volatility separately. Within this conditional context, other studies have considered alternate measures of private information flow. For instance, Jiang and Lo (2011) use a Markov-Switching framework to identify the probability of private information flow (also from the return equation), and find that volatility is high when this probability is high. Jones *et al.* (1994) identify private and public information by comparing periods with and without trading.

Several strands of literature touch related subjects. First, Chordia *et al.* (2005) analyze the comovement of returns, volatility, liquidity and order flow. One of their main findings is that volatility shocks are informative for predicting shifts in liquidity, highlighting the interrelation between the two. We extend their work by investigating specifically the role of private information on volatility. We take order flow as the private information measure to be able to rely on economic theory that predicts causality from order flow to prices (as in Kyle (1985) and Glosten and Milgrom (1985)). Second, our analysis of private information effects on volatility relates to the literature on the effects of buying- and selling-pressure in options markets (Bollen and Whaley, 2004; Garleanu *et al.*, 2009). The focus in this literature is typically on implied volatility rather than option prices directly,

as a way of controlling for price differences due to option contract specification (such as strike and maturity).

Obviously there is a relation between order flow and volume. Order flow is defined as the difference between buyer-initiated and seller-initiated trades, while volume is the sum of all trades. Literature on the relation between volume and volatility dates back to Clark (1973) and Tauchen and Pitts (1983), who develop the mixture of distribution hypothesis for the joint movement of return and volume based on an unobserved information arrival variable. Andersen (1996) modifies this framework to include both information asymmetry and liquidity needs, and shows it is useful for analyzing the economic factors behind volatility clustering. Further empirical evidence on the volume-volatility relation is provided by Karpoff (1987), who also surveys the literature. For our purpose it is important to realize that private information is measured not by the total amount of trading that takes place, but by how this volume can be attributed to either buying- or selling-pressure. In our robustness analysis we control for volume to examine the differential impact of the two on volatility.

Finally, there is also a link between private information and liquidity. For example, Green (2004) finds that information asymmetry increases when liquidity is high while Brandt and Kavajecz (2004) find the opposite. In addition, Jiang and Lo (2011) argue that heterogeneous private information is followed by low trading volume, low total market depth and hidden depth. If we only include (unanticipated) order flow in our volatility specification, we discard the effect of liquidity (shocks) on volatility. Jiang and Lo (2011) control for trading volume and the bid-ask spread to differentiate between private information and liquidity. We follow their approach and include beyond volume the bid-ask spread and the illiquidity ratio of Amihud (2002) as control variables in our analysis.

3.3 Data and summary statistics

We combine two data sets to study the public and private information sources of volatility. In the following three subsections we discuss the 30-year U.S. Treasury bond futures data, the variables that we use as proxies for the presence of private information in the market, and the public information variables as constructed from expectations and announcements

of macroeconomic fundamentals. In the final subsection, we provide an illustration for the effects of public and private information on Treasury futures returns and volatility, focusing on the announcements of Nonfarm Payroll Employment.

3.3.1 U.S. Treasury bond futures data

We employ a high-frequency data set of intraday transaction prices and volumes of 30-year U.S. Treasury bond futures contracts over the period from January 1, 2004 until December 31, 2009. The contract initially trades on the Chicago Board of Trade (CBOT), and after the merger of the Chicago Mercantile Exchange (CME) with the CBOT in 2007 on the CME. During our sample period trading takes place both on the trading floor and electronically. Floor trading occurs in a pit through open-outcry from 8:20 a.m. to 3:00 p.m. Eastern Time (ET). Electronic trading occurs through Globex from 6.30 p.m. to 5.00 p.m. from Sunday to Friday. Since volume has gradually shifted from pit trading to electronic trading we focus on the volume generated through the electronic venue. In addition, we restrict ourselves to day trading (that is, from 8:00 a.m. to 5:00 p.m.) as this is where volume concentrates. The 30-year Treasury futures trade in the March quarterly cycle, that is, contracts mature in March, June, September or December. At each point in time the next three consecutive contracts can be traded.⁵ We construct a single time series of transaction prices and volumes using the most nearby contract, which is the most intensely traded and is a close substitute for the underlying spot instrument. We roll over to the next contract when its daily tick volume exceeds the daily tick volume of the most nearby contract. This generally occurs between five to three days before expiration of the nearest-to-maturity contract. Our data set, obtained from Tickdata Inc., records the timestamp (in seconds), price and volume for each transaction. We aggregate the data to 5-minute intervals.⁶ For computing returns we use the last observed transaction price in each intraday interval. Overnight returns are excluded.

⁵The last trading day of a given contract is the seventh business day preceding the last business day of the delivery month.

⁶As a robustness check, we repeat our complete empirical analysis with 15-minute intervals. The main results do not change.

3.3.2 Private information variables

It is generally not possible to directly measure private information. In our study we use two variables to proxy private information, namely order flow and the market sidedness measure of Sarkar and Schwartz (2009). We use the first one for our main analysis, while the second proxy is used in the robustness section.

Order flow, or net buying pressure, is defined as the difference between the volume in trades initiated by a buyer and the volume in trades initiated by a seller during a certain interval. When there is a large positive order flow this could indicate investors are initiating trades based on having (the belief that) private information that indicates that the price of the asset is relatively low, and vice versa. As such a large (positive or negative) value may mark a situation with high information asymmetry among participants. Taking N_t as the number of trades in interval t and $v_{t,j}$ as the volume of the j -th trade in this interval, we first calculate our measure of order flow, denoted OF_t , as

$$OF_t = \sum_{j=1}^{N_t} q_{t,j} v_{t,j}, \quad (3.1)$$

where $q_{t,j} = 1$ if the j -th trade is initiated by a buyer and -1 if it is initiated by a seller. We scale order flow by its empirical standard deviation. Related to the definition of order flow OF_t as given above, this measure is also referred to as the ‘signed’ trading volume. We follow Pasquariello and Vega (2007) by using the unanticipated portion of aggregate order flow. In practice, we estimate an appropriate ARMA model to clean order flow from autocorrelations that are (partly) due to microstructure imperfections causing lagged effects in the observed order flow, see Hasbrouck (2004a).⁷ The residuals of this model represent the aggregate unanticipated order flow OF_t^* over each time interval t .⁸ While our empirical analysis is based on unanticipated order flow, we refer to order flow and unanticipated order flow interchangeably.

Our data set does not include the ‘sign’ variable $q_{t,j}$, which indicates whether a trade is initiated by the buying or selling party in (3.1). Hasbrouck (2004b) proposes to estimate

⁷An AR(1) filter was sufficient to capture the autocorrelation in order flow.

⁸We perform the analysis also by using order flow as defined in (3.1). This does not affect the main results.

the sign following sequential trade models such as Roll (1984). In these models the transaction price series is decomposed into an unobserved efficient price and a deviation from this efficient price which is equal to the (half-)spread times the trade sign. The trade sign is thus also an unobserved series that can be filtered from the data. This filtering can be done through Monte Carlo techniques (as Hasbrouck (2004b) proposes), or by state space estimation (Van der Wel *et al.*, 2009). To apply this method on our data set consisting of on average 15,562 daily transactions over 6 years, we use the latter approach.⁹

Our second measure for private information is the market sidedness, introduced by Sarkar and Schwartz (2009). This measure is estimated by the correlation between the number of buyer-initiated trades and seller-initiated trades in brief intervals. Sarkar and Schwartz (2009) use the market sidedness measure to provide insights into the motives for trading. They distinguish two possibilities: trading when some investors are better informed than others, and trading when investors have different information or interpret the same information differently. In the first case trading is more one-sided (a decrease in the correlation between the number of buyer- and seller-initiated trades), while in the second case trading is more two-sided (an increase in the correlation). As the second case coincides with the definition of private information in the treasury market on being about endowments or the interpretation of macroeconomic news (see e.g. Green (2004) and our discussion in the introduction), we use the sidedness measure as alternative proxy for private information.¹⁰ We do so in a robustness analysis based on data at the daily level, as the sidedness measure cannot be straightforwardly constructed at higher-frequencies.

Table 3.1 shows summary statistics of the data. All numbers are based on aggregated 5-minute data for the full sample period 2004-2009. The data set covers 1,420 full days with observations from 8:00 a.m. to 5:00 p.m. and 69 ‘half days’ with observations from

⁹For robustness, we also use the tick-test (Lee and Ready, 1991), which determines the sign based on the relation of the current trade price to the previous trade price. If a trade occurs at a price higher (lower) than the previous trade a trade is called an up-tick (down-tick), and is assumed to be initiated by the buying (selling) party. The main results do not change when we use order flow based on the tick-test.

¹⁰This definition of private information is broader than that of Sarkar and Schwartz (2009), as we focus on a treasury market setting and contrast public information (information available to everybody) with private information (information distributed asymmetrically among market participants). Sarkar and Schwartz (2009) further disentangle private information as being either some investors having *better* information or simply investors having *differential* information. They label the former category information asymmetry, in the traditional equity market sense where private information is about possible inside information (surrounding mergers). As such situation is unlikely in a Treasury market setting we do not make this further distinction.

Table 3.1: Summary statistics

This table shows the mean, standard deviation, skewness, and kurtosis, as well as the minimum and maximum values, over the full 2004-2009 sample period for the variables in our analysis. We show return (in %, denoted with R_t) which is defined as 100 times the difference between the logarithm of the last observed transaction price in each intraday interval, the total number of trades in each interval (N_t), order flow (OF_t) with unanticipated order flow (OF_t^*), the total trading volume of the trades in each interval (Qnt_t), and the spread (Spr_t) computed as the difference between the volume weighted buy price and the average sell price during each intraday interval. All numbers are based on 157,500 aggregated 5-minute observations, the sample goes from January 2, 2004 through December 31, 2009.

		Mean	St Dev	Skew	Kurt	Min	Max
Return (in %)	R_t	0.00	0.058	0.599	118.0	-2.21	3.37
Number of transactions	N_t	144	157	3.527	32.3	0	3,924
Order flow	OF_t	-1	1,249	-0.039	11.7	-16,754	16,351
Unanticipated order flow	OF_t^*	0	1,249	-0.041	11.6	-16,750	16,385
Trading volume	Qnt_t	2,335	2,578	3.362	24.1	0	63,924
Spread (in \$)	Spr_t	0.0186	0.0170	1.660	15.7	0	0.3250

8:00 until 1:00 p.m. The latter occurs when the market is partly closed in the afternoon, for example when the preceding day was Thanksgiving. Mean futures returns are small or zero, as is expected for the 5-minute observation frequency. The volatility is 5.8 basis points (bps) at the 5-minute level, which roughly translates into 53 bps daily volatility. The table also shows the high activity of the 30-year futures market as there are on average 144 trades in each 5-minute interval. In addition, the difference in the standard deviations of order flow and unanticipated order flow is rather small, showing that the AR(1) coefficient is close to zero or, in other words, the autocorrelation in order flow is rather limited.

3.3.3 Public information variables

An advantage of studying the Treasury market is that public information affecting the prices and volatility is readily identified: these are the scheduled releases of macroeconomic variables. We use data on expectations and announcements of 24 key U.S. macro variables. These data are obtained from Econoday, and are derived from information that is published on Bloomberg. The data on the actual released values comes from Haver Analytics. We derive our proxy for the market expectation from analysts' forecasts, which stem from Market News International and Thomson Financial. To obtain the analysts' forecasts the aforementioned companies hold a survey among a number of analysts for the announcements that are to come in the following week. The data set records the me-

dian of these forecasts, which provides an estimate of the market consensus. In addition, we have the lowest and highest analysts' forecasts for each announcement from June 2007 onwards.

We follow the existing literature by considering the surprise in each announcement. This surprise is constructed as the difference between the actual released value and the consensus analysts' forecast. Since units of measurement vary widely across macroeconomic variables, we standardize the surprises by dividing by their sample standard deviation. Hence the surprise $S_{k,t}$ in the announcement of variable k at time t is

$$S_{k,t} = \frac{A_{k,t} - M_{k,t}}{\sigma_k}, \quad (3.2)$$

where $A_{k,t}$ denotes the announced value, $M_{k,t}$ is the median of the analysts forecasts, and σ_k is the sample standard deviation of their difference.

Table 3.2 lists the 24 macroeconomic variables in our data set.¹¹ Most of these are released at a monthly frequency. The exceptions are GDP, which is announced quarterly, and initial claims for unemployment insurance, which is released on a weekly basis. We have in total 1,825 announcements during the sample period 2004-2009. Out of the 1,489 days that are covered by our dataset, 1,000 days contain at least one announcement. The actual number of surprises that we use in our analysis is somewhat smaller than the total number of releases as the consensus forecast is unavailable for some announcements or the market is (partly) closed on the day of the announcement. On average, for each (monthly) macro variable we are able to use 70 of the 72 announcements. For the later part of the sample, since June 2007, we also have the minimum and maximum analysts' forecasts available. We will use this information to calculate a dispersion measure that we use in one of the extensions in Section 3.5. Table 3.2 shows for each variable the number of minimum and maximum forecasts that are available.

Table 3.2: Macroeconomic announcements

This table describes scheduled macroeconomic announcements from 2004 to 2009. The data is from Econoday, the table is modeled after Anderson, Bollerslev, Diebold and Vega (2003, Table 1, p.43). The number of surprises is equal to the total number of announcements over the sample period minus missing consensus and missing trading days.

	Time (ET)	Number of announcements	Surprises in our sample	Number of dispersions ^a
<i>Quarterly</i>				
1. GDP advance	8:30 a.m.	24	24	10
2. GDP preliminary	8:30 a.m.	24	24	11
3. GDP final	8:30 a.m.	24	24	12
<i>Monthly</i>				
Real Activity				
4. Nonfarm payroll employment	8.30 a.m.	72	69	31
5. Retail sales	8.30 a.m.	72	70	31
6. Industrial production	9.15 a.m.	72	69	31
7. Capacity utilization	9.15 a.m.	72	69	31
8. Personal income	8.30 a.m.	72	71	33
9. Consumer credit	3.00 p.m.	72	63	31
Consumption				
10. Personal consumption exp.	8.30 a.m.	72	71	33
11. New home sales	10.00 a.m.	72	69	32
Investment				
12. Durable goods orders	8.30 a.m.	72	70	32
13. Construction spending	10.00 a.m.	72	72	31
14. Factory orders	10.00 a.m.	72	72	31
15. Business inventories	10.00 a.m. ^b	72	69	30
Government Purchases				
16. Government budget deficit	2.00 p.m. ^c	72	65	29
Net Exports				
17. Trade balance	8.30 a.m.	72	71	31
Prices				
18. Producer price index	8.30 a.m.	72	69	31
19. Consumer price index	8.30 a.m.	72	71	32
Forward-looking				
20. Consumer confidence index	10.00 a.m.	72	72	33
21. NAPM index	10.00 a.m. ^d	72	72	33
22. Housing starts	8.30 a.m.	72	71	30
23. Index of leading indicators	10.00 a.m.	73	68	32
<i>Weekly</i>				
24. Initial unemployment claims	8.30 a.m.	312	304	135

^a Dispersions available only after June 2007.

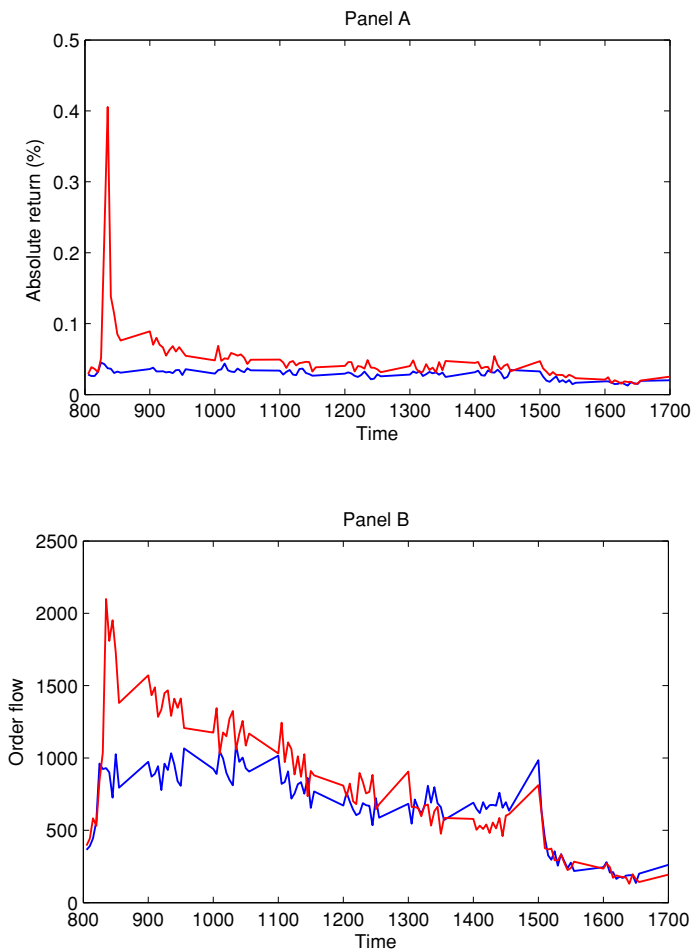
^b Earlier part of sample often at 8:30.

^c Announcement at 1:30 on October 14, 2005; at 1:00 on October 14, 2008 and at 3:30 on October 16, 2009.

^d Latest part of sample at 9:45 (since January 2007).

Figure 3.1: Order flow and volatility during announcement and non-announcement days

This figure shows time series of the volatility and order flow of 30-year U.S. Treasury bond futures on announcement days and non-announcement days. Each series is obtained by the average over 69 trading days where no announcements were made (blue line), and 69 trading days when an 8:30 Nonfarm Payroll Announcement was made (red line). Panel A denotes the intraday volatility, measures by the absolute returns. Panel B depicts the unanticipated order flow on each 5-minute interval.



3.3.4 Public and private information effects: A first impression

Figure 3.1 gives a first impression of the effect of public and private information on volatility. The chart considers patterns in the volatility and order flow during all (69) days with scheduled Nonfarm Payroll Employment announcements at 8:30 during our sample period. For this purpose, we compute the average of these variables across these announcement days (where we use the absolute value of order flow) as well as corresponding averages across 69 randomly selected trading days without any macro announcement. Figure 3.1 depicts the average volatility and order flow during these announcement and non-announcement days. The graphs provide an important insight. Volatility spikes during the interval after the announcement time, reflecting the initial reaction of the market to the announcement. A spike immediately after the announcement is also seen for unanticipated order flow, though less pronounced. After the announcement, volatility does not immediately revert to its pre-announcement level. Similarly, unanticipated order flow also remains at a higher level for some time after the announcement. It is this pattern in volatility and order flow that encourages our research question whether and how private information affects volatility.

3.4 Modeling the response of Treasury futures to public and private information

In this section we describe the methodology that we adopt to examine the effects of private information on the volatility of the 30-year Treasury bond futures and provide our main results. We recognize that it is crucial to control for the effects of public and private information on returns as well as the effects of public information on volatility. To accomplish this, we propose a model specifying the dynamics of both returns and volatility in such a way that these can be estimated jointly. We develop our general modeling framework in different stages. First, we consider the model specification for 5-minute intraday returns. Second, we augment this with a specification for (conditional) volatility.

¹¹We do not include FOMC announcements of the federal funds target rate since the surprises for this variable are mostly equal to zero during our sample period.

3.4.1 Public and private information matter for returns

We adopt the approach of Andersen, Bollerslev, Diebold and Vega (2003); Andersen *et al.* (2007), ABDV hereafter. We assume a linear model specification for the 5-minute returns on the 30-year Treasury bond futures, including I autoregressive terms and J lags of the announcement surprises of each of the K macroeconomic fundamentals. In addition, we include unexpected order flow as the proxy for private information, such that the model reads

$$R_t = \beta_0 + \beta_a D_a + \sum_{i=1}^I \beta_i R_{t-i} + \sum_{k=1}^K \sum_{j=0}^J \beta_{kj} S_{k,t-j} + \beta_{OF} OF_t^* + \varepsilon_t, \quad (3.3)$$

for $t = 1, \dots, T$, where R_t is the 5-minute return from time $t - 1$ to time t , $S_{k,t}$ is the standardized surprise of announcement k at time t as defined in (3.2) and D_a is a dummy variable that equals one on announcement days. The disturbance term ε_t in (3.3) is likely to be heteroscedastic. We subsequently specify a separate model equation for the conditional volatility of ε_t , which allows us to examine the effects of public and private information on the Treasury bond futures volatility explicitly. First, however, we document the effects on returns by estimating (3.3) only, accounting for the heteroscedasticity (and any remaining serial correlation not captured by the autoregressive terms) by using Newey-West standard errors.

Table 3.3 reports results based on estimating the model in (3.3) using the full sample period from January 1, 2004 to December 31, 2009. We estimate the model both with and without order flow. We choose the number of lags of the explanatory variables by means of the Akaike and Schwarz information criteria. Both criteria suggest $I = 3$ and $J = 0$. Note that, while the macroeconomic surprises are only included contemporaneously with the return during the first 5-minutes following the announcements, they do affect subsequent returns through the autoregressive terms.

In line with earlier research (e.g., ABDV 2007, Table 5A) we find that many of the fundamentals have a significant effect on Treasury bond futures returns. The magnitude and significance of the coefficients are not affected by the inclusion of the private information variable. We explain the reaction of the bond market to macroeconomic news mainly in terms of revisions of inflationary expectations, which is in accordance with commentaries

Table 3.3: Effect of public and private information on returns

This table reports the estimation results of the following regression model:

$$R_t = \beta_0 + \beta_a D_a + \sum_{i=1}^I \beta_i R_{t-i} + \sum_{k=1}^K \beta_k S_{k,t} + \beta_{OF} OF_t^* + \varepsilon_t,$$

where R_t is the 5-minute log return of 30-year U.S. Treasury bond futures from time period $t-1$ to t , $S_{k,t}$ is the standardized surprise of announcement type k , $k = 1, \dots, 24$ and OF_t^* denotes the unanticipated order flow from period $t-1$ to t . We include a dummy variable that equals one on announcement days. The first column presents results of the regression where only public information is included. The second column provide results of the regression that contains both public and private information. The superscripts ***, ** and * indicate significance at the 1%, 5% and 10% level respectively, where the significance is assessed using Newey-West standard errors. The sample goes from January 2, 2004 through December 31, 2009.

Dependent variable: 30-year futures returns		
	(1)	(2)
Private Information		
Order flow		0.025***
Public Information		
1. GDP advance	-0.069	-0.056
2. GDP preliminary	-0.036	-0.030
3. GDP final	0.008	0.002
Real Activity		
4. Nonfarm payroll employment	-0.306***	-0.290***
5. Retail sales	-0.105***	-0.097***
6. Industrial production	0.003	0.002
7. Capacity utilization	-0.016	-0.010
8. Personal income	-0.023	-0.019
9. Consumer credit	-0.003	-0.002
Consumption		
10. Personal consumption expenditures	-0.020	-0.017
11. New home sales	-0.046***	-0.040***
Investment		
12. Durable goods orders	-0.070***	-0.053***
13. Construction spending	-0.021	-0.019
14. Factory orders	-0.017	-0.022
15. Business inventories	0.017	0.014
Government Purchases		
16. Government budget deficit	0.006	0.008
Net Exports		
17. Trade balance	-0.026**	-0.020**
Prices		
18. Producer price index	-0.054**	-0.041**
19. Consumer price index	-0.036	-0.037
Forward Looking		
20. Consumer confidence index	-0.063***	-0.054***
21. NAPM index	-0.020	-0.022
22. Housing starts	-0.015	-0.020
23. Index of leading indicators	-0.047***	-0.039***
24. Initial unemployment claims	0.041***	0.036***
R^2	0.019	0.212
Nr. Observations	157,500	157,500

in the financial press. In line with the view of the Phillips curve, inflation should be positively correlated with economic activity. Higher inflation leads to higher interest rates, hence the returns on Treasuries decline. The estimation results support this interpretation. We find that procyclical variables such as Nonfarm Payroll Employment, Retail Sales and New Home Sales, indeed affect bond prices negatively, while countercyclical variables such as the Initial Unemployment Claims have a positive impact on bond prices. Further, expected inflation also is a key focal point for fixed-income investors since news associated with the Producer Price Index (PPI) significantly affects bond prices. Comparing the magnitudes of the coefficients across surprises, we observe that new public information in Nonfarm Payroll Employment announcements are most important in economic terms. A one standard deviation surprise in this variable implies a change of 30 bps in the futures returns.

The estimation results in the rightmost column of Table 3.3 demonstrate that unanticipated order flow significantly influences returns, which is consistent with prior literature (Evans and Lyons, 2002; Green, 2004). Including unanticipated order flow increases the regression R^2 from 1.9% to 21.2%. A decomposition of the explained return variance assigns 91% to (orthogonalized) order flow and 9% to public information variables. Hence order flow accounts for roughly 20% of the price variance. This number is in line with the literature, as Evans and Lyons (2008, Table 6) report a corresponding number of 22%. The coefficient of unanticipated order flow is equal to 0.025 and significantly different from zero. Hence, on a 5-minute basis, a one standard deviation positive shock of order flow increases the return by two and a half bps. Although this number is lower than for example the corresponding value of Nonfarm Payroll Employment, we should take into account that this macroeconomic announcement is released only once per month, while order flow is available at each time interval. Note that a positive shock to order flow can be due either to an increase in buyer-initiated trades or a reduction in seller-initiated trades. In both cases, this signals positive private information, which increases bond returns.

Considering the importance of macroeconomic announcements in terms of explanatory power for the 5-minute returns, the R^2 of the specification (3.3) is rather small at 1.9% when order flow is not included. This is somewhat misleading though, since these announcements occur relatively rarely, in the sense that the number of observations for

Table 3.4: Effect of public information on returns during announcements

This table reports the estimation results of the following regression model:

$$R_t = \alpha_k + \beta_k S_{k,t} + \varepsilon_t,$$

where R_t is the 5-minute log return of the 30-year U.S. Treasury bond futures from period $t - 1$ to t and $S_{k,t}$ is the standardized surprise of announcement type k , $k = 1, \dots, 24$. We include only the pairs $(R_t, S_{k,t})$ when an announcement of fundamental k was made at time t . The superscripts ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The sample goes from January 2, 2004 through December 31, 2009.

Dependent variable: 30-year futures returns			
	β_k	R^2	Nr. Obs
1. GDP advance	-0.068	0.094	24
2. GDP preliminary	-0.038	0.082	24
3. GDP final	0.008	0.005	24
Real Activity			
4. Nonfarm payroll employment	-0.297***	0.268	69
5. Retail sales	-0.110***	0.223	70
6. Industrial production	-0.011	0.009	68
7. Capacity utilization	-0.014	0.022	69
8. Personal income	-0.019	0.021	71
9. Consumer credit	-0.003	0.003	63
Consumption			
10. Personal consumption expenditures	-0.019	0.022	71
11. New home sales	-0.054***	0.237	69
Investment			
12. Durable goods orders	-0.073***	0.135	70
13. Construction spending	-0.020	0.013	72
14. Factory orders	-0.017	0.019	72
15. Business inventories	0.014	0.011	69
Government Purchases			
16. Government budget deficit	0.007	0.011	65
Net Exports			
17. Trade balance	-0.029**	0.057	71
Prices			
18. Producer price index	-0.071***	0.132	69
19. Consumer price index	-0.031	0.017	71
Forward Looking			
20. Consumer confidence index	-0.067***	0.285	72
21. NAPM index	-0.016	0.014	72
22. Housing starts	-0.009	0.003	71
23. Index of leading indicators	-0.046***	0.271	68
24. Initial unemployment claims	0.039***	0.075	304

which a surprise actually occurs is only a tiny fraction of the total sample size. To highlight the importance of macro news during the announcement periods, we estimate the simplified model

$$R_t = \alpha_k + \beta_k S_{k,t} + \varepsilon_t, \quad (3.4)$$

using only those observations when an announcement of variable k was made at time t . The estimation results for the 24 different announcements are shown in Table 3.4. The magnitude of the least squares coefficient estimates as well as their standard errors are rather similar to those obtained with the general model specification in (3.3). The R^2 values are considerably higher for most announcements, exceeding 20% for the Consumer Confidence Index, Nonfarm Payroll Employment, the Conference Board Index of Leading Indicators, New Home Sales, and Retail Sales.

3.4.2 Public and private information also matter for volatility

ABDV (2003, 2007) focus on modeling the effects of macro announcements on returns. To account for heteroskedasticity, the return equation as given in (3.3) is estimated by means of a two-step Feasible Weighted Least Squares (FWLS) procedure, where the weights are inversely related to estimates of the volatility of the returns. It is well-documented that the main determinants of volatility of high-frequency returns (apart from its mere persistence) are macroeconomic news announcements and a pronounced deterministic pattern related to the trading activity during different parts of the day, see Bollerslev *et al.* (2000), among many others. For this reason, in the ‘ABDV approach’ a linear specification for the time-varying volatility of the intraday unexpected returns ε_t in (3.3) is proposed that includes the standardized surprises, autoregressive terms and a flexible Fourier series capturing the intraday pattern of volatility, that is,

$$\begin{aligned} |\hat{\varepsilon}_t| = & \beta_0 + \sum_{i=1}^{I'} \beta_i |\hat{\varepsilon}_{t-i}| + \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{kj'} |S_{k,t-j}| \\ & + \sum_{q=1}^Q \left(\delta_q \cos\left(\frac{q2\pi t}{N}\right) + \phi_q \sin\left(\frac{q2\pi t}{N}\right) \right) + u_t, \end{aligned} \quad (3.5)$$

where $|\hat{\varepsilon}_t|$ is the absolute value of the residual of equation (3.3), and the sine and cosine terms aim to capture the intraday volatility pattern. As we consider 5-minute returns during the period from 8:00 a.m. to 5:00 p.m., we set N equal to 108 (9 hours \times 12 five-minute intervals). The specification in (3.5) is estimated by Ordinary Least Squares (OLS) and the inverse of the fitted values of the dependent variable are used as weights to perform a FWLS estimation of the return equation in (3.3).

Although the above specification is flexible and easy to estimate, a numerical problem may occur: the conditional volatility equation can produce negative fitted values. This happens when, e.g., the Fourier series attains its minimum value and a macro announcement does not occur during this interval. Such a negative value implies negative weights for the FWLS step of the return equation, which invalidates the approach.¹² Note that negative fitted values need not necessarily occur to create problems. Even if the fitted value of (3.5) remains positive but becomes very small, the corresponding observations receive a disproportionately large weight in the FWLS estimation of the return specification (3.3), which may not be desirable. A second issue is that the ABDV two-step approach neglects parameter uncertainty of the conditional mean equation when estimating the volatility equation. As we are particularly interested in the effects of public and private information on volatility, it is important to take this uncertainty into account.

We overcome both econometric issues discussed above by using a GARCH-type approach. We adopt the basic idea of the Engle and Rangel (2008) Spline-GARCH model, and decompose the conditional volatility in two (multiplicative) parts. One of these components is a standard GARCH term (but normalized to have unconditional mean equal to one) to capture high-frequency movements in volatility. In Engle and Rangel (2008), the other component captures low-frequency movements in volatility, which is achieved by using a quadratic spline function of time t . For the purpose of our analysis, we specify the second component such that it contains the intraday volatility pattern (by means of a Fourier series), dummy variables and the effects of public and private information. Combined with the return equation from the ABDV approach, i.e. (3.3), the complete model

¹²Estimating (3.5) for the 5-minute Treasury futures returns using the full sample period, the fitted value becomes negative for 70 observations.

specification is given by

$$\begin{aligned}
 R_t &= \beta_{0,m} + \beta_{a,m} D_a + \sum_{i=1}^I \beta_i R_{t-i} + \sum_{k=1}^K \sum_{j=0}^J \beta_{kj} S_{k,t-j} + \beta_{OF,m} OF_t^* + \varepsilon_t, \\
 \varepsilon_t &= \sqrt{g_t \tau_t} u_t, \\
 g_t &= (1 - \alpha - \beta_{GA}) + \alpha \left(\frac{\varepsilon_{t-1}^2}{\tau_{t-1}} \right) + \beta_{GA} g_{t-1}, \\
 \tau_t &= \exp \left(\beta_{0,v} + \beta_{a,v} D_a + \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{kj'} |S_{k,t-j'}| + \beta_{OF,v} |OF_t^*| \right. \\
 &\quad \left. + \sum_{q=1}^Q \left(\delta_q \cos\left(\frac{q2\pi t}{N}\right) + \phi_q \sin\left(\frac{q2\pi t}{N}\right) \right) \right),
 \end{aligned} \tag{3.6}$$

where $u_t \sim N(0, 1)$, the conditional variance $E[\varepsilon_t^2 | \mathcal{I}_{t-1}] = g_t \tau_t$ with \mathcal{I}_t the information set at time t , g_t represents the unit-GARCH term and τ_t denotes the joint public and private news component. The exponential specification of τ_t obviously avoids the problem whereby the conditional variance can become negative by construction.

Both De Goeij and Marquering (2006) and Brenner *et al.* (2009) model volatility effects of macroeconomic announcements through GARCH models. We differ in set-up not only by in addition considering private information, but also by use of the Spline-GARCH model. An important difference of our set-up compared to the augmented-GARCH specifications of these two papers is that the latter model implies by definition a persistent effect of surprises on volatility. This is the result of, in effect, introducing a dummy variable that accounts for the announcements in the GARCH equation. In this case, a high impact of macro announcements through the GARCH coefficients (that is, the usual α and β from, e.g., a GARCH(1,1) specification) always models a persistent effect of public information on volatility.

The Spline-GARCH model allows controlling for the length of the effect of the surprises by simply adding lags of surprises in the specification of τ_t , which does not influence g_t . Hence we are able to distinguish between the effect of private information on volatility and the effect of macroeconomic surprises on volatility.

Table 3.5 provides estimation results for the GARCH model in (3.6) based on the complete sample period from January 2004 until December 2009. We set $Q = 5$ and $J' = 1$ in

Table 3.5: Public and private information effects

This table reports the estimation results of the Spline-GARCH model:

$$\begin{aligned}
 R_t &= \beta_{0,m} + \beta_{a,m} D_a + \sum_{i=1}^I \beta_i R_{t-i} + \sum_{k=1}^K \sum_{j=0}^J \beta_{kj} S_{k,t-j} + \beta_{OF,m} OF_t^* + \varepsilon_t, \\
 \varepsilon_t &= \sqrt{\tau_t} g_t u_t, \\
 g_t &= (1 - \alpha - \beta_{GA}) + \alpha \left(\frac{\varepsilon_{t-1}^2}{\tau_{t-1}} \right) + \beta_{GA} g_{t-1}, \\
 \tau_t &= \exp \left(\beta_{0,v} + \beta_{a,v} D_a + \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{kj'} |S_{k,t-j}| + \beta_{OF,v} |OF_t^*| \right. \\
 &\quad \left. + \sum_{q=1}^Q \left(\delta_q \cos\left(\frac{q2\pi t}{N}\right) + \phi_q \sin\left(\frac{q2\pi t}{N}\right) \right) \right),
 \end{aligned}$$

where R_t is the 5-minute log return of the 30-year U.S. Treasury bond futures from time period $t-1$ to t , D_a is a dummy variable that is 1 on announcement days and 0 else, $S_{k,t}$ is the standardized news announcement for $k = 1, \dots, 24$ and OF_t^* denote the unanticipated order flow from period $t-1$ to t . Further, $I = 3$, $J = 0$, $J' = 1$ and $Q = 5$. We show only the effects of public and private information variables on returns and volatility. In Panel A we report parameter estimates corresponding with private news for two various specifications, whereas panel B provides the public news coefficients of the last specification. In Panel B the first column denotes the contemporaneous effect while the second column denotes the one-period lagged effect of public information on volatility. The superscripts ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. The sample goes from January 2, 2004 through December 31, 2009.

Panel A: Private information		
Dependent variable: 30-year futures returns		
	(1)	(2)
Conditional mean equation		
Order flow		0.022***
Macro announcements	yes	yes
Conditional volatility equation		
Order flow		0.431***
Macro announcements	yes	yes
Log-likelihood	257,562	284,311
Nr. Observations	157,500	157,500

(continued from previous page)

Panel B: Public information			
	Cond Mean	Cond Volatility	
1. GDP advance	−0.134***	1.412***	0.204
2. GDP preliminary	−0.049	1.237***	−0.159
3. GDP final	−0.001	0.614***	0.069
Real Activity			
4. Nonfarm payroll employment	−0.277***	3.124***	0.848***
5. Retail sales	−0.088***	1.027***	0.551***
6. Industrial production	−0.015	0.411*	0.085
7. Capacity utilization	−0.033	0.149	0.149
8. Personal income	−0.025	0.387	−0.062
9. Consumer credit	−0.001	−0.102	−0.167
Consumption			
10. Personal consumption expenditures	−0.037*	0.593***	0.053
11. New home sales Investment	−0.076***	0.840***	0.342**
Investment			
12. Durable goods orders	−0.051**	1.097***	−0.038
13. Construction spending	−0.011	1.271***	0.544***
14. Factory orders	−0.008	0.995***	0.436***
15. Business inventories	−0.003	0.147	−0.070
Government Purchases			
16. Government budget deficit	0.004	−0.090	−0.053
Net Exports			
17. Trade balance	−0.028***	0.079	0.100
Prices			
18. Producer price index	−0.045*	1.067***	0.224*
19. Consumer price index	−0.132***	1.170***	0.005
Forward Looking			
20. Consumer confidence index	−0.072***	0.624***	0.200*
21. NAPM index	−0.017	0.647***	0.131
22. Housing starts	−0.031**	0.702***	−0.059
23. Index of leading indicators	−0.023**	0.072	0.130
24. Initial unemployment claims	0.039***	0.355***	−0.059

the Spline-GARCH model and estimate all parameters in the model (3.6) simultaneously, by (quasi) maximum likelihood. Panel A reports results of two different specifications. The first specification does not include private information. The other specification adds private information to the model. The order flow impact of the return is significant, though somewhat lower than when only the return equation is estimated: compared to Table 3.3 the coefficient decreases from 0.025 to 0.022. Interestingly, Panel A of the table indicates that (unanticipated) order flow has a significant positive effect on volatility, after controlling for public information. The positive coefficient on volatility suggests that the learning process of the agents to clear the market increases uncertainty about the bond returns. The market aggregates private information which in turn affects the returns and uncertainty on the futures market.

Panel B of Table 3.5 reports the estimated coefficients for the surprises in macro announcements corresponding with the second specification of Panel A.¹³ The results of the conditional mean equation indicate that 13 out of the 24 announcement surprises significantly affect subsequent returns, where in general the same reasoning holds as described in the discussion about Table 3.3, i.e. procyclical variables cause a negative bond return whereas countercyclical variables positively influence bond returns.¹⁴ In addition, the impact of public information is short-lived, since the return equation requires no lag of the macro variables (guided by a Likelihood-Ratio test). Thus the price discovery process is quick. This is consistent with earlier studies, see for example ABDV (2007). We find that 16 out of the 24 macro surprises affect volatility in the first five minutes after the announcement, whereas some variables also have a significant impact on volatility during the next five minutes. The impact of public information on volatility is less short-lived compared to the impact on returns, although a Likelihood-Ratio test indicates only one lag of the macro announcement variables is required.

¹³The estimated public information coefficients corresponding to the other two specifications are of similar order.

¹⁴As a separate analysis, we include dummy variables to capture the effect that the mere presence of an announcement boosts volatility, apart from the size of the associated surprise, see ABDV (2003). Results suggest that indeed the presence of an announcement matters for volatility. In line with previous literature, we consider the specification with surprises rather than dummies for our main analysis. As another additional analysis, we introduce dummy variables to interact with surprises of low and high size, allowing for a different effect of low surprises on volatility. This effect is not present in our data. For both analyses our main results on private information and volatility are qualitatively similar.

Figure 3.2: Intraday volatility effects

This figure depicts the intraday effects that capture the high-frequency pattern of deviations of intraday volatility from its daily average, as estimated in the Spline-GARCH model defined in (3.6). In addition, the figure depicts 95% confidence bounds.

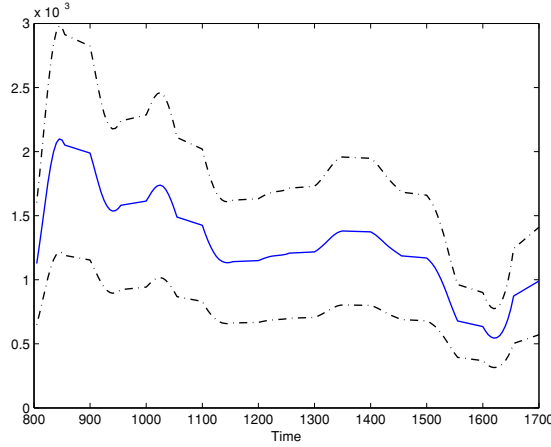


Table 3.6: Economic significance of public and private information on volatility

This table provides economic significance of the volatility part of the Spline-GARCH model:

$$\begin{aligned}
 R_t &= \beta_{0,m} + \beta_{a,m}D_a + \sum_{i=1}^I \beta_i R_{t-i} + \sum_{k=1}^K \sum_{j=0}^J \beta_{kj} S_{k,t-j} + \beta_{m,OF} OF_t^* + \varepsilon_t, \\
 \varepsilon_t &= \sqrt{\tau_t} g_t u_t, \\
 g_t &= (1 - \alpha - \beta_{GA}) + \alpha \left(\frac{\varepsilon_{t-1}^2}{\tau_{t-1}} \right) + \beta_{GA} g_{t-1}, \\
 \tau_t &= \exp \left(\beta_{0,v} + \beta_{a,v} D_a + \sum_{q=1}^Q \left(\delta_q \cos\left(\frac{q2\pi t}{N}\right) + \phi_q \sin\left(\frac{q2\pi t}{N}\right) \right) + \right. \\
 &\quad \left. \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{kj'} |S_{k,t-j'}| + \beta_{v,OF} |OF_t^*| + \beta_{v,OFD} D_{OF} |OF_t^*| \right),
 \end{aligned}$$

where R_t is the 5-minute log return of the 30-year U.S. Treasury bond futures from time period $t-1$ to t , D_a is a dummy variable that is 1 on announcement days and 0 else, D_{OF} represents a dummy variable that is 1 during 15 minutes after an announcement has been made, $S_{k,t}$ is the standardized news announcement for $k = 1, \dots, 24$ and OF_t^* denotes unanticipated order flow from period $t-1$ to t . Further, $I = 3$, $J = 0$, $J' = 1$ and $Q = 5$. We measure the economic significance by taking the partial derivative of the conditional volatility ($\sqrt{h_t}$) with respect to the variable S_{kt} and OF_t^* . The numbers represent the partial effect on volatility of a one standard deviation shock to the public or private variable. The sample goes from January 2, 2004 through December 31, 2009.

Economic significance in basis points	
Public information	
GDP advance	4.1
Nonfarm payroll employment	9.5
Consumer price index	3.5
Private information	
Order flow	2.7

Figure 3.2 depicts the Fourier series that captures the intraday deviations from the daily volatility. Typically, when floor trading starts, volatility is relatively high, whereas during lunch and after closing time (15:00) volatility is relatively low. Incorporating this deterministic pattern contributes substantially to explaining the intraday behavior of volatility. The remaining estimated parameters of the model are the constants, dummy variables and the GARCH parameters. Both $\beta_{0,m}$ and $\beta_{a,m}$ are almost zero and not significantly different from zero. In contrast, the corresponding values in the volatility equation ($\beta_{0,v}$ and $\beta_{a,v}$) are statistically significant. Moreover, $\beta_{a,v}$ is positive, indicating that on announcement days the overall level of volatility is higher compared to non-announcement days. Finally the GARCH parameters indicate high persistence in volatility, since β_{GA} and α are equal to 0.96 and 0.03 respectively.

We assess the economic significance of the public and private information measures by analyzing the partial derivatives of the volatility specification with respect to these variables. Multiplying this partial derivative with the sample standard deviation of the variable under review provides the economic significance. To get a feeling for the economic significance of public information we consider the three news announcements that have the strongest effect on volatility (GDP advance, Nonfarm Payroll Employment and the CPI). We consider the economic significance of order flow only during the first fifteen minutes after a news announcement to deal with the difference that order flow is available at each of the 5-minute intervals but announcements only in one 5-minute interval each month.¹⁵ Table 3.6 shows the outcome of this analysis. The table shows that the economic significance of order flow is substantial. A one standard deviation change in unanticipated order flow implies on average an increase of 2.7 bps of the volatility, while a similar change in the most influential surprise, Nonfarm Payroll Employment, increases volatility by 9.5 bps. Hence the importance of order flow is about one-fourth of the economic significance of Nonfarm Payroll Employment. The economic significance of the CPI and GDP advance announcements is slightly higher than the economic significance

¹⁵More specifically, we include a new variable in our volatility equation (3.6) which interacts (unanticipated) order flow with a dummy variable that takes the value one during the first fifteen minutes after an announcement.

of order flow.¹⁶ For the remaining 21 macro announcements (which are not displayed in the table), order flow significance exceeds the economic significance.

Our result of the statistically and economically significant impact of order flow on volatility confirms the finding of Berger *et al.* (2009). However, one of the key differences between the two types of analyses is that we control for the effect of order flow on the first moment (returns), whereas Berger *et al.* (2009) solely consider (realized) volatility. Ignoring the first moment has the risk that return effects are attributed to the second moment. In fact, running the Spline-GARCH model from (3.6) but setting all mean effects to 0 (thus taking $R_t = \varepsilon_t$) would overestimate the economic significance of order flow by 6 bps. An additional key difference is the inclusion of public information in the form of macro announcements when analyzing the impact of order flow on volatility. The effect of this incorporation is twofold: it identifies the impact of order flow while controlling for the effect of public information, and allows comparison of the relative contribution of public and private information.

In summary, the results suggest that private information, measured by order flow, is a statistically and economically significant explanatory variable for volatility. Higher absolute values of order flow signal the presence of a higher level of private information, which in turn increases uncertainty on the Treasury market.

3.5 Robustness checks and extensions

We consider several modifications and extensions of the Spline-GARCH model from the previous section to examine the robustness of our result that private information matters for volatility. First, we control for liquidity effects by including the bid-ask spread, the volume of trades and the illiquidity ratio (ILR) of Amihud (2002) as explanatory variables in our model. Second, we consider the market sidedness measure of Sarkar and Schwartz (2009) as an alternative proxy for private information. Third, we relate private

¹⁶A lower bound of the economic significance is given by considering only the first five minutes after an announcement. In this case, a one standard deviation change in order flow increases volatility on average by 1.9 bps from which 0.6 bps is caused by the extra effect of order flow on volatility during announcement times, while 1.3 bps represents the average effect of order flow in each 5-minute interval in our sample. A fair upper bound of the economic significance is the daily level of economic significance, in which case the economic significance is 14 bps (conservatively ignoring the immediate impact of the announcement and scaling up the 2.73 bps effect per 5-minute interval to the daily level).

information effects on volatility to periods of crises versus normal periods. Finally, we consider whether the dispersion of beliefs among analysts influences the importance of private information for both returns and volatility.

3.5.1 Controlling for liquidity effects

So far we have argued that order flow represents private information. However, there is also a link between private information and liquidity, as discussed in Section 3.2. We isolate the private information effect of order flow by including several liquidity proxies. Specifically, we consider the bid-ask spread, the illiquidity ratio (ILR) of Amihud (2002) and the volume of trades. We measure the intraday bid-ask spread as the difference between the (volume weighted) average buy and sell prices during each 5-minute interval, see also Manaster and Mann (1996). The illiquidity ratio is defined as the absolute (percentage) price change per dollar of trading volume during a particular time period. We calculate this liquidity proxy at the daily level as the mean of the ratio of intraday absolute return and intraday dollar volume. We use the volume of trades as a third proxy of liquidity, and consider similar to order flow the unanticipated part of volume.¹⁷ The control for liquidity effects changes the joint public and private news component τ_t of (3.6) as follows:

$$\begin{aligned} \tau_t = \exp \left(\beta_{0,v} + \beta_{a,v} D_a + \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{kj'} |S_{k,t-j}| + \beta_{OF,v} |OF_t^*| + \beta_{Spr} Spr_t \right. \\ \left. + \beta_{Qnt} Qnt_t^* + \beta_{ILR} ILR_t + \sum_{q=1}^Q \left(\delta_q \cos\left(\frac{q2\pi t}{N}\right) + \phi_q \sin\left(\frac{q2\pi t}{N}\right) \right) \right), \end{aligned} \quad (3.7)$$

where Qnt_t denotes the volume of trades from period $t-1$ to t , Spr_t is the bid-ask spread and ILR_t represents the illiquidity ratio. The specifications for R_t , ε_t and g_t are again as in (3.6).¹⁸

Table 3.7 show results of estimating the modified Spline-GARCH model when we control for liquidity. We introduce the liquidity variables stepwise in 5 specifications.

¹⁷We apply an ARMA(2,1) filter on volume after controlling for the intraday effects of volume via a Fourier series.

¹⁸We leave the return equation unaffected. As returns can be positive and negative, it only makes sense to include variables that include a similar possible difference in sign. This allows for including order flow and the surprise, but not variables such as the spread, trade volume, and the illiquidity ratio.

Table 3.7: Controlling for liquidity effects

This table reports the estimation results of the Spline-GARCH model:

$$\begin{aligned}
 R_t &= \beta_{0,m} + \beta_{a,m} D_a + \sum_{i=1}^I \beta_i R_{t-i} + \sum_{k=1}^K \sum_{j=0}^J \beta_{kj} S_{k,t-j} + \beta_{OF,m} OF_t^* + \varepsilon_t, \\
 \varepsilon_t &= \sqrt{\tau_t} g_t u_t, \\
 g_t &= (1 - \alpha - \beta_{GA}) + \alpha \left(\frac{\varepsilon_{t-1}^2}{\tau_{t-1}} \right) + \beta_{GA} g_{t-1}, \\
 \tau_t &= \exp \left(\beta_{0,v} + \beta_{a,v} D_a + \sum_{q=1}^Q \left(\delta_q \cos\left(\frac{q2\pi t}{N}\right) + \phi_q \sin\left(\frac{q2\pi t}{N}\right) \right) + \right. \\
 &\quad \left. \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{kj'} |S_{k,t-j'}| + \beta_{OF,v} |OF_t^*| + \beta_{Qnt} Qnt_t^* + \beta_{Spr} Spr_t + \beta_{ILLR} ILLR_t \right),
 \end{aligned}$$

where R_t is the 5-minute log return of the 30-year U.S. Treasury bond futures from time period $t-1$ to t , D_a is a dummy variable that is 1 on announcement days and 0 else, $S_{k,t}$ is the standardized news announcement for $k = 1, \dots, 24$, OF_t^* and Qnt_t^* denote the unanticipated order flow and unanticipated trading volume respectively, Spr_t represents the bid-ask spread and $ILLR_t$ is the illiquidity ratio of Amihud (2002) from period $t-1$ to t . Further, $I = 3$, $J = 0$, $J' = 1$ and $Q = 5$. We show only the effects of private information variables on returns and volatility. The superscripts ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. The sample goes from January 2, 2004 through December 31, 2009.

Dependent variable: 30-year futures returns					
	(1)	(2)	(3)	(4)	(5)
Conditional mean equation					
Order flow	0.022***	0.022***	0.022***	0.022***	0.021***
Macro ann.	yes	yes	yes	yes	yes
Conditional volatility equation					
Order flow	0.431***	0.414***	0.256***	0.430***	0.235***
Bid-ask spread		0.088***			0.112***
Trading volume			0.327***		0.340***
Illiquidity ratio				-0.005	0.034**
Macro ann.	yes	yes	yes	yes	yes
Log-likelihood	284,311	286,276	289,425	285,341	291,756
Nr. Observations	157,500	157,500	157,500	157,500	157,500

First, in specification (1) we repeat the estimates of our analysis without liquidity controls for comparison purposes. Specifications (2) through (4) each add one of the three liquidity controls, while specification (5) includes all three. Our proxy for private information remains significant, irrespective of whether we consider volume, bid-ask spread and/or the illiquidity ratio as additional explanatory variables for volatility. The order flow effect for returns is largely unchanged. The coefficient of order flow in the volatility equation does decline, from 0.431 to 0.235 when liquidity proxies are included, but remains statistically significant. Thus, while liquidity effects seem to play some role in the relation between volatility and private information that we document, they do not provide a complete alternative explanation for these results.¹⁹

3.5.2 An additional proxy for private information

We use order flow to measure of private information. To examine the robustness of our results for the effects of private information on volatility, we repeat our analysis using market sidedness as alternative proxy. Market sidedness is calculated as the correlation between the number of buyer-initiated trades and seller-initiated trades in brief intervals (see Section 3.3.2). As it concerns the correlation between variables it is only available at a lower frequency than order flow itself; we follow Sarkar and Schwartz (2009) and use it at a daily level. Because we use this proxy at the daily level, in this robustness exercise we consider all variables at the daily level. To examine the relation between market sidedness and volatility on a daily basis, we rewrite (3.6) for the daily level and include it as additional explanatory variable to obtain the joint public and private news component τ_t :

$$\tau_t = \exp \left(\beta_{0,v} + \beta_{a,v} D_a + \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{kj'} |S_{k,t-j}| + \beta_{OF,v} |OF_t^*| + \beta_{Side} Side_t + \beta_{Spr} Spr_t + \beta_{Qnt} Qnt_t^* + \beta_{ILR} ILR_t \right), \quad (3.8)$$

¹⁹The table also confirms the positive relation between volatility and volume. This result is consistent with the literature on the volume-volatility relation that is mentioned in Section 3.2.

Table 3.8: Private information at the daily level

This table reports the estimation results of the Spline-GARCH model:

$$\begin{aligned}
 R_t &= \beta_{0,m} + \beta_{a,m} D_a + \sum_{k=1}^K \sum_{j=0}^J \beta_{kj} S_{k,t-j} + \beta_{OF,m} OF_t^* + \varepsilon_t, \\
 \varepsilon_t &= \sqrt{\tau_t} g_t u_t, \\
 g_t &= (1 - \alpha - \beta_{GA}) + \alpha \left(\frac{\varepsilon_t^2}{\tau_{t-1}} \right) + \beta_{GA} g_{t-1}, \\
 \tau_t &= \exp \left(\beta_{0,v} + \beta_{a,v} D_a + \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{kj'} |S_{k,t-j}| + \right. \\
 &\quad \left. \beta_{OF,v} |OF_t^*| + \beta_{Side} Side_t + \beta_{Spr} Spr_t + \beta_{Qnt} Qnt_t^* + \beta_{ILR} ILR_t \right),
 \end{aligned}$$

where R_t is the daily log return of the 30-year U.S. Treasury bond futures at day t , D_a is a dummy variable that is 1 on announcement days and 0 else, $S_{k,t}$ is the standardized news announcement for $k = 1, \dots, 24$, OF_t^* and $Side_t$ denote the unanticipated order flow and the market sidedness measure of Sarkar and Schwartz (2009) respectively. In addition, Qnt_t^* denotes unanticipated trading volume, Spr_t represents the bid-ask spread and ILR_t is the illiquidity ratio of Amihud (2002) at day t . Further, $J = 0$ and $J' = 0$. We show only the effects of private information variables and liquidity measures on returns and volatility. The superscripts ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. The sample goes from January 2, 2004 through December 31, 2009.

Dependent variable: 30-year futures returns							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Conditional mean equation							
Order flow	0.206***	0.180***	0.191***	0.191***	0.191***	0.191***	0.170***
Macro ann.	yes	yes	yes	yes	yes	yes	yes
Conditional volatility equation							
Order flow	0.224***		0.215***	0.215***	0.124***	0.210***	0.127***
Sidedness		0.421***	0.407***	0.407***	0.150***	0.408***	0.148***
Spread				0.093**			0.044
Trading vol.					0.488***		0.499***
Illiq. ratio						-0.062	0.042
Macro ann.	yes	yes	yes	yes	yes		yes
Log-likelihood	-971	-954	-938	-935	-873	-937	-872
Nr. Observations	1420	1420	1420	1420	1420	1420	1420

where $Side_t$ denotes the market sidedness at day t . Following the analysis in Section 3.5.1, we control for liquidity effects by including the volume, bid-ask spread and the illiquidity ratio as additional explanatory variables. We measure the daily bid-ask spread Spr_t by averaging all its intraday values. Note that the Fourier series disappears since we estimate the model at the daily frequency.

Table 3.8 shows results of estimating the (modified) Spline-GARCH model at the daily level. Because of the change of frequency, we also repeat our analysis where we use order flow as the private information proxy (specification (1) of the table). A striking result is that order flow is still significant at the daily level. Next, we include sidedness both instead of order flow (specification (2)) and with order flow (specification (3)). The second

specification shows that market sidedness is highly significant in explaining volatility. The coefficient of sidedness on volatility is positive at 0.421, implying that a more two-sided Treasury market increases uncertainty. Sarkar and Schwartz (2009) argue that trading is more one-sided when some investors are better informed than others, and more two-sided when investors have different information or interpret the same information differently. The positive coefficient of sidedness in explaining volatility, coupled with the positive coefficient of order flow (more private information leads to higher uncertainty), provides indeed evidence on private information in the Treasury market being more about differential information (two-sided markets) than better information (one-sided markets). Furthermore, specification (3) shows that both order flow and sidedness are significant in explaining volatility. This is consistent with order flow and sidedness capturing different elements of the distribution of order flow and private information. Their joint significance strengthens our main result that private information matters for volatility. Finally, we add the liquidity controls to these two private information proxies. Similar to Section 3.5.1, the table indicates liquidity effects play some role in the relation between volatility and private information as the coefficients of order flow and market sidedness reduce from 0.22 and 0.41 to 0.13 and 0.15 respectively. However, both coefficients remain statistically significant.

3.5.3 Conditioning on the state of the economy

We investigate how the effects of public and private information on both returns and volatility depend on the state of the economy. Previous studies, including Beber and Brandt (2010) and ABDV (2007), find that the impact of macro announcements depends on whether the economy is in expansion or recession. We perform a similar analysis, but instead label the different economic regimes as 'normal' and 'crisis period' as the recession period in our sample (2008/2009) is usually referred to as a crisis rather than merely a recession. We introduce a dummy variable D_c that takes the value one during the 2008-2009 crisis. An indicator of normal times is then obtained as $D_n = 1 - D_c$. Using these dummy variables, the five-minute frequency Spline-GARCH model is modified as

follows:

$$\begin{aligned}
R_t = & \beta_{0,m} + \beta_{a,m}D_a + \beta_{n,m}D_n + \sum_{i=1}^I \beta_i R_{t-i} + \sum_{k=1}^K \sum_{j=0}^J \beta_{n,kj} D_n S_{k,t-j} \\
& + \sum_{k=1}^K \sum_{j=0}^J \beta_{c,kj} D_c S_{k,t-j} + \beta_{n,OF,m} D_n OF_t^* + \beta_{c,OF,m} D_c OF_t^* + \varepsilon_t, \\
\varepsilon_t = & \sqrt{g_t \tau_t} u_t, \\
g_t = & (1 - \alpha - \beta_{GA}) + \alpha \left(\frac{\varepsilon_{t-1}^2}{\tau_{t-1}} \right) + \beta_{GA} g_{t-1}, \\
\tau_t = & \exp \left(\beta_{0,v} + \beta_{a,v} D_a + \beta_{n,v} D_n + \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{n,kj'} D_n |S_{k,t-j'}| \right. \\
& + \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{c,kj'} D_c |S_{k,t-j'}| + \beta_{n,OF,v} D_n |OF_t^*| + \beta_{c,OF,v} D_c |OF_t^*| \\
& \left. + \sum_{l=1}^3 \beta_{n,l} D_n LIQ_{l,t} + \sum_{l=1}^3 \beta_{c,l} D_c LIQ_{l,t} + \sum_{q=1}^Q \left(\delta_q \cos\left(\frac{q2\pi t}{N}\right) + \phi_q \sin\left(\frac{q2\pi t}{N}\right) \right) \right),
\end{aligned} \tag{3.9}$$

where $LIQ_t = \{Qnt_t^*, Spr_t, ILLR_t\}$. Again, $u_t \sim N(0, 1)$ and τ_t denotes the public and private news component, whereas g_t represents the unit-GARCH term.

Table 3.9 reports coefficient estimates of the adapted Spline-GARCH model, together with Wald tests for the null hypothesis that the return and volatility impact of the public and private information variables is the same in the two economic regimes. Panel A shows results for the return equation, and suggests that private information has a statistically significant different effect on returns in the (non-)crisis period. The impact of order flow on returns is more than three times higher during the crisis 2008-2009. The table also reports the effect of public information on returns. Surprises generally have a larger effect during normal times compared to the crisis period. For most macroeconomic announcements the differences are small however, and not significant according to the Wald statistic.

Panel B of the table shows that also the effect of private information on volatility is significantly different between normal and crisis periods. Again, order flow is more important for volatility during the crises. This difference is also economically significant as the coefficient increases from 0.19 in normal times to 0.40 in 2008-2009. Regarding public information, panel B suggests that for most announcements the effect on volatility

Table 3.9: Crises versus non-crises periods

This table reports the estimation results of the Spline-GARCH model:

$$\begin{aligned}
 R_t &= \beta_{0,m} + \beta_{a,m}D_a + \beta_{n,m}D_n + \sum_{i=1}^I \beta_i R_{t-i} + \sum_{k=1}^K \sum_{j=0}^J \beta_{n,kj} D_n S_{k,t-j} + \\
 &\quad \sum_{k=1}^K \sum_{j=0}^J \beta_{c,kj} D_c S_{k,t-j} + \beta_{n,OF} D_n OF_t^* + \beta_{c,OF} D_c OF_t^* + \varepsilon_t, \\
 \varepsilon_t &= \sqrt{g_t \tau_t} u_t, \\
 g_t &= (1 - \alpha - \beta_{GA}) + \alpha \left(\frac{\varepsilon_{t-1}^2}{\tau_{t-1}} \right) + \beta_{GA} g_{t-1}, \\
 \tau_t &= \exp \left(\beta_{0,v} + \beta_{a,v} D_a + \beta_{n,v} D_n + \sum_{q=1}^Q \left(\delta_q \cos \left(\frac{q2\pi t}{N} \right) + \phi_q \sin \left(\frac{q2\pi t}{N} \right) \right) + \right. \\
 &\quad \beta_{n,OF} D_n |OF_t^*| + \beta_{c,OF} D_c |OF_t^*| + \sum_{l=1}^3 \beta_{n,l} D_n LIQ_{l,t} + \sum_{l=1}^3 \beta_{c,l} D_c LIQ_{l,t} \\
 &\quad \left. + \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{n,kj'} D_n |S_{k,t-j}| + \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{c,kj'} D_c |S_{k,t-j}| \right),
 \end{aligned}$$

where R_t is the 5-minute log return of the 30-year U.S. Treasury bond futures from time period $t-1$ to t , D_a is a dummy variable that is 1 on announcement days and 0 else, $S_{k,t}$ is the standardized news announcement for $k = 1, \dots, 24$, OF_t^* denotes the unanticipated order flow and LIQ_t stacks the volume of trades, the bid-ask spread and the illiquidity ratio from time $t-1$ to t ($LIQ_t = \{Qnt_t^*, Spr_t, ILR_t\}$). The dummy variables D_c and D_n denote a crises and non-crises period. Further, $I = 3$, $J = 0$, $J' = 1$ and $Q = 5$. Panel A contains estimated coefficients regarding the return part of the model whereas panel B contains the volatility part. We show only the contemporaneous effect and omit the coefficients corresponding with the first lag. We perform a Wald test to test on equal impact during expansions and recessions. The superscripts ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The sample goes from January 2, 2004 through December 31, 2009.

Panel A: Conditional Return			
Dependent variable: 30-year futures returns			
	Non-Crises	Crises	Wald test
Private information			
Order flow	0.019***	0.061***	3,876***
Public information			
1. GDP advance	-0.079	0.038	2.22
2. GDP preliminary	-0.039	-0.034	0.00
3. GDP final	0.006	0.009	0.01
Real Activity			
4. Nonfarm payroll employment	-0.270***	-0.025	2.80*
5. Retail sales	-0.090***	-0.060**	0.72
6. Industrial production	-0.016	0.043***	7.83***
7. Capacity utilization	-0.039*	-0.042**	0.01
8. Personal income	-0.040*	-0.052**	0.09
9. Consumer credit	-0.005	0.013	2.58
Consumption			
10. Personal consumption expenditures	-0.028	-0.025**	0.01
11. New home sales Investment	-0.061***	-0.084**	0.29
Investment			
12. Durable goods orders	-0.058**	-0.037	0.22
13. Construction spending	-0.024	-0.004	0.40
14. Factory orders	-0.017	-0.023	0.13
15. Business inventories	-0.003	-0.008	0.10
Government Purchases			
16. Government budget deficit	-0.004	0.010**	2.46
Net Exports			
17. Trade balance	-0.030***	-0.026	0.02
Prices			
18. Producer price index	-0.060**	-0.005	1.92
19. Consumer price index	-0.084**	-0.033*	1.31
Forward Looking			
20. Consumer confidence index	-0.078***	-0.036**	3.37*
21. NAPM index	-0.028*	0.021	3.08*
22. Housing starts	-0.026	0.020	1.76
23. Index of leading indicators	-0.017**	-0.037***	2.13
24. Initial unemployment claims	0.037***	0.038***	0.01

(continued from previous page)

Panel B: Conditional Volatility			
	Normal times	Crises period	Wald test
Private information			
Order flow	0.189***	0.403***	173***
Control variables			
Trading volume	0.287***	0.758***	2,053***
Bid-ask spread	0.119***	0.080***	36.7***
Illiquidity ratio	0.025***	0.094***	10.5***
Public information			
1. GDP advance	1.129***	0.379	2.59
2. GDP preliminary	1.187**	0.428	1.91
3. GDP final	0.586**	−0.667	4.98**
Real Activity			
4. Nonfarm payroll employment	2.303***	0.978***	14.7***
5. Retail sales	0.956***	0.232	8.04***
6. Industrial production	0.539**	−0.579	7.87***
7. Capacity utilization	−0.110	0.438	1.44
8. Personal income	0.734**	0.705*	0.01
9. Consumer credit	−0.187	−0.472	1.39
Consumption			
10. Personal consumption expenditures	0.418*	−1.970**	8.20***
11. New home sales Investment	0.540***	−0.017	1.08
Investment			
12. Durable goods orders	0.862***	−0.004	9.77***
13. Construction spending	1.037***	0.035	15.5***
14. Factory orders	1.062***	0.205	9.48***
15. Business inventories	0.206	−0.135	2.19
Government Purchases			
16. Government budget deficit	−0.309	−0.360	0.01
Net Exports			
17. Trade balance	−0.063	0.185	1.18
Prices			
18. Producer price index	0.956***	−0.165	18.4***
19. Consumer price index	1.328***	−0.087	34.6***
Forward Looking			
20. Consumer confidence index	0.572***	−0.023	7.05**
21. NAPM index	0.573***	0.119	3.24*
22. Housing starts	0.555***	0.349	0.27
23. Index of leading indicators	0.034	−0.342	2.03
24. Initial unemployment claims	0.388***	0.121	5.02**
Log-likelihood	295,996		
Nr. observations	157,500		

is larger in normal times. Based on the Wald test, most of these differences are statistically significant. Finally, the table highlights again the importance of controlling for liquidity effects. These effects are much stronger during the crises, except the bid-ask spread. This is not surprising, as 2008-2009 was characterized by a general lack of liquidity in many financial markets, including the U.S. Treasury market.

In sum, private information matters for volatility, irrespective of the occurrence of a crisis. The effect of order flow on returns and volatility is much stronger during the 2008-2009 crisis. In addition, public information has a larger influence on volatility during normal times. Particularly in times of the crisis, it is important to consider private information when modeling volatility.

3.5.4 Conditioning on the dispersion of beliefs

We interpret private information in the Treasury market as information about risk preferences and endowments (see the Introduction). In particular, when a public signal enters the market (e.g. a news announcement) order flow captures the aggregation of differential information processing skills. The value of this aggregated private information depends on the degree of information heterogeneity among market participants. When there is little consensus in the market, order flow is more informative. This can be beneficial for traders who observe order flow. As a final additional analysis we investigate the impact of private information on both returns and volatility for different levels of dispersion of beliefs among analysts. This extends Pasquariello and Vega (2007), who study this issue for returns only.

We measure the dispersion of beliefs using the range of analysts' forecasts for the 24 listed macro announcements. This is slightly different from Pasquariello and Vega (2007), who take the standard deviation of analysts' forecasts to estimate the dispersion. We follow their methodology for incorporating the dispersion of beliefs into the model. Specifically, we first convert the weekly and quarterly dispersions to a monthly frequency. For the weekly announcements of Initial Unemployment Claims this conversion is done by simply averaging the range across four weeks. For the three quarterly announcements in our data set, GDP Advance, Preliminary, and Final, we assume that the dispersion of

beliefs is constant throughout the quarter. The monthly proxy for information heterogeneity is defined as a sum of monthly (scaled) dispersions across announcements,

$$SRA_{P,t} = \sum_{k=1}^K \frac{RA_{k,t} - \mu(RA_{k,t})}{\sigma(RA_{k,t})}, \quad (3.10)$$

where $RA_{k,t}$ is the highest minus the lowest professional forecast of announcement k at time t and $\mu(RA_{k,t})$ and $\sigma(RA_{k,t})$ are its sample mean and standard deviation, respectively. Given the monthly dispersion estimates, we divide the empirical distribution function of $SRA_{P,t}$ in a low (0 – 30%), medium (30 – 70%) and high (70 – 100%) dispersion regime. Then we interact order flow with three dummy variables, which are constructed on these three regimes. In principle, it would also be possible to interact the standardized surprises with these dummy variables to allow for different effects of public information for different levels of dispersion. However, the range of the analysts' forecasts used to construct the dispersion measure is available only from June 2007 onwards. This implies that the subsample that can be used in this analysis contains only a limited number of announcements (30 for the monthly variables and 12 for the quarterly GDP figures), making it difficult to obtain reliable results. Due to the limited sample period, the number of observations for this analysis drops from 157,500 to 68,904.

Table 3.10 reports estimation results for the Spline-GARCH model conditioning the effects of private information on the level of the dispersion of analysts' forecasts. We do not report the estimates for the macro announcements, which are similar to those obtained for the complete sample period as shown in Table 3.5. In addition, for brevity we omit the estimated coefficients corresponding with the controls for liquidity effects. The informativeness of order flow is considerably higher for months with highest dispersion in analyst forecasts, both in the return and volatility equation. We find a monotonic increase in the coefficients, indicating that the higher the dispersion, the more informative order flow is for return and volatility. This result confirms the findings of Pasquariello and Vega (2007), but also provides new insight in explaining the conditional volatility. Order flow has a considerably stronger impact on uncertainty of Treasury futures in times of high dispersion of beliefs among traders about macro fundamentals – a result that is consistent

Table 3.10: Order flow and information heterogeneity

This table reports the estimation results of the following Spline-GARCH model:

$$\begin{aligned}
R_t &= \beta_{0,r} + \beta_{a,r}D_a + \beta_{m,r}D_m + \beta_{h,r}D_h + \sum_{i=1}^I \beta_i R_{t-i} + \sum_{k=1}^K \sum_{j=0}^J \beta_{kj} S_{k,t-j} + \\
&\quad \gamma_{OF,l} OF_t^* D_l + \gamma_{OF,m} OF_t^* D_m + \gamma_{OF,h} OF_t^* D_h + \gamma_{OF,OF_D,l} OF_t^* D_{OF} D_l + \\
&\quad \gamma_{OF,OF_D,m} OF_t^* D_{OF} D_m + \gamma_{OF,OF_D,h} OF_t^* D_{OF} D_h + \varepsilon_t, \\
\varepsilon_t &= \sqrt{\tau_t} g_t u_t, \\
g_t &= (1 - \alpha - \beta_{GA}) + \alpha \left(\frac{\varepsilon_{t-1}^2}{\tau_{t-1}} \right) + \beta_{GA} g_{t-1}, \\
\tau_t &= \exp \left(\beta_{0,v} + \beta_{a,v} D_a + \beta_{m,v} D_m + \beta_{h,v} D_h + \sum_{q=1}^Q \left(\delta_q \cos\left(\frac{q2\pi t}{108}\right) + \phi_q \sin\left(\frac{q2\pi t}{108}\right) \right) + \right. \\
&\quad \sum_{k=1}^K \sum_{j'=0}^{J'} \beta_{kj'} |S_{k,t-j}| + \gamma_{|OF|,l} |OF_t^*| D_l + \gamma_{|OF|,m} |OF_t^*| D_m + \gamma_{|OF|,h} |OF_t^*| D_h + \\
&\quad \gamma_{|OF|,OF_D,l} |OF_t^*| D_{OF} D_l + \gamma_{|OF|,OF_D,m} |OF_t^*| D_{OF} D_m + \gamma_{|OF|,OF_D,h} |OF_t^*| D_{OF} D_h + \\
&\quad \left. \sum_{d=1}^3 \sum_{l^*=l}^3 \gamma_{d,l^*} D_d LIQ_{l^*,t} \right),
\end{aligned}$$

where R_t is the 5-minute log return of the 30-year U.S. Treasury bond futures from time period $t - 1$ to t , D_a is a dummy variable that is 1 on announcement days and 0 else, $S_{k,t}$ is the standardized news announcement for $k = 1, \dots, 24$, and OF_t^* denotes the unanticipated order flow from period $t - 1$ to t . In addition, D_{OF} represents a dummy variable that is 1 during 15 minutes after an announcement has been made. Further, LIQ_t stacks the volume of trades, the bid-ask spread and the illiquidity ratio from time $t - 1$ to t ($LIQ_t = \{Qnt_t^*, Spr_t, LLR_t\}$). We introduce three additional dummy variables, D_l , D_m , and D_h (or D_d ($d = l, m, h$)) which represent a low, medium or high level of information heterogeneity respectively. Results in the table correspond with $I = 3$, $J = 0$, $J' = 1$ and $Q = 5$. In the table, we report the coefficients corresponding with private information, that is $\gamma_{OF,d} + \gamma_{OF,OF_D,d}$ and $\gamma_{|OF|,d} + \gamma_{|OF|,OF_D,d}$ ($d = l, m, h$). The superscripts ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. The sample goes from June 2, 2007 through December 31, 2009.

Dependent variable: 30-year futures returns		
	Cond Mean	Cond Vol
Private information		
Order flow (low)	0.033***	0.234***
Order flow (medium)	0.034***	0.355***
Order flow (high)	0.139***	0.373***
Control variables		yes
Macro announcements	yes	yes
Log-likelihood	114,579	
Nr. observations	68,904	

with the notion that a sizeable proportion of macro news is transmitted via order flow (see Evans and Lyons 2008, for evidence of this in the DM/\$ exchange rate market).

To shed light on the economic significance of this result, the partial effect of a one standard deviation shock in order flow during 15 minutes after a news announcement in months with high dispersion on the conditional volatility is equal to 1.8 bps. As a comparison, the corresponding values in times of medium and low dispersion are equal to 1.7 and 1.1 respectively. Hence order flow is more important when the dispersion of beliefs is high compared to when it is low.

3.6 Conclusion

We study the impact of private information on volatility in financial markets. We design a unified framework, inspired by the Spline-GARCH model of Engle and Rangel (2008), to study the relationship between public and private information and prices and volatility simultaneously. We apply the model to 5-minute returns and volatility for the 30-year U.S. Treasury futures over the period 2004-2009. We use surprises in 24 key macroeconomic variables to capture public information, and use order flow as our proxy for private information.

Our main finding is that private information significantly affects volatility, with this type of information being important compared to public information in this respect. Indeed, the effect of a shock to order flow on volatility is one-fourth of the effect of a surprise of the same magnitude in the most influential macroeconomic news announcement. We extend our main finding by controlling for liquidity effects and including an additional private information proxy at the daily level. In addition, we show that the effect of private information is stronger in crises periods compared to normal times. Third, when we take the dispersion of beliefs about macroeconomic announcements among analysts into account, the effect of private information on volatility is larger when there is high dispersion compared to low dispersion. Our results imply that risk managers, portfolio managers and regulators should take into account private information variables as a determinant of volatility.

Chapter 4

Predicting covariance matrices with financial conditions indexes

Joint work with Dick van Dijk and Michel van der Wel.

4.1 Introduction

Volatilities and correlations of financial asset returns are crucial for many different purposes in financial management. Forming an investment portfolio using the mean-variance framework or related approaches requires the covariance matrix of the constituent assets returns, and so does calculation of Value-at-Risk for such portfolios. Pricing options depending on various underlying assets is highly sensitive to the correlations between their returns.

A key feature of volatilities and correlations of asset return volatilities is that they vary over time. Not surprisingly then, there is a substantial literature that attempts to link volatility and correlation dynamics to macroeconomic and financial fundamentals. For volatility most studies consider measures of aggregate market volatility in equity, bond and foreign exchange markets. Concentrating on the equity market, which is most relevant for this chapter, it has been well documented that, in-sample, variables like bond returns, short term interest rates, producer prices and industrial production growth contain information for volatility (Schwert, 1989; Andersen, Bollerslev, Diebold and Vega, 2003; Andersen *et al.*, 2007, among many others). Out-of-sample, volatility forecasts can

be improved by including short term interest rates (Glosten *et al.*, 1993), and credit market variables including the commercial paper-to-Treasury spread, default returns, default spread and proxies for funding liquidity (Paye, 2012; Christiansen *et al.*, 2012).

Most of the existing studies on the macroeconomic and financial drivers of return correlations focus on correlations between international equity returns or between stocks and bonds. International stock return correlations are positively associated with world market volatility and a trend; while they are negatively related to exchange rate volatility, term structure differentials across markets, real interest rate differentials, and the return on a world market index (Bracker and Koch, 1999). Bond and stock return correlations are only little explained by macro-economic fundamentals, but other factors, especially liquidity proxies, play a more important role (Baele *et al.*, 2010; Christiansen and Rinaldo, 2007). Sheppard (2008) studies return correlations within an asset market, and using monthly covariances of the six Fama-French equity portfolio returns finds widespread evidence of statistically significant variation in conditional covariances attributable to financial and macro-economic state variables.

In this chapter, we study the impact of financial conditions on both volatility and correlations. This is motivated by a substantial body of recent evidence that particularly financial conditions are an important driver of the economy at large. Financial conditions impact monetary policy and macroeconomic variables, such as inflation and real GDP (see Goodhart and Hofmann, 2008; Guichard and Turner, 2008; Hatzius *et al.*, 2010). In this literature, financial conditions are defined as the current state of financial variables that influence economic behavior and the future state of the economy (cf. Hatzius *et al.*, 2010). From the literature on general economic conditions, it is known that high uncertainty causes an increase in volatility as investors react more strongly to new information (Pastor and Veronesi, 2008; Hamilton and Lin, 1996; Perez-Quiros and Timmermann, 2001), and asset returns may be more closely connected during bear markets than during bull markets (Longin and Solnik, 2001; Ang and Chen, 2002). So far little is known on how volatility and comovement of asset returns respond to changes in financial conditions. A notable exception is the work of Boudt *et al.* (2012), who document that financial conditions affect transition probabilities between volatility and correlation states.¹

¹We turn to a more detailed comparison with our analysis later on in the introduction.

We propose several models to study the effects of financial conditions on asset volatilities and correlations in a multivariate GARCH framework. Specifically, we consider two approaches that both model the volatilities and correlations separately, which has become popular since the introduction of the DCC model of Engle (2002). Together, the models for volatility and correlation define the covariance matrix. The advantage of this set-up is that the effect of financial conditions on volatility and correlation can be disentangled. In the first approach, we extend the Spline-GARCH model of Engle and Rangel (2008) to include a measure of financial conditions as an explanatory variable. This model is combined with various specifications for the correlation dynamics. Specifically, we start with a simple linear relationship between (quasi) conditional correlations and a measure of financial conditions. This specification allows for flexibility in the correlation structure by allowing series specific impact of financial conditions. In addition, we adapt the cDCC model of Engle (2002) and Aelli (2013). In the second approach, we extend the Factor-Spline-GARCH model of Rangel and Engle (2012) by allowing financial conditions to affect the conditional variances of both the common market factor and the idiosyncratic returns. Indirectly, this also implies effects of financial conditions on the correlations between asset returns.

We conduct an empirical analysis using daily equity returns of the largest U.S. head-quartered bank holding companies for the period January 1, 1994, through December 30, 2011. As a measure of financial conditions we use the Bloomberg Financial Condition Index (FCI), which is an equally weighted sum of money market indicators, bond market indicators and equity market indicators.² Our analysis consists of two parts. First, we study the in-sample fit of the models. We compare model specifications with and without financial conditions, and test for significance of the impact of the FCI on both volatilities and correlations. Second, in an out-of-sample framework we consider whether Value-at-Risk predictions improve when financial conditions are included. The empirical application is inspired by Boudt *et al.* (2012), who use weekly returns of the same financial

²In recent years, several Financial Conditions Indexes (FCIs) (also sometimes called Financial Stress Indexes) have been developed, both by institutions including the IMF, World Bank and several branches of the Federal Reserve, as well as by major investment banks including Citi and Goldman Sachs. The various FCIs have in common that they summarize key indicators of financial market conditions into a single number. We opt for the Bloomberg FCI for two reasons: First, it is available at a daily frequency, and second, it comprises different financial markets.

institutions to study the impact of the St. Louis Fed financial stress indicator on regime switching probabilities of a regime-switching GARCH model with two (equi-)correlation regimes.³

Our results show that financial conditions indeed do affect both the volatilities and the correlations of large U.S. bank holding companies. Specifically, worse financial conditions are associated with higher volatilities and higher correlations. This result is both statistically and economically significant. For example, the FCI boosts the level of the correlation by 0.15 on average during the 2007-2008 crisis. In addition, Value-at-Risk (VaR) estimates improve significantly compared to models without the FCI, such that less violations are made and the unconditional coverage match more closely to the nominal frequency. Finally, we find that our main result is not only driven by the VIX index - which is also captured by the Bloomberg FCI - but also by the other components of the Bloomberg FCI.

The remainder of this chapter is organized as follows. Section 4.2 presents our modeling framework including our new proposed dynamic correlation model. In Section 4.3, we describe the data. Sections 4.4 and 4.5 discuss the in-sample and out-of-sample results, respectively. Section 4.6 investigates the role of the VIX index in the Bloomberg FCI. Section 4.7 concludes.

4.2 The modeling framework

We use two different approaches to examine the role of financial conditions in shaping the volatilities and correlations of asset returns. First, we assume that volatilities can be described by a variation of the Spline-GARCH model of Engle and Rangel (2008) and combine this with variations of the dynamic conditional correlation (DCC) model of Engle (2002). Second, we extend the Factor-Spline-GARCH model of Rangel and Engle (2012), for volatilities and correlations simultaneously. In both cases we extend

³In contrast, we analyze daily returns and interact financial conditions directly with volatilities and (nonequi-)correlations. We use the Bloomberg FCI instead of the St. Louis Fed Indicator, since the Bloomberg index is available on a daily basis. We turn to data at the daily level as daily information has an important role for financial conditions during crises.

the models to incorporate the effects of financial conditions. We explain both modeling frameworks and the estimation of the models in the following subsections.

4.2.1 The Spline-GARCH model for volatility with DCC for correlation

Let $r_{i,t}$ denote the (excess) return for asset i in period t . We consider the volatilities and correlations for K assets, with their returns collected in the K -dimensional vector $r_t \equiv (r_{1,t}, \dots, r_{K,t})'$, for $t = 1, \dots, T$. In our first modeling approach we assume that the conditional mean of r_t is constant⁴, while its conditional covariance matrix is time-varying. That is, we have

$$\begin{aligned} E_{t-1}[r_t] &= \mu, \\ E_{t-1}[(r_t - \mu)(r_t - \mu)'] &= H_t, \end{aligned} \quad (4.1)$$

where $\mu = (\mu_1, \dots, \mu_K)'$, E_{t-1} denotes the expectation given the information set through time $t - 1$, and H_t the time-varying conditional covariance matrix. We disentangle the conditional correlations from the conditional variances by writing H_t as

$$H_t = D_t^{1/2} R_t D_t^{1/2}, \quad (4.2)$$

where $D_t = \text{diag}(h_{i,t})$, $i = 1, \dots, K$ is a diagonal matrix with the conditional variances of the stock returns on the diagonal, and R_t is the conditional correlation matrix.

We assume that the conditional variances can be described by a variant of the Spline-GARCH model of Engle and Rangel (2008). In the standard Spline-GARCH model the volatility is decomposed into two components: a fast-moving GARCH term to capture typical high-frequency behavior, and a slow-moving spline function to capture more gradual movements due to, e.g., macroeconomic fluctuations. We adapt the model by replacing the second component with a function of a time-varying explanatory variable, such as a

⁴The assumption of a fixed expected return is made purely for ease of exposition. It can be relaxed in a straightforward manner. In fact, in our empirical application we considered the possibility of time-varying expected returns by including autoregressive terms and possible explanatory variables such as the Bloomberg FCI in (4.1). All corresponding coefficients were insignificant, even at the 10% level.

financial conditions index. The resulting model specification is given by

$$\begin{aligned} h_{i,t} &= g_{i,t}\tau_{i,t}, \\ g_{i,t} &= \omega_i + \alpha_i \frac{(r_{i,t-1} - \mu_i)^2}{\tau_{i,t-1}} + \gamma_i \frac{(r_{i,t-1} - \mu_i)^2}{\tau_{i,t-1}} I[(r_{i,t-1} - \mu_i) < 0] + \delta_i g_{i,t-1}, \\ \tau_{i,t} &= \exp(\kappa_{i,0} + \kappa_{i,1} X_{t-1}). \end{aligned} \quad (4.3)$$

with $g_{i,t}$ the fast-moving GARCH component and $h_{i,t}$ the slower-moving volatility component linked to the explanatory variable X_{t-1} , which is scaled to have mean zero. In the GARCH term $g_{i,t}$ in (4.3), $I[A]$ denotes an indicator function taking the value one when its argument A holds true and zero otherwise. The corresponding term allows for different effects of positive and negative shocks on volatility, as in the GJR-model of Glosten, Jagannathan and Runkle (1993). For identification purposes, we impose the restriction $\omega_i = (1 - \alpha_i - \delta_i - \gamma_i/2)$, such that the unconditional mean of $g_{i,t}$ is equal to one. Accordingly, $g_{i,t}$ is referred to as the unit GARCH component. In the $\tau_{i,t}$ component the coefficient $\kappa_{i,0}$ essentially is a scaling parameter determining the unconditional mean of $h_{i,t}$, while $\kappa_{i,1}$ measures the impact of the exogenous variable X_t on the volatility for asset i . We label this model the Spline-GARCH-X model, to highlight its dependence on the explanatory variable X . The main advantage of this approach over simply including the explanatory variable into a GARCH specification directly, is that the unit GARCH term $g_{i,t}$ in (4.3) does not change and that the variance $h_{i,t}$ is always positive because of the exponent term in $\tau_{i,t}$ in (4.3).

In the spirit of the DCC model of Engle (2002), we model the correlation matrix R_t in (4.2) by means of a quasi-correlation matrix Q_t . By appropriately scaling this quasi-correlation matrix the correlation matrix is obtained. This approach guarantees that the correlation matrix will have ones on the diagonal, and values between -1 and 1 in the off-diagonal elements. We start with a simple linear relationship between the quasi-correlation matrix $Q_t \equiv [q_{ij,t}]$, $i, j = 1, \dots, K$ and the exogenous variable X_t :

$$Q_t = \bar{Q} + \gamma_{DC} X_{t-1} J, \quad (4.4)$$

with J a $(K \times K)$ matrix of ones and \bar{Q} the unconditional covariance matrix of the standardized returns $\varepsilon_t = D_t^{-1/2}(r_{i,t} - \mu_i)$, estimated by its sample analogue: $\bar{Q} =$

$\frac{1}{T} \sum_{t=1}^T \varepsilon_t \varepsilon_t'$. In this model the innovation in the correlations stems solely from the X_{t-1} term. The impact of this explanatory variable is measured by γ_{DC} , and is equal across asset correlations. Finally, we scale the Q_t matrix to ensure a valid time-varying conditional correlation matrix R_t :

$$R_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2}. \quad (4.5)$$

We label this specification the DC-X model, for Dynamic Correlation with eXplanatory variables.

The specification in (4.4) may be considered restrictive, for the two properties just mentioned: the dynamics in the correlations are purely driven by the single variable X_{t-1} , while its impact is identical for all correlations. Both characteristics can be relaxed in a relatively straightforward manner. First, we may allow for a series specific impact of X_t on the correlations by defining γ_{DC} as a $(N \times 1)$ vector and specifying Q_t as

$$Q_t = \bar{Q} + \gamma_{DC} \gamma_{DC}' X_{t-1}, \quad (4.6)$$

which resembles the generalized DCC [G-DCC] model of Cappiello *et al.* (2006) and Hafner and Franses (2009). Note that all elements of the $(K \times K)$ matrix $\gamma_{DC} \gamma_{DC}'$ are positive, such that the direction of the effect of the X_{t-1} variable is still the same across all correlations but its magnitude may obviously differ. We label this more flexible variation as the DC-X-F model.

Second, richer time-varying correlation patterns may be obtained by including the exogenous variable in the DCC model of Engle (2002). We start from the corrected DCC (cDCC) model proposed by Aelli (2013), given by

$$Q_t = \bar{Q} + \alpha_{cDCC} (Q_{t-1}^{*1/2} \varepsilon_{t-1} \varepsilon_{t-1}' Q_{t-1}^{*1/2} - \bar{Q}) + \beta_{cDCC} (Q_{t-1} - \bar{Q}), \quad (4.7)$$

with $Q_t^* = \text{diag}(q_{ii,t})$, $i = 1, \dots, K$ and α_{cDCC} and β_{cDCC} scalar parameters. Following van Dijk *et al.* (2011), we extend this specification by replacing the unconditional (quasi-)correlation matrix \bar{Q} with a time-varying matrix \bar{Q}_t driven by the explanatory variable

X. This results in the model

$$\begin{aligned} Q_t &= \bar{Q}_t + \alpha_{cDCC}(Q_{t-1}^{*1/2} \epsilon'_{t-1} \epsilon_{t-1}^{*1/2} - \bar{Q}_{t-1}) + \beta_{cDCC}(Q_{t-1} - \bar{Q}_{t-1}) \\ \bar{Q}_t &= \bar{Q} + \gamma_{cDCC} X_{t-1} J, \end{aligned} \quad (4.8)$$

In this setting, the scalar coefficient γ_{cDCC} has the same interpretation as in (4.4), namely the total effect of the variable X_{t-1} on Q_t . We label this model as cDCC-X.⁵

4.2.2 The Factor-Spline-GARCH model

As a second approach to model volatilities and correlations and their relations with financial conditions, we adapt the Factor-Spline-GARCH model of Rangel and Engle (2012). This model extends the standard Spline-GARCH model in two dimensions. First, the asset returns are assumed to be adequately described by a one-factor version of the arbitrage pricing theory asset pricing model of Ross (1976). Second, time-varying volatilities and correlations of the asset returns are obtained jointly by specifying Spline-GARCH and DCC models for the market factor and the idiosyncratic returns. Specifically, the first extension of the Spline-GARCH model boils down to the assumption that the excess return of asset i , $i = 1, \dots, K$ depends linearly on a single market factor:

$$r_{i,t} = \xi_i + \beta_i r_{m,t} + u_{i,t}, \quad (4.9)$$

where $r_{m,t}$ denotes the market excess return, ξ_i and β_i are parameters to be estimated, and $u_{i,t}$ are the residuals or idiosyncratic returns. Allowing for cross-correlation between the idiosyncratic returns of different assets and for correlation between the market return and the residuals (that is, $E_{t-1}[u_{i,t} u_{j,t}] \neq 0$ for some $i \neq j$ and $E_{t-1}[r_{m,t} u_{i,t}] \neq 0$ for some i), Rangel and Engle (2012) show that the conditional correlation between assets i and j at

⁵A natural question arises whether it is guaranteed that Q_t is positive definite (pd) in our proposed models. In the standard cDCC model, Q_t is positive definite when $\alpha_{cDCC} \geq 0$, $\beta_{cDCC} \geq 0$ and $\alpha_{cDCC} + \beta_{cDCC} \leq 1$. When exogenous variables enter the specification these conditions are no longer sufficient. In case $\gamma_{cDCC} X_{t-1}$ is positive, Q_t consists of the sum of a positive definite and a positive semi-definite matrix. As long as the effect of the latter matrix is not dominant, Q_t is still positive definite. The same holds for the DC-X and DC-X-F models. In general it is however not guaranteed that Q_t is positive definite. In the estimation procedure we use a pragmatic rule to ensure it is positive definite, through allowing γ_{cDCC} to be negative, as long as the Q_t matrix is positive definite for every t . In our empirical application this did not generate problems for both in the in-sample estimation of Section 4.4 and the forecasting exercise in Section 4.5.

time t is given by

$$\begin{aligned} \rho_{ij,t} = & (\beta_j \beta_i \mathbf{V}_{t-1}[r_{m,t}] + \beta_j \mathbf{E}_{t-1}[r_{m,t} u_{i,t}] + \beta_i \mathbf{E}_{t-1}[r_{m,t} u_{j,t}] + \mathbf{E}_{t-1}[u_{i,t} u_{j,t}]) \\ & \times (\beta_i^2 \mathbf{V}_{t-1}(r_{m,t}) + \mathbf{V}_{t-1}[u_{i,t}] + 2\beta_i \mathbf{E}_{t-1}[r_{m,t} u_{i,t}])^{-1/2} \\ & \times (\beta_j^2 \mathbf{V}_{t-1}(r_{m,t}) + \mathbf{V}_{t-1}[u_{j,t}] + 2\beta_j \mathbf{E}_{t-1}[r_{m,t} u_{j,t}])^{-1/2} \end{aligned} \quad (4.10)$$

where \mathbf{V}_{t-1} denotes the variance given the information set through time $t - 1$. Thus, as a result of the one-factor APT set-up of (4.9), the conditional correlation $\rho_{ij,t}$ is driven by various factors, including the variances of the market returns and the idiosyncratic returns, the β 's, and the possible correlation between the market return and the idiosyncratic returns.

The idea of Rangel and Engle (2012) for the second departure of the spline-GARCH model is to now model the time-varying volatilities and correlations of the market factor, as well as and the N idiosyncratic returns $u_{i,t}$. The Spline-GARCH model is used for all $(N+1)$ volatilities to discriminate between long- and short-term volatility effects. Here we adapt the Spline-GARCH model in a similar way as in the previous section, and replace the slow-moving spline by the explanatory variable X_t . This results in the following specification for the market return $r_{m,t}$:

$$\begin{aligned} r_{m,t} = & \mu_m + \sqrt{g_{m,t} \tau_{m,t}} \varepsilon_{m,t}, \quad \text{where } \varepsilon_{m,t} | I_{t-1} \sim (0, 1), \\ g_{m,t} = & \omega_m + \alpha_m \frac{(r_{m,t-1} - \mu_m)^2}{\tau_{m,t-1}} + \delta_m g_{m,t-1} \\ & + \gamma_m \frac{(r_{m,t-1} - \mu_m)^2}{\tau_{m,t-1}} I[(r_{m,t-1} - \mu_m) < 0], \\ \tau_{m,t} = & \exp(\kappa_{m,0} + \kappa_{m,1} X_{t-1}), \end{aligned} \quad (4.11)$$

with $\omega_m = (1 - \alpha_m - \delta_m - \gamma_m/2)$ for identification purposes. Likewise, the specification of the volatility of the idiosyncratic part of the returns reads:

$$\begin{aligned} u_{i,t} = & \sqrt{g_{i,t} \tau_{i,t}} \varepsilon_{i,t}, \quad \text{where } \varepsilon_{i,t} | I_{t-1} \sim (0, 1), \\ g_{i,t} = & \omega_i + \alpha_i \frac{(r_{i,t-1} - \xi_i - \beta_i r_{m,t-1})^2}{\tau_{i,t-1}} + \delta_i g_{i,t-1} \\ & + \gamma_i \frac{(r_{i,t-1} - \xi_i - \beta_i r_{m,t-1})^2}{\tau_{i,t-1}} I[(r_{i,t-1} - \xi_i - \beta_i r_{m,t-1}) < 0], \\ \tau_{i,t} = & \exp(\kappa_{i,0} + \kappa_{i,1} X_{t-1}), \end{aligned} \quad (4.12)$$

with $\omega_i = (1 - \alpha_i - \delta_i - \gamma_i/2)$. Finally, we follow Rangel and Engle (2012) by assuming a cDCC process for the innovations of (4.11) and (4.12), collected as $\varepsilon_t^{tot} = (\varepsilon_t^m, \varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{K,t})'$:

$$Q_t = (1 - \alpha_{cDCC} - \beta_{cDCC})\bar{Q} + \alpha_{cDCC}(Q_{t-1}^{*1/2} \varepsilon_{t-1}^{tot} \varepsilon_{t-1}^{tot'} Q_{t-1}^{*1/2}) + \beta_{cDCC} Q_{t-1}, \quad (4.13)$$

with $Q_t^* = \text{diag}(q_{mm,t}, q_{11,t}, \dots, q_{KK,t})$ and scale the matrix Q_t using (4.5). We label this model as Factor-Spline-GARCH model.

The main difference between the modeling approaches of the previous section and the one presented here, is that the first case volatilities and correlations of the equity returns are modeled directly through the Spline-GARCH-X model and the (c)D(C)C-X model, and also the impact of the exogenous variable X_t is directly incorporated. In contrast, in the second case the link between X_t and between volatilities and correlation is indirect. An increase of the explanatory variable in the market factor increases both the volatility and correlation of all assets, and likewise an increase of the explanatory variable in the idiosyncratic part increase both the volatility and relevant pair-wise correlations. An advantage of the Factor-Spline-GARCH model is computational elegance: as the cDCC model for the remaining idiosyncratic terms $\varepsilon_{i,t}$ does not contain extra variables all matrix are by definition positive definite.

4.2.3 Estimation

This last subsection briefly discusses the estimation of the models put forward in the previous sections. For both modeling approaches, we follow the two-estimation procedure advocated by Engle (2002). In the first step, we estimate the parameters in the Spline-GARCH models for volatilities of the (idiosyncratic) returns of all K assets separately, as well as for the market return in case of the Factor-Spline-GARCH model.⁶ Here we follow Rangel and Engle (2012) by assuming a conditional Student- t distribution for the standardized returns. In the second step, we estimate the parameters of the correlation models using the standardized residuals from the Spline-GARCH volatility models. For

⁶For the Factor-Spline-GARCH model, we obviously first estimate the coefficients δ_i and β_i in (4.9) by means of least squares.

the cDCC models we estimate these parameters by maximizing the (quasi) Composite Likelihood (CL) of Engle, Shephard and Sheppard (2008), as standard likelihood-based optimization implies biased estimates of α_{cDCC} for large K (see Engle and Sheppard, 2005). The composite likelihood sums up likelihoods for each pair (i, j) of assets:

$$CL(\theta_{corr}) = \sum_{t=1}^T \sum_{i=1}^K \sum_{i>j} \log L_{i,j,t}, \quad (4.14)$$

where $\log L_{i,j,t}$ denotes the log-likelihood corresponding to the correlation part of asset i and j of the relevant correlation model and θ_{corr} stacks the correlation parameters. In practice, we follow the suggestion of Engle, Shephard and Sheppard (2008) and reduce the number of pairs by only considering randomly chosen contiguous pairs $i, i+1$. This does not affect the parameter estimates. We do not use the composite likelihood for the DC-X(-F) correlation models in (4.4) and (4.6), since the only innovation term is the X_{t-1} variable. Using the regular likelihood in this case reduces the computational costs considerably.

For the Factor-Spline-GARCH model it is useful to note that estimation may be facilitated by first rewriting the factor setup as follows:

$$r_t^{tot} = \xi + Bu_t^{tot}, \quad (4.15)$$

where $r_t^{tot} = (r_{m,t}, r_{1,t}, r_{2,t}, \dots, r_{K,t})'$ is the vector of excess returns at time t , $u_t^{tot} = (r_{m,t}, u_{1,t}, u_{2,t}, \dots, u_{K,t})'$ such that it contains the market factor as well as the idiosyncratic returns, $B = \begin{pmatrix} 1 & 0_{1 \times N} \\ \beta & I_{N \times N} \end{pmatrix}$, $\beta = (\beta_1, \beta_2, \dots, \beta_K)'$ and ξ is a vector of intercepts. Then we have that $r_t^{tot} | I_{t-1} \sim (\xi, H_t)$, such that

$$V_{t-1} = H_t = BD_t R_t D_t B', \quad (4.16)$$

with $D_t = \text{diag}(\sqrt{\tau_{k,t} g_{k,t}})$, ($k = m, 1, \dots, K$), R_t the conditional correlation matrix as in (4.5) with Q_t denoted in (4.13). The corresponding correlation matrix of r_t^{tot} is then the covariance matrix scaled by the volatilities on its main diagonal: $\text{diag}(H_t)^{-1/2} H_t \text{diag}(H_t)^{-1/2}$.

4.3 Data

While financial conditions may affect the volatilities and correlations of equity returns of all kinds of firms, this should in particular be the case for financial institutions. For this reason we apply the models discussed in the previous section to returns on stocks of the largest U.S. headquartered bank holding companies. We consider daily returns, over the period from January 1, 1994 to December 30, 2011. Following Boudt, Danielsson, Koopman and Lucas (2012), the banks in our sample are selected based on a ranking in terms of U.S. domestic deposits constructed by the Federal Deposit Insurance Corporation.⁷ We select those banks that have been among the fifteen largest banks for at least three consecutive years during the sample period. Table 4.1 lists the 16 bank holding companies that satisfy this requirement and are included in our sample. The penultimate column in the table indicates the years for which the banks belonged to the top 15 of largest deposit holding companies in the United States. Out of the 16 banks, six have been in the top list through the complete sample period, including Wells Fargo and Bank of America. Further, several banks have merged during the sample period (and some more than once), such as Nationsbanks Corp. and Chemical Banking Corp. We download daily price index returns from CRSP for all 16 banks. The last column provides the exact sample for which daily data is available. We do not have a balanced panel since the daily returns of Wachovia Corp. and National City Corp. are only available until December 30, 2008, as they merged in 2009 with Wells Fargo & Co and PNC Financial Services, respectively.

Figure 4.1 provides an impression of the average level of the correlations between the bank returns by showing the sample correlations for the period with an unbalanced sample (November 11, 1994 until December 30, 2011). The correlations of Wachovia Corp. and National City Corp. are based on the sample November 11, 1994, until December 30, 2008. The horizontal axis in the graph indicates the 16 banks (in the same order as in Table 4.1), and for each bank we plot the 15 sample correlations with the other banks. All correlations are positive and sizeable, ranging between 0.40 and 0.75 with an average close to 0.60. The maximum correlation is 0.78, between Wells Fargo & Co. and National

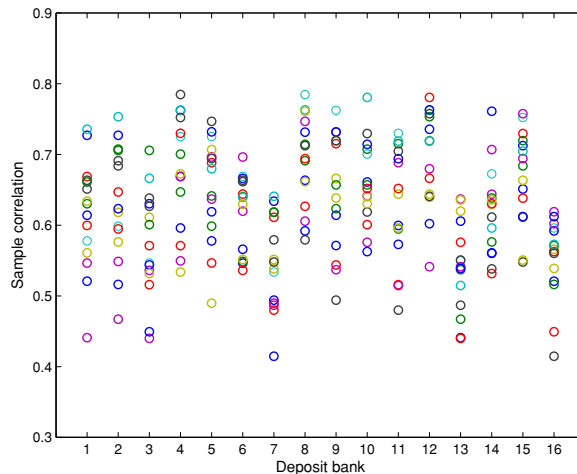
⁷<http://www2.fdic.gov/sod/index.asp>

City Corp. It is also interesting to note that Capital One Financial Corp. consistently has the lowest correlation with each of the other bank holding companies.

The correlations between the different bank stock returns are not constant over time, as shown by Figure 4.2. This graph plots the average daily correlation of the Bank of America Corp. with the remaining 15 banks, as obtained by means of the Riskmetrics Model of Morgan (1994).⁸ We observe substantial time-variation in this average correlation. Most notably, it increases considerable to a level around 0.8 during the financial crisis (2007-2008), which is particularly pronounced compared to the substantial decline to around 0.1 at the end of 2006. A similar pattern occurs in 2011. The opposite pattern is observed in 2003-4, where the average correlation drops even to negative values after a period of high values around 0.8. For the other banks we find similar patterns.

Figure 4.1: Sample correlations of daily returns of financial institutions

This figure depicts sample correlations of daily excess stock returns from 14 U.S. bank holding companies from November 11, 1994 until December 30, 2011 (4312 observations) and correlations of Wachovia Corp. and National Citi Corp. (nr. 3 and 7) from November 11, 1994 until December 30, 2008 (3556 observations). Each column i of small circles represents the sample correlations of stock i with the 15 other constituents.



⁸This model specifies the covariance matrix Σ_t for day t as $\Sigma_t = \lambda \Sigma_{t-1} + (1 - \lambda) r_{t-1} r'_{t-1}$. We use the default value of $\lambda = 0.94$.

This table lists 16 U.S. bank holding companies, together with the first and last year for which they belong to the top 15 of largest deposit banks in the United States. Permo denotes the CRSRP identifier. The table is modeled after Boudt, Danielsson, Koopman and Lucas (2012).

Table 4.1: Description of U.S. bank holding companies

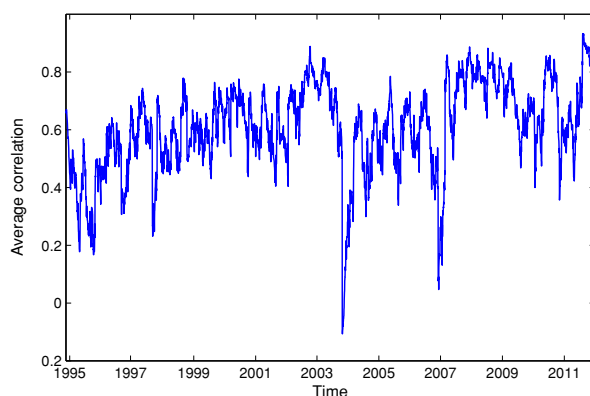
	Permo	Name	In top 15	Sample
1.	34746	Fifth Third Bancorp	2001 - 2011	1994 - 2011
2.	35044	Regions Financial Corp	2005 - 2011	1994 - 2011
3.	36469	First Union Corp, Wachovia Corp	1994 - 2008	1994 - 2008 ^a
4.	38703	Norwest Corp, Wells Fargo & Co	1994 - 2011	1994 - 2011
5.	47896	Chemical Banking Corp, Chase Manhattan Corp, JP Morgan Chase & Co	1994 - 2011	1994 - 2011
6.	49656	Bank of New York Mellon Corp	2008 - 2011	1994 - 2011
7.	56232	National City Corp	1996 - 2008	1994 - 2008 ^b
8.	59408	Nationsbank Corp, Bankamerica Corp, Bank of America Corp	1994 - 2011	1994 - 2011
9.	60442	PNC Bank Corp, PNC Financial Services GRP Inc	1994 - 2011	1994 - 2011
10.	64995	Keycorp	1994 - 2011	1994 - 2011
11.	66157	US Bancorp	1998 - 2011	1994 - 2011
12.	68144	Suntrust Banks Inc	1994 - 2011	1994 - 2011
13.	69032	Morgan Stanley	2009 - 2011	1994 - 2011
14.	70519	Citigroup	1999 - 2011	1994 - 2011
15.	71563	Southern National Corp NC, BB&T Corp	2000 - 2011	1994 - 2011
16.	81055	Capital One Financial Corp	2006 - 2011	1995 - 2011

^a Wells Fargo & Co and Wachovia announced on October 3, 2008 they had agreed to merge. The purchase has been completed on December 31, 2008.

^b PNC Financial Services announced October 24, 2008, its purchase of National City.

Figure 4.2: Average daily correlations of the Bank of America Corp. with other banks

This figure depicts average daily correlations between the Bank of America Corp. and the other U.S. deposit bank returns from November 11, 1994 until December 30, 2011 (4312 observations), estimated with Riskmetrics. The correlations between the Bank of America Corp. and Wachovia Corp. and National City Corp. are based on the sample November 11, 1994 until December 30, 2008 (3556 observations).



In the models of Section 4.2 we extend volatility and correlation models with a financial conditions index. In our empirical application we use the Bloomberg FCI. This daily index tracks the overall conditions in the U.S. money market, bond market and equity market. The FCI is an equal-weighted sum of three sub-indexes for each of these markets, see Rosenberg (2009). Table 4.2 shows the series of underlying indicators that form each sub-index, which are also equally-weighted. The underlying indicators as well as the FCI itself are presented as a Z-score, such that the index value should be interpreted as the number of standard deviations above or below its average, by construction over the period 1994-2008. Figure 4.3 shows the evolution of the Bloomberg FCI over time. The recent crisis in 2008 clearly stands out, but also other episodes of relatively low FCI values (below -2, say) can be associated with turmoil such as the Russian crisis and the LTCM bailout both in 1998, 9/11 in 2001, the WorldCom bankruptcy in September 2002 and the Lehman collapse in 2008.

Figure 4.3: The Bloomberg Financial Conditions Index

This figure depicts the daily Bloomberg FCI from January 3, 1994, through December 30, 2011. The horizontal line denotes the sample mean of the indicator over the period 1994-2008. A positive value of the Bloomberg FCI indicates better (future) financial conditions of the U.S. economy than on average while negative values means worsening of the financial conditions. The scale should be interpreted as the number of standard deviations above or below its average over the index 1994-2008.

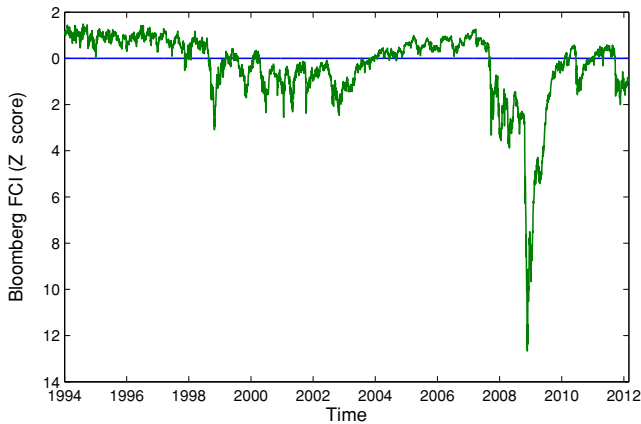


Table 4.2: Bloombergs' U.S. FCI components and weights

This table lists the major sub-indexes and their underlying indicators that form the Bloomberg FCI. The table is modeled after Rosenberg (2009).

Category/item	Weight	
Money Market		
Ted Spread	11.1 %	
Commercial Paper/ T-Bill Spread	11.1 %	
Libor-OIS Spread Spread	11.1 %	+
	33.3 %	
Bond Market		
Baa/Treasury Spread	6.7 %	
Muni/Treasury Spread	6.7 %	
Swaps/Treasury Spread	6.7 %	
High Yield/Treasury Spread	6.7 %	
Agency/Treasury Spread	6.7 %	+
	33.3 %	
Equity Market		
S&P 500 Share Prices	16.7 %	
VIX Index	16.7 %	+
	33.3 %	
		+
Total		100%

4.4 In-sample results

This section analyzes in-sample estimation results of the volatility and correlation models, applied to daily deposit bank excess returns. The in-sample period goes from November 11, 1994, through December 30, 2011 (4,312 observations). We consider a balanced sample, such that we can also maximize the regular likelihood as discussed in Section 4.2.3, and therefore remove Wachovia Corp. and National City Corp (see Table 4.1).

We first turn to the results for the volatility models. Table 4.3 provides estimation results for the Spline-GARCH model of (4.3). To preserve space we report estimates of Morgan Stanley and Citigroup. See Table 4.A.1 in Appendix 4.A.2 for results of the other banks. In Panel A we report estimates of the Spline-GARCH model with and without the FCI included as regressor. Specifically, the columns labeled (1) refer to the specification which does not include a financial conditions index, while the columns with label (2) add the Bloomberg FCI in the volatility specification. A key result from the table is that the coefficient on the FCI, κ_1 , is significant at the 1% level. The coefficient is negative, thus a decrease in financial conditions causes volatility on the assets to increase. The effect is stronger for Morgan Stanley than for Citigroup. An decrease of the FCI by 1 point, increases Morgan Stanleys volatility by 1.05 percentage points. Further, the leverage effect, captured by the γ parameter, reduces considerably for both banks when we include the Bloomberg FCI in the volatility specification. Hence a fraction of this effect is translated to economic fundamentals. Nevertheless, the α parameter indicates that the leverage effect is still stronger than the effect of the squared demeaned excess returns.

In Table 4.4 we report estimation results for the Factor Spline-GARCH model. In Panel A we report estimates of the volatility model of the idiosyncratic returns, as in (4.12), and in Panel B on the S&P 500 market return from our second model approach, as in (4.11). For comparability to Table 4.3 we focus on Morgan Stanley and Citigroup as well. The estimates of the remaining financial institutions are given in Table 4.A.2 of Appendix 4.A.2. The estimates of the idiosyncratic return as reported in Panel A are very similar to those reported in Panel A of Table 4.3. The impact of financial conditions on the volatility is somewhat weaker for the market factor (reported in Panel B), compared to the idiosyncratic return volatilities (-0.32 vs. -0.40). Again, a part of the leverage effect

translates to the FCI. Note that for the market return the leverages effect together with the FCI are the source of innovations for the volatility, while the squared demeaned excess market return does not influence the volatility.

Figures 4.4 and 4.5 depicts the in-sample volatilities of both banks for the period 1994 - 2011 according to the two different model set-ups. The upper part of Figure 4.4 represents estimates of the square root of the unit GARCH term g_t of the Spline-GARCH model without the FCI, while presented in (4.3), while the middle and the bottom part shows estimates of the decomposed volatility terms from the Spline-GARCH-X model with the Bloomberg FCI included as exogenous variable: the square root of the unit GARCH term g_t (middle) and the square root of the “spline” component τ_t (bottom) containing the Bloomberg FCI. The figure indicates that including the Bloomberg FCI in the “spline” component τ_t results in less severe spikes of the unit GARCH term, as they are now incorporated in the τ component. For both banks the estimated volatility of the Spline-GARCH-X model is substantially higher during certain turmoil periods, especially in 1998 (Asia crisis) and the second half of 2007 (credit crunch due to subprimes).

Figure 4.5 visualizes again in-sample volatilities of Morgan Stanley and Citigroup, but now corresponding with the factor Spline-GARCH approach from Section 4.2.2. In this figure the top part corresponds to the Factor-Spline-GARCH without the FCI included, and the bottom part to the Factor-Spline-GARCH model with the FCI included. In this case the differences between the volatility patterns is small. The reason could be due to the S&P500 factor, which is now subtracted from the excess returns. At the same time, the S&P 500 is also part of the Bloomberg FCI (see Table 4.2). Hence it could be that this factor dominates the other variables that are part of the Bloomberg FCI. Nevertheless, some peaks are higher, especially at the two aforementioned times (1998 and 2007).

The bottom part of Table 4.3 shows estimation results of alternative correlation models of Section 4.2.1, using daily returns of all 14 deposit banks. All models use standardized returns corresponding to specification where the volatilities are “corrected” for the impact of the FCI on volatility. For both the DC-X and cDCC-X correlation models, the FCI has a significant impact on the correlations between asset returns. A lower value of the FCI means worse financial conditions and increases correlations. This finding is in line with

Table 4.3: In-sample estimation results - Spline-GARCH Models

This table reports the estimation results of the cDCC(-X) and DC-X model. The volatility is formulated as a Spline-GARCH-X model:

$$\begin{aligned} r_{i,t} &= \mu_i + \sqrt{g_{i,t}\tau_{i,t}}\varepsilon_{i,t}, \\ g_{i,t} &= (1 - \alpha_i - \delta_i - \gamma_i/2) + \alpha_i \frac{(r_{i,t-1} - \mu_i)^2}{\tau_{i,t-1}} + \gamma_i \frac{(r_{i,t-1} - \mu_i)^2}{\tau_{i,t-1}} I[(r_{i,t-1} - \mu_i) < 0] + \delta_i g_{i,t-1}, \\ \tau_{i,t} &= \exp(\kappa_{i,0} + \kappa_{i,1} FCI_{t-1}), \end{aligned}$$

and the correlation part is given by the DC-X model:

$$Q_t = \bar{Q} + \gamma_{DC} FCI_{t-1} J,$$

or by the cDCC(-X) model:

$$\begin{aligned} Q_t &= \bar{Q}_t + \alpha_{cDCC} (Q_{t-1}^{*1/2} \epsilon_{t-1}' \epsilon_{t-1} Q_{t-1}^{*1/2} - \bar{Q}_{t-1}) + \beta_{cDCC} (Q_{t-1} - \bar{Q}_{t-1}), \\ \bar{Q}_t &= \bar{Q} + \gamma_{cDCC} FCI_{t-1} J, \end{aligned}$$

with J a matrix of ones. We scale Q_t from each model to obtain a valid correlation matrix:

$$R_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2},$$

with $Q_t^* = \text{diag}(q_{11,t}, \dots, q_{KK,t})$. Further, $r_{i,t}$ the daily excess return of holding bank i , ($i = 1, \dots, 14$), and FCI_t represents the Bloomberg FCI. The first part of the table shows Maximum Likelihood estimates of the Spline-GARCH(-X) model, assuming a conditional Student- t distribution for ε_t . We show the volatilities corresponding to Morgan Stanley and Citigroup. Panel B shows estimation results of maximizing the Composite likelihood and normal likelihood, using all 14 financial institutions. The log-likelihood values of the DCC and cDCC-X models correspond with their regular likelihood values, obtained by plugging in the optimized Composite likelihood estimates. Standard errors are in parentheses. The sample goes from November 11, 1994, through December 30, 2011 (4,312 observations).

Panel A: Volatility part				
	Morgan Stanley		Citigroup	
	(1)	(2)	(1)	(2)
μ	0.029 (0.027)	0.021 (0.030)	0.015 (0.019)	0.006 (0.033)
α	0.025 (0.007)	0.029 (0.008)	0.036 (0.007)	0.031 (0.007)
δ	0.923 (0.009)	0.939 (0.012)	0.928 (0.009)	0.948 (0.009)
γ	0.094 (0.014)	0.046 (0.014)	0.069 (0.012)	0.038 (0.014)
κ_0	2.368 (0.309)	1.891 (0.201)	2.167 (0.399)	1.844 (0.347)
κ_1		-0.403 (0.050)		-0.323 (0.089)
ν	7.203 (0.667)	7.460 (0.759)	7.151 (0.608)	7.072 (0.613)
Log-lik	-9900	-9877	-9228	-9211

Panel B: Correlation part			
	DC-X	cDCC	cDCC-X
γ_{DC}	-0.214 (0.005)		
α_{cDCC}		0.026 (0.001)	0.019 (0.001)
β_{cDCC}		0.964 (0.002)	0.970 (0.002)
γ_{cDCC}			-0.220 (0.013)
Log-lik	20835	20296	21023
Comp-lik		11748	11904

Table 4.4: In-sample estimation results - Factor-Spline-GARCH models

This table reports the estimation results of the Factor-Spline-GARCH model. The volatility is formulated as a Spline-GARCH-X model:

$$\begin{aligned} r_{i,t} &= \mu_{i,t} + \sqrt{g_{i,t}\tau_{i,t}}\varepsilon_{i,t}, \\ g_{i,t} &= (1 - \alpha_i - \delta_i - \gamma_i/2) + \alpha_i \frac{(r_{i,t-1} - \mu_{i,t})^2}{\tau_{t-1}} + \gamma_i \frac{(r_{i,t-1} - \mu_{i,t})^2}{\tau_{t-1}} I[(r_{i,t-1} - \mu_{i,t}) < 0] + \delta_i g_{i,t-1}, \\ \tau_{i,t} &= \exp(\kappa_{i,0} + \kappa_{i,1} FCI_{t-1}), \end{aligned}$$

and the correlation part is given by the cDCC model:

$$Q_t = (1 - \alpha_{cDCC} - \beta_{cDCC})\bar{Q} + \alpha_{cDCC}(Q_{t-1}^{*1/2}\epsilon_{t-1}'\epsilon_{t-1}Q_{t-1}^{*1/2}) + \beta_{cDCC}Q_{t-1},$$

where we scale Q_t from each model to obtain a valid correlation matrix:

$$R_t = Q_t^{*-1/2}Q_tQ_t^{*-1/2},$$

with $Q_t^* = \text{diag}(q_{11,t}, \dots, q_{KK,t})$. Further, $r_{i,t}$ the daily excess return of holding bank i , ($i = m, 1, \dots, 16$) and the S&P 500 market return. In addition, $\mu_{i,t} = \xi_i + \beta_i r_{m,t}$ with $r_{m,t}$ the excess market return at day t . Otherwise, $\mu_{i,t} = \mu_m$ in case of the market return. FCI_t represents the Bloomberg FCI and $Q_t^* = \text{diag}(q_{mm,t}, q_{11,t}, \dots, q_{KK,t})$. The first part of the table shows Maximum Likelihood estimates of the Spline-GARCH(-X) model, assuming a conditional Student- t distribution for ε_t . We show the volatilities corresponding to Morgan Stanley and Citigroup. Panel B shows similar results, but for the daily S&P500 returns. Panel C shows estimation results of maximizing the Composite likelihood (CI) using the standardized returns of all assets including the market return. The log-likelihood value of the cDCC model corresponds with the regular likelihood, obtained by plugging in the optimized Composite likelihood estimates. Standard errors are in parentheses. The sample goes from November 11, 1994 through December 30, 2011 (4312 observations).

Panel A: Idiosyncratic volatility				
	Morgan Stanley		Citigroup	
	(1)	(2)	(1)	(2)
β	1.605 (0.023)	1.612 (0.023)	1.301 (0.020)	1.306 (0.020)
α	0.028 (0.007)	0.018 (0.005)	0.065 (0.012)	0.054 (0.012)
δ	0.943 (0.009)	0.976 (0.007)	0.914 (0.012)	0.929 (0.013)
γ	0.050 (0.011)	0.010 (0.009)	0.036 (0.014)	0.026 (0.012)
κ_0	1.649 (0.315)	1.203 (0.278)	1.588 (0.442)	1.241 (0.361)
κ_1		-0.401 (0.039)		-0.401 (0.053)
ν	5.537 (0.393)	5.918 (0.453)	5.987 (0.445)	6.154 (0.485)
Log-lik	-8362	-8327	-7739	-7715

Panel B: Market volatility		
	(1)	(2)
μ	0.042 (0.012)	0.036 (0.013)
α	0.000 (0.009)	0.000 (0.008)
δ	0.923 (0.012)	0.924 (0.012)
γ	0.133 (0.017)	0.114 (0.017)
κ_0	0.037 (0.336)	-0.101 (0.112)
κ_1		-0.316 (0.064)
ν	9.194 (1.144)	9.546 (1.293)
Log-lik	-6126	-6116

Panel C: Correlations	
	cDCC
α_{cDCC}	0.010 (0.001)
β_{cDCC}	0.987 (0.001)
Log-lik	8478
Comp-lik	4051

Figure 4.4: In-sample volatilities - Spline-GARCH models

This figure depicts daily estimated volatilities of Morgan Stanley and Citigroup from November 11, 1994, through December 30, 2011. The top part denotes estimates of the square root of the unit GARCH term g_t from the Spline-GARCH model without exogenous variables included. The middle and bottom part shows estimates of the GARCH term $\sqrt{g_t}$ (middle part) and the evolution of the 'spline' component $\sqrt{\tau_t}$ (bottom part) from the Spline-GARCH-X model with the Bloomberg FCI included as exogenous variable.

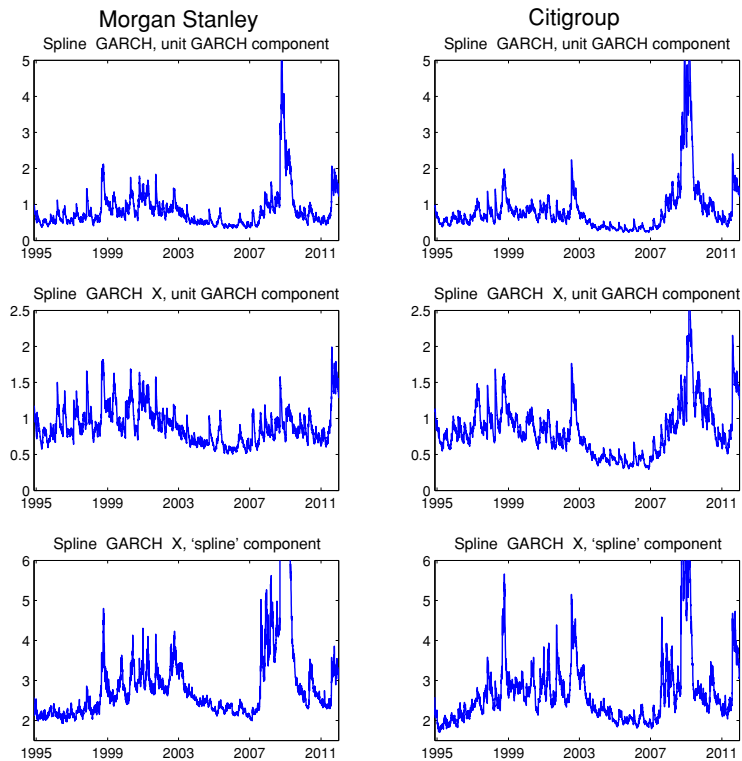
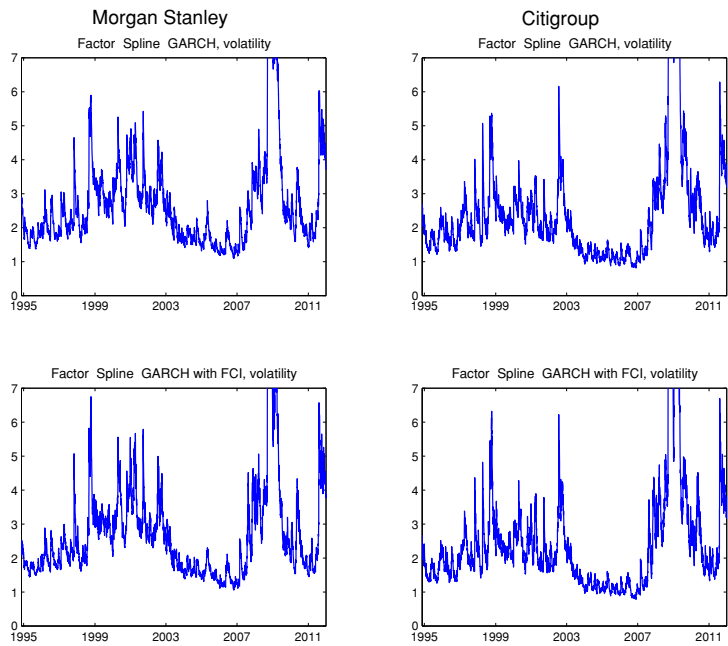


Figure 4.5: In-sample volatilities - Factor-Spline-GARCH models

This figure depicts daily estimated volatilities of Morgan Stanley and Citigroup from November 11, 1994, through December 30, 2011. The top part denotes estimates of the conditional volatility $\sqrt{h_t}$ from the Factor-Spline-GARCH model without exogenous variables included. The bottom part shows estimates of conditional volatility from the Factor-Spline-GARCH model with the Bloomberg FCI included as exogenous variable.



the literature (Longin and Solnik, 2001; Ang and Chen, 2002) that documents a stronger connection between asset returns during bear markets than during bull markets.

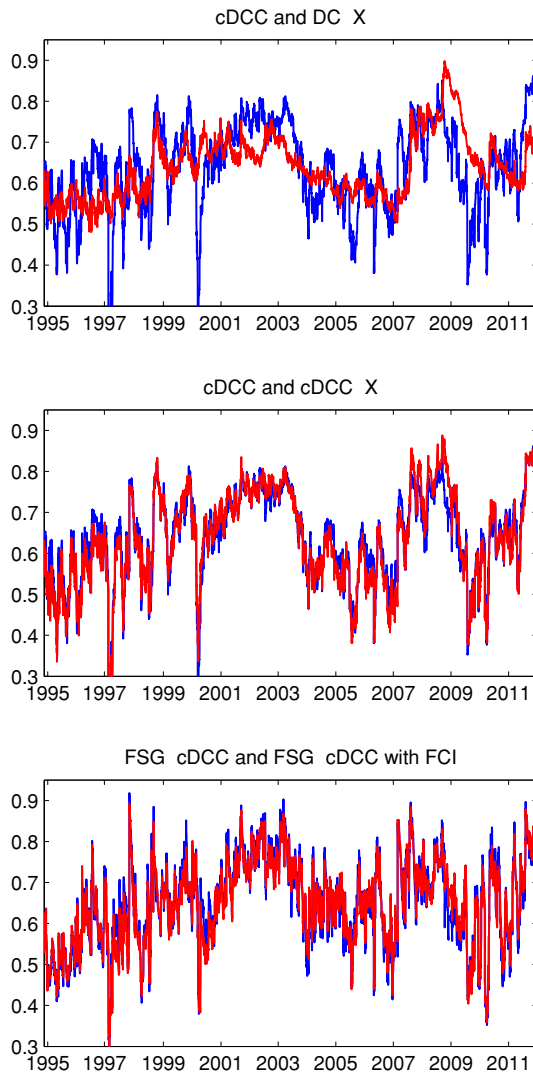
Figure 4.6 highlights the effect of including the FCI into correlation models by plotting the estimated correlation between Morgan Stanley and Citigroup. The three subgraphs correspond to the DC-X, cDCC, the cDCC-X and the Factor-Spline-GARCH model. As ‘true correlation’ is not observable, it is not possible to judge which of these correlation pattern is better.⁹ Nevertheless the plots highlight some interest in differences in the patterns, from which we draw three main conclusions. First, the DC-X model is less volatile than the Factor-Spline-GARCH and the cDCC(-X) models while during historical events like the 2007-2008 crisis the DC-X model produces a higher correlation than the remaining models. The average difference in correlation between the DC-X model and the cDCC(-X) model during the period October 1 through December 31, 2008, is equal to 0.15. All cDCC type of models shows a considerable drop in correlations during 1997, 2000, and 2009, while the DC-X model does not. A second finding is the similarity between the Factor-Spline-GARCH correlations and the cDCC(-X) correlations, although the former is more volatile and exhibits more spikes. Hence it seems the one-factor approach with cDCC correlations for the idiosyncratic returns captures the same dynamics as the cDCC-X correlations for the excess returns. Third, the inclusion of a financial conditions index in the cDCC model or in the Factor-Spline-GARCH model does play a role, although the effect is more visible in the cDCC model than in the Factor-Spline-GARCH model. Similar as in the volatilities, it seems that the market return factor captures a part of the effect of the Bloomberg FCI.

In summary, the results suggest that the Bloomberg FCI has a statistically significant impact on both volatilities and correlations of bank returns. Lower financial conditions will increase the volatilities of bank returns as well as the correlation between financial institutions, in particular during crises periods.

⁹In the next section, we analyse the economic impact of the different correlation patterns by means of estimating portfolio VaRs.

Figure 4.6: In-sample correlations

This figure depicts the estimated correlation between daily excess returns of Morgan Stanley and Citigroup from November 11, 1994, through December 30, 2011 (4,312 observations) according to five different correlation models. The upper part shows in-sample correlations from the cDCC (blue line) and the DC-X model of (4.4) (red line), while the middle corresponds with estimated correlations of again the cDCC model (blue line) and the cDCC-X model (red line). The bottom part provides the estimated correlations of the Factor-Spline-GARCH model (4.9)-(4.12) with (blue line) and without (red line) including the Bloomberg FCI.



4.5 Estimating portfolio Value-at-Risk

We assess the economic value of including the Bloomberg FCI into correlations and volatilities by considering portfolio Value-at-Risk estimates. The VaR is a particular quantile of the conditional portfolio distribution which is of interest for risk-management in general. Based on this number, financial institutions for example have to decide how much capital should be hold for possible losses.

In the context of our models presented in section 4.2, we have assumed a particular specification for the conditional mean and variance. It follows then that the $q\%$ portfolio VaR at time t can be expressed as

$$VaR_t^q = \mu_{P,t} + z_{1-q} \sqrt{w_t' H_t w_t}, \quad (4.17)$$

with $\mu_{P,t}$ the portfolio conditional mean, w_t the portfolio weights, and H_t the (forecasted) portfolio covariance matrix based on the Factor-Spline-GARCH (with and without the FCI included), the cDCC(-X) and the DC-X models, such that $\sqrt{w_t' H_t w_t}$ denotes the portfolio standard deviation at time t . Finally z_{1-q} represents the $(1 - q)$ -th quantile of the standardized portfolio distribution.

In the analysis, we consider a equally-weighted portfolio, such that $w_t = w = 1/K$, with K the number of assets. In addition, $\mu_{P,t} = \mu_P = \mu' w$ with μ defined in (4.1) in case of the cDCC-X and DC-X models and $\mu = \alpha + \alpha_m \beta$ in case of the Factor-Spline-GARCH model (see (4.9)-(4.13)). We use an expanding-window to estimate the volatility/correlation parameters, starting with a sample of 1500 observations, which corresponds to the years 1995-2000.¹⁰ We re-estimate the parameters repeatedly after each 50 observations, which corresponds roughly to two months. At each point in time, we compute a 99%, 95% and 90% 1-day portfolio VaR. In addition, we consider 95% and 90% 10-day portfolio VaRs and estimate a suitable ARMA process to forecast the Bloomberg FCI 10 steps ahead.¹¹ Finally, we create 10 different sub-series in order to circumvent

¹⁰Inclusion of the FCI makes most sense in an expanding window framework, as the FCI is quite calm in non-crises periods and volatile in crises periods. A moving window could therefore contain no crises, and hence the inclusion of the FCI results in little forecasting power.

¹¹We do not forecast 10-step ahead at once, as the daily dynamics of the FCI fits better to an ARMA process than the 10-day dynamics.

possible serial correlation in the 10-day portfolio VaRs. Hence sub-series j contains the estimates $\{VaR_j^q, VaR_{j+10}^q, VaR_{j+20}^q, \dots\}$ for $j = 1, \dots, 10$.

Let us continue by focusing on the quantile z_{1-q} in (4.17) of the standardized portfolio returns. We emphasize the importance of this number as it has a direct impact on the VaR. Given the non-normality of our financial institution excess returns, the use of the traditional values corresponding with the standard-normal distribution is not desirable. We therefore estimate z_{1-q} using two different approaches. The first approach considers the empirical quantiles corresponding with the in-sample standardized portfolio returns which is most reliable as one does not make a parametric judgement about the distribution. The second approach estimates a Student- t distribution on the in-sample standardized portfolio returns such that $z_{1-q} = t_{\nu_P}^{-1}(1-q)$. By doing so, we avoid a computationally involved estimation of our multivariate models assuming a conditional Multivariate Student- t distribution.¹²

Finally, we backtest our estimated portfolio VaR using the unconditional and conditional coverage tests and independence test proposed by Kupiec (1995) and Christoffersen (1998).¹³ We follow the suggestion of Diebold *et al.* (1998) by using Bonferroni bounds for the 10 sub-series in our multi-step ahead VaR estimates. That is, we assume that the VaR series has autocorrelation up to and including lag 9, whereas each sub-series should have correct coverage and independent VaR violations. We therefore backtest each sub-series separately with a size of α/J . Rejecting the null hypothesis of unconditional coverage/independence occurs when the null is rejected for *any* of the 10 sub-series.

Tables 4.5 and 4.6 shows results of the 1-day and 10-day portfolio VaR estimates. In the former table, we report the test results for the 99%, 95% and 90% VaRs of the various volatility and correlation models, while the latter table shows only the tests corresponding with the 95% and 90% VaR estimates. In the left hand part we show the results where the quantiles are calculated using empirical quantiles, and in the right part the results with

¹²A second argument is the fact that if univariate returns $r_{i,t}$ ($i = 1, \dots, J$) are Student- $t(\nu_i)$ distributed, the weighted average of these K return vectors may not follow a Student- t distribution.

¹³See Appendix 4.A.1 for more detailed information about these tests. As indicated by Santos *et al.* (2013), these tests do not rank the performance of each model. Statistical tests that do take this into account are for example the CPA test of Giacomini and White (2006) or the encompassing test of Giacomini and Komunjer (2005). However, as these tests require an in-sample window of a fixed length we do not use these tests. In addition, as it is our aim to investigate whether including the FCI leads to a better coverage and/or independence a ranking of the various methods is not a necessity.

Table 4.5: Evaluation of 1-day portfolio Value-at-Risk estimates

This table reports the performance of various models in estimating the $q\%$ 1-day portfolio VaR ($q = 99\%, 95\%, 90\%$), using an equally weighted portfolio over the period 2001-2011. The first column denotes the (correlation) model, where FCI means that this variable is included in the conditional variance specification of the individual assets (market factor and idiosyncrasies). FSG is the abbreviation of the Factor-Spline-GARCH model of Section 4.2.2. $V(\%)$ denotes the number of violations (percentage w.r.t. the total number of estimates). p_{uc} , p_{ind} and p_{cc} are p -values of the (un)conditional coverage test and the independence test of Christoffersen (1998). The number of estimated VaRs is equal to 3,025.

Model	V(%)	p_{uc}	p_{ind}	p_{cc}	V(%)	p_{uc}	p_{ind}	p_{cc}
	z_{1-q}	using emp. quantile			z_{1-q}	$t_{\nu_P}^{-1}(1-q)$		
99% VaR								
cDCC	34 (1.12)	0.502	0.379	0.542	34 (1.12)	0.502	0.399	0.559
FCI-cDCC	33 (1.09)	0.620	0.393	0.615	30 (0.99)	0.964	0.438	0.740
FCI-cDCC-X	33 (1.09)	0.620	0.393	0.615	33 (1.09)	0.620	0.393	0.615
FCI-DC-X	35 (1.16)	0.397	0.365	0.464	35 (1.16)	0.397	0.365	0.464
FSG	32 (1.06)	0.751	0.351	0.615	32 (1.06)	0.751	0.351	0.615
FCI-FSG	32 (1.06)	0.751	0.408	0.675	32 (1.06)	0.751	0.351	0.615
95% VaR								
cDCC	187 (6.18)	0.004	0.892	0.016	172 (5.69)	0.090	0.788	0.229
FCI-cDCC	167 (5.52)	0.196	0.789	0.418	161 (5.32)	0.421	0.615	0.637
FCI-cDCC-X	164 (5.42)	0.294	0.700	0.535	169 (5.59)	0.146	0.850	0.341
FCI-DC-X	170 (5.62)	0.125	0.628	0.274	179 (5.92)	0.024	0.653	0.071
FSG	197 (6.51)	0.000	0.731	0.001	181 (5.98)	0.016	0.785	0.053
FCI-FSG	177 (5.85)	0.036	0.834	0.109	161 (5.32)	0.421	0.615	0.637
90% VaR								
cDCC	330 (10.91)	0.100	0.446	0.193	331 (10.94)	0.088	0.423	0.170
FCI-cDCC	326 (10.78)	0.159	0.872	0.366	319 (10.55)	0.321	0.654	0.553
FCI-cDCC-X	319 (10.55)	0.321	0.946	0.610	323 (10.68)	0.219	0.637	0.420
FCI-DC-X	329 (10.88)	0.113	0.969	0.284	340 (11.24)	0.026	0.748	0.078
FSG	361 (11.93)	0.001	0.371	0.002	354 (11.70)	0.002	0.247	0.005
FCI-FSG	329 (10.88)	0.113	0.360	0.187	328 (10.84)	0.127	0.381	0.212

the Student- t distribution. From the table it is clear that including the FCI leads to an improvement, especially regarding the 95% and 90% VaR estimates. For example, when the empirical quantile is used for z_{1-q} , the cDCC and the Factor-Spline-GARCH model both fail on the correct frequency in case of the 95% VaR, while the cDCC-X, the DC-X and cDCC model with the FCI in the volatilities do not. This result holds also for all aforementioned models except the DC-X model when the student- t distribution is used for z_{1-q} and one takes a significance level of 10%. Even the Factor-Spline-GARCH model with the FCI included in the variances fails in its coverage in case of 95% VaR estimates. It seems that including the FCI in both the volatility as well as in the correlations, i.e. the cDCC-X model, delivers the best coverage, although the DC-X model fails when one uses an Student- t distribution for the 95% and 90% VaR estimates. This could be due to parameter uncertainty around estimating ν , or due to the fact that the standardized portfolio returns does not closely match a Student- t distribution.

The influence of the FCI becomes smaller for 10-day VaR estimates, as indicated by Table 4.6. This is not surprising, as forecasting the Bloomberg FCI 10-steps ahead introduces more uncertainty than forecasting the same quantity one step ahead. Moreover, the effect of FCIs could be short-lived, similar to the effect of realized measures in (multivariate) volatility models, see for example Noreldin *et al.* (2012). In addition, the statistical tests have less power as the number of expected violations are lower, compared to the same tests applied to the 1-day VaR estimates. Nevertheless, the 95% and 90% VaR estimates corresponding with the Factor-Spline-GARCH model fails for the coverage test at the 1% level when using the empirical distribution for z_{1-q} . Even if one uses the Student- t distribution, the null of correct coverage is rejected at a 5% level. At the same time, the cDCC model is not statistically outperformed.

All in all, we conclude that including the FCI into volatility and correlation modeling improves the VaR estimates in case of unconditional coverage, especially at the short horizon. For the 10-day horizon, the evidence is less strong, although the Factor-Spline-GARCH model is still outperformed by the same model extended with the FCI.

Table 4.6: Evaluation of 10-day portfolio Value-at-Risk estimates

This table reports the performance of various models in estimating the $q\%$ 10-day portfolio VaR ($q = 95\%, 90\%$), using an equally weighted portfolio over the period 2001-2011. The first column denotes the (correlation) model, whether FCI means that this variable is included in the conditional variance specification of the individual assets (market factor and idiosyncrasies), FSG is the abbreviation of the Factor-Spline-GARCH model of Section 4.2.2. For each model, we have 10 different sub-series of VaR estimates. We report (test) results of the sub-series that has the lowest p -value of the test on unconditional coverage of the VaR estimates. $V(\%)$ denotes the number of violations (percentage w.r.t. the total number of VaR estimates), p_{uc} , p_{ind} and p_{cc} are the p -values of the tests on (un)conditional coverage and independence of the VaR estimates. The number of estimated VaRs is equal to 303.

Model	$V(\%)$		p_{uc}		p_{ind}		p_{cc}		$V(\%)$		$p_{uc} = t_{\nu_P}^{-1}(1-q)$		p_{ind}		p_{cc}	
	z_{1-q}		using emp. quantile								z_{1-q}					
95% VaR																
cDCC	21 (6.93)	0.144	0.667	0.314	0.667	0.314	0.667	0.314	20 (6.60)	0.222	0.754	0.452	0.754	0.452	0.754	0.452
FCI-cDCC	11 (3.63)	0.251	0.362	0.341	0.362	0.341	0.362	0.341	11 (3.63)	0.251	0.362	0.341	0.362	0.341	0.362	0.341
FCI-cDCC-X	11 (3.64)	0.256	0.361	0.346	0.361	0.346	0.361	0.346	11 (3.64)	0.256	0.361	0.346	0.361	0.346	0.361	0.346
FCI-DC-X	20 (6.62)	0.217	0.796	0.451	0.796	0.451	0.796	0.451	20 (6.62)	0.217	0.796	0.451	0.796	0.451	0.796	0.451
FSG	31 (10.26)	0.000	0.289	0.001	0.289	0.001	0.289	0.001	27 (8.94)	0.004	0.102	0.005	0.102	0.005	0.102	0.005
FCI-FSG	21 (6.95)	0.140	0.707	0.314	0.707	0.314	0.707	0.314	20 (6.62)	0.217	0.172	0.184	0.172	0.184	0.172	0.184
90% VaR																
cDCC	38 (12.58)	0.149	0.916	0.350	0.916	0.350	0.916	0.350	38 (12.58)	0.149	0.916	0.350	0.916	0.350	0.916	0.350
FCI-cDCC	24 (7.92)	0.212	0.942	0.458	0.942	0.458	0.942	0.458	24 (7.92)	0.212	0.942	0.458	0.942	0.458	0.942	0.458
FCI-cDCC-X	23 (7.59)	0.146	0.842	0.341	0.842	0.341	0.842	0.341	24 (7.92)	0.212	0.942	0.458	0.942	0.458	0.942	0.458
FCI-DC-X	36 (11.92)	0.279	0.659	0.505	0.659	0.505	0.659	0.505	36 (11.92)	0.279	0.659	0.505	0.659	0.505	0.659	0.505
FSG	50 (16.56)	0.000	0.354	0.001	0.354	0.001	0.354	0.001	50 (16.56)	0.000	0.354	0.001	0.354	0.001	0.354	0.001
FCI-FSG	38 (12.58)	0.149	0.863	0.347	0.863	0.347	0.863	0.347	37 (12.25)	0.206	0.759	0.429	0.759	0.429	0.759	0.429

4.6 The role of the VIX in the FCI

This chapter shows that including the FCI into volatility and correlation models gives a better in-sample fit as well as better VaR estimates. The FCI is composed of various components, including money market spreads, bond market spreads and S&P500 share prices and the VIX index (see Table 4.2). The question arises which of these elements drives our results. As shown in the literature, e.g. Koopman *et al.* (2005), the VIX is a powerful predictor of future volatility. To examine to which extent this measure drives the results on the FCI index and to what level the other components of the Bloomberg FCI have explaining/forecasting power beyond the VIX, we re-consider our in-sample and VaR analyses, now including both the FCI and VIX index.

First, we rewrite the Spline-GARCH specification in our two modeling approaches. We change the τ_t component (omitting the subscript i) as follows to include also the VIX index:

$$\tau_t = \exp(\kappa_0 + \kappa_1 FCI_{t-1} + \kappa_2 VIX_{t-1}). \quad (4.18)$$

Note that the entire specification of the conditional variance of both model approaches is given in (4.3) and (4.9)-(4.12). If the significance of the FCI_{t-1} vanishes, then the remaining components of the Bloomberg FCI do not have impact on the volatilities of the excess bank returns, market factor or idiosyncratic bank returns. Table 4.7 reports estimation results of the model parameters corresponding with Morgan Stanley, Citigroup and the market factor. In the Spline-GARCH specifications the FCI variable is significant for both financial institutions. In addition, the FCI has also statistical impact on the market volatility beyond the VIX. Using the Factor-Spline-GARCH specification, the VIX seems to take away the significance in case of Citigroup.¹⁴ Nevertheless, as the FCI is significant for the market factor and Morgan Stanley, we conclude the FCI contains information beyond the VIX that impacts the volatility of asset returns. Finally, we analyse the possible influence of the VIX on the correlation between our deposit banks. The results (not shown here) suggest that the VIX does not have statistical influence on the correlations between the banks. Hence the other components of the Bloomberg FCI are additional drivers of

¹⁴With regards to the other financial institutions of our study, the result is qualitatively similar.

Table 4.7: In-sample estimation results with the VIX included

This table reports the estimation results of the (Factor-) Spline-GARCH-X model:

$$\begin{aligned}
 h_t &= g_t \tau_t, \\
 g_t &= (1 - \alpha - \delta - \gamma/2) + \alpha \frac{(r_{t-1} - E_{t-1}(r_{t-1}))^2}{\tau_{t-1}} + \delta g_{t-1} \\
 &\quad + \gamma \frac{(r_{t-1} - E_{t-1}(r_{t-1}))^2}{\tau_{t-1}} I[(r_{t-1} - E_{t-1}(r_{t-1})) < 0], \\
 \tau_t &= \exp(\kappa_0 + \kappa_1 FCI_{t-1} + \kappa_2 VIX_{t-1}),
 \end{aligned}$$

with r_t the excess daily return or the idiosyncratic return of Morgan Stanley and Citigroup or the S&P 500 market return r_m . Further, E_{t-1} denotes the conditional expectation using all information up to and including time $t-1$, FCI_t represents the Bloomberg FCI, VIX_t is the VIX at time t . The table shows Maximum Likelihood estimates, assuming a conditional Student- t distribution for the standardized returns $(r_t - E_{t-1}(r_t))/\sqrt{h_t}$. (F)-Spl-GARCH denotes the (Factor)-Spline-GARCH model. Standard errors are in parentheses. The sample goes from November 11, 1994, through December 30, 2011 (4,312 observations).

	Morgan Stanley		Citigroup		Market factor
	Spl-GARCH	F-Spl-GARCH	Spl-GARCH	F-Spl-GARCH	S&P 500 ret.
α	0.029 (0.009)	0.015 (0.004)	0.023 (0.006)	0.054 (0.013)	0.000 (0.014)
β	0.940 (0.019)	0.977 (0.005)	0.965 (0.008)	0.929 (0.014)	0.917 (0.016)
δ	0.029 (0.013)	0.011 (0.007)	0.020 (0.008)	0.024 (0.012)	0.077 (0.020)
κ_0	0.632 (0.286)	0.682 (0.323)	0.759 (0.376)	0.427 (0.415)	-1.590 (0.195)
κ_1	-0.172 (0.061)	-0.257 (0.060)	-0.093 (0.065)	-0.196 (0.077)	-0.092 (0.049)
κ_2	0.551 (0.113)	0.285 (0.091)	0.519 (0.103)	0.402 (0.109)	0.707 (0.093)
ν	7.600 (0.811)	5.953 (0.466)	7.153 (0.646)	6.134 (0.495)	10.611 (1.620)
Log-lik	-9865	-8322	-9200	-7709	-6096

our main result that financial conditions have impact on correlations.

Second, we perform VaR estimates in the same spirit as the previous section. We consider the Factor-Spline-GARCH model, the cDCC model and the DC-X model with only the VIX included in the conditional variance specification and investigate whether similar results are made as the models that contains the Bloomberg FCI in the corresponding specification. Table 4.8 shows that including only the VIX does not lead to correct coverage. Regarding the 1-day 95% and 90% VaR, the cDCC-X, DC-X and Factor-Spline-GARCH model, all significantly overestimate the number of violations. Only the cDCC model with the VIX does not fail when a Student- t distribution is used. Inspecting the 10-day VaR leads to the finding that all models perform well, except the Factor-Spline-GARCH model. The minimum p -value takes the value 0.009 for the 95% VaR estimates, hence we reject the null hypotheses of unconditional coverage at the 10% significance level. We conclude that the other components of the FCI also have a significant impact on the VaR estimates, in particular at the short-horizon.

Table 4.8: Evaluation of 1- and 10-day portfolio Value-at-Risk estimates with the VIX index

This table reports the performance of various models in estimating the $q\%$ 1- (Panel A) and 10-day (Panel B) portfolio VaR ($q = 99\%, 95\%, 90\%$ for 1-day VaRs and $q = 95\%, 90\%$ for 10-day VaRs), using an equally weighted portfolio over the period 2001-2011. All models contain the VIX index in its conditional variance specification. FSG is the abbreviation of the Factor-Spline-GARCH model of Section 4.2.2. $V(\%)$ denotes the number of violations (percentage w.r.t. the number of estimates). p_{uc} , p_{ind} and p_{cc} are p -values of the (un)conditional coverage test and the independence test of Christoffersen (1998). For the 10-day portfolio VaRs, we have 10 different sub-series. In Panel B, we report (test) results corresponding with the sub-series that contains the lowest p -value on the unconditional coverage test. The number of estimated VaRs is equal to 3025 (1-day) and 303 (10-day) respectively.

Panel A: 1-day VaR estimates								
Model	$V(\%)$	p_{uc}	p_{ind}	p_{cc}	$V(\%)$	p_{uc}	p_{ind}	p_{cc}
	z_{1-q} using emp. quantile				$z_{1-q} = t_{\hat{V}_P}^{-1}(1 - q)$			
99% VaR								
cDCC	31 (1.02)	0.891	0.327	0.613	30 (0.99)	0.964	0.305	0.590
cDCC-X	33 (1.09)	0.620	0.375	0.596	34 (1.12)	0.502	0.399	0.559
DC-X	38 (1.26)	0.173	0.503	0.316	41 (1.36)	0.062	0.586	0.152
FSG	33 (1.09)	0.620	0.375	0.596	30 (0.99)	0.964	0.305	0.590
95% VaR								
cDCC	179 (5.92)	0.024	0.653	0.071	168 (5.55)	0.169	0.573	0.332
cDCC-X	175 (5.79)	0.053	0.543	0.128	171 (5.65)	0.106	0.911	0.270
DC-X	188 (6.21)	0.003	0.829	0.012	191 (6.31)	0.001	0.984	0.006
FSG	212 (7.01)	0.000	0.969	0.000	194 (6.41)	0.001	0.656	0.003
90% VaR								
cDCC	334 (11.04)	0.060	0.465	0.131	327 (10.81)	0.142	0.403	0.240
cDCC-X	339 (11.21)	0.030	0.579	0.081	340 (11.24)	0.026	0.552	0.069
DC-X	352 (11.64)	0.003	0.477	0.011	368 (12.17)	0.000	0.409	0.000
FSG	366 (12.10)	0.000	0.146	0.000	353 (11.67)	0.003	0.264	0.006
Panel B: 10-day VaR estimates								
	$V(\%)$	p_{uc}	p_{ind}	p_{cc}	$V(\%)$	p_{uc}	p_{ind}	p_{cc}
95% VaR								
cDCC	21 (6.95)	0.140	0.223	0.160	11 (3.63)	0.251	0.362	0.341
cDCC-X	21 (6.93)	0.144	0.221	0.163	11 (3.64)	0.256	0.361	0.346
DC-X	22 (7.26)	0.090	0.077	0.050	21 (6.93)	0.144	0.221	0.163
FSG	26 (8.61)	0.009	0.855	0.032	24 (7.95)	0.030	0.946	0.094
90% VaR								
cDCC	37 (12.21)	0.213	0.449	0.346	36 (11.88)	0.288	0.370	0.380
cDCC-X	36 (11.88)	0.288	0.370	0.380	23 (7.59)	0.146	0.842	0.341
DC-X	39 (12.91)	0.105	0.969	0.268	37 (12.25)	0.206	0.759	0.429
FSG	44 (14.57)	0.013	0.538	0.037	42 (13.91)	0.032	0.722	0.093

4.7 Conclusion

We study the impact of Financial Conditions Indexes (FCIs) on volatilities and correlations of equity returns. We propose extensions of (factor-)GARCH models for volatility and DCC models for correlation that incorporate an index to measure financial conditions. In our empirical application, we use daily excess stock returns of U.S. deposit banks during the period 1994–2011 and proxy the financial conditions by the Bloomberg FCI. This daily index summarizes the money, bond and equity market.

Our results show that financial conditions affect both the volatilities and the correlations of large U.S. bank holding companies. Specifically, worse financial conditions are associated with higher volatilities and higher correlations. This result is both statistically and economically significant. During crisis periods, the inclusion of the FCI results in an increase in the estimated correlation of 0.15. A forecasting exercise shows the economic gain of including the Bloomberg FCI into our volatility and correlation models. Specifically, we consider portfolio VaRs using an equal-weighted portfolio and conduct 1- and 10-day ahead VaRs at various quantiles. We find that including the FCI in both the volatility and correlation specification improves the VaR estimates at the short horizon, such that less violations are made and hence the unconditional coverage match more closely to the nominal quantile. Our results imply that risk managers and portfolio managers of financial institutions should take into account financial conditions as a predictor of future volatility and correlations.

4.A Appendix

4.A.1 VaR backtests

The Unconditional Coverage test

The aim of the unconditional coverage test is to investigate whether the fraction of observations inside the (estimated/predicted) interval is equal to the nominal coverage probability. Given τ observations, define I_t , $t = 1, \dots, \tau$ as an indicator function which takes the value 1 if the VaR lies outside the estimated interval and 0 otherwise. Christoffersen

(1998) and Kupiec (1995) propose the following null hypothesis about correct unconditional coverage:

$$H_0 : Pr[I_{t+1} = 1] = q,$$

where q is the nominal coverage probability. This probability is equal to the expectation of the indicator function I_t . Assuming independence of I_1, \dots, I_τ , the likelihood function for interval estimates with coverage probability $p = Pr[I_{t+1} = 1]$ is given by

$$L(p; I_\tau, I_{\tau-1}, \dots, I_1) = p^{T_1} (1 - p)^{T_0},$$

where T_1 and T_0 represent the amount of violations and non-violations respectively. The likelihood under the null hypothesis given above with $p = q$ is compared with the likelihood under the alternative hypothesis where p equals the failure rate f . This rate is estimated by $\hat{f} = \widehat{Pr}(I_{t+1} = 1) = T_1 / (T_0 + T_1)$. The LR statistic is then given by

$$LR_{uc} = -2 \log \left(\frac{q^M (1 - q)^{T-M}}{\hat{f}^M (1 - \hat{f})^{T-M}} \right) \quad (4.A.1)$$

and is asymptotically distributed as a χ^2 distribution with one degrees of freedom.

The Independence test

The independence test advocated by Christoffersen (1998) investigates whether the occurrences violations are spread out over the sample instead of appearing in clusters. This ‘independence’ is tested against the specific alternative of a first-order Markov chain. Using the same notation as in the previous section, this boils down to

$$H_0 : Pr[I_{t+1} = 1 | I_t] = Pr[I_{t+1}] \quad t = 1, \dots, \tau.$$

If I_{t+1} follows a first-order Markov chain with transition probability matrix, then

$$\Pi_1 = \begin{pmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix}$$

where $\pi_{ij} = Pr[I_{t+1} = j | I_t = i]$. The likelihood is equal to

$$L(\Pi_1; I_\tau, I_{\tau-1}, \dots, I_1) = (1 - \pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_{10}} \pi_{11}^{T_{11}},$$

where T_{ij} is equal to the number of observations out of $(\tau - 1)$ such that $I_{t+1} = j$ and $I_t = i$. The maximum likelihood estimate of Π_1 is given by

$$\hat{\Pi}_1 = \begin{pmatrix} \frac{T_{00}}{T_{00}+T_{01}} & \frac{T_{01}}{T_{00}+T_{01}} \\ \frac{T_{10}}{T_{10}+T_{11}} & \frac{T_{11}}{T_{10}+T_{11}} \end{pmatrix}.$$

Under the null hypothesis of independence, $\pi_{01} = \pi_{11} \equiv \pi_2$, which leads to the corresponding likelihood function

$$L(\pi_2; I_\tau, I_{\tau-1}, \dots, I_1) = (1 - \pi_2)^{T_{00}+T_{01}} \pi_2^{T_{01}+T_{11}}.$$

Then by estimating π_2 with $\hat{\pi}_2 = (T_{00} + T_{01}) / (T_{00} + T_{01} + T_{10} + T_{11})$ the likelihood ratio test of independence is computed as

$$LR_{ind} = -2 \log \left(\frac{L(\hat{\pi}_2; I_\tau, I_{\tau-1}, \dots, I_1)}{L(\hat{\Pi}_1; I_\tau, I_{\tau-1}, \dots, I_1)} \right) \xrightarrow{asy} \chi^2 \quad (4.A.2)$$

under the null hypothesis. Finally, as shown by Christoffersen (1998), the unconditional coverage and the independence test can be combined to test *correct conditional coverage* by adding the two likelihood ratio test statistics. Under H_0 this statistic follows asymptotically a χ^2 distribution with two degrees of freedom.

4.A.2 In-sample volatilities of all assets

Table 4.A.1: In-sample estimation results - Spline-GARCH Models - All assets

This table reports the estimation results of the Spline-GARCH(-X) model:

$$\begin{aligned} r_{i,t} &= \mu_i + \sqrt{g_{i,t}\tau_{i,t}}\varepsilon_{i,t}, \\ g_{i,t} &= (1 - \alpha_i - \delta_i - \gamma_i/2) + \alpha_i \frac{(r_{i,t-1} - \mu_i)^2}{\tau_{i,t-1}} + \gamma_i \frac{(r_{i,t-1} - \mu_i)^2}{\tau_{i,t-1}} I[(r_{i,t-1} - \mu_i) < 0] + \delta_i g_{i,t-1}, \\ \tau_{i,t} &= \exp(\kappa_{i,0} + \kappa_{i,1} FCI_{t-1}), \end{aligned}$$

with $r_{i,t}$ the daily excess return of holding bank i , ($i = 1, \dots, 16$), FCI_t represents the Bloomberg FCI. Panel A of the table shows Maximum Likelihood estimates of the Spline-GARCH model without any added exogenous variable, assuming a conditional Student- t distribution for ε_t . Panel B shows parameter estimates of the Spline-GARCH-X model with the Bloomberg FCI included as exogenous variable. Nr in the first column corresponds with the order of Table 4.1. Standard errors are in parentheses. The sample goes from November 11, 1994, through December 30, 2011 (4,312 observations) for all numbers except nr. 3 and 7 (Wachovia Corp. and National City Corp.). For these two financial institutions, the sample goes from November 11, 1994, through December 30, 2008 (3,556 observations).

Panel A: Spline-GARCH							
Nr.	μ	α	δ	γ	κ_0	ν	
1.	0.007 (0.014)	0.054 (0.011)	0.901 (0.013)	0.078 (0.016)	1.889 (0.340)	6.405 (0.517)	
2.	0.012 (0.016)	0.039 (0.010)	0.911 (0.011)	0.088 (0.014)	1.472 (0.309)	6.105 (0.471)	
3.	0.059 (0.026)	0.052 (0.013)	0.924 (0.008)	0.049 (0.015)	-0.414 (0.027)	7.947 (0.024)	
4.	0.021 (0.023)	0.024 (0.007)	0.931 (0.008)	0.086 (0.012)	2.174 (0.415)	8.521 (0.902)	
5.	0.032 (0.022)	0.024 (0.007)	0.933 (0.008)	0.080 (0.012)	1.990 (0.334)	7.889 (0.766)	
6.	0.014 (0.032)	0.022 (0.009)	0.922 (0.011)	0.104 (0.017)	2.397 (0.042)	5.737 (0.082)	
7.	0.012 (0.060)	0.062 (0.016)	0.883 (0.017)	0.084 (0.025)	1.504 (0.345)	5.858 (0.559)	
8.	0.029 (0.020)	0.046 (0.009)	0.919 (0.009)	0.064 (0.013)	1.963 (0.435)	7.150 (0.620)	
9.	0.014 (0.017)	0.037 (0.010)	0.920 (0.013)	0.079 (0.016)	1.987 (0.413)	6.446 (0.545)	
10.	0.025 (0.024)	0.031 (0.008)	0.937 (0.009)	0.057 (0.012)	1.575 (0.334)	7.568 (0.740)	
11.	0.037 (0.018)	0.044 (0.010)	0.919 (0.010)	0.068 (0.013)	1.528 (0.363)	7.499 (0.677)	
12.	0.023 (0.018)	0.037 (0.008)	0.933 (0.009)	0.050 (0.012)	1.321 (0.286)	7.014 (0.616)	
13.	0.029 (0.027)	0.025 (0.007)	0.923 (0.009)	0.094 (0.014)	2.368 (0.309)	7.203 (0.667)	
14.	0.015 (0.016)	0.036 (0.007)	0.928 (0.009)	0.069 (0.012)	2.166 (0.399)	7.151 (0.609)	
15.	0.017 (0.014)	0.028 (0.010)	0.922 (0.012)	0.095 (0.014)	1.961 (0.437)	6.864 (0.591)	
16.	0.046 (0.031)	0.027 (0.007)	0.945 (0.009)	0.051 (0.011)	2.254 (0.334)	5.535 (0.359)	

Panel B: Spline-GARCH-X							
Nr.	μ	α	δ	γ	κ_0	κ_1	ν
1.	0.000 (0.017)	0.052 (0.012)	0.910 (0.015)	0.058 (0.015)	1.578 (0.255)	-0.354 (0.067)	6.415 (0.544)
2.	0.006 (0.015)	0.037 (0.010)	0.917 (0.012)	0.073 (0.013)	1.275 (0.242)	-0.359 (0.063)	6.072 (0.489)
3.	0.028 (0.022)	0.070 (0.014)	0.883 (0.016)	0.048 (0.019)	1.042 (0.156)	-0.436 (0.062)	7.989 (0.958)
4.	0.014 (0.031)	0.020 (0.006)	0.944 (0.008)	0.069 (0.011)	2.009 (0.392)	-0.282 (0.048)	8.667 (0.957)
5.	0.026 (0.020)	0.021 (0.006)	0.948 (0.008)	0.055 (0.012)	1.755 (0.295)	-0.272 (0.051)	8.139 (0.846)
6.	0.006 (0.012)	0.021 (0.011)	0.936 (0.020)	0.072 (0.020)	1.926 (0.362)	-0.337 (0.055)	5.679 (0.463)
7.	0.010 (0.020)	0.050 (0.014)	0.900 (0.024)	0.048 (0.021)	1.029 (0.146)	-0.476 (0.050)	6.405 (0.671)
8.	0.024 (0.020)	0.041 (0.009)	0.937 (0.010)	0.039 (0.012)	1.635 (0.376)	-0.327 (0.052)	7.334 (0.663)
9.	0.008 (0.017)	0.024 (0.010)	0.950 (0.015)	0.044 (0.014)	1.536 (0.357)	-0.369 (0.045)	6.306 (0.554)
10.	0.021 (0.029)	0.024 (0.008)	0.958 (0.009)	0.030 (0.010)	1.293 (0.264)	-0.336 (0.046)	7.952 (0.829)
11.	0.029 (0.019)	0.038 (0.010)	0.932 (0.011)	0.052 (0.012)	1.293 (0.295)	-0.354 (0.053)	7.658 (0.734)
12.	0.019 (0.019)	0.029 (0.008)	0.952 (0.009)	0.029 (0.010)	1.089 (0.237)	-0.317 (0.047)	7.091 (0.635)
13.	0.021 (0.030)	0.029 (0.008)	0.939 (0.012)	0.046 (0.014)	1.891 (0.201)	-0.403 (0.050)	7.460 (0.759)
14.	0.006 (0.033)	0.031 (0.007)	0.948 (0.010)	0.038 (0.014)	1.846 (0.035)	-0.323 (0.089)	7.081 (0.094)
15.	0.010 (0.014)	0.016 (0.009)	0.945 (0.012)	0.074 (0.012)	1.671 (0.400)	-0.317 (0.049)	7.038 (0.642)
16.	0.035 (0.027)	0.014 (0.005)	0.972 (0.006)	0.026 (0.007)	1.880 (0.319)	-0.317 (0.041)	5.575 (0.368)

Table 4.A.2: In-sample estimation results - Factor-Spline-GARCH models - All assets

This table reports the estimation results of the Factor-Spline-GARCH model. The volatility is formulated as a Spline-GARCH-X model:

$$\begin{aligned} r_{i,t} &= \mu_{i,t} + \sqrt{g_{i,t}\tau_{i,t}}\varepsilon_{i,t}, \\ g_{i,t} &= (1 - \alpha_i - \delta_i - \gamma_i/2) + \alpha_i \frac{(r_{i,t-1} - \mu_{i,t})^2}{\tau_{t-1}} + \gamma_i \frac{(r_{i,t-1} - \mu_{i,t})^2}{\tau_{t-1}} I[(r_{i,t-1} - \mu_{i,t}) < 0] + \delta_i g_{i,t-1}, \\ \tau_{i,t} &= \exp(\kappa_{i,0} + \kappa_{i,1} FCI_{t-1}), \end{aligned}$$

with $r_{i,t}$ the daily excess return of holding bank i , ($i = m, 1, \dots, 16$) and the S&P 500 market return. In addition, $\mu_{i,t} = \xi_i + \beta_i r_{m,t}$ with $r_{m,t}$ the excess market return at day t . Otherwise, $\mu_{i,t} = \mu_m$ in case of the market return. FCI_t represents the Bloomberg FCI. Panel A of the table shows Maximum Likelihood estimates of the Spline-GARCH model without any added exogenous variable, assuming a conditional Student- t distribution for ε_t . Panel B shows parameter estimates of the Spline-GARCH-X model with the Bloomberg FCI included as exogenous variable. Nr in the first column corresponds with the order of Table 4.1. Standard errors are in parentheses. The sample goes from November 11, 1994, through December 30, 2011 (4,312 observations) for all numbers except nr. 3 and 7 (Wachovia Corp. and National City Corp.). For these two financial institutions, the sample goes from November 11, 1994, through December 30, 2008 (3,556 observations).

Panel A: Spline-GARCH							
Nr.	β	α	δ	γ	κ_0		ν
Panel A							
1.	0.980 (0.022)	0.082 (0.016)	0.897 (0.017)	0.027 (0.016)	1.508 (0.384)		5.352 (0.370)
2.	0.844 (0.018)	0.053 (0.012)	0.910 (0.012)	0.065 (0.014)	1.303 (0.378)		5.473 (0.382)
3.	0.977 (0.019)	0.106 (0.019)	0.864 (0.020)	0.037 (0.021)	1.073 (0.379)		5.856 (0.509)
4.	0.957 (0.018)	0.051 (0.010)	0.929 (0.010)	0.037 (0.013)	1.828 (0.509)		6.759 (0.572)
5.	1.327 (0.019)	0.028 (0.006)	0.952 (0.006)	0.036 (0.009)	1.332 (0.370)		5.597 (0.384)
6.	1.241 (0.020)	0.043 (0.012)	0.923 (0.014)	0.063 (0.016)	2.035 (0.462)		4.812 (0.295)
7.	0.930 (0.018)	0.120 (0.019)	0.871 (0.018)	0.000 (0.020)	1.518 (0.039)		4.809 (0.066)
8.	1.011 (0.019)	0.058 (0.011)	0.924 (0.011)	0.032 (0.013)	1.579 (0.493)		5.893 (0.438)
9.	1.060 (0.018)	0.059 (0.015)	0.933 (0.017)	0.008 (0.015)	1.232 (0.474)		5.144 (0.351)
10.	1.045 (0.018)	0.045 (0.010)	0.944 (0.010)	0.018 (0.012)	1.158 (0.415)		5.253 (0.352)
11.	1.005 (0.018)	0.079 (0.013)	0.907 (0.011)	0.016 (0.014)	1.072 (0.377)		5.888 (0.450)
12.	0.995 (0.017)	0.066 (0.014)	0.920 (0.015)	0.016 (0.013)	0.833 (0.346)		5.717 (0.413)
13.	1.605 (0.023)	0.028 (0.007)	0.943 (0.009)	0.050 (0.011)	1.648 (0.317)		5.538 (0.394)
14.	1.301 (0.020)	0.065 (0.012)	0.914 (0.012)	0.036 (0.014)	1.588 (0.447)		5.984 (0.445)
15.	0.909 (0.016)	0.075 (0.018)	0.911 (0.016)	0.024 (0.016)	1.625 (0.554)		5.168 (0.348)
16.	1.310 (0.031)	0.033 (0.005)	0.967 (0.003)	0.000 (0.007)	-4.193 (0.076)		4.595 (0.133)
Panel B: Spline-GARCH-X							
Nr.	β	α	θ	γ	κ_0	κ_1	ν
1.	0.986 (0.022)	0.077 (0.018)	0.898 (0.021)	0.023 (0.015)	1.158 (0.241)	-0.376 (0.054)	5.438 (0.401)
2.	0.844 (0.018)	0.041 (0.012)	0.924 (0.013)	0.059 (0.012)	1.086 (0.316)	-0.318 (0.055)	5.563 (0.399)
3.	0.981 (0.019)	0.103 (0.019)	0.865 (0.024)	0.020 (0.021)	0.675 (0.219)	-0.418 (0.062)	6.243 (0.594)
4.	0.961 (0.018)	0.033 (0.009)	0.946 (0.010)	0.041 (0.011)	1.632 (0.445)	-0.339 (0.046)	6.871 (0.602)
5.	1.338 (0.019)	0.015 (0.004)	0.971 (0.005)	0.026 (0.007)	1.160 (0.311)	-0.339 (0.039)	5.962 (0.445)
6.	1.246 (0.020)	0.032 (0.013)	0.934 (0.018)	0.061 (0.016)	1.746 (0.482)	-0.356 (0.048)	4.790 (0.312)
7.	0.934 (0.019)	0.066 (0.027)	0.910 (0.043)	0.016 (0.020)	0.695 (0.228)	-0.464 (0.057)	5.295 (0.461)
8.	1.013 (0.019)	0.048 (0.010)	0.939 (0.010)	0.023 (0.011)	1.277 (0.418)	-0.310 (0.049)	6.068 (0.469)
9.	1.062 (0.018)	0.021 (0.007)	0.971 (0.007)	0.014 (0.007)	1.047 (0.377)	-0.362 (0.037)	5.121 (0.349)
10.	1.047 (0.018)	0.034 (0.009)	0.959 (0.010)	0.010 (0.010)	0.940 (0.338)	-0.320 (0.047)	5.500 (0.386)
11.	1.008 (0.018)	0.062 (0.013)	0.918 (0.012)	0.025 (0.014)	0.806 (0.301)	-0.352 (0.042)	6.188 (0.510)
12.	0.998 (0.017)	0.043 (0.015)	0.944 (0.017)	0.017 (0.009)	0.569 (0.286)	-0.315 (0.049)	5.846 (0.434)
13.	1.612 (0.023)	0.018 (0.005)	0.976 (0.007)	0.010 (0.009)	1.201 (0.292)	-0.401 (0.040)	5.921 (0.457)
14.	1.306 (0.020)	0.054 (0.012)	0.929 (0.013)	0.026 (0.012)	1.240 (0.369)	-0.401 (0.053)	6.150 (0.484)
15.	0.911 (0.016)	0.062 (0.020)	0.919 (0.019)	0.029 (0.013)	1.357 (0.505)	-0.322 (0.058)	5.210 (0.366)
16.	1.383 (0.027)	0.017 (0.005)	0.973 (0.007)	0.017 (0.006)	1.589 (0.330)	-0.289 (0.038)	4.444 (0.237)

Chapter 5

Improving density forecasts and Value-at-Risk estimates by combining densities

Joint work with Dick van Dijk and Michel van der Wel.

5.1 Introduction

Value-at-Risk (VaR) is a commonly used measure of downside risk for investments. Financial institutions are allowed by regulation (i.e. the Basel accords) to report VaR estimates for their asset portfolios obtained from their own “internal” model. An important related issue in this estimation is *model uncertainty*, as each model has its prespecified known form and takes no account of possible uncertainty regarding the model structure. In addition, given the availability of a considerable number of different risk-management methods, based on academic literature and/or his expertise, it is a difficult task for a decision-maker to choose the “best” model. Moreover, each model is an incomplete description of reality. Hence relying upon a single model is dangerous to construct a VaR, i.e. a density forecast in the left tail, as any model is “wrong” in some sense.

In this chapter, we investigate the usefulness of combining density forecasts with the focus on a particular region of the density. This is motivated in the first place by well

known advantages of combining point or density forecasts.¹ We aim to obtain more realistic and more accurate VaR estimates and density forecasts in the left tail. This motivates the investigation of combining density forecasts based on their behavior in the left tail as using the whole density does not necessarily lead to the same quality of forecasts as when we focus purely on the left tail of a density. Therefore, we develop a density forecast combination method that extends the method of Geweke and Amisano (2011), which uses the *whole* density, by considering the censored likelihood (*csl*) scoring rule of Diks *et al.* (2011) that focuses on a region of the densities' support of particular interest, such as the left tail.

We use our novel methodology in an empirical application involving several recently developed univariate volatility models. Hence, as a second contribution to the literature, we make a comparison between these models with respect to their predictive ability in terms of density forecasts. In particular, beyond the traditional GARCH model (Bollerslev, 1986), we consider the Heavy model (Shephard and Sheppard, 2010) and the Realized GARCH model (Hansen *et al.*, 2012) that include realized measures, as well as the GAS model (Creal *et al.*, 2013). All models are applied to daily returns on the S&P 500, DJIA, FTSE and Nikkei stock market indexes from 2000 until 2013.

We evaluate the added value of combining density forecasts both statistically and economically. First, we test equal predictive accuracy in the left tail of a combined density forecast based on our new method and three alternatives: (i) the method based on the whole density, (ii) a benchmark that consists of equal weights, and (iii) the density forecast of each individual model. Second, we compare 1- and 5-day VaR estimates based on these methods using the Unconditional Coverage test and the Independence test of Christoffersen (1998). In addition we test on equal accuracy based on an asymmetric tick-loss function using the test procedure of Giacomini and White (2006).

Our results show statistically that density forecasts in the tail are more accurate if one pools density forecasts using the *csl* scoring rule than using the aforementioned method, using equal weights or using the density forecast of any individual volatility model. 90% and 95% 1-day VaR estimates improve significantly compared to the other pooling methods or the individual models, such that less violations are made and the unconditional

¹We discuss this literature in more detail below.

coverage matches more closely to the nominal value. Moreover, the accuracy of the VaR estimates improves significantly upon using equal weights or any individual model according to the asymmetric tick-loss function. In addition, we show that the combination weights based on the *csl* scoring rule differ considerably from the weights obtained by using the whole density. Hence, a certain volatility model could get no or less weight in the method of Geweke and Amisano (2011), but may be useful in our new method.

We contribute to the literature on combining forecasts, see Timmermann (2006) for a survey. Starting with the seminal work of Bates and Granger (1969), combining point forecasts appears to be a successful forecasting strategy, improving upon individual forecasts. Timmermann (2006) shows from a theoretical point of view why forecast combinations could work well. This is confirmed by numerous empirical applications in different areas including macroeconomic and financial forecasting. For example, forecasting output growth using individual predictors typically delivers forecasts that are unstable over time. Combining forecasts offers more stable forecasts which improve upon autoregressive forecasts (Stock and Watson, 2004). Rapach *et al.* (2010) provide similar evidence in the context of equity premium prediction, by showing that combining forecasts leads to statistically and economically significant out-of-sample gains relative to the historical average return.

Although the literature shows the usefulness of combining point forecasts, point forecasts themselves are not very informative if there is no indication of their uncertainty (see Granger and Pesaran, 2000; Garratt *et al.*, 2003). This finding has led to a growing interest in *density* forecasts, which represent a full predictive distribution of a random variable and hence provide the most complete measure of this uncertainty. It is a natural step forward to bring together the concepts of forecast combinations and density forecasts. The literature on combining density forecasts is yet scarce, although the interest in this topic of research grows with applications to for example macro-economics (Jore *et al.*, 2010; Aastveit *et al.*, 2011). Wallis (2005) considers a finite mixture distribution, which takes a weighted linear combination of multiple density forecasts. Hall and Mitchell (2007) address the issue how to choose the weights assigned to each competing density. They propose a methodology with the aim to obtain the most accurate density forecast from a statistical point of view. This boils down to using the logarithmic scoring rule, which

takes the log of the predictive density evaluated at the observed value of the variable of interest. Closely related is the work of Geweke and Amisano (2011), who use the logarithmic scoring rule to obtain weights to form optimal linear combinations of predictive densities. We extend this approach, by substituting the log score rule by the censored likelihood scoring rule.

The remainder of this chapter is organized as follows. Section 5.2 puts forward our methodology of combining density forecasts using the *csl* scoring rule. In Section 5.3, we provide an overview of the univariate volatility models and the related assumed conditional density functions, which are used in the empirical application (Section 5.4). Section 5.5 concludes.

5.2 Combining density forecasts

Suppose a decision maker has n different models for a variable of interest y . Conditional on information available up to and including time $t - 1$, the predictive density corresponding with a particular model at time t is of the form $p_t(y_t | I_{t-1}, \theta_{A_i}, A_i)$, where I_{t-1} indicate the information set up to and including time $t - 1$, A_i denotes the particular model i , ($i = 1, \dots, n$) and θ_{A_i} the estimated parameters of model A_i given I_{t-1} . Suppose further that the decision maker aims to choose the best predictive density at time $T + 1$, given the available density forecasts from time $t = 1, \dots, T$. An often used approach is to make use of scoring rules. A scoring rule measures the quality of density forecasts by assigning a numerical score.² Typically, this rule is a objective function that depends on the density forecast and the actually observed value, such that a higher score is associated with a “better” density forecast. According to Gneiting and Raftery (2007), a scoring rule is *proper* if it satisfies the condition that incorrect density forecasts do not receive a higher average score than the true density. This property is important and a natural requirement for any rational decision maker.

A well founded scoring rule is the log score function (see Mitchell and Hall, 2005; Amisano and Giacomini, 2007). This function for a particular model A_i at a specific time

²Note that we use the term ‘score’ twice: (i) in the GAS models to indicate the derivative of the logarithm of the density with respect to a certain parameter and (ii) a number that is assigned to measure density forecasts.

t is defined as

$$S^l(y_t; A_i) = \log p_t(y_t | I_{t-1}, A_i), \quad (5.1)$$

with S^l the abbreviation of the log scoring rule, which simply takes the logarithm of the predictive density evaluated at y_t . This scoring rule is closely related to information theoretic goodness-of-fit measures such as the Kullback-Leibler Information Criterion (KLIC) associated with the density forecast $p_t(y_t | I_{t-1}, A_i)$. It can be shown that a higher value of the logarithmic score coincides with a lower value of the KLIC. Put differently, maximizing the logarithmic score is equivalent with minimizing the KLIC.

Geweke and Amisano (2011) argue that it is highly unlikely that one model is the true model for constructing a predictive density. They propose therefore to combine the predictive densities using the log score function of (5.1). In particular, they consider predictive densities of the form

$$\sum_{i=1}^n w_i p_t(y_t | I_{t-1}, A_i), \quad (5.2)$$

for $i = 1, \dots, n$ and weights w_i , restricted such that they are positive and sum to one to ensure that (5.2) is a valid probability density function. It is natural to choose the weights in such a way that the log score function in (5.1) is maximized (and hence the KLIC is minimized):

$$S^l(Y_T, C) = \sum_{t=1}^T \log \left[\sum_{i=1}^n w_i p_t(y_t | I_{t-1}, A_i) \right], \quad (5.3)$$

with $Y_T = y_1, \dots, y_T$, and C representing the fact that a combination of models is evaluated instead of a single model A_i . Following Bacharach (1974), linear combinations of (subjective) probability distributions are known as *linear opinion pools*. We use the term *pooling* and *linear opinion pools* interchangeably in this chapter.

The main idea of this chapter is to extend the approach of Geweke and Amisano (2011) by focusing on a particular region of interest of the predictive density. In order to do so, we consider a scoring rule based on the censored likelihood (*csl*), advocated by Diks *et al.*

(2011). They prove that this scoring rule is proper and show the usefulness of this scoring rule if one is interested in the accuracy of density forecasts in a specific region. In this study, the focus is on the left tail, which is important for risk management purposes. The *csl* score function for a specific region B_t for model A_i at time t reads

$$S^{csl}(y_t|A_i) = I[y_t \in B_t] \log p_t(y_t|I_{t-1}, A_i) + I[y_t \in B_t^c] \log \left(\int_{B_t^c} p_t(y|I_{t-1}, A_i) dy \right) \quad (5.4)$$

with B_t^c the complement of B_t and $I[\cdot]$ an indicator function that takes the value 1 if the argument is true. The first part of this scoring rule focuses on the behavior of the density forecast in the region of interest B_t . The second part computes the cdf of the density in the region outside B_t .³ Hence any observation outside B_t ignores the shape of $p_t(y_t|I_{t-1}, A_i)$ outside B_t . Note that (5.4) simplifies to the log scoring rule of (5.1) if B_t represents the full sample space.

The next step is combine the predictive densities based on the *csl* scoring rule. That is, we consider again predictive densities as defined in (5.2), however with the weights obtained by optimizing the corresponding censored likelihood score function over the values $Y_T = y_1, \dots, y_T$:

$$S^{csl}(Y_T, C) = \sum_{t=1}^T \log \left[\sum_{i=1}^n w_i \left(I[y_t \in B_t] p_t(y_t|I_{t-1}, A_i) + I[y_t \in B_t^c] \int_{B_t^c} p_t(y|I_{t-1}, A_i) dy \right) \right] \quad (5.5)$$

We end this section by a brief comment about the optimization of the weights w_t in (5.3) and (5.5). Although (numerical) constrained optimization techniques may be used, we consider the algorithm of Conflitti *et al.* (2012). This iterative algorithm is easy to implement works well even when the number of forecasts to combine gets large. See Appendix 5.A.1 for more details.

³To interpret this second part, if B_t is the left tail $y_{t+1} < r$ (with r a certain quantile of the cdf of y_t), the second part of (5.4) ensures that the tail probability implied by $p_t(y_t|I_{t-1}, A_i)$ matches with the frequency at which tail observations actually occur.

5.3 Models and distributions

This study focuses on density forecasting in the context of univariate volatility models. We consider several classes of models, including the standard GARCH model of Bollerslev (1986), the HEAVY model of Shephard and Sheppard (2010), the Realized GARCH model of Hansen *et al.* (2012) and the GAS model of Creal *et al.* (2013). All models are based on the following general specification for y_t , the return for a financial asset at day t :

$$y_t = \mu + \sqrt{h_t} z_t, \quad \text{with } z_t | I_{t-1} \sim D(0, 1), \quad (5.6)$$

where μ denotes the conditional mean of the returns, h_t the conditional variance and z_t the standardized unexpected return following a certain conditional distribution $D(\cdot)$ with mean zero and unit variance. Further, I_t denotes the information set up to and including time t .⁴ The following subsections differentiate between various specifications for the dynamics of h_t and possible choices for the conditional return density function $D(\cdot)$.

5.3.1 Univariate volatility models

The first model we consider is the traditional GARCH(1,1) model (Bollerslev, 1986) for the conditional variance h_t :

$$h_t = \omega + \alpha(y_{t-1} - \mu)^2 + \beta h_{t-1}, \quad (5.7)$$

with $\omega > 0$, $\alpha > 0$ and $\beta > 0$ to ensure a positive variance. The past squared demeaned return in this model is the innovation for the conditional variance. Many extensions of the GARCH model are proposed (e.g. the EGARCH and GJR GARCH models Nelson, 1991; Glosten *et al.*, 1993), however we stick to the basis specification as given in (5.7). We restrict also the other considered model classes in this study to the basis specification, although many variants/extensions are possible. The reason is that the aim is to compare model *classes* combined with distributions, and not models *within* a specific class.

⁴For ease of exposition, we assume the conditional mean fixed, although it could easily be extended to a time-varying mean μ_t .

Creal *et al.* (2013) develop a broader set of models which also includes the GARCH model of (5.7), namely the Generalized Autoregressive Score (GAS) models. The key property of these models is that innovations for time-varying parameters are based on the score of the probability density function at time t . In terms of our univariate volatility models, the time-varying parameters are the conditional variances h_t . The GAS(1,1) model proposes the following structure for h_t :

$$\begin{aligned} h_t &= \omega + \alpha s_{t-1} + \beta h_{t-1}, \\ s_t &= Q_t \nabla_t, \\ \nabla_t &= \frac{\partial p(y_t | h_t, I_{t-1}, \theta)}{\partial h_t}, \end{aligned} \tag{5.8}$$

with $p(y_t | h_t, I_{t-1}, \theta)$ the conditional return density, θ the parameter vector, ∇_t the score and Q_t a scale factor. We follow Creal *et al.* (2013) and define the scale factor as $1/E_{t-1}[\nabla_t^2]$, where E_t denotes the expectation with respect to the return density $p(y_t | h_t, I_{t-1}, \theta)$. For example, when the returns y_t follow a conditional Normal distribution, the GAS model corresponds exactly to the GARCH(1,1) model of (5.7).⁵ In case of a fat-tailed Student- t distribution for y_t , the score based volatility model reads

$$h_t = \omega + \alpha(1 + 3/\nu) \frac{\nu + 1}{(\nu - 2) + \frac{(y_{t-1} - \mu)^2}{h_{t-1}}} (y_{t-1} - \mu)^2 + \beta h_{t-1}, \tag{5.9}$$

and will be labeled as the GAS- t model. The specification downweights the more extreme observations, in the sense that if the distribution is more heavy tailed, it is less likely that an extreme observation is due to an increase in volatility. Note that this is a function of ν ; when $\nu \rightarrow \infty$, (5.9) converges to the GARCH(1,1) model of (5.7). We again impose $\omega > 0, \alpha > 0$ and $\beta > 0$ in the estimation of the parameters.

The third and fourth model classes in this study include realised measures to describe the dynamics of daily volatility. A realised measure is a high-frequency estimator of the variance of a particular asset return during the times the asset is traded on an exchange. For example, the realised variance (RV) for a particular day sums the squared returns during a specific intra-day period. The intuition is that realised measures are a more accurate

⁵When $y_t \sim N(0, h_t)$, $\nabla_t = -0.5h_t^{-1} + 0.5h_t^{-2}y_t^2$ and $Q_t = 2h_t^2$. Hence the GAS model becomes $h_t = \omega + \alpha(y_t^2 - h_t) + \beta h_t$, which is equivalent with the GARCH model of (5.7).

estimate of daily volatility than the squared daily return, as used in the GARCH models (see Andersen, Bollerslev, Diebold and Labys, 2003).

A recently developed model that explicitly introduces high-frequency estimators in daily volatility models is the HEAVY model of Shephard and Sheppard (2010). In particular, this model assumes the following structure for the conditional variance h_t and the expectation of the realised measure $\xi_t = E[RM_t|I_{t-1}]$:

$$h_t = \omega + \alpha RM_{t-1} + \beta h_{t-1}, \quad (5.10)$$

$$\xi_t = \omega_R + \alpha_R RM_{t-1} + \beta_R \xi_{t-1}. \quad (5.11)$$

All parameters should be positive to avoid negative values of h_t and ξ_t . The Heavy model is seen to consist of a GARCH structure for both h_t and ξ_t , with RM_t as innovation term. One may also include the squared (demeaned) daily return in (5.10), however in practice the estimate of the corresponding parameter is generally close to zero and insignificant, as noted by Shephard and Sheppard (2010). Equation (5.11) “completes” the system, in the sense that without this equation one can only perform one-step ahead forecasts of the conditional variance h from (5.10) since future values of the realised measure are unknown at time t .

A second model that relates conditional volatility with realised measures is the Realized GARCH model (RGARCH) of Hansen *et al.* (2012). The basic specification is given by:

$$h_t = \omega + \alpha RM_{t-1} + \beta h_{t-1}, \quad (5.12)$$

$$RM_t = \delta + \phi h_t + \tau(z_t) + u_t, \quad (5.13)$$

with $\tau(z_t)$ the leverage function, defined in the basic form as $\tau_1 z_t + \tau_2(z_t^2 - 1)$. This function allows for the empirical finding that negative and positive shocks may have a different impact on the volatility. Except τ_1 , which is typically negative, all parameters are restricted to be positive. The dynamics for h_t are similar for both the HEAVY and RGARCH model, however the difference arises in the specification of (the expectation of) RM_t . The HEAVY model proposes a GARCH structure for $E[RM_t|I_{t-1}]$, while the

RGARCH model explicitly relates RM_t to the conditional variance at time t and introduces additionally a leverage component.

5.3.2 Conditional distributions

We consider four possible distributions $D(\cdot)$ of z_t in (5.6), which corresponds with the conditional density of the returns y_t . The starting point is the conditional Normal distribution, since this distribution is simple and often used. However, to take into account possible conditional non-normality, skewness, and excess kurtosis, we also allow the return y_t to follow a Student- t distribution with mean μ , variance h_t and ν degrees of freedom. That is,

$$f(y_t|\mu, h_t, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{h_t(\nu-2)}\pi} \left(1 + \frac{(y_t - \mu)^2}{h_t(\nu-2)}\right)^{-\frac{\nu+1}{2}}. \quad (5.14)$$

The degrees of freedom ν is treated as an unknown parameter and is estimated together with the volatility parameters. In addition, $\nu > 2$ is required to ensure a existing variance. The excess kurtosis of the Student- t distribution is equal to $6/(\nu-4)$, hence it is only defined if $\nu > 4$. In general, a lower value of ν implies a more fat-tailed distribution. Third, we consider the Laplace distribution, which also exhibits fatter tails than the Normal distribution, but does not involve additional parameters:

$$f(y_t|\mu, h_t) = \frac{1}{\sqrt{2}h_t} \exp\left(-\sqrt{2}\frac{|y_t - \mu|}{\sqrt{h_t}}\right) \quad (5.15)$$

with again mean μ and variance h_t . Finally the Skewed- t distribution of Hansen (1994) enables returns to be distributed asymmetrically, in contrast to the three symmetric distributions discussed above. For a zero mean and unit variance variable $z_t = (y_t - \mu)/\sqrt{h_t}$, the distribution reads

$$f(z_t; \lambda, \nu) = \begin{cases} bc \left(1 + \frac{1}{\nu-2} \left(\frac{bz_t+a}{1-\lambda}\right)^2\right)^{-\frac{\nu+1}{2}} & \text{if } z_t < -\frac{a}{b} \\ bc \left(1 + \frac{1}{\nu-2} \left(\frac{bz_t+a}{1+\lambda}\right)^2\right)^{-\frac{\nu+1}{2}} & \text{if } z_t \geq -\frac{a}{b} \end{cases} \quad (5.16)$$

with

$$a = 4\lambda c \frac{\nu - 2}{\nu - 1}, \quad b^2 = 1 + 3\lambda^2 - a^2, \text{ and } c = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma(\frac{\nu}{2})}$$

such that $f(y_t|\mu, h_t, \nu, \lambda) = 1/h_t f(z_t; \lambda, \nu)$. Further, λ is the skewness parameter and ν again represents the degrees of freedom. A (positive) negative value of λ indicates (positive) negative skewness.

Table 5.1 summarizes the various choices for the dynamics of the conditional variance h_t and a conditional distribution $D(\cdot)$, both defined in the general specification for the daily return of (5.6). For the GARCH(1,1), HEAVY and RGARCH models, we estimate their parameters in combination with the assumption of the four described conditional distributions of 5.3.2 (i.e. Normal, Student- t , Laplace and Skewed- t). Further, we assume a Student- t and Laplace distribution for the GAS models. This delivers 14 models in total. We estimate all models by Maximum Likelihood. This is not a computationally involved step, since we are dealing with univariate models with a maximum of 8 parameters (RGARCH models) to be estimated. In addition, we can estimate the HEAVY parameters of (5.10) and (5.11) separately, see Shephard and Sheppard (2010) for more details.

Table 5.1: Overview of volatility models and conditional distributions

This table reports the various choices for the dynamics of the conditional variance h_t and the possible conditional distributions $D(\cdot)$, both apparent in the general specification of (5.6). An “x” (“-”) denotes that an particular specification together with a conditional distribution is (not) chosen.

	Normal	Student- t	Laplace	Skewed- t
GARCH(1,1)	x	x	x	x
GAS(1,1)	x ^a	x	x	- ^b
Heavy	x	x	x	x
RGARCH	x	x	x	x

^a The GAS(1,1) model with Normal distributed errors is the same as the GARCH(1,1) model with Normal errors.

^b We leave the GAS(1,1) with Skewed- t distributed errors as a topic of further research.

5.4 Application

This section contains an application of our new method of combining density forecasts, applied in the context of univariate volatility models. In the following subsections, we discuss the data and implementation details, the evaluation of the density forecasts and finally the results.

5.4.1 Data and implementation details

We apply the volatility models of Section 5.3 to daily returns from four major stock market indexes: S&P 500, DJIA, Nikkei and the FTSE. The sample period goes from January 3, 2000 until June 28, 2013. Daily returns as well as their corresponding realized measures are obtained from the Oxford-Man Institute's "realised library".⁶ We follow Shephard and Sheppard (2010) and use the realised kernel (see Barndorff-Nielsen *et al.*, 2008) as the realised measure at time t (RM_t). When the exchange is closed, days are deleted from the sample.⁷ Figure 5.1 shows the dynamics of the S&P 500 index and Japanese equity index, together with the square root of the realised kernel estimate of the daily variance. The dynamics of both indexes are quite similar, however the Nikkei index contains more downward spikes (e.g. the 2011 Tohoku earthquake). Nevertheless, both return graphs clearly show the presence of conditional heteroskedasticity, since calm periods and periods of high volatility occur in an alternating pattern.

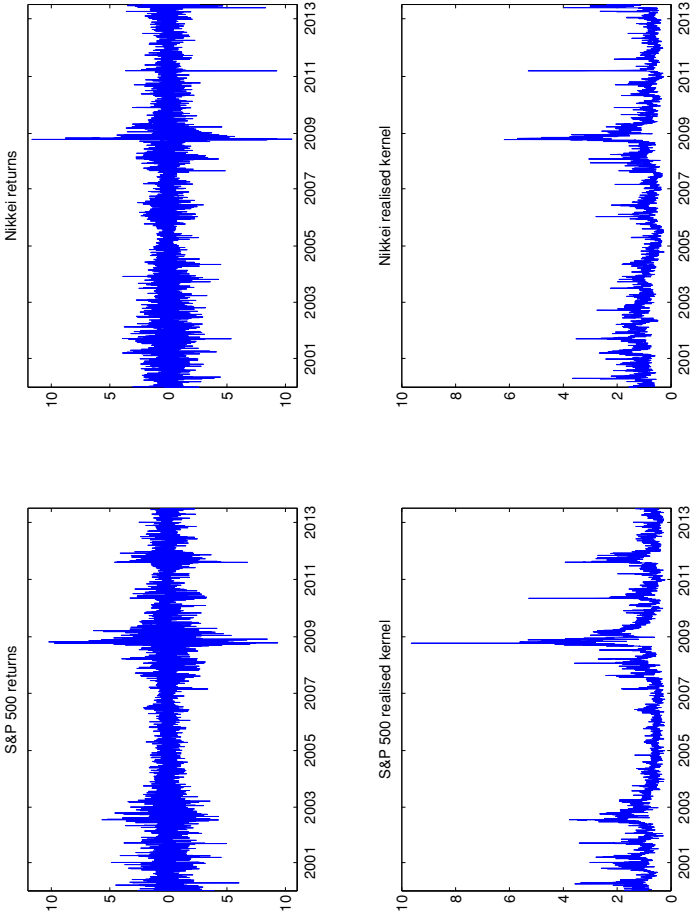
We apply a rolling window scheme to estimate the model parameters and construct density forecasts. More specifically, we use an estimation window of approximately 3 years ($T_{est} = 750$ observations), estimate the model parameters and construct 1- until 5-step ahead forecasts of h_t at each time t ($t = T_{est}, T_{est} + 1, \dots, T - 5$). Given these forecasts, we also construct the corresponding 1- and 5-step ahead density forecasts. After 750 subsequent density forecasts ($T_w = 750$) have been obtained for each model, we optimize (5.3) and (5.5) to obtain w_t . In case of the *csl* score function, we define the region B_t as the left tail $y_t < \hat{r}_t^\kappa$ with \hat{r}_t^κ the κ th quantile of the empirical CDF of the 750 returns corresponding with the estimation window T_{est} . We repeat also this optimization

⁶See <http://realized.oxford-man.ox.ac.uk/>

⁷We have to delete 1-1.5% of the daily returns on the S&P 500, DJIA and FTSE index. The Nikkei index loses 3% of its daily returns.

Figure 5.1: Daily returns and realised measures

This figure depicts the daily (close-to-close) returns on the S&P 500 index and the Nikkei index (upper part) and a realised kernel estimate of the corresponding daily (open-to-close) volatility (bottom part) from January 3, 2000, through June 27, 2013 (3,364 and 3,206 observations respectively). Both daily returns and volatilities are given in percentages.



by means of a rolling window scheme with a window of T_w density forecasts evaluations at each time t ($t = T_{est} + T_w + 1, T_{est} + T_w + 2, \dots, T - 5$). We choose κ equal to 0.15 and 0.25 respectively. We choose $T_{est} = 750$ such that there is a sufficient number of observations for parameter estimation of the models. Further, we emphasize the trade-off in the choice of κ . Given our interest in the left tail, we should take a small value of κ . However, the corresponding number of observations in the region of interest becomes very low, such that the variation in the *csl* scores of the different models declines.⁸ Similarly, there is a trade-off in the choice of T_w . On the one hand, one would choose T_w as high as possible in order to use the largest amount of available observations to compute the weights w_t . But on the other hand, if the relative performance of different models varies through time, one should take this into account and choose a smaller value of T_w . In addition, T_w and κ are related in the sense that a low value of κ combined with a small window results in a small amount of observations within the region B_t . Hence given these trade-offs and the relation between those two variables, we choose T_w and κ as 750 and 0.15 (0.25) such that there are 112 (187) observations in the left tail.

5.4.2 Evaluation

We assess the accuracy of our (combined) density forecasts in two ways. First, we focus purely on the predictive density in the left tail and investigate statistically whether pooling based on censored densities adds any value. Following Diks *et al.* (2011), we test the null hypothesis of equal performance of two density forecasts $p_t(y_t; I_{t-1}, A_i)$ and $p_t(y_t; I_{t-1}, A_j)$ based on the scoring rule of (5.4).⁹ That is, given a sample of density forecasts and corresponding realizations for m periods, define the relative score d_t as

$$d_t = S^{csl}(y_t; A_i) - S^{csl}(y_t; A_j) \quad (5.17)$$

⁸Recall that if y_t is outside the region B_t with B_t the left tail $y_t < \hat{r}_t^\kappa$, the *csl* score is the cdf of y_t in the complement of the region.

⁹In case we consider density forecasts using combinations, the density forecast is given by $\sum_{i=1}^n w_{it} p_t(y_t; I_{t-1}, A_i)$.

with corresponding null-hypothesis $H_0 : E[d_t] = 0$ for all m periods. The resulting Diebold and Mariano (1995) test-statistic is then given by

$$t_m = \frac{\bar{d}_m}{\sqrt{\hat{\sigma}_m^2/m}}, \quad (5.18)$$

with \bar{d}_m the sample average of the score differences and $\hat{\sigma}_m^2$ a HAC-consistent variance estimator of the true variance σ_m^2 of d_t . A positive value means that the density forecasts in the tail of model A_i are more accurate than the corresponding density forecasts of model A_j . This test allows for parameter estimation uncertainty and fits the framework of Giacomini and White (2006), who show that the use of a rolling window of m past observations for parameter estimation simplifies the asymptotic theory of tests of equal predictive accuracy. Moreover, the test allows to compare density forecasts of both nested and non-nested models.

The second way to explore the additional value of using censored densities in this study is based on 1- and 5-day Value-at-Risk (VaR) estimates. For the individual models considered in this study, the 1-day VaR estimate reads

$$VaR_t^{1-q} = \mu + z_q \sqrt{h_t}, \quad (5.19)$$

with μ the estimated conditional mean return, h_t the (forecasted) conditional variance, and z_q represents the q -th quantile of the assumed cdf. However, we cannot apply (5.19) when our predictive distribution is a combination of individual distributions.¹⁰ This also holds for the h -day ($h \geq 2$) VaR estimates if the assumed distribution is non-Normal. We use simulation techniques to overcome this issue. That is, we simulate daily returns from each individual model/distribution according to the assigned weight (and conditional variance) to obtain the required quantile of the total distribution to compute the $(1 - q)\%$ VaR.

Finally, we test the accuracy of the VaR estimates by focusing on two aspects. First, we assess the frequency of the VaR violations with the unconditional coverage (UC) of Kupiec (1995) and Christoffersen (1998). These tests compare the actual with the expected number of violations. In addition, we test whether the violations occur in clus-

¹⁰The VaR of a mixture of densities is not equal to the weighted average of each individual VaR.

ters by means of the Independence test (Ind) of Christoffersen (1998). In order to apply both tests on the estimated 5-day VaRs, we create first 5 different sub-series to avoid any overlap. Thus, sub-series j contains the estimates $\{VaR_j^{1-q}, VaR_{j+5}^{1-q}, VaR_{j+10}^{1-q}, \dots\}$ for $j = 1, \dots, 5$. According to the suggestion of Diebold *et al.* (1998), we use Bonferroni bounds for the 5 sub-series. That is, we assume that the VaR series has autocorrelation up to and including lag 4, whereas each sub-series should have correct coverage and independent VaR violations. Hence we therefore backtest each sub-series separately with a size of $\alpha/5$, with α the used significance level. Rejecting the null hypothesis of unconditional coverage/independence occurs when the null is rejected for *any* of the 5 sub-series. Second, we compare the 1-day VaR estimates of two different methods/models using the following asymmetric linear (tick) loss function of order q , which is also used in the CPA test of Giacomini and White (2006):

$$L_{A_i}^q(e_t) = (q - I[e_t < 0])e_t, \quad (5.20)$$

where $q = 5\%$ and 10% and $e_t = y_t - VaR_t^{1-q}$. The loss function is asymmetric in the sense that if there occurs a violation (i.e. $e_t < 0$) the negative number $q - 1$ is multiplied by the magnitude of the violation e_t , resulting in a penalization of $(1 - q) \times e_t$. In contrast to this, if there is no violation, the loss is equal to $q \times e_t$, which is considerable lower.¹¹ Hence a model A_i is more penalized when a VaR violation is observed. The larger the magnitude of this violation, the larger the penalization. Similar to the density forecasts, we define the relative loss as

$$d_t^q = L_{A_i}^q(e_t) - L_{A_j}^q(e_t) \quad (5.21)$$

and consider again a Diebold and Mariano (1995) type statistic as given in (5.18). A negative value of the unconditional mean of d_t^q means that on average the VaR estimates of model A_i are better than the corresponding estimates of model A_j .

¹¹Suppose the 95% 1-day VaR of model A and B are equal to -5% and -8% respectively, while the actual return is -6%. The loss associated with model A is equal to $(0.05 - 1)(-1) = 0.95$, while the loss of model B is equal to $(0.05 - 0)(-2) = 0.10$.

5.4.3 Results

In this subsection, we present both the statistical and economic results. In order to understand these results, we first present the weights which are obtained by optimizing the log score function of (5.3) and the *csl* score function of (5.5). Figure 5.2 shows the result of the iterative process of optimizing weights according to both score functions. The sub-graphs depict the dynamics of the weights using daily returns from the DJIA index according to the 14 models listed in Table 5.1.¹² The top part of the figure corresponds with the log score function, while the bottom part corresponds with the *csl* score function with $\kappa = 0.25$. The top part shows that using the log score function results in a large weight for the Heavy model with Skewed-*t* distributed errors until 2009. Moreover, it gets the full weight until 2008. Subsequently, the weight of the Heavy Skewed-*t* model declines to zero and there is room for the Heavy N model and the Heavy model with Laplace distributed returns. The latter gets almost a weight of 0.6 in 2011. Nevertheless, the Heavy Skewed-*t* model appears again and dominates from 2012 onwards.

A rather different dynamic pattern arises from the lower part of Figure 5.2, i.e. when the *csl* score function is optimized. Although the graph is similar in the sense that (i) the Heavy Skewed-*t* model dominates the other models during 2006-2007 and since 2012 and (ii) the Heavy model with Laplace distributed errors dominates during 2009-2012, the years 2008-2012 show two main differences. First, the GARCH Skewed-*t* model has more impact in case of the *csl* score function, reaching a maximum weight of 0.65 at the end of 2008. Second, the RGARCH model class gets considerably more weight, either combined with the Laplace distribution (2009) or the Normal distribution (2010-2011). Hence the Heavy N model is replaced by the RGARCH N model during the period 2010-2012.

To ease the interpretation of this finding, Figure 5.3 sums up the weights according to each model class (upper part) and distribution (lower part), for both types of scoring rules. It seems that in case of the log score function, the Heavy model with Skewed-*t* distributed errors dominates all the remaining models for most years. Only during 2009-2010, the RGARCH model has some influence, while the Skewed-*t* distribution is replaced by the Normal and Laplace distribution. In contrast to this, focusing of the left tail of the distri-

¹²Figure 5.A.1 in Appendix 5.A.2 provides weights corresponding to the other three stock market indexes.

bution does lead to more influence of the GARCH and RGARCH class of models. Furthermore, the Laplace distribution is more apparent during the years 2009-2012, with a climax at the start of 2012. Finally, both the Laplace and Normal distribution characterize 2012, while the log score function allocates the most weight to the Skewed- t distribution in that year.

Statistical results

Table 5.2 provides the importance of pooling of censored densities by showing results of the t -test on equal predictive accuracy of (5.18). In more detail, we test equal accuracy of the combined density forecasts based on the *csl* score function and based on the log score function. In addition, we test the accuracy of the individual censored density of each competing model. Panel A reports HAC-based t -statistics of the test of equal accuracy of density forecasts made by means of combination, using the *csl* or log score function of Section 5.2. As a benchmark, we consider also the case of equal weights assigned to each competing density. A positive number corresponds with more accurate density forecasts of the forecast method based on the *csl* score function. The table suggests that (except for the FTSE returns), combined forecasts based on the *csl* score function statistically outperform the density forecasts based on the log scoring rule. In addition, the benchmark forecasts are also improved, especially when $\kappa = 0.25$, as indicated by the t -statistics 1.65 and 1.77 (DJIA), 3.85 and 4.62 (FTSE) and 1.66 (Nikkei). However for the S&P 500 returns this improvement is not statistically significant.

Panel B shows test results of using combined forecasts based on the left tail and the individual censored density forecasts. In general, using the *csl* score function results statistically in better density forecasts. This holds in particular when $\kappa = 0.25$.¹³ Even if the null hypothesis cannot be rejected for a particular model, this result is not consistent for all data sets. For example, the Heavy Lap model performs well in case of the U.S. stock market indexes, but is statistically beaten in case of the FTSE returns. In general, there is no striking difference between the 1-step and the 5-step ahead density forecasts.

¹³Table 5.A.1 in Appendix 5.A.2 provides results where the weights are based on the log score function. The results indicate that the pooled density forecasts do not add any value in case of the S&P 500, DJIA and Nikkei indexes. Only in case of the FTSE index, there is evidence that the combined density forecasts statistically outperform the individual density forecasts.

Figure 5.2: Pooling weights DJIA index

This figure depicts the evolution of weights based on optimizing the logarithmic score function (upper part) of (5.3) or the *csl* score function (bottom part) of (5.5) with a moving window of $T = 750$ one-step ahead evaluated density forecasts using daily returns of the DJIA index. In case of the *csl* score function, B_t represents the left tail $y_t < \hat{r}^{0.25}$ with $\hat{r}^{0.25}$ the 0.25th quantile of the empirical CDF of the moving estimation window of 750 returns. The labels refer to the models that have the highest weight at a given period. The abbreviations “ST”, “Lap” and “N” stand for Skewed-*t*, Laplace and Normal respectively.

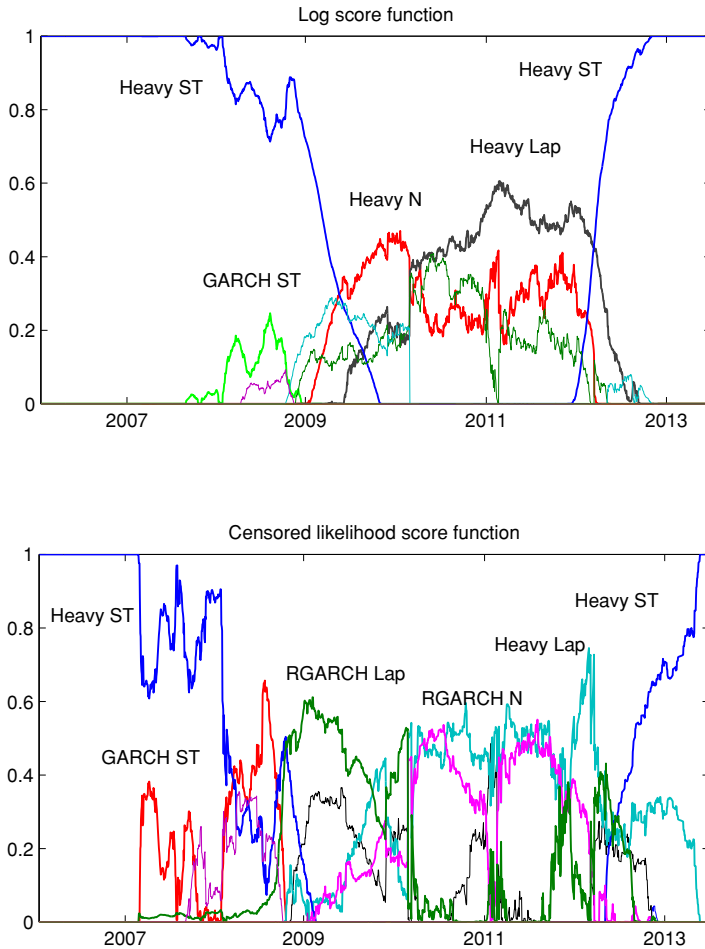


Figure 5.3: Pooling weights per model class and distribution

This figure sums up the optimized weights per model class (top panels) and distribution (bottom panels) based on optimizing the logarithmic score function (left part) of (5.3) or the *csl* score function of (5.5) (right part) with a moving window of $T = 750$ one-step ahead evaluated density forecasts using daily returns of the DJIA index. In case of the *csl* scoring function, B_t represents the left tail $y_t < \hat{p}^{0.25}$ with $\hat{p}^{0.25}$ the 0.25th quantile of the empirical CDF of the moving estimation window of 750 returns.

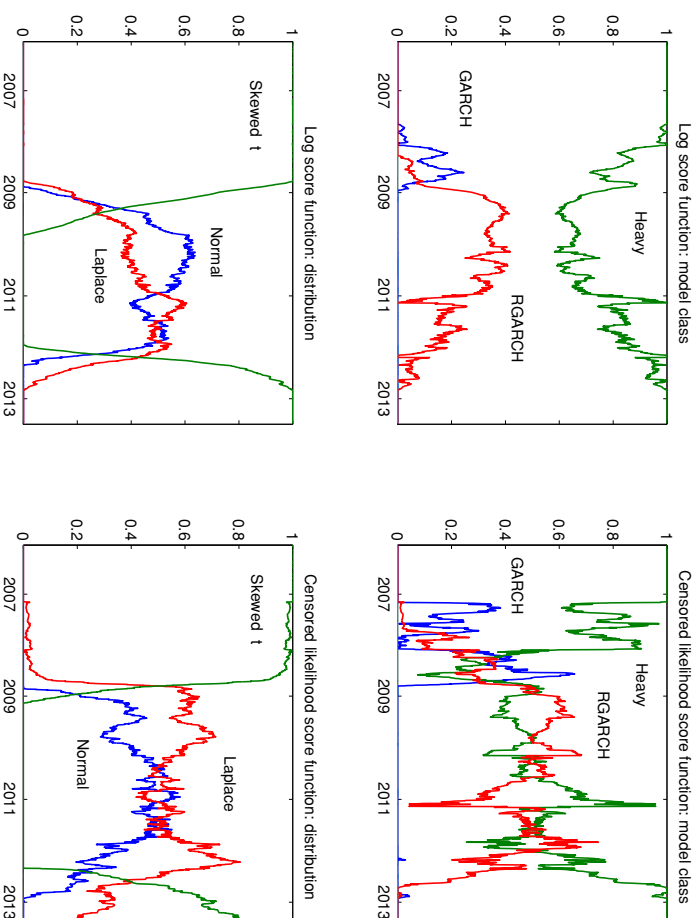


Table 5.2: Evaluation of 1- and 5-day ahead censored density forecasts

This table reports results of testing equal predictive accuracy using the censored likelihood scoring rule of (5.4), with B_t the left tail $y_t < \hat{r}_t^\kappa$ with \hat{r}_t^κ the κ th quantile of the empirical CDF of the in-sample returns. We set κ equal to 0.15 and 0.25 respectively. The weights are repeatedly optimized based on a moving window of 750 evaluated density forecasts. We focus on 1- and 5-step ahead density forecasts. The test statistic is given in (5.18). Panel A compares combined density forecasts where the weights are based on the *csl* score function with (i) weights based on the log score function and (ii) each competing model gets the same weight. In Panel B, we test equal predictive accuracy of combined density forecasts based on the *csl* score function and density forecasts of each competing model, which are listed in Table 5.1. All models are estimated with a moving window of 750 daily returns from the S&P 500, DJIA, FTSE and Nikkei index through the period January, 2000 - June, 2013. The test statistics are based on HAC-based standard errors and 1864 (S&P 500), 1866 (DJIA), 1882(FTSE) and 1766 (Nikkei) out-of-sample observations respectively.

Panel A: Csl score function vs. log score function and equal weighted								
	S&P500	DJIA	FTSE	Nikkei	S&P500	DJIA	FTSE	Nikkei
	1-step ahead forecasts				5-step ahead forecasts			
	$\kappa = 0.15$							
csl vs log	3.54***	3.60***	−0.09	2.69**	3.76***	3.78***	−1.15	2.36**
csl vs eqw	1.14	1.08	2.36**	0.89	0.74	1.83*	3.17***	1.03
	$\kappa = 0.25$							
csl vs log	3.06***	3.34***	−0.75	2.06**	3.34***	3.36***	−0.96	1.93*
csl vs eqw	1.32	1.65*	3.85***	1.66*	0.73	1.77*	4.62***	1.31
Panel B: Pooled (csl score function) vs. individual								
	$\kappa = 0.15$							
GARCH N	3.22***	3.30***	4.17***	2.27**	2.34**	2.43**	3.74***	2.00**
GARCH T	3.24***	2.96***	4.25***	2.26**	1.56	1.84*	4.41***	1.60
GARCH Lap	1.59	1.44	4.46***	2.65***	0.15	0.80	4.42***	1.34
GARCH ST	6.09***	5.64***	2.92***	3.71***	6.12***	5.96***	2.58***	3.50***
HEAVY N	1.41	1.50	3.13***	1.34	1.58	1.64	3.64***	1.57
HEAVY T	0.92	1.16	2.95***	0.29	0.26	0.75	3.96***	1.25
HEAVY Lap	−0.24	−0.13	3.42***	0.24	−0.99	−0.34	3.78***	0.16
HEAVY ST	5.41***	4.88***	−0.23	3.32***	5.50***	5.01***	0.23	2.88***
RGARCH N	1.44	1.53	3.57***	2.53**	3.09***	3.33***	5.62***	3.58***
RGARCH T	1.22	1.07	3.35***	2.04**	2.82***	3.77***	5.49***	3.44***
RGARCH Lap	0.21	0.26	3.62***	3.10***	1.36	2.38**	5.06***	3.63***
RGARCH ST	5.51***	5.14***	−0.60	5.64***	6.76***	7.32***	1.66*	6.31***
GAS T	2.95***	2.71***	4.27***	2.07**	1.53	1.65*	4.47***	1.78*
GAS Lap	1.45	1.42	4.43***	2.52**	0.09	0.70	4.45***	1.30
	$\kappa = 0.25$							
GARCH N	3.70***	3.89***	5.80***	2.53**	2.53**	2.64***	4.79***	2.09**
GARCH T	3.65***	3.72***	5.88***	3.07***	1.58	1.96*	5.78***	2.00**
GARCH Lap	1.89*	2.11**	5.52***	3.33***	0.11	0.82	5.39***	1.68*
GARCH ST	5.98***	5.60***	2.58***	3.26***	5.63***	5.58***	2.40**	3.13***
HEAVY N	2.12**	2.39**	5.16***	1.52	2.13**	2.26**	5.07***	1.65*
HEAVY T	1.66*	2.43**	4.86***	1.29	0.94	1.65*	5.62***	1.54
HEAVY Lap	0.18	0.81	4.59***	1.40	−0.80	0.20	4.94***	0.85
HEAVY ST	5.00***	4.50***	−1.25	2.35**	5.34***	4.55***	−0.35	2.52**
RGARCH N	2.52**	2.81***	5.62***	2.93***	3.98***	4.10***	7.05***	3.83***
RGARCH T	2.33**	2.31**	5.25***	3.05***	3.68***	4.27***	7.09***	3.93***
RGARCH Lap	0.66	0.86	4.78***	3.86***	1.30	2.33**	6.08***	4.12***
RGARCH ST	6.00***	5.88***	−1.19	7.56***	8.71***	9.16***	2.32*	8.57***
GAS T	3.36***	3.46***	5.89***	2.83***	1.57	1.82*	5.87***	2.19**
GAS Lap	1.75*	2.05**	5.48***	3.30***	0.02	0.69	5.44***	1.74*

Interestingly, considering the S&P 500 and DJIA indexes in the upper part of Panel B, the RGARCH model with Normal or Student- t distributed errors produces accurate 1-step ahead density forecasts, while forecasting 5 steps ahead results in inaccurate forecasts compared to pooled density forecasts with weights based on the 0.15th quantile of the individual densities.

Table 5.3 reports additional evidence of the added value of pooling using the *csl* score function, by providing the *csl* score over the out-of-sample period:

$$\sum_{t=1}^T \log \left[\sum_{i=1}^n w_{i,t-1}^* \left(I[y_t \in B_t] \log p_t(y_t; I_{t-1}, A_i) + I[y_t \in B_t^c] \int_{B_t^c} p_t(y; I_{t-1}, A_i) dy \right) \right] \quad (5.22)$$

where $w_{i,t-1}^*$ is the optimized weight for model A_i at the end of trading day $t - 1$, based on the evaluated density forecasts at time $t - T_w$ through $t - 1$. In addition, we provide corresponding values of the individual models with bold numbers representing the maximum *csl* score over the competing models. The pooled *csl* scores are higher than the *csl* scores of most of the individual models. If this is not the case, the differences are small, with a maximum difference of 9 points (S&P 500, 5-step ahead forecasts with $\kappa = 0.25$). Further, the *csl* score of our pooling method are higher than the scores of the remaining 13 models (12 in case of the FTSE index), with differences that can be quite substantial. For example, if one favours the best performing individual model, i.e. the Heavy Lap model, this results in a loss of 43 or 46 points with respect to the pooled *csl* score based with $\kappa = 0.15$ in case of the FTSE data set. Finally, there is quite some positive difference between the *csl* scores of pooling with weights based on the *csl* score function and simply using equal weights. For the U.S. and Japanese indexes, the difference is on average around 8 points, however in case of the FTSE index, the difference increases to 20 ($\kappa = 0.15$) or 45 ($\kappa = 0.25$) points respectively. Note that the table relates to Table 5.2, in the sense that a negative t -stat of a particular model corresponds with a higher *csl* score of that model than the pooled *csl* score. We refer to Table 5.A.2 in Appendix 5.A.2 for similar type of results regarding the log scores.

Figure 5.4: Censored likelihood scores w.r.t. individual models

This figure depicts the cumulative sum of the difference of the censored likelihood score corresponding with one-step ahead density forecasts of the pooled densities and the *csl* score of the three best competing individual models according to Table 5.3. The weights of the pooled densities are based on maximizing the *csl* score function of (5.5) with a moving window of 750 evaluated density forecasts, using daily returns of the DJIA index. Further, B_t the left tail $y_t < \hat{r}^{0.25}$ with $\hat{r}^{0.25}$ the 0.25th quantile of the empirical CDF of the in-sample returns.

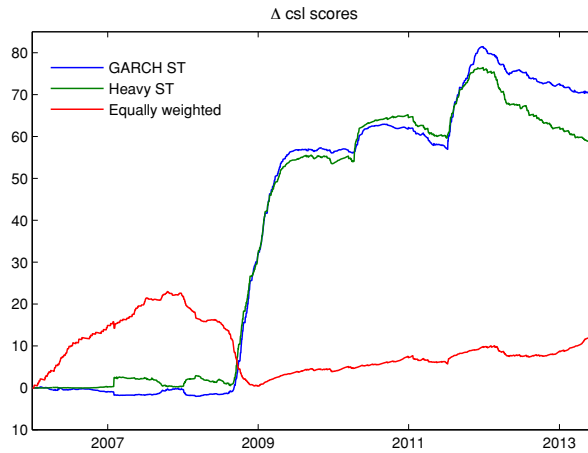


Table 5.3: Censored likelihood scores

This table reports censored likelihood scores corresponding with individual models and combined models, where the weights are based on optimizing the *csl* score function of (5.5), with B_t the left tail $y_t < \hat{r}^\kappa$ with \hat{r}^κ the κ th quantile of the empirical CDF of the in-sample returns. We set κ equal to 0.15 and 0.25. The weights are repeatedly optimized based on a moving window of 750 evaluated density forecasts. In addition, *csl* scores are reported of combined models using equal weights (eqw). The bold numbers represent the maximum of all models per data set. All models are estimated with a moving window of 750 daily returns from the S&P500, DJIA, FTSE and Nikkei index through the period January, 2000 - June, 2013. The number of out-of-sample observations are equal to 1864 (S&P 500), 1866 (DJIA), 1882(FTSE) and 1766 (Nikkei)respectively.

	S&P500	DJIA	FTSE	Nikkei	S&P500	DJIA	FTSE	Nikkei
	1-step ahead forecasts				5-step ahead forecasts			
	$\kappa = 0.15$							
GARCH N	-1052	-1014	-1020	-1002	-1081	-1031	-1050	-1083
GARCH T	-1030	-991	-1002	-937	-1039	-993	-1013	-956
GARCH Lap	-1019	-980	-1001	-934	-1029	-985	-1009	-952
GARCH ST	-1091	-1043	-964	-961	-1105	-1049	-973	-986
HEAVY N	-1026	-990	-988	-979	-1055	-1005	-1027	-1060
HEAVY T	-1013	-976	-983	-919	-1030	-985	-1007	-962
HEAVY Lap	-1004	-967	-990	-919	-1019	-977	-1004	-944
HEAVY ST	-1075	-1028	-946	-950	-1093	-1037	-959	-995
RGARCH N	-1022	-983	-992	-1001	-1079	-1027	-1051	-1096
RGARCH T	-1014	-975	-987	-936	-1050	-1010	-1023	-994
RGARCH Lap	-1008	-970	-991	-953	-1040	-998	-1017	-999
RGARCH ST	-1080	-1039	-944	-985	-1114	-1074	-969	-1055
GAS T	-1031	-990	-1005	-941	-1039	-992	-1015	-964
GAS Lap	-1019	-980	-1001	-935	-1028	-984	-1010	-951
pooled csl	-1006	-968	-947	-917	-1027	-979	-958	-943
eqw	-1013	-974	-968	-922	-1032	-988	-984	-950
	$\kappa = 0.25$							
GARCH N	-1480	-1443	-1364	-1338	-1508	-1458	-1398	-1421
GARCH T	-1454	-1419	-1343	-1274	-1461	-1417	-1358	-1295
GARCH Lap	-1440	-1404	-1336	-1273	-1447	-1405	-1348	-1291
GARCH ST	-1506	-1460	-1270	-1281	-1521	-1468	-1286	-1312
HEAVY N	-1453	-1421	-1332	-1313	-1486	-1438	-1376	-1398
HEAVY T	-1436	-1405	-1324	-1253	-1455	-1414	-1353	-1300
HEAVY Lap	-1422	-1390	-1324	-1256	-1437	-1399	-1342	-1283
HEAVY ST	-1485	-1439	-1244	-1265	-1512	-1456	-1265	-1318
RGARCH N	-1454	-1416	-1337	-1340	-1516	-1463	-1402	-1435
RGARCH T	-1441	-1403	-1328	-1274	-1481	-1441	-1370	-1334
RGARCH Lap	-1427	-1391	-1325	-1290	-1460	-1420	-1356	-1338
RGARCH ST	-1502	-1466	-1244	-1337	-1555	-1518	-1285	-1427
GAS T	-1455	-1418	-1346	-1278	-1462	-1416	-1360	-1303
GAS Lap	-1439	-1404	-1336	-1274	-1446	-1404	-1349	-1291
pooled csl	-1420	-1381	-1251	-1243	-1446	-1397	-1268	-1276
eqw	-1430	-1394	-1295	-1255	-1451	-1409	-1317	-1287

Figure 5.4 illustrates the evolution of the cumulative gain in the *csl* scores of Table 5.3 through time. In particular, it shows the cumulative difference of the *csl* scores corresponding with the combined density forecasts relative to the *csl* scores of the GARCH and Heavy models with a Skewed-*t* distribution and the benchmark (i.e. pooling based on equal weights). The figure shows two different patterns. First, the gain of pooling with respect to the benchmark occurs mainly during the first years, decreases during the crisis period and increases slowly from 2009 onwards. Second, pooling does not add much value with respect to the Skewed-*t* distribution during the first years, regardless whether the GARCH or Heavy model class is used. However, the gain becomes striking at the end of 2008 and in 2011. This result is related with Figure 5.2, as the Heavy model with Skewed-*t* distributed returns dominates all the remaining models until the end of 2009.

Economic results

Tables 5.4 and 5.5 shed light on the economic impact of pooling (censored) density forecasts by shedding VaR estimates. For each data set, we first compare the frequency and independence of the VaR violations corresponding with the combined densities based on pooling, either using the *csl* or log scoring rule or using equally weights. The latter can be seen as a benchmark. We report results of the approach using equal weights by using simulation corresponding with the actual weight ($eqw(1)$), and the approach that simply takes the average of all individual VaR's ($eqw(2)$), as done in Giacomini and Komunjer (2005). Second, we show results of each individual model per data set. Furthermore, we compare the accuracy of the 1-day VaR estimates based on the *csl* scoring rule with VaR estimates from any other pooling method or from any individual model by applying a *t*-test based on the asymmetric tick loss function of (5.21). A negative number indicates that the pooled *csl* based VaR estimates are more accurate. Apart from this test, both tables contain the same type of results, although Table 5.4 focuses on the 1-day VaR estimates, while Table 5.5 provides results of the 5-day VaR estimates.

Three main conclusions are apparent from Table 5.4. First, the VaR estimates corresponding with our new proposed technique outperform the benchmark of equal weights, both regarding the frequency of violations and the test on equal accuracy. Using equal weights leads to rejection of the nominal frequency of 5% for the S&P 500 and FTSE

Table 5.4: Evaluation of 1-day Value-at-Risk estimates

This table provides the accuracy of 1-day VaR estimates. For each data set, the table reports results based on combined density forecasts using the log scoring rule of (5.1), the *csl* scoring rule of (5.4) and using equal weights applied on 14 volatility models using daily returns from the S&P 500, DJIA, FTSE and Nikkei index over the period January, 2000 - June, 2013. In case of using equal weights, we report the approach by means of simulation (eqw(1)) and the approach that takes simply the average of all individual VaR estimates (eqw(2)). Further, we report results based on VaR estimates of the individual models. The columns represent for both 95% and 90% VaRs the number of violations, the percentage of violations with respect to the total number of VaR estimates in parentheses, the *p*-values of the Unconditional Coverage (UC) and Independence (Ind) test of Christoffersen (1998) and finally HAC-based *t*-statistics of the unconditional test on predictive ability of the combination method/individual model and the combined density forecasts with weights based on the *csl* score function, using the tick-loss function of (5.21). Bold numbers represent those models which have *p*-values for the UC and Ind test above 5% for both horizons. The number of estimated VaRs for each series is equal to 1864 (S&P 500), 1866 (DJIA), 1882(FTSE) and 1766 (Nikkei) respectively.

S&P500								
Model/Sc. rule	V(%)	p_{uc}	p_{ind}	t -stat	V(%)	p_{uc}	p_{ind}	t -stat
		95% VaR				90% VaR		
csl	111 (5.97)	0.063	0.484		200 (10.75)	0.284	0.056	
log	115 (6.18)	0.024	0.375	0.44	199 (10.70)	0.320	0.032	1.19
eqw(1)	123 (6.61)	0.002	0.089	-2.50**	213 (11.45)	0.041	0.043	-1.90*
eqw(2)	120 (6.45)	0.006	0.117	-2.38**	208 (11.18)	0.095	0.042	-1.39
GARCH N	124 (6.67)	0.002	0.192	-3.50***	192 (10.32)	0.644	0.127	-2.41**
GARCH T	132 (7.10)	0.000	0.385	-3.58***	229 (12.31)	0.001	0.021	-2.98***
GARCH Lap	125 (6.72)	0.001	0.177	-3.35***	229 (12.31)	0.001	0.021	-2.81***
GARCH ST	123 (6.61)	0.002	0.209	-3.50***	207 (11.13)	0.110	0.084	-2.57**
HEAVY N	125 (6.72)	0.001	0.177	-1.92*	191 (10.27)	0.700	0.020	1.21
HEAVY T	133 (7.15)	0.000	0.360	-1.93*	215 (11.56)	0.028	0.371	-0.56
HEAVY Lap	116 (6.24)	0.018	0.165	-0.69	220 (11.83)	0.010	0.166	0.29
HEAVY ST	113 (6.08)	0.039	0.210	-0.47	194 (10.43)	0.539	0.058	0.92
RGARCH N	113 (6.08)	0.039	0.427	-1.92*	185 (9.95)	0.938	0.041	-0.53
RGARCH T	123 (6.61)	0.002	0.209	-2.15**	208 (11.18)	0.095	0.042	-2.40**
RGARCH Lap	108 (5.81)	0.119	0.577	-1.86*	214 (11.51)	0.034	0.038	-2.00**
RGARCH ST	103 (5.54)	0.295	0.750	-0.96	189 (10.16)	0.817	0.168	-1.23
GAS T	130 (6.99)	0.000	0.044	-3.28***	223 (11.99)	0.005	0.024	-2.67***
GAS Lap	122 (6.56)	0.003	0.032	-3.03***	223 (11.99)	0.005	0.012	-2.51**
DJIA								
csl	116 (6.23)	0.019	0.744		202 (10.85)	0.228	0.024	
log	120 (6.44)	0.006	0.900	1.19	211 (11.33)	0.060	0.031	0.65
eqw(1)	120 (6.44)	0.006	0.792	-1.06	210 (11.28)	0.071	0.112	-0.79
eqw(2)	118 (6.34)	0.011	0.567	-0.81	209 (11.22)	0.083	0.123	-0.73
GARCH N	130 (6.98)	0.000	0.456	-3.08***	198 (10.63)	0.366	0.074	-1.92*
GARCH T	137 (7.36)	0.000	0.479	-2.95***	221 (11.87)	0.009	0.059	-2.43**
GARCH Lap	116 (6.23)	0.019	0.364	-2.47**	223 (11.98)	0.006	0.204	-2.35**
GARCH ST	124 (6.66)	0.002	0.647	-2.62***	212 (11.39)	0.051	0.150	-2.10**
HEAVY N	130 (6.98)	0.000	0.713	0.12	201 (10.79)	0.258	0.161	1.81*
HEAVY T	138 (7.41)	0.000	0.690	-0.47	214 (11.49)	0.035	0.074	-0.08
HEAVY Lap	103 (5.53)	0.300	0.876	1.18	215 (11.55)	0.029	0.037	0.55
HEAVY ST	112 (6.02)	0.051	0.597	1.39	203 (10.90)	0.200	0.042	1.15
RGARCH N	117 (6.28)	0.014	0.783	-1.15	187 (10.04)	0.951	0.069	0.46
RGARCH T	124 (6.66)	0.002	0.767	-1.69*	202 (10.85)	0.228	0.048	-1.02
RGARCH Lap	103 (5.53)	0.300	0.876	-1.03	199 (10.69)	0.328	0.117	-0.49
RGARCH ST	101 (5.42)	0.407	0.798	-0.33	191 (10.26)	0.712	0.045	-0.67
GAS T	133 (7.14)	0.000	0.201	-2.79***	216 (11.60)	0.024	0.059	-1.97**
GAS Lap	113 (6.07)	0.040	0.220	-2.40**	219 (11.76)	0.013	0.074	-2.00**

(continued from previous page)

FTSE								
Sc. rule/Model	V(%)	p_{uc}	p_{ind}	t -stat	V(%)	p_{uc}	p_{ind}	t -stat
		95% VaR				90% VaR		
csl	112 (5.96)	0.063	0.246		208 (11.08)	0.126	0.483	
log	108 (5.75)	0.144	0.328	0.94	206 (10.97)	0.167	0.398	1.44
eqw(1)	115 (6.12)	0.031	0.195	-1.49	208 (11.08)	0.126	0.232	-1.24
eqw(2)	114 (6.07)	0.039	0.211	-1.54	209 (11.13)	0.109	0.213	-1.34
GARCH N	122 (6.50)	0.004	0.737	-2.52**	203 (10.81)	0.248	0.145	-1.98**
GARCH T	129 (6.87)	0.000	0.502	-2.61***	222 (11.82)	0.010	0.975	-2.39**
GARCH Lap	105 (5.59)	0.248	0.714	-2.26**	224 (11.93)	0.007	0.721	-2.48**
GARCH ST	106 (5.64)	0.209	0.678	-2.40**	210 (11.18)	0.093	0.422	-1.93*
HEAVY N	124 (6.60)	0.002	0.405	-0.11	202 (10.76)	0.280	0.691	1.25
HEAVY T	135 (7.19)	0.000	0.179	-0.76	219 (11.66)	0.019	0.433	0.19
HEAVY Lap	97 (5.17)	0.744	0.985	-0.05	227 (12.09)	0.003	0.336	-0.56
HEAVY ST	103 (5.48)	0.342	0.786	0.75	199 (10.60)	0.393	0.808	0.99
RGARCH N	128 (6.82)	0.001	0.308	-1.98**	201 (10.70)	0.315	0.105	-0.98
RGARCH T	136 (7.24)	0.000	0.316	-2.22**	220 (11.71)	0.016	0.285	-1.56
RGARCH Lap	101 (5.38)	0.457	0.861	-0.93	225 (11.98)	0.005	0.186	-1.96*
RGARCH ST	112 (5.96)	0.063	0.246	-0.53	200 (10.65)	0.353	0.190	-0.53
GAS T	126 (6.71)	0.001	0.354	-2.60***	218 (11.61)	0.023	0.963	-2.34**
GAS Lap	93 (4.95)	0.924	0.416	-2.53**	224 (11.93)	0.007	0.721	-2.26**
Nikkei								
csl	90 (5.11)	0.836	0.161		178 (10.10)	0.887	0.001	
log	89 (5.05)	0.922	0.175	-0.03	172 (9.76)	0.738	0.004	-0.72
eqw(1)	84 (4.77)	0.652	0.248	-0.62	167 (9.48)	0.462	0.007	-1.51
eqw(2)	86 (4.88)	0.818	0.221	-1.05	164 (9.31)	0.328	0.011	-1.78*
GARCH N	104 (5.90)	0.091	0.042	-0.85	166 (9.42)	0.414	0.047	-1.69*
GARCH T	108 (6.13)	0.035	0.027	-1.04	189 (10.73)	0.314	0.030	-1.51
GARCH Lap	88 (4.99)	0.991	0.182	-0.66	181 (10.27)	0.704	0.038	-1.64
GARCH ST	100 (5.68)	0.202	0.062	-0.72	178 (10.10)	0.887	0.026	-1.50
HEAVY N	91 (5.16)	0.752	0.142	1.73*	162 (9.19)	0.254	0.014	1.42
HEAVY T	99 (5.62)	0.242	0.072	1.70*	176 (9.99)	0.987	0.002	2.46**
HEAVY Lap	74 (4.20)	0.113	0.150	1.15	175 (9.93)	0.924	0.002	2.02**
HEAVY ST	88 (4.99)	0.991	0.189	1.93*	172 (9.76)	0.738	0.004	1.90*
RGARCH N	85 (4.82)	0.733	0.566	-2.51**	144 (8.17)	0.008	0.110	-3.56***
RGARCH T	90 (5.11)	0.836	0.417	-2.71***	160 (9.08)	0.192	0.094	-2.81***
RGARCH Lap	73 (4.14)	0.089	0.993	-2.88***	151 (8.57)	0.041	0.217	-3.32***
RGARCH ST	76 (4.31)	0.176	0.889	-3.11***	152 (8.63)	0.050	0.200	-3.48***
GAS T	106 (6.02)	0.058	0.034	-1.48	191 (10.84)	0.246	0.011	-1.52
GAS Lap	86 (4.88)	0.818	0.213	-1.60	183 (10.39)	0.591	0.002	-2.19**

Table 5.5: Evaluation of 5-day Value-at-Risk estimates

This table provides the accuracy of 5-day VaR estimates. For each data set, the table reports results based on combined density forecasts using the log scoring rule of (5.1), the *csl* scoring rule of (5.4) and using equal weights applied on 14 volatility models using daily returns from the S&P 500, DJIA, FTSE and Nikkei index over the period January, 2000 - June, 2013. In case of using equal weights, we report the approach by means of simulation (eqw(1)) and the approach that takes simply the average of all individual VaR estimates (eqw(2)). Further, we report results based on VaR estimates of the individual models. For each combination method/model, we have 5 different sub-series of VaRs. The table reports for both 95% and 90% VaRs the sub-series that has the lowest *p*-value of the test on unconditional coverage of the VaR estimates. The columns represents the corresponding number of violations, the percentage of violations with respect to the total number of VaR estimates in parentheses and the *p*-values of the Unconditional Coverage (UC) and Independence (Ind) test of Christoffersen (1998). The number of estimated VaRs for each series is equal to 372 (S&P 500), 372 (DJIA), 375(FTSE) and 352 (Nikkei) respectively.

S&P500						
Model/sc. rule	V(%)	<i>p_{uc}</i>	<i>p_{m,ind}</i>	V(%)	<i>p_{uc}</i>	<i>p_{ind}</i>
	95% VaR			90% VaR		
csl	24 (6.45)	0.218	0.712	43 (11.56)	0.327	0.615
log	24 (6.45)	0.218	0.712	44 (11.83)	0.252	0.703
eqw(1)	24 (6.45)	0.218	0.615	46 (12.37)	0.141	0.888
eqw(2)	24 (6.45)	0.218	0.615	44 (11.83)	0.252	0.913
GARCH N	26 (6.99)	0.096	0.479	45 (12.10)	0.190	0.461
GARCH T	28 (7.53)	0.037	0.390	55 (14.78)	0.004	0.360
GARCH Lap	25 (6.72)	0.147	0.544	49 (13.17)	0.051	0.237
GARCH ST	24 (6.45)	0.218	0.615	46 (12.37)	0.141	0.396
HEAVY N	27 (7.26)	0.060	0.455	43 (11.56)	0.327	0.993
HEAVY T	28 (7.53)	0.037	0.531	45 (12.10)	0.190	0.821
HEAVY Lap	26 (6.99)	0.096	0.889	47 (12.63)	0.102	0.648
HEAVY ST	15 (4.03)	0.376	0.278	43 (11.56)	0.327	0.993
RGARCH N	23 (6.18)	0.312	0.729	30 (8.06)	0.199	0.802
RGARCH T	24 (6.45)	0.218	0.712	43 (11.56)	0.327	0.993
RGARCH Lap	23 (6.18)	0.312	0.729	33 (8.89)	0.470	0.971
RGARCH ST	13 (3.49)	0.160	0.427	29 (7.80)	0.142	0.337
GAS T	28 (7.53)	0.037	0.363	51 (13.71)	0.023	0.652
GAS Lap	25 (6.72)	0.147	0.320	49 (13.17)	0.051	0.237
DJIA						
csl	25 (6.72)	0.147	0.579	47 (12.63)	0.102	0.648
log	25 (6.72)	0.147	0.579	47 (12.63)	0.102	0.648
eqw(1)	26 (6.99)	0.096	0.844	48 (12.90)	0.073	0.567
eqw(2)	28 (7.53)	0.037	0.492	48 (12.90)	0.073	0.567
GARCH N	32 (8.60)	0.004	0.830	48 (12.90)	0.073	0.567
GARCH T	29 (7.80)	0.022	0.570	50 (13.44)	0.035	0.196
GARCH Lap	28 (7.53)	0.037	0.492	48 (12.90)	0.073	0.284
GARCH ST	28 (7.53)	0.037	0.492	49 (13.17)	0.051	0.237
HEAVY N	26 (6.99)	0.096	0.511	49 (13.17)	0.051	0.492
HEAVY T	27 (7.26)	0.060	0.933	53 (14.25)	0.010	0.494
HEAVY Lap	23 (6.18)	0.312	0.729	51 (13.71)	0.023	0.360
HEAVY ST	24 (6.45)	0.218	0.652	48 (12.90)	0.073	0.567
RGARCH N	23 (6.18)	0.312	0.729	30 (8.04)	0.194	0.763
RGARCH T	25 (6.72)	0.147	0.579	31 (8.31)	0.264	0.782
RGARCH Lap	15 (4.03)	0.376	0.132	30 (8.04)	0.194	0.763
RGARCH ST	14 (3.76)	0.253	0.099	30 (8.04)	0.194	0.763
GAS T	28 (7.53)	0.037	0.492	46 (12.37)	0.141	0.733
GAS Lap	27 (7.26)	0.060	0.418	46 (12.37)	0.141	0.733

(continued from previous page)

FTSE						
Sc. rule/Model	V(%)	p_{uc}	p_{ind}	V(%)	p_{uc}	p_{ind}
	95% VaR			90% VaR		
csI	24 (6.38)	0.237	0.701	33 (8.80)	0.430	0.214
log	23 (6.12)	0.336	0.616	32 (8.53)	0.333	0.171
eqw(1)	24 (6.38)	0.237	0.059	34 (9.07)	0.541	0.085
eqw(2)	23 (6.12)	0.336	0.043	32 (8.53)	0.333	0.047
GARCH N	24 (6.40)	0.232	0.232	34 (9.07)	0.541	0.098
GARCH T	25 (6.65)	0.162	0.080	41 (10.90)	0.564	0.029
GARCH Lap	22 (5.87)	0.453	0.145	33 (8.80)	0.430	0.074
GARCH ST	24 (6.40)	0.232	0.232	33 (8.80)	0.430	0.074
HEAVY N	24 (6.38)	0.237	0.059	33 (8.80)	0.430	0.214
HEAVY T	25 (6.65)	0.162	0.788	41 (10.90)	0.564	0.073
HEAVY Lap	22 (5.85)	0.460	0.144	33 (8.80)	0.430	0.214
HEAVY ST	23 (6.12)	0.336	0.184	33 (8.80)	0.430	0.909
RGARCH N	26 (6.91)	0.106	0.375	32 (8.53)	0.333	0.430
RGARCH T	27 (7.18)	0.068	0.409	33 (8.80)	0.430	0.214
RGARCH Lap	25 (6.65)	0.162	0.312	31 (8.27)	0.250	0.774
RGARCH ST	26 (6.91)	0.106	0.375	31 (8.27)	0.250	0.774
GAS T	24 (6.40)	0.232	0.232	41 (10.90)	0.564	0.029
GAS Lap	21 (5.60)	0.601	0.901	33 (8.80)	0.430	0.192
Nikkei						
csI	13 (3.69)	0.239	0.317	21 (5.95)	0.006	0.804
log	14 (3.98)	0.362	0.281	20 (5.67)	0.003	0.890
eqw(1)	12 (3.40)	0.144	0.414	20 (5.67)	0.003	0.432
eqw(2)	11 (3.12)	0.082	0.341	19 (5.38)	0.002	0.979
GARCH N	14 (3.97)	0.356	0.575	24 (6.80)	0.034	0.767
GARCH T	13 (3.68)	0.234	0.492	27 (7.65)	0.126	0.957
GARCH Lap	12 (3.40)	0.144	0.414	24 (6.80)	0.034	0.767
GARCH ST	13 (3.68)	0.234	0.492	26 (7.37)	0.085	0.951
HEAVY N	13 (3.69)	0.239	0.317	21 (5.95)	0.006	0.804
HEAVY T	14 (3.98)	0.362	0.576	24 (6.80)	0.034	0.767
HEAVY Lap	9 (2.56)	0.021	0.491	19 (5.38)	0.002	0.979
HEAVY ST	14 (3.98)	0.362	0.281	21 (5.95)	0.006	0.804
RGARCH N	12 (3.41)	0.147	0.357	18 (5.10)	0.001	0.009
RGARCH T	12 (3.41)	0.147	0.415	19 (5.38)	0.002	0.002
RGARCH Lap	10 (2.84)	0.044	0.275	17 (4.82)	0.000	0.006
RGARCH ST	12 (3.40)	0.144	0.005	17 (4.82)	0.000	0.006
GAS T	21 (5.97)	0.419	0.802	28 (7.93)	0.181	0.590
GAS Lap	11 (3.12)	0.082	0.341	26 (7.37)	0.085	0.431

indexes, while this is not the case for VaR estimates based on the *csl* score function. Furthermore, the 90% VaR estimates using the *csl* scoring rule are closer to its nominal values using optimized weights. According to the *t*-statistics of equal accuracy of the VaR estimates, using the *csl* score function produces significantly better VaR estimates than the benchmark in case of the S&P 500 index (both 90% and 95%) and the Nikkei index (90%).

Second, pooling based on the *csl* scoring rule generally outperforms pooling based on the log scoring rule, but only from the perspective of the nominal frequency of the VaR violations. The differences between the *csl* and log scoring rules arise mainly in the case of U.S. stock market indexes. For example, considering the S&P 500, using the whole density implies a violation frequency that is significantly different from 5% (using a significance level of 5%), while it is not significant using the *csl* scoring rule. A similar view arises in case of the DJIA returns, although the *csl* scoring rule produces still too many violations of the 95% VaR. Nevertheless, the unconditional coverage corresponding with the *csl* scoring rule is closer to its nominal value, especially for the 90% VaR estimates (10.85% vs. 11.33%). Finally, according to the *t*-statistics, there is no significant difference between the accuracy of the VaR of both pooling methods, although all numbers are negative.

Third, there is no single model that consistently outperforms our method of combining density forecasts. Each model fails at least once in the frequency of violations or in the test of equal accuracy. The best competitors are the HEAVY and RGARCH model classes. The RGARCH model with Skewed-*t* distributed errors outperforms the combined approach in case of the DJIA returns with regards to the frequency of violations, however the unconditional coverage of 10% is borderline significant in case of the Nikkei data set. Moreover, using the same data set, the corresponding *t*-statistics on equal accuracy favour significantly our combined method.

The differences between our various methods to estimate a VaR become much smaller if we put attention to the 5-day estimated VaRs, as indicated by Table 5.5. Using the Bonferroni bound corresponding with a 5% significance level, we conclude that using the *csl* scoring rule or the log scoring rule to obtain weights does not make a clear difference in the VaR estimates. In addition, the individual models perform also well. This could

be explained partly by the decreasing power of the tests when the number of exceptions decreases. Although there is a gain of using the *csl* scoring rule over the other pooling methods in case of 90% 5-day VaR estimates of the daily Nikkei returns, the difference boils down to one exception.

To summarize, short-horizon VaR estimates improve when using combined density forecasts based on the *csl* score function, either with respect the nominal size and/or with respect to the statistical accuracy using the asymmetric tick-loss function of (5.21).

5.5 Conclusion

We investigate the benefits of combining density forecasts based on a specific region of interest. We develop a new density forecast method that combines density forecasts of different models based on the censored likelihood scoring rule (Diks *et al.*, 2011). Using daily returns from the S&P 500, DJIA, FTSE and Nikkei stock market indexes from 2000 until 2013, we apply our technique on recently developed univariate volatility models, including the HEAVY, GAS and Realized GARCH models.

Our results show that density forecasts in the tail are statistically more accurate if one pools density forecasts using the censored likelihood scoring rule than using density forecasts based on the log score rule, using the benchmark of equal weights or density forecasts of any individual volatility model. Second, we show that the 1-day 95% and 90% VaR estimates improve significantly compared to the benchmark forecasting method or the method based on the log scoring rule. Moreover, the VaR estimates of each individual is beaten, either with respect to the nominal frequency of the VaR violations, or with respect to a statistical test on equal accuracy of the VaR estimates. Our results imply that risk managers and portfolio managers should not rely on one single model if they are interested in the left tail. Instead, they should make combinations of density forecasts using the *csl* scoring rule.

5.A Appendix

5.A.1 Optimizing weights

We follow Conflitti *et al.* (2012) to optimize the weights according to the log or *csl* score function of (5.3) and (5.5) respectively. We provide here only an outline of the algorithm.

Define $\mathbf{p}(y_{t+1})$ as the vector of n density forecasts $p_i(y_{t+1}) = p_{t+1}(y_{t+1}; Y_t, A_i)$ ($i = 1, \dots, n$) of the variable y_{t+1} at time t over a one-day horizon. The combined density is then equal to:

$$p(y_{t+1}) = \mathbf{w}' \mathbf{p}(y_{t+1}) = \sum_{i=1}^n w_i p_i(y_{t+1}), \quad (5.A.1)$$

with the assumption that the weights are positive and sum to one. For both scoring rules, we have to maximize the logarithm of the combined (censored) density over a given time period:¹⁴

$$\Phi(\mathbf{w}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \log p(y_{t+1}). \quad (5.A.2)$$

Note that we omitted the factor $\frac{1}{T-1}$ in this chapter. This does not change the result as it is a constant. Define the $(T-1) \times n$ matrix \hat{P} with non-negative elements $P_{ti} = p_i(y_{t+1})$. Now, (5.A.2) can be rewritten as $\frac{1}{T-1} \sum_{t=1}^{T-1} \log(P\mathbf{w}_t)$. Denote \mathbf{w}_{opt} as the maximum of $\Phi(\mathbf{w})$ subject to the weight constraints. Further, the Lagrange multiplier is introduced to take into account these constraints:

$$\Phi_\lambda(\mathbf{w}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \log(P\mathbf{w}_t) - \lambda \sum_{i=1}^N w_i. \quad (5.A.3)$$

¹⁴For the log scoring rule, it is indeed the log of the combined density. For the *csl* scoring rule, it is the log of the first part (corresponding with the region B_t) or the second part (corresponding with the region outside B_t) of (5.5).

Instead of optimizing (5.A.3), Conflitti *et al.* (2012) consider the following “surrogate” function, which depends on a vector \mathbf{a} of arbitrary weights:

$$\Psi_\lambda(\mathbf{w}; \mathbf{a}) = \frac{1}{T-1} \sum_{t=1}^{T-1} \sum_{i=1}^n b_{ti} \log \left(\frac{w_i}{a_i} \sum_{l=1}^n \log P_{tl} a_l \right) - \lambda \sum_{i=1}^n w_i. \quad (5.A.4)$$

with $b_{ti} = \frac{P_{ti} a_i}{\sum_{l=1}^n P_{tl} a_l}$. Further, the function has the properties $\Psi_\lambda(\mathbf{a}; \mathbf{a}) = \Psi_\lambda(\mathbf{a})$ for any \mathbf{a} and $\Psi_\lambda(\mathbf{w}; \mathbf{a}) \leq \Psi_\lambda(\mathbf{w})$ for any \mathbf{a} and \mathbf{w} .

The iterative algorithm is now defined as

$$\mathbf{w}_\lambda^{(k+1)} = \arg \max_w \Psi_\lambda(\mathbf{w}; \mathbf{w}_\lambda^{(k)}) \quad (5.A.5)$$

which yields a monotonic increase of Ψ_λ , according to the two aforementioned properties. Setting the derivatives of $\Psi_\lambda(\mathbf{w}; \mathbf{w}_\lambda^{(k)})$ with respect to w_i equal to zero leads to the maximum $w_{\lambda,i} = (1/\lambda) \sum_{t=1}^{T-1} b_{ti}$. Using the constraint that the weights should sum up to one, it holds that $\lambda = T - 1$. This changes (5.A.5) into

$$w_i^{(k+1)} = w_i^{(k)} \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{P_{ti}}{\sum_{l=1}^n P_{tl} w_l^{(k)}} \quad (5.A.6)$$

where we replace a_i by $w_i^{(k)}$ in the expression of b_{ti} . We start the algorithm with equal weights, that is $w_i^0 = 1/n$ and use as a stopping criterion a tolerance of $1e^{-6}$ of the sum of the absolute deviation of two successive iterates.

5.A.2 Pooling results of the log score function

Figure 5.A.1: Pooling weights of the S&P 500, FTSE and Nikkei index

This figure depicts the evolution of weights based on optimizing the logarithmic score function (left part) of (5.3) or the *csl* score function (right part) of (5.5) with a moving window of $T = 750$ one-step ahead evaluated density forecasts using daily returns of the S&P500, FTSE and Nikkei indexes. In case of the *csl* score function, B_t the left tail $y_t < \hat{r}^{0.25}$ with $\hat{r}^{0.25}$ the 0.25th quantile of the empirical CDF of the in-sample returns. The labels refer to the models that have the highest weight at a given period. The abbreviations “ST”, “Lap” and “N” stand for Skewed-*t*, Laplace and Normal respectively.

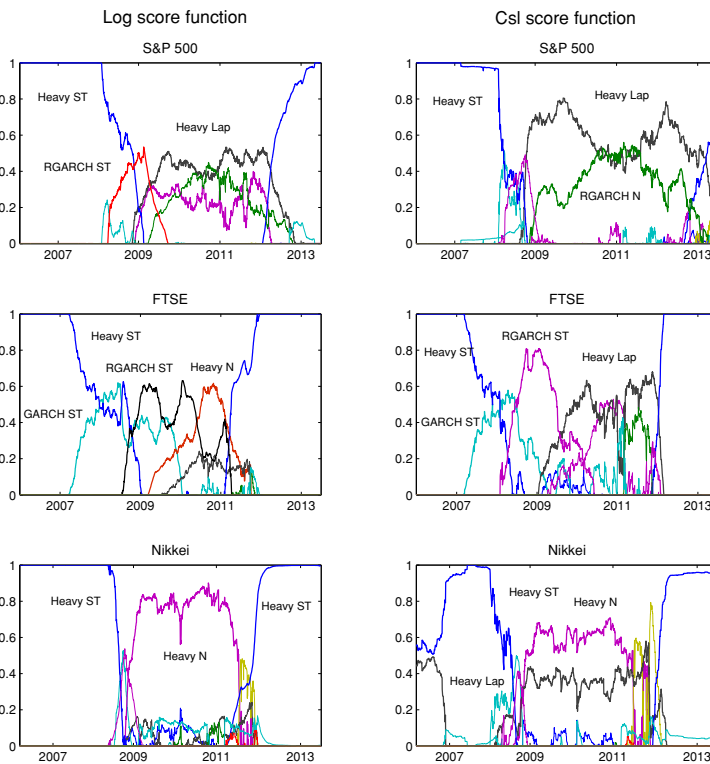


Table 5.A.1: Evaluation of 1- and 5-day ahead censored density forecasts based on the log scoring rule

This table reports results of testing equal predictive accuracy using the censored likelihood scoring rule of (5.4), with B_t the left tail $y_t < \hat{r}^\kappa$ with \hat{r}^κ the κ th quantile of the empirical CDF of the in-sample returns. We set κ equal to 0.15 and 0.25 respectively. The weights are repeatedly optimized based on a the log score function of (5.3), using a moving window of 750 evaluated density forecasts. We focus on 1- and 5-step ahead density forecasts. The test statistic is given in (5.18) and compares censored density forecast with weights based on the log score function and density forecasts of each competing model, which are listed in Table 5.1. All models are estimated with a moving window of 750 daily returns from the S&P500, DJIA, FTSE and Nikkei index through the period January, 2000 - June, 2013. The test statistics are based on HAC-based standard errors and 1864 (S&P 500), 1866 (DJIA), 1882(FTSE) and 1766 (Nikkei) out-of-sample observations respectively.

Pooled (log score function) vs. individual								
	S&P500	DJIA	FTSE	Nikkei	S&P500	DJIA	FTSE	Nikkei
	1-step ahead forecasts				5-step ahead forecasts			
	$\kappa = 0.15$							
GARCH N	0.70	0.64	4.12***	1.86*	0.80	0.73	3.71***	1.82*
GARCH T	−0.78	−0.86	4.15***	0.35	−1.63	−1.50	4.44***	−0.63
GARCH Lap	−1.59	−1.71*	4.38***	0.11	−2.42**	−2.20**	4.64***	−0.92
GARCH ST	6.34***	5.12***	3.37***	3.37***	5.85***	4.98***	3.18***	3.31***
HEAVY N	−0.78	−0.63	3.03***	0.99	−0.19	−0.40	3.62***	1.38
HEAVY T	−2.05**	−2.00**	2.85***	−1.80*	−2.27**	−2.12**	3.95***	0.07
HEAVY Lap	−2.75***	−2.73***	3.34***	−1.50	−3.25***	−2.90***	3.98***	−1.64
HEAVY ST	5.92***	4.90***	−0.31	2.86***	5.39***	4.65***	0.79	2.62***
RGARCH N	−1.11	−1.30	3.45***	2.07**	1.01	0.84	5.54***	3.45***
RGARCH T	−2.06**	−2.08**	3.23***	0.22	−0.77	−0.21	5.53***	2.32**
RGARCH Lap	−2.42**	−2.37**	3.54***	1.64	−1.60	−1.17	5.35***	2.41**
RGARCH ST	5.39***	4.27***	−0.64	5.08***	7.18***	8.19***	2.08**	6.01***
GAS T	−0.66	−0.90	4.20***	0.64	−1.50	−1.47	4.50***	0.29
GAS Lap	−1.59	−1.67*	4.37***	0.19	−2.42**	−2.21**	4.67***	−1.03
	$\kappa = 0.25$							
GARCH N	1.84*	1.76*	5.97***	2.33**	1.44	1.34	5.00***	2.02**
GARCH T	0.72	0.72	5.95***	2.04**	−0.51	−0.45	5.98***	0.86
GARCH Lap	−0.26	−0.24	5.49***	1.96*	−1.48	−1.22	5.46***	0.38
GARCH ST	6.06***	5.21***	3.52***	3.11***	5.36***	4.78***	3.32***	2.87***
HEAVY N	0.47	0.62	5.21***	1.35	0.81	0.69	5.21***	1.56
HEAVY T	−0.56	−0.22	4.87***	0.23	−0.91	−0.62	5.74***	0.97
HEAVY Lap	−1.46	−1.16	4.54***	0.38	−2.21**	−1.66*	4.93***	−0.27
HEAVY ST	5.11***	4.17***	−1.46	2.03**	5.03***	3.92***	0.07	2.44**
RGARCH N	0.56	0.45	5.69***	2.70***	2.37**	2.12**	7.26***	3.82***
RGARCH T	−0.22	−0.33	5.26***	2.02**	0.96	1.18	7.25***	3.18***
RGARCH Lap	−1.11	−1.08	4.73***	2.99***	−0.58	−0.25	6.08***	3.12***
RGARCH ST	6.18***	5.58***	−1.05	7.26***	9.38***	9.89***	2.82***	8.25***
GAS T	0.74	0.63	6.00***	2.13**	−0.43	−0.47	6.06***	1.50
GAS Lap	−0.32	−0.25	5.46***	2.07**	−1.52	−1.28	5.48***	0.41

Table 5.A.2: Log scores

This table reports log scores corresponding with individual models and combined models, where the weights are based on optimizing the log score function of (5.3). The weights are repeatedly optimized based on a moving window of 750 evaluated density forecasts. The bold numbers represent the maximum of all models per data set. All models are estimated with a moving window of 750 daily returns from the S&P500, DJIA, FTSE and Nikkei index through the period January, 2000 - June, 2013. The number of out-of-sample observations are equal to 1864 (S&P 500), 1866 (DJIA), 1882(FTSE) and 1766 (Nikkei)respectively.

	S&P500	DJIA	FTSE	Nikkei	S&P500	DJIA	FTSE	Nikkei
	1-step ahead forecasts				5-step ahead forecasts			
GARCH N	-2687	-2609	-2338	-2500	-2735	-2648	-2400	-2642
GARCH T	-2651	-2574	-2314	-2434	-2673	-2588	-2348	-2491
GARCH Lap	-2642	-2568	-2338	-2454	-2662	-2584	-2364	-2499
GARCH ST	-2653	-2507	-1948	-2377	-2711	-2564	-2017	-2480
HEAVY N	-2622	-2552	-2269	-2459	-2698	-2611	-2350	-2603
HEAVY T	-2603	-2533	-2263	-2399	-2660	-2578	-2321	-2490
HEAVY Lap	-2603	-2535	-2309	-2430	-2646	-2572	-2348	-2483
HEAVY ST	-2542	-2402	-1812	-2301	-2683	-2518	-1907	-2466
RGARCH N	-2636	-2564	-2281	-2517	-2780	-2691	-2420	-2742
RGARCH T	-2622	-2549	-2276	-2453	-2728	-2647	-2373	-2584
RGARCH Lap	-2619	-2546	-2315	-2493	-2703	-2627	-2384	-2598
RGARCH ST	-2634	-2541	-1853	-2671	-2898	-2802	-2065	-2987
GAS T	-2655	-2573	-2317	-2439	-2678	-2591	-2353	-2495
GAS Lap	-2643	-2567	-2339	-2457	-2662	-2582	-2366	-2494
pooled log	-2488	-2374	-1867	-2310	-2601	-2472	-1959	-2412

Nederlandse samenvatting

(Summary in Dutch)

In de laatste decennia is volatiliteit één van de meest actieve en succesvolle onderzoeksgebieden in de financiële (tijdreeks) econometrie. Volatiliteit wordt informeel beschouwd wel aangeduid als de beweeglijkheid van financiële instrumenten, zoals aandelen of obligaties. Het is de meest gebruikte maatstaf voor het kwantificeren van het risico van financiële instrumenten, en heeft derhalve een cruciale rol in portfolio management, risico management en het prijzen van derivaten zoals opties. Private investeerders, institutionele beleggers zoals pensioenfondsen, centrale bankiers en beleidsmakers hebben een grote interesse in volatiliteitsmaatstaven. Het Basel-akkoord is een duidelijk voorbeeld waarin regels zijn vastgelegd die de hoeveelheid kapitaal die banken dienen aan te houden voor het afdekken van de risico's van hun hypotheek, leningen en andere producten beschrijft. De geschatte volatiliteit van de financiële producten van de bank is van grote invloed op de hoeveelheid aan te houden kapitaal. De financiële crisis in 2007-2008 leert ook dat de invloed van volatiliteit op financiële markten niet beperkt blijft tot de financiële sector, maar dat volatiele omstandigheden een gehele economie kunnen schaden. Extreme volatiliteit op de aandelenmarkt heeft een negatieve invloed op het investeringsgedrag van ondernemingen, in het bijzonder als deze ondernemingen investeringen financieren door het uitgeven van aandelen.

Dit proefschrift richt zich op verscheidene aspecten van volatiliteit. Allereerst schenken we aandacht aan het schatten van parameters van multivariate volatiliteitsmodellen. Dit is in het algemeen problematisch wanneer het aantal beschouwde financiële instrumenten toeneemt. Ten tweede gaan we dieper in op de vraag *wat* nu precies volatiliteit op financiële markten veroorzaakt. We maken hierbij onderscheid tussen publieke en pri-

vate informatie die een rol spelen in de hoogte van de volatiliteit van de markt. Tevens vestigen we de aandacht op het modelleren van volatiliteit door reeds bestaande modellen aan te passen, zodanig dat exogene variabelen invloed uitoefenen op de volatiliteit van een financieel instrument. Ten slotte onderzoeken we voorspellingen van volatiliteit, die worden gebruikt bij het combineren van kansverdelingsvoorspellingen van rendementen. In meer detail omvatten de hoofdstukken van dit proefschrift het volgende.

Hoofdstuk 2 is gebaseerd op Hoogerheide *et al.* (2012) en introduceert een adaptieve samplingmethode om te simuleren uit een niet-standaard kansverdeling. Deze methode maakt gebruik van gewogen EM stappen om efficiënt een combinatie (mixture) van Student- t kansdichtheden te construeren. Deze combinatie geeft een accurate benadering van een door de gebruiker ten doel gestelde niet-standaard kansverdeling (target-verdeling). In de Bayesiaanse econometrie is dit bijvoorbeeld een posterior kansverdeling, waar alleen de kern van de verdeling gevraagd is (geen constante multiplicatoren), zodat de Kullback-Leibler divergentie tussen de target-verdeling en de mixture wordt geminimaliseerd. We noemen deze methode MitISEM. The geconstrueerde mixture kan worden gebruikt als een kandidaat-verdeling voor Importance Sampling (IS) of de Metropolis-Hasting (MH) methode. We introduceren ook een extensie van MitISEM, hetwelk we aanduiden met Sequential MitISEM. Deze methode past de basismethode herhaaldelijk toe over tijd, zodat de kandidaat-verdeling wordt bijgewerkt wanneer er nieuwe data verkrijgbaar is. Als resultaat vinden we dat de nieuwe methoden de hoeveelheid rekentijd substantieel verkleint wanneer we parameters schatten van een multivariaat GARCH model en van een mixture GARCH model.

In Hoofdstuk 3 bestuderen we de invloed van private informatie op volatiliteit. We ontwikkelen een omvattend raamwerk om deze relatie te onderzoeken. We controleren in dit raamwerk voor het effect van zowel publieke als private informatie op rendementen en het effect van publieke informatie op volatiliteit. We passen dit raamwerk toe op hoge-frequentie data van de Amerikaanse obligatiemarkt (om precies te zijn: futures van de 30-jaar Amerikaanse Treasury bond) en vinden dat private informatie, gemeten door middel van de order flow, zowel statistisch als economisch significant is voor het verklaren van volatiliteit. Het gevonden resultaat blijft intact wanneer we liquiditeitseffecten

toevoegen, een andere maatstaf voor private informatie (market sidedness) hanteren en we data gebruiken met een dagelijkse frequentie.

Het centrale thema in Hoofdstuk 4 betreft de relatie tussen financiële condities en de volatiliteit en correlaties tussen aandelenrendementen. We stellen een uitbreiding voor van (factor)-GARCH modellen voor de volatiliteit en DCC modellen voor de correlatie zodanig dat de variabelen die de financiële condities meten in deze modellen kunnen worden toegevoegd. In een empirische toepassing beschouwen we de dagelijkse aandelenrendementen van Amerikaanse banken gedurende de periode 1994-2011. We benaderen de Amerikaanse financiële condities met de Bloomberg Financial Conditions Index (FCI), dat elke dag de geld-, obligatie- en aandelenmarkt samenvat in één getal. We constateren dat slechtere financiële condities samengaan met een hogere volatiliteit en een hogere gemiddelde correlatie tussen aandelenrendementen. Dit effect is in het bijzonder aanwezig tijdens crises, waarbij door toevoeging van de FCI in onze modellen de gemiddelde correlatie toeneemt met 0.15. Tevens vinden we dat de Value-at-Risk schattingen verbeteren, in het bijzonder de VaR schattingen op de korte termijn.

Hoofdstuk 5 onderzoekt de toegevoegde waarde van het combineren van kansdichtheidsvoorspellingen voor het voorspellen van aandeelrendementen in een bepaald, vooraf gespecificeerd gedeelte van de kansverdeling. We ontwikkelen een nieuwe techniek die modelonzekerheid meeneemt door gewichten toe te kennen aan individuele voorspelverdelingen, gebruikmakend van een scoremaatstaf die gebaseerd is op de aannemelijkheid van het rendement in een specifiek gedeelte van de kansverdeling. We passen onze methode toe op recent ontwikkelde univariate volatiliteitsmodellen (inclusief het HEAVY model en het Realized GARCH model), gebruikmakend van de dagelijkse rendementen op de S&P 500, DJIA, FTSE en Nikkei index van januari 2000 tot en met juni 2013. We stellen vast dat de gecombineerde kansdichtheidsvoorspellingen gebaseerd op de nieuwe techniek statistisch significant beter zijn dan kansdichtheidsvoorspellingen van (i) een bestaande combinatietechniek waarbij de gewichten bepaald worden aan de hand van de *gehele* voorspelverdeling, (ii) de combinatietechniek die elk individueel model een gelijk gewicht toekent, en (iii) de kansdichtheidsvoorspelling van elk individueel model. De door ons toegepaste methode verbetert eveneens de korte-termijn Value-at-Risk (VaR) schattingen.

Bibliography

- Aastveit, K.A., K.R. Gerdrup, A.S. Jore and L.A. Thorsrud (2011), Nowcasting GDP in real-time: A density combination approach, Working Paper.
- Aelli, G.P. (2013), Dynamic conditional correlation: On properties and estimation, *Journal of Business and Economic Statistics* **31**, 282–299.
- Akaike, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **19**, 716–723.
- Albuquerque, R. and C. Vega (2009), Economic news and international stock market co-movement, *Review of Finance* **13**, 401–465.
- Albuquerque, R., G.H. Bauer and M. Schneider (2009), Marketwide private information in stocks: Forecasting currency returns, *Journal of Finance* **63**, 2297–2343.
- Amihud, Y. (2002), Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* **5**, 31–56.
- Amisano, G. and R. Giacomini (2007), Comparing density forecasts via weighted likelihood ratio tests, *Journal of Business and Economic Statistics* **25**, 177–190.
- Andersen, T.G. (1996), Return volatility and trading volume: An information flow interpretation of stochastic volatility, *Journal of Finance* **51**, 169–204.
- Andersen, T.G. Bollerslev, T., P.F. Christoffersen and F.X. Diebold (2006), Volatility and correlation forecasting, *Handbook of economic forecasting* **1**, 779–865.

- Andersen, T.G., T. Bollerslev, F.X. Diebold and C. Vega (2003), Micro effects of macro announcements: Real-time price discovery in foreign exchange, *American Economic Review* **93**, 38–62.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and C. Vega (2007), Real-time price discovery in global stock, bond and foreign exchange markets, *Journal of International Economics* **73**, 251–277.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003), Modeling and forecasting realized volatility, *Econometrica* **71**, 529–626.
- Ang, A. and J. Chen (2002), Asymmetric correlations of equity portfolios, *Journal of Financial Economics* **63**, 443–494.
- Bacharach, J. (1974), Bayesian dialogues, Working Paper.
- Baele, L., G. Bekaert and K. Inghelbrecht (2010), The determinants of stock and bond return comovements, *Review of Financial Studies* **23**, 2374–2428.
- Balduzzi, P., E.J. Elton and T.C. Green (2001), Economic news and bond prices: Evidence from the U.S. Treasury market, *Journal of Financial and Quantitative Analysis* **36**, 523–543.
- 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.
- Bartolini, L., L. Goldberg and A. Sacarny (2008), How economic news moves markets, *Current issues in economics and Finance* **14**. Federal Reserve Bank of New York.
- Bates, J.M. and C.W.J. Granger (1969), The combination of forecasts, *Operational Research* **20**, 451–468.
- Beber, A. and M.W. Brandt (2010), When it cannot get better or worse: The asymmetric impact of good and bad news on bond returns in expansions and recessions, *Review of Finance* **14**, 119–155.

- Berger, D., A. Chaboud and E. Hjalmarsson (2009), What drives volatility persistence in the foreign exchange market?, *Journal of Financial Economics* **94**, 192–213.
- Berger, J. O. and L. R. Pericchi (1996), The intrinsic Bayes factor for model selection and prediction, *Journal of the American Statistical Association* **91**, 109–122.
- Berry, T.D. and K.M. Howe (1992), Public information arrival, *Journal of Finance* **49**, 1331–1346.
- Bollen, N.P.B. and R.E. Whaley (2004), Does net buying pressure affect the shape of implied volatility functions?, *Journal of Finance* **59**, 711–753.
- Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307–327.
- Bollerslev, T., J. Cai and F.M. Song (2000), Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the U.S. Treasury bond market, *Journal of Empirical Finance* **7**, 37–55.
- Boudt, K., J. Danielsson, S.J. Koopman and A. Lucas (2012), Regime switches in volatility and correlation of financial institutions, Working Paper.
- Boyd, J.H., J. Hu and R. Jagannathan (2005), The stock market's reaction to unemployment news: Why bad news is usually good for stocks, *Journal of Finance* **60**, 649–672.
- Bracker, K. and P. Koch (1999), Economic determinants of the correlation structure across international equity markets, *Journal of Economics and Business* **51**, 443–471.
- Brandt, M.W. and K.A. Kavajecz (2004), Price discovery in the U.S. Treasury market: The impact of orderflow and liquidity on the yield curve, *Journal of Finance* **59**, 2623–2654.
- Brenner, M., P. Pasquariello and M. Subrahmanyam (2009), On the volatility and comovement of U.S. financial markets around macroeconomic news announcements, *Journal of Financial and Quantitative Analysis* **44**, 1265–1289.
- Campbell, J.Y., A.W. Lo and A.C. MacKinlay (1997), *The econometrics of financial markets*, Princeton: Princeton University Press.

- Cappé, O., R. Douc, A. Guillin, J. M. Marin and C. P. Robert (2008), Adaptive importance sampling in general mixture classes, *Statistics and Computing* **18**, 447–459.
- Cappiello, L., R.F. Engle and K. Sheppard (2006), Asymmetric dynamics in the correlations of global equity and bond returns, *Journal of Financial Econometrics* **4**, 537–572.
- Chopin, N. (2002), A sequential particle filter method for static models, *Biometrika* **89**, 539–551.
- Chordia, T., A. Sarkar and A. Subrahmanyam (2005), An empirical analysis of stock and bond market liquidity, *Review of Financial Studies* **18**, 85–129.
- Christiansen, C. and A. Rinaldo (2007), Realized bond stock correlation: Macroeconomic announcement effects, *Journal of Futures Markets* **27**, 439–469.
- Christiansen, C., M. Schmeling and A. Schrimpf (2012), A comprehensive look at financial volatility prediction by economic variables, *Journal of Applied Econometrics* **27**, 956–977.
- Christoffersen, P. (1998), Evaluating interval forecasts, *International Economic Review* **39**, 841–862.
- Clark, P.K. (1973), A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* **41**, 135–155.
- Conflitti, C., C. De Mol and D. Giannone (2012), Optimal combination of survey forecasts, Working Paper.
- Cornuet, J. M., J. M. Marin, A. Mira and C. P. Robert (2012), Adaptive multiple Importance Sampling, *Scandinavian Journal of Statistics* **39**, 798–812.
- Creal, D., S.J. Koopman and A. Lucas (2013), Generalized autoregressive score models with applications, *Journal of Applied Econometrics* **28**, 777–795.
- Cutler, D.M., J.M. Poterba and L.H. Summers (1989), What moves stock prices?, *Journal of Portfolio Management* **15**, 4–12.

- De Goeij, P. and W. Marquering (2006), Macroeconomic announcements and asymmetric volatility in bond returns, *Journal of Banking and Finance* **30**, 2659–2680.
- Dempster, A. P., N. M. Laird and D. B. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society B* **39**, 1–38.
- Diebold, F.X. and R.S. Mariano (1995), Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**, 253–263.
- Diebold, F.X., T.A. Gunther and A.S. Tay (1998), Evaluating density forecasts with applications to financial risk management, *International Economic Review* **39**, 863–883.
- Diks, C., V. Panchenko and D. van Dijk (2011), Likelihood-based scoring rules for comparing density forecasts in tails, *Journal of Econometrics* **163**, 215–230.
- Doucet, A., N. De Freitas and N. J. Gordon (2001), *Sequential Monte Carlo methods in practice*, Springer-Verlag, New York.
- Ederington, L. and J. Lee (1993), How markets process information: News releases and volatility, *Journal of Finance* **45**, 1161–1191.
- Eklund, J. and S. Karlsson (2007), Forecast combination and model averaging using predictive measures, *Econometric Reviews* **26**, 329–363.
- Engle, R.F. (1982), Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**, 987–1007.
- Engle, R.F. (2002), Dynamic conditional correlation - A simple class of multivariate GARCH models, *Journal of Business and Economic Statistics* **20**, 339–350.
- Engle, R.F. and J. Rangel (2008), The Spline-GARCH model for low-frequency volatility and its global macroeconomic causes, *Review of Financial Studies* **21**, 1188–1222.
- Engle, R.F. and K. Sheppard (2005), Evaluating the specifications of covariance models for large portfolios, Working Paper.
- Engle, R.F. and K.F. Kroner (1995), Multivariate simultaneous generalized ARCH, *Econometric Theory* **11**, 122–150.

- Engle, R.F., N. Shephard and K. Sheppard (2008), Fitting vast dimensional time-varying covariance models, Working Paper.
- Evans, M. and R. Lyons (2002), Order flow and exchange rate dynamics, *Journal of Political Economy* **110**, 170–180.
- Evans, M. and R. Lyons (2008), How is macro news transmitted to exchange rates?, *Journal of Financial Economics* **88**, 26–50.
- Faust, J., J.H. Rogers, S.Y.B. Wang and J.H. Wright (2007), The high-frequency response of exchange rates and interest rates to macroeconomic announcements, *Journal of Monetary Economics* **54**, 1051–1068.
- Fleming, M.J. and E.M. Remolona (1999), Price formation and liquidity in the U.S. Treasury market: the response to public information, *Journal of Finance* **54**, 1901–1915.
- French, K.R. and R. Roll (1986), Stock return variances: the arrival of information and the reaction of traders, *Journal of Financial Economics* **17**, 5–26.
- Frühwirth-Schnatter, S. and H. Wagner (2008), Marginal likelihoods for non-Gaussian models using auxiliary mixture sampling, *Computational Statistics & Data Analysis* **52**, 4608–4624.
- Garleanu, N., L.H. Pedersen and A. Poteshman (2009), Demand-based option pricing, *Review of Financial Studies* **22**, 4259–4299.
- Garratt, A., K. Lee, M.H. Peseran and Y. Shin (2003), Forecast uncertainties in macroeconomic modelling: An application to the UK economy, *Journal of the American Statistical Association* **98**, 829–838.
- Gelfand, A.E. and D.K. Dey (1994), Bayesian model choice: Asymptotics and exact calculations, *Journal of the Royal Statistical Society Series B* **56**(3), 501–514.
- Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin (2003), *Bayesian data analysis*, London: Chapman and Hall.
- Geweke, J. (1989), Bayesian inference in econometric models using Monte Carlo integration, *Econometrica* **57**, 1317–1339.

- Geweke, J. and G. Amisano (2011), Optimal prediction pools, *Journal of Econometrics* **164**, 130–141.
- Geyer, C. J. (1991), Markov chain Monte Carlo maximum likelihood, in E.M. Keramidas (ed.), *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pp. 156–163.
- Giacomini, R. and H. White (2006), Tests of conditional predictive ability, *Econometrica* **74**, 1545–1578.
- Giacomini, R. and I. Komunjer (2005), Evaluation and combination of conditional quantile forecasts, *Journal of Business and Economic Statistics* **23**, 416–431.
- Giordani, P. and R. Kohn (2009), Appendix to ‘Adaptive independent Metropolis-Hastings by fast estimation of mixtures of normals’, available at website of the *Journal of Computational and Graphical Statistics*.
- Giordani, P. and R. Kohn (2010), Adaptive independent Metropolis-Hastings by fast estimation of mixtures of normals, *Journal of Computational and Graphical Statistics* **19**, 243–259.
- Glosten, L., R. Jagannathan and D. Runkle (1993), On the relationship between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* **48**, 1779–1801.
- Glosten, L.R. and P. Milgrom (1985), Bid, ask and transaction prices in a specialist market with heterogeneously informed agents, *Journal of Financial Economics* **14**, 71–100.
- Gneiting, T. and A.E. Raftery (2007), Strictly proper scoring rules, prediction and estimation, *Journal of the American Statistical Association* **102**, 359–378.
- Goodhart, C. and B. Hofmann (2008), Asset prices, financial conditions, and the transmission of monetary policy, Working Paper.
- Granger, C.W.J. and M.H. Pesaran (2000), Economic and statistical measures of forecast accuracy, *Journal of Forecasting* **19**, 537–560.

- Green, T.C. (2004), Economic news and the impact of trading on bond prices, *Journal of Finance* **59**, 1201–1233.
- Guichard, S. and D. Turner (2008), Quantifying the effect of financial conditions on U.S. activity, Working Paper.
- Haario, H., E. Saksman and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli* **7**, 223–242.
- Hafner, C. and P.H. Franses (2009), A generalized dynamic conditional correlation model: Simulation and application to many assets, *Econometric Reviews* **28**, 612–631.
- Hall, S.G. and J. Mitchell (2007), Combining density forecasts, *Journal of Forecasting* **23**, 1–13.
- Hamilton, J.D. and G. Lin (1996), Stock market volatility and the business cycle, *Journal of Applied Econometrics* **11**, 573–593.
- Hammersley, J. M. and D. C. Handscomb (1964), *Monte Carlo methods*, 1st edition, Methuen, London.
- Hansen, B.E. (1994), Autoregressive conditional density estimation, *International Economic Review* **35**, 705–730.
- Hansen, P.R., Z. Huang and H.H. Shek (2012), Realized GARCH: A joint model for returns and realized measures of volatility, *Journal of Applied Econometrics* **27**, 877–906.
- Hasbrouck, J. (1991), Measuring the information content of stock trades, *Journal of Finance* **46**, 179–207.
- Hasbrouck, J. (2004a), *Empirical market microstructure: The institutions, economics, and econometrics of securities trading*, Oxford: Oxford University Press.
- Hasbrouck, J. (2004b), Liquidity in the futures pit: Inferring market dynamics from incomplete data, *Journal of Financial and Quantitative Analysis* **39**, 305–326.

- Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**, 97–109.
- Hatzius, J., P. Hooper, F.S. Mishkin, K.L. Schoenholtz and M. Watson (2010), Financial conditions indexes: A fresh look after the financial crisis, U.S. Monetary Policy Forum.
- He, Y., H. Lin, J. Wang and C. Wu (2009), Price discovery in the round-the-clock U.S. Treasury market, *Journal of Financial Intermediation* **18**, 464–490.
- Hoogerheide, L. F., A. Opschoor and H. K. Van Dijk (2012), A class of adaptive EM-based Importance Sampling algorithms for efficient and robust posterior and predictive simulation, *Journal of Econometrics* **171**, 101–120.
- Hoogerheide, L. F. and H. K. Van Dijk (2010), Bayesian forecasting of Value-at-Risk and Expected Shortfall using adaptive Importance Sampling, *International Journal of Forecasting* **26**, 231–247.
- Hoogerheide, L. F., J. F. Kaashoek and H. K. Van Dijk (2007), On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: An application of flexible sampling methods using neural networks, *Journal of Econometrics* **139**(1), 154–180.
- Hu, W. (2005), *Calibration of multivariate generalized hyperbolic distributions using the EM algorithm, with applications in risk management, portfolio optimization and portfolio credit risk*, PhD thesis, Florida State University, College of Arts and Sciences.
- Hukushima, K. and K. Nemoto (1996), Exchange Monte Carlo and application to spin glass simulations, *Journal of the Physical Society of Japan* **65**, 1604–1608.
- Jiang, G.J. and I. Lo (2011), Private information flow and price discovery in the U.S. Treasury market, Working Paper.
- Jones, C.M., G. Kaul and M.L. Lipson (1994), Information, trading, and volatility, *Journal of Financial Economics* **36**, 127–154.
- Jore, A.S., J. Mitchell and S. P. Vahey (2010), Combining forecast densities from VARs with uncertain instabilities, *Journal of Applied Econometrics* **25**, 621–634.

- Karpoff, J.M. (1987), The relation between price changes and trading volume: a survey, *Journal of Financial and Quantitative Analysis* **22**, 109–126.
- Keith, J. M., D. P. Kroese and G. Y. Sofronov (2008), Adaptive independence samplers, *Statistics and Computing* **18**, 409–420.
- Kloek, T. and H. K. Van Dijk (1978), Bayesian estimates of equation system parameters: An application of integration by Monte Carlo, *Econometrica* **46**, 1–20.
- Koopman, S.J., B. Jungbacker and E. Hol (2005), Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements, *Journal of Empirical Finance* **12**, 445–475.
- Kou, S. C., Q. Zhou and W. H. Wong (2006), Equi-energy sampler with applications in statistical inference and statistical mechanics, *The Annals of Statistics* **34**, 1581–1619.
- Kullback, S. and R. A. Leibler (1951), On information and sufficiency, *The Annals of Mathematical Statistics* **22**, 79–86.
- Kupiec, P.H. (1995), Techniques for verifying the accuracy of risk measurement models, *Journal of Derivatives* **3**, 73–82.
- Kyle, A.S. (1985), Continuous auctions and insider trading, *Econometrica* **53**, 1315–1335.
- Lange, K. L., R. J. A. Little and J. M. G. Taylor (1989), Robust statistical modeling using the t distribution, *Journal of the American Statistical Association* **84**, 881–896.
- Lee, C. and M. Ready (1991), Inferring the trade direction from intraday data, *Journal of Finance* **46**, 733–746.
- Li, H., J. Wang, C. Wu and Y. He (2009), Are liquidity and information risks priced in the Treasury bond market?, *Journal of Finance* **64**, 467–503.
- Liu, J. S. and R. Chen (1998), Sequential Monte Carlo methods for dynamic systems, *Journal of the American Statistical Association* **93**, 1032–1044.
- Longin, F. and B. Solnik (2001), Extreme correlations of international equity markets, *Journal of Finance* **56**, 676–649.

- Manaster, S. and S.C. Mann (1996), Life in the pits: Competitive market making and inventory control, *Review of Financial Studies* **9**, 953–975.
- Mandelbrot, B. (1963), The variation of certain speculative prices, *Journal of Business* **36**, 394–419.
- Menkveld, A.J., A. Sarkar and M. Van der Wel (2012), Customer flow, intermediaries, and discovery of the equilibrium riskfree rate, *Journal of Financial and Quantitative Analysis* **47**, 821–849.
- Metropolis, N., A. W. Rosenbluth, A. H. Rosenbluth, M. N. Teller and E. Teller (1953), Equation of state calculations by fast computing machines, *The Journal of Chemical Physics* **21**, 1087–1092.
- Mitchell, J. and S.G. Hall (2005), Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR fan charts of inflation, *Oxford Bulletin of Economics and Statistics* **67**, 995–1033.
- Morgan, JP (1994), Riskmetrics, Second Edition, J.P. Morgan.
- Nelson, D.B. (1991), Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**, 347–370.
- Noureldin, D., N. Shephard and K. Sheppard (2012), Multivariate high-frequency-based volatility (HEAVY) models, *Journal of Applied Econometrics* **27**, 907–933.
- O'Hagan, A. (1995), Fractional Bayes factors for model comparison, *Journal of the Royal Statistical Society Series B* **57**(1), 99–138.
- Pasquariello, P. and C. Vega (2007), Informed and strategic order flow in the bond market, *Review of Financial Studies* **20**, 1975–2019.
- Pastor, L. and P. Veronesi (2008), Learning in financial markets, Working Paper.
- Paye, B. (2012), Déjà vol: Predictive regressions for aggregate stock market volatility using macroeconomic variables, *Journal of Financial Economics* **106**, 527–546.

- Peel, D. and G. J. McLachlan (2000), Robust mixture modelling using the t distribution, *Statistics and Computing* **10**, 339–348.
- Perez-Quiros, G. and A. Timmermann (2001), Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities, *Journal of Econometrics* **103**, 259–306.
- Pitt, M. K., R. S. Silva, P. Giordani and R. Kohn (2010), Auxiliary particle filtering within adaptive Metropolis-Hastings sampling. <http://arxiv.org/abs/1006.1914>.
- Rangel, J.G. and R.F. Engle (2012), The factor-spline-GARCH model for high and low frequency correlations, *Journal of Business and Economic Statistics* **30**, 109–124.
- Rapach, D.E., J.K. Strauss and G. Zhou (2010), Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* **23**, 821–862.
- Ritter, C. and M. A. Tanner (1992), Facilitating the Gibbs sampler: The Gibbs stopper and the griddy-Gibbs sampler, *Journal of the American Statistical Association* **87**, 861–868.
- Roberts, G. O. and J. S. Rosenthal (2001), Optimal scaling for various Metropolis-Hastings algorithms, *Statistical Science* **16**, 351–367.
- Roberts, G. O. and J. S. Rosenthal (2009), Examples of adaptive MCMC, *Journal of Computational and Graphical Statistics* **18**, 349–367.
- Roll, R. (1984), A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* **39**, 1127–1139.
- Rosenberg, M. (2009), Financial conditions watch, Bloomberg.
- Ross, S. (1976), The arbitrage theory of capital asset pricing, *Journal of Economic Theory* **13**, 341–360.
- Rubin, D. B. (1983), Iteratively reweighted least squares, *Encyclopedia of Statistical Sciences*, Vol. 4, Wiley, New York, pp. 272–275.

- Santos, A.A.P., F.J. Nogales and E. Ruiz (2013), Comparing univariate and multivariate models to forecast portfolio Value-at-Risk, *Journal of Financial Econometrics* **11**, 400–441.
- Sarkar, A. and R.A. Schwartz (2009), Market sidedness: Insights into motives for trade initiation, *Journal of Finance* **64**, 375423.
- Schwarz, G. (1978), Estimating the dimension of a model, *Annals of Statistics* **6**, 461–464.
- Schwert, G.W. (1989), Why does stock market volatility change over time?, *Journal of Finance* **44**, 1115–1153.
- Shephard, N. and K. Sheppard (2010), Realising the future: Forecasting with high-frequency-based volatility (heavy) models, *Journal of Applied Econometrics* **25**, 197–231.
- Sheppard, K. (2008), Economic factors and the covariance of equity returns, Working Paper.
- Stock, J. and M. Watson (2004), Combination forecasts of output growth in a seven-country data set, *Journal of Forecasting* **23**, 405–430.
- Tauchen, G.E. and M. Pitts (1983), The price variability-volume relationship on speculative markets, *Econometrica* **51**, 485–505.
- Timmermann, A. (2006), Forecast combinations, *Handbook of economic forecasting* **1**, 135–196.
- Van der Wel, M., A.J. Menkveld and A. Sarkar (2009), Are market makers uninformed and passive?, Working Paper.
- van Dijk, D., H. Munandar and C. Hafner (2011), The euro introduction and noneuro currencies, *Applied Financial Economics* **21**, 95–116.
- Van Dijk, H. K. and T. Kloek (1980), Further experience in Bayesian analysis using Monte Carlo integration, *Journal of Econometrics* **14**, 307–328.

Van Dijk, H. K. and T. Kloek (1984), Experiments with some alternatives for simple importance sampling in Monte Carlo integration, in J. M. Bernardo, M. J. Degroot, D. Lindley and A. F. M. Smith (eds.), *Bayesian Statistics*, Vol. 2, Amsterdam, North Holland.

Wallis, K.F. (2005), Combining density and interval forecasts: A modest approach, *Oxford Bulletin of Economics and Statistics* **67**, 983–994.


Zeevi, A. J. and R. Meir (1997), Density estimation through convex combinations of densities: Approximation and estimation bounds, *Neural Networks* **10**(1), 99–106.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

523. P. JANUS, *Developments in Measuring and Modeling Financial Volatility*
524. F.P.W. SCHILDER, *Essays on the Economics of Housing Subsidies*
525. S.M. MOGHAYER, *Bifurcations of Indifference Points in Discrete Time Optimal Control Problems*
526. C. ÇAKMAKLI, *Exploiting Common Features in Macroeconomic and Financial Data*
527. J. LINDE, *Experimenting with new combinations of old ideas*
528. D. MASSARO, *Bounded rationality and heterogeneous expectations in macroeconomics*
529. J. GILLET, *Groups in Economics*
530. R. LEGERSTEE, *Evaluating Econometric Models and Expert Intuition*
531. M.R.C. BERSEM, *Essays on the Political Economy of Finance*
532. T. WILLEMS, *Essays on Optimal Experimentation*
533. Z. GAO, *Essays on Empirical Likelihood in Economics*
534. J. SWART, *Natural Resources and the Environment: Implications for Economic Development and International Relations*
535. A. KOTHIYAL, *Subjective Probability and Ambiguity*
536. B. VOOGT, *Essays on Consumer Search and Dynamic Committees*
537. T. DE HAAN, *Strategic Communication: Theory and Experiment*
538. T. BUSER, *Essays in Behavioural Economics*
539. J.A. ROSERO MONCAYO, *On the importance of families and public policies for child development outcomes*
540. E. ERDOGAN CIFTCI, *Health Perceptions and Labor Force Participation of Older Workers*

541. T.WANG, *Essays on Empirical Market Microstructure*
542. T. BAO, *Experiments on Heterogeneous Expectations and Switching Behavior*
543. S.D. LANSDORP, *On Risks and Opportunities in Financial Markets*
544. N. MOES, *Cooperative decision making in river water allocation problems*
545. P. STAKENAS, *Fractional integration and cointegration in financial time series*
546. M. SCHARTH, *Essays on Monte Carlo Methods for State Space Models*
547. J. ZENHORST, *Macroeconomic Perspectives on the Equity Premium Puzzle*
548. B. PELLOUX, *The Role of Emotions and Social Ties in Public On Good Games: Behavioral and Neuroeconomic Studies*
549. N. YANG, *Markov-Perfect Industry Dynamics: Theory, Computation, and Applications*
550. R.R. VAN VELDHUIZEN, *Essays in Experimental Economics*
551. X. ZHANG, *Modeling Time Variation in Systemic Risk*
552. H.R.A. KOSTER, *The internal structure of cities: the economics of agglomeration, amenities and accessibility*
553. S.P.T. GROOT, *Agglomeration, globalization and regional labor markets: micro evidence for the Netherlands*
554. J.L. MÖHLMANN, *Globalization and Productivity Micro-Evidence on Heterogeneous Firms, Workers and Products*
555. S.M. HOOGENDOORN, *Diversity and Team Performance: A Series of Field Experiments*
556. C.L. BEHRENS, *Product differentiation in aviation passenger markets: The impact of demand heterogeneity on competition*
557. G. SMRKOLJ, *Dynamic Models of Research and Development*
558. S. PEER, *The economics of trip scheduling, travel time variability and traffic information*
559. V. SPINU, *Nonadditive Beliefs: From Measurement to Extensions*
560. S.P. KASTORYANO, *Essays in Applied Dynamic Microeconometrics*
561. M. VAN DUIJN, *Location, choice, cultural heritage and house prices*
562. T. SALIMANS, *Essays in Likelihood-Based Computational Econometrics*
563. P. SUN, *Tail Risk of Equidity Returns*

564. C.G.J. KARSTEN, *The Law and Finance of M&A Contracts*
565. C. OZGEN, *Impacts of Immigration and Cultural Diversity on Innovation and Economic Growth*
566. R.S. SCHOLTE, *The interplay between early-life conditions, major events and health later in life*
567. B.N.KRAMER, *Why dont they take a card? Essays on the demand for micro health insurance*
568. M. KILIÇ, *Fundamental Insights in Power Futures Prices*
569. A.G.B. DE VRIES, *Venture Capital: Relations with the Economy and Intellectual Property*
570. E.M.F. VAN DEN BROEK, *Keeping up Appearances*
571. K.T. MOORE, *A Tale of Risk: Essays on Financial Extremes*
572. F.T. ZOUTMAN, *A Symphony of Redistributive Instruments*
573. M.J. GERRITSE, *Policy Competition and the Spatial Economy*



Financial market volatility is central to financial economics. Since it is the most common measure of the risk involved in investments in traded securities, it plays a crucial role in portfolio management and risk management. This thesis contributes to the volatility literature by investigating several aspects of volatility. First, we focus from a Bayesian point of view on the parameter estimation of multivariate volatility models. Second, we consider the question what exactly causes financial market volatility? In this context, we relate volatility with various types of information. In addition, we pay attention to modeling volatility, by adapting volatility models such that they allow for including possible exogenous variables. Finally, we turn to forecasting techniques of volatility, with the focus on the combination of density forecasts.

Anne Opschoor (1987) obtained his master's degree in econometrics with honors from Erasmus University Rotterdam in November 2009. In that same month he started as a PhD-student at the Tinbergen Institute and Econometric Institute at the Erasmus University, with volatility modeling and forecasting as his primary fields of interest. The article version of Chapter 2 is published in the Journal of Econometrics. Currently, he works as a post doctoral researcher at VU University Amsterdam.

