

Modeling and Forecasting of Multivariate Stock Market Volatility



Inaugural-Dissertation zur Erlangung des akademischen Grades eines Doktors der
Wirtschafts- und Sozialwissenschaften der Wirtschafts- und Sozialwissenschaftlichen
Fakultät der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Diplom-Volkswirt Bastian Gribisch
aus Bad Oldesloe

Kiel, 2012

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan: Professor Horst Raff, Ph.D.

Erstberichterstattender: Professor Dr. Roman Liesenfeld

Zweitberichterstattender: Professor Dr. Thomas Lux

Tag der Abgabe der Arbeit: 04.12.2012

Tag der mündlichen Prüfung: 19.02.2013

Für meine Eltern

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Statistik und Ökonometrie der Christian-Albrechts-Universität zu Kiel.

Mein besonderer Dank gilt meinem Doktorvater Professor Dr. Roman Liesenfeld. Er hat durch seine zahlreichen Anregungen, seine stets konstruktive Kritik und immerwährende Diskussionsbereitschaft ein produktives Arbeitsklima entstehen lassen, welches entscheidend zum Gelingen dieser Arbeit beigetragen hat. Herrn Professor Dr. Thomas Lux danke ich herzlich für die Übernahme des Zweitgutachtens. Dank schulde ich auch Herrn Professor Dr. Vasyl Golosnoy für die konstruktive und stets erkenntnisreiche Zusammenarbeit im Rahmen gemeinsamer wissenschaftlicher Projekte. Des Weiteren möchte ich allen (ehemaligen) Kollegen am Institut für die vielen hilfreichen Kommentare und die ausgesprochen angenehme Zusammenarbeit danken, insbesondere jedoch Herrn Markus Pape, Herrn Dr. Jan Roestel und Herrn Dr. Christian Assmann. Für die Hilfe in Computerfragen und die laufende Bereitstellung erforderlicher Rechenkapazitäten danke ich Herrn Albrecht Mengel.

Schließlich gilt mein Dank meinen Eltern, welche mit ihrem mir entgegengebrachten Vertrauen und ihrer steten Unterstützung diese Arbeit erst möglich gemacht haben.

Kiel, am 30.11.2012

Bastian Gribisch

Contents

1	Introduction	13
1.1	Object of Investigation	13
1.1.1	Multivariate GARCH Models	14
1.1.2	Multivariate Stochastic Volatility Models	16
1.1.3	The Realized Volatility Approach	19
1.2	Outline	24
2	Multivariate Wishart Stochastic Volatility and Changes in Regime	27
2.1	Introduction	27
2.2	Model Specification, Inference and Diagnostics	29
2.2.1	The Basic WMSV Model	29
2.2.2	The Markov Switching WMSV Model	31
2.2.3	Estimation and Diagnostics	32
2.3	Empirical Application	34
2.3.1	Data	34
2.3.2	Estimation Results	34
2.3.3	Forecasting Results	37
2.4	Summary	41
2.5	Technical Details	42
2.5.1	Simulation Based Bayesian Inference	42
2.5.2	Full Conditional Distributions: Basic WMSV Model	43
2.5.3	Full Conditional Distributions: Markov Switching MWSV Model	47
2.5.4	A Particle Filter Algorithm for the Basic WMSV and the MS WMSV Model	48
3	The Conditional Autoregressive Wishart Model	69
3.1	Introduction	69
3.2	Model Specification, Inference and Diagnostics	71
3.2.1	The CAW(p,q) Model	71
3.2.2	Stochastic Properties of the CAW(p,q) Model	73
3.2.3	Estimation and Diagnostics	76
3.2.4	Extensions of the Baseline CAW Model	77

3.3	Empirical Application	79
3.3.1	Data	79
3.3.2	Estimation Results	80
3.3.3	Forecasting Results	83
3.4	Summary	85
3.5	Technical Details	87
4	Intra-Daily Volatility Spillovers between the US and German Stock Markets	101
4.1	Introduction	101
4.2	Trading Times, Data and Adjustments	104
4.3	Model Specification, Inference and Diagnostics	106
4.3.1	Period 1: Germany-US Trading Overlap	107
4.3.2	Periods 2 and 3: US-Only and Germany-Only Trading	108
4.3.3	Model Properties, Estimation and Diagnostics	109
4.4	Impulse-Response Analysis	111
4.5	Empirical Application	112
4.5.1	Estimation Results	112
4.5.2	Impulse-Response Analysis	114
4.5.3	Impact of the Subprime Crisis	115
4.6	Summary	117
5	A Latent Dynamic Factor Approach to Multivariate Stock Market Volatility	129
5.1	Introduction	129
5.2	Model Specification, Inference and Diagnostics	131
5.2.1	The Matrix Logarithm	131
5.2.2	The Dynamic Factor Model	132
5.2.3	Estimation and Diagnostics	134
5.3	Empirical Application	135
5.3.1	Data	135
5.3.2	Estimation Results	137
5.3.3	Forecasting Results	140
5.4	Summary	143
5.5	Technical Details	144
6	Conclusion	173
	Bibliography	177

List of Tables

2.1	Descriptive Statistics for the Daily Index Log Returns	50
2.2	Estimation Results	51
2.3	Model Diagnostics Results: Pearson Residuals	52
2.4	Distributional Diagnostics	53
2.5	Simulation Results for the Fitted MS WMSV Model	54
2.6	VaR Forecasting Results	55
3.1	Descriptive Statistics for the Realized Variances and Covariances	90
3.2	Optimized Log-likelihood and Diagnostics for the Baseline CAW Model	91
3.3	ML-Parameter Estimates for the Baseline CAW(2,2) model	92
3.4	Optimized Log-likelihood and Diagnostics for the MIDAS-CAW and HAR- CAW Model	93
3.5	ML-Parameter Estimates for the MIDAS-CAW(2,2) Model	94
3.6	Evaluation of Forecasting Accuracy Prior to the Subprime Crisis	95
3.7	Evaluation of Forecasting Accuracy During the Subprime Crisis	96
4.1	Descriptive Statistics for Detrended Realized Variances and Covariance	118
4.2	ML-Parameter Estimates for the BIC Selected Sequential Three-Phase CAW Model	119
4.3	ML-Estimates of the Marginal Effects for the BIC Selected Sequential Three- Phase Model	120
4.4	Shock Scenario of the Impulse Response Analysis	121
5.1	NYSE Traded Stocks of Data Set 1 and Data Set 2	147
5.2	Descriptive Statistics	148
5.3	Portmanteau Diagnostic Test Results	149
5.4	P-values of Residual F-tests	150
5.5	Estimation Results: 2 Factor Model with Idiosyncratic AR(1) Dynamics, Data Set 1	151
5.6	Estimation Results: HAR Parameter Estimates for 3 Factor Model with Id- iosyncratic AR(1) Dynamics, Data Set 2	152
5.7	Statistical Evaluation of Forecasting Accuracy: Data Set 1	153
5.8	Statistical Evaluation of Forecasting Accuracy: Data Set 2, Phase 1	154

List of Figures

2.1	Simulated Means of Covariances	56
2.2	Simulated Means of Correlation	57
2.3	Simulated Autocorrelation Functions for σ_{11}	58
2.4	Simulated Cross-Correlation Functions	59
2.5	Return Series	60
2.6	Smoothed Volatility Estimates and Corresponding Return Series: Basic WMSV Model	61
2.7	Smoothed Correlation Estimates: Basic WMSV Model	62
2.8	Sample Autocorrelation Functions of Squared Residual Series: Basic WMSV Model	63
2.9	Residual QQ-Plots: Basic WMSV Model	64
2.10	Smoothed Volatility and Markov State Estimates	65
2.11	Sample Autocorrelation Functions of Squared Residual Series: MS WMSV Model	66
2.12	Residual QQ-Plots: MS WMSV Model	67
3.1	Time Series and ACFs of Realized Variances, Covariances and Residual Series	97
3.2	Predicted (Co)Variances, Correlations and Secular Components obtained under the MIDAS-CAW(2,2) Model	98
3.3	Sample Autocorrelation Function of Standardized Residuals from Unrestricted MIDAS-CAW(2,2) Model	99
4.1	Daily Trading Hours for the US (DJ) and the German (DAX) Stock Markets	122
4.2	Time Series of Realized (Co)Variances and Long-Run Component for DAX and Dow Jones Stock Indices	123
4.3	Time series of Detrended Realized (Co)Variances for DAX and Dow Jones Stock Indices	124
4.4	Estimates of the Direct Marginal Effects of Intra-Day Variances on the Next-Period Variance at Home and Abroad	125
4.5	Estimated Impulse Response Functions	126
4.6	Estimates of Direct Marginal Effects during Non-Crisis Periods and during the Subprime Crisis	127

4.7	Estimated IR Functions during Non-Crisis Periods and during the Subprime Crisis	128
5.1	Time Series of Daily Realized Variances and Covariances: Data Set 1	155
5.2	Sample Autocorrelation Function of Daily Realized Variances and Covariances: Data Set 1	156
5.3	Time Series of the Matrix Logarithm of Daily Realized Variances and Covariances: Data Set 1	157
5.4	Sample Autocorrelation Function of the Matrix Logarithm of Daily Realized Variances and Covariances: Data Set 1	158
5.5	Logarithmic Variances and Correlations vs. Matrix-Logarithmic Approximations	159
5.6	Sample Autocorrelation Function of Standardized Residuals: 2-Factor Model, Data Set 1	160
5.7	Fraction of Total Variance Explained by Factors: 2-Factor Model, Data Set 1	161
5.8	Filtered Common Factors: 2-Factor Model, Data Set 1	162
5.9	Filtered Idiosyncratic Factors: 2-Factor Model, Data Set 1	163
5.10	Point Estimates and 95% Posterior Confidence Regions for Factor Loadings and Measurement Error Variances of Diagonal Matrix-Log Elements: 3-Factor Model, Data Set 2	164
5.11	Point Estimates and 95% Posterior Confidence Regions for Factor Loadings and Measurement Error Variances of Off-Diagonal Matrix-Log Elements: 3-Factor Model, Data Set 2	165
5.12	Point Estimates and 95% Posterior Confidence Regions for Persistency of Idiosyncratic Factors: 3-Factor Model, Data Set 2	166
5.13	Fraction of Total Variance Explained by Factors: 3-Factor Model, Data Set 2	167
5.14	Filtered Common Factors: 3-Factor Model, Data Set 2	168
5.15	Mean-Variance Plots: Data Set 1, Phase I	169
5.16	Mean-Variance Plots: Data Set 1, Phase II	170
5.17	Mean-Variance Plots: Data Set 2, Phase I	171

Chapter 1

Introduction

1.1 Object of Investigation

This thesis is concerned with the modeling and forecasting of daily and intra-daily multivariate stock market volatility. Time-varying volatilities and correlations of asset returns constitute a stylized fact in financial statistics and understanding the serial and cross-sectional dynamics of variances and covariances is fundamental in order to obtain precise (co)variance forecasts which are important ingredients for many fields in financial econometrics. Typical applications are the forecasting of optimal portfolio weight vectors in mean/variance portfolio optimization, asset pricing, hedging and the computation of Value-at-Risk (VaR) measures in risk management applications e.g. required by the Basel Committee on Banking Supervision. Multivariate volatility models allow us to study the relations between variances and covariances of several markets, thereby addressing important questions in financial econometrics: Does the volatility of one market lead the volatility of another market? How is the volatility of an asset transmitted to other assets - directly through the variance, or indirectly through a covariance channel? Does cross-asset volatility transmission increase in periods of high market volatility, indicating volatility contagion and reinforcing global financial crises? Do asset correlations increase in periods of turmoil indicating that diversification benefits vanish when they are needed most?

Multivariate volatility modeling focuses on the covariance matrix comprising the variances and covariances of asset returns. This covariance matrix is not directly observable and most existing models treat it either as measurable given past observations, such as multivariate generalized autoregressive conditional heteroscedasticity (GARCH) models introduced by Bollerslev et al. (1988), or as an inherently latent quantity, such as multivariate stochastic volatility (SV) models introduced by Harvey et al. (1994). An alternative approach of covariance estimation and modeling, which has attracted substantial interest in recent years, uses high-frequency return data to construct realized variances and covariances as precise estimates for the variances and covariances of low-frequency returns (see e.g., Andersen et al., 2003, and Barndorff-Nielsen and Shephard, 2004). As such, the observed realized variances and covariances can be modeled directly as advocated, for example, by Andersen

et al. (2003). Multivariate volatility models should satisfy two important requirements, namely that the predicted covariance matrices remain positive definite, and, second, that the specification is parsimoniously parameterized yet empirically realistic with the ability to account for complex serial- and cross-sectional dynamics typically observed for realized variances and covariances.

In order to provide an overview on multivariate volatility modeling the following sections give a short summary on multivariate GARCH models, multivariate SV models and the realized volatility approach. This overview lays the foundation for illustrating the contribution of the present thesis.

1.1.1 Multivariate GARCH Models

Multivariate GARCH (MGARCH) models represent multivariate extensions of the univariate GARCH model of Bollerslev (1986) and have been developed in the late 1980s and the first half of the 1990s.¹ MGARCH models assume an observable covariance process given past return observations. The approach is illustrated best considering the basic *VEC*-GARCH specification of Bollerslev et al. (1988). Consider a stochastic k -dimensional mean-adjusted asset return vector ξ_t at time period t ($t = 1, \dots, T$). Conditioning on the history $\mathcal{F}_{t-1} = \{\xi_{t-1}, \xi_{t-2}, \dots\}$, we write

$$\xi_t = H_t^{1/2} u_t, \quad (1.1)$$

where $H_t^{1/2}$ is the lower-triangular Cholesky factor of the $k \times k$ conditional covariance matrix $H_t = \text{Var}[\xi_t | \mathcal{F}_{t-1}]$. The $k \times 1$ innovation vector u_t is typically assumed to follow a zero-mean multivariate normal or Student- t distribution with $\text{Var}[u_t] = I_k$, where I_k denotes the k -dimensional identity matrix. The GARCH approach captures serial and cross-correlation in variances and covariances by assuming that volatility depends on lagged squared (mean-adjusted) returns and return cross-products (ARCH terms) and lagged volatilities (GARCH terms). The *VEC*(p, q) model is then given by

$$\text{vech}(H_t) = c + \sum_{i=1}^p G_i \text{vech}(H_{t-i}) + \sum_{j=1}^q A_j \text{vech}(\xi_{t-j} \xi_{t-j}'), \quad (1.2)$$

where $\text{vech}(\cdot)$ denotes the operator that stacks the lower triangular portion of a $k \times k$ matrix including the diagonal into a $k(k+1)/2 \times 1$ vector. The $k(k+1)/2 \times 1$ dimensional parameter vector c drives the overall volatility level and G_i and A_j are square parameter matrices of dimension $k(k+1)/2$ allowing for flexible serial and cross-sectional dependence patterns in variances and covariances. Cross-asset volatility dependence is typically referred to as “volatility spillover” (see e.g. Hamao et al., 1990). Initiated by the seminal work of Engle

¹See the excellent overview on MGARCH models of Bauwens et al., 2006.

et al. (1990) and Hamao et al. (1990) the presence of volatility spillovers on international financial markets has been subject to tremendous research during the past two decades. This huge interest is motivated by the fact that cross-market volatility dependencies tend to reinforce the impact of financial crisis like the East Asian currency crisis in late 1997 or the subprime crisis starting in 2008. MGARCH models are widely employed in order to investigate volatility transmission effects and are typically estimated by means of Maximum Likelihood.

The $VEC(p, q)$ model defined by Equations (1.1) and (1.2) suffers from the fact that the dimension of the parameter vector increases quadratically in the number of assets k . This is known as the “curse of dimensionality” of multivariate volatility modeling and prevents practical applications of the basic model to high-dimensional asset portfolios. To overcome this problem Bollerslev et al. (1988) suggest to restrict the autoregressive parameter matrices to the diagonal case, thereby excluding volatility spillovers. Nevertheless, even under this restriction the system is still heavily parameterized. Furthermore the VEC -GARCH model requires strong restrictions on the model parameters in order to secure positive-definiteness of covariance matrix forecasts. To overcome this problem Engle and Kroner (1995) proposed the BEKK-GARCH approach. The model uses an alternative parametrization for H_t which automatically secures positive definiteness without parametric restrictions. The BEKK-GARCH model is often used in applied financial econometrics.

In the early 1990s factor GARCH models have been proposed in order to facilitate high-dimensional (co)variance estimation. The models exploit the idea that co-movements of stock returns are driven by a small number of common factors, which in turn feature GARCH dynamics (see e.g. Engle et al., 1990b, Lin, 1992, and Bollerslev and Engle, 1993). The factor model implied covariance matrix is always positive-definite. Factor models alleviate the curse of dimensionality in multivariate volatility modeling but impose restrictions on (co)variance dynamics: the model assumes that variances and covariances are jointly driven by the dynamics of a low-dimensional factor process.

An alternative popular class of MGARCH models being applicable to high-dimensional (co)variance estimation is based on a separate specification of conditional volatilities and conditional correlations. A positive-definite (co)variance process is ensured by restricting the variances to positivity and requiring a positive-definite correlation matrix. The constant conditional correlation (CCC) model of Bollerslev (1990) assumes univariate GARCH dynamics for the conditional variances and constant correlation coefficients for all asset-pairs. The model greatly reduces the number of unknown parameters and simplifies estimation. Since the assumption of time-invariant correlations is restrictive and generally rejected by the data, Christodoulakis and Satchell (2002), Engle (2002) and Tse and Tsui (2002) propose to generalize the CCC model to dynamic conditional correlations (DCC) resulting in DCC-type GARCH models, which are now widely applied in order to forecast (co)variances for large asset return vectors. The DCC-GARCH approach mitigates the curse of dimensionality

by assuming that the entire correlation process is driven by only two parameters.

1.1.2 Multivariate Stochastic Volatility Models

Multivariate SV (MSV) models represent multivariate extensions of the univariate SV model originally proposed by Taylor (1982, 1986).² While GARCH-type models assume a measurable (co)variance process given past return observations (observation driven models), SV models assume a “purely” random and therefore latent volatility process (parameter-driven models). The randomness of the SV process comes as a consequence of an additional error term in the volatility equation, which makes SV models more flexible than GARCH-type models. Kim et al. (1998) find that simple SV models typically fit the daily asset return data as well as more heavily parameterized GARCH models. Yet the enhanced flexibility does not come without a cost. SV models are more difficult to estimate than GARCH-type models. The likelihood function involves high-dimensional integrals which cannot be solved analytically. In practice Method of Moment estimation or simulation based classical or Bayesian inference are applied. Compared to GARCH-type models the SV approach features a closer link to continuous time models which are typically applied in the asset pricing framework of finance including generalizations of the Black-Scholes option pricing formula (see e.g. Hull and White, 1987).

Consider a k -dimensional vector of log-prices S . A general continuous time diffusion model for S is given by (see Asai et al., 2006)

$$dS(t) = \Sigma^{1/2}(t) dW^{(1)}(t) \quad (1.3)$$

$$df(\text{vech}(\Sigma(t))) = a(\text{vech}(\Sigma(t))) dt + b(\text{vech}(\Sigma(t))) dW^{(2)}(t), \quad (1.4)$$

where $W^{(1)}(t)$ and $W^{(2)}(t)$ are two vectors of Brownian motions with potentially cross-correlated elements, $\Sigma(t) = \Sigma^{1/2}(t)\Sigma^{1/2}(t)'$ is the spot covariance matrix and f , a and b are known functions³. Empirical versions of SV models are formulated in discrete time. A general discrete time MSV model for the log-return $\xi_t = S_t - S_{t-1}$ is obtained by applying the Euler method:

$$\xi_t = \Sigma_t^{1/2} u_t, \quad u_t \sim N(0, I_k) \quad (1.5)$$

$$f(\text{vech}(\Sigma_t)) = a(\text{vech}(\Sigma_{t-1})) + f(\text{vech}(\Sigma_{t-1})) + b(\text{vech}(\Sigma_{t-1}))\eta_{t-1} \quad (1.6)$$

$$\eta_{t-1} \sim N(0, \Sigma_\eta),$$

where $\Sigma_t = (\sigma_{ij,t}) = \int_{t-1}^t \Sigma(s) ds$ is the so-called integrated covariance matrix, which is a measure of the ex-post covariation of $S(t)$ over the time interval $[t-1, t]$. The methodological

²See the excellent overview on MSV models of Asai et al., 2006.

³Since volatility models focus on second order dynamics, Eq. (1.3) assumes zero drift components for the log-price diffusion. This assumption is made for expositional purposes and is easily relaxed.

difference between the GARCH and the SV approach is illustrated by Eq. (1.6): The SV-implied volatility in period t features the random innovation η_{t-1} . This innovation is not measurable given past return observations.

The general MSV model of Eqs. (1.5) and (1.6) does not guarantee a positive definite covariance matrix. The latent nature of Σ_t makes positive definiteness even more difficult to achieve than in MGARCH models (see Asai et al., 2006). Empirically applied MSV models simplify Eq. (1.6) in order to ensure positivity of Σ_t and to address the trade-off between flexible (co)variance dynamics and the curse of dimensionality.

The standard univariate SV model of Taylor (1986) assumes a conditional log-normal distribution for the volatility process. The log-normal distribution is difficult to extend to the multivariate case and proposed MSV models, e.g. employed by Harvey et al. (1994), Danielsson (1998) and Smith and Pitts (2006), therefore typically feature vectors of log-volatilities interacting through a constant correlation structure. The MSV model of Harvey et al. (1994) is given by

$$\xi_t = Q_t^{1/2} u_t \tag{1.7}$$

$$Q_t^{1/2} = \text{diag}(\exp(h_{1,t}/2), \dots, \exp(h_{k,t}/2)) = \text{diag}(\exp(h_t/2)) \tag{1.8}$$

$$h_{t+1} = \phi_0 + \phi_1 \odot h_t + \eta_t, \tag{1.9}$$

$$\begin{pmatrix} u_t \\ \eta_t \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P_u & 0 \\ 0 & \Sigma_\eta \end{pmatrix} \right], \tag{1.10}$$

where $h_t = (h_{1,t}, \dots, h_{k,t})'$ is a $k \times 1$ vector of latent log-volatilities, ϕ_0 and ϕ_1 are $k \times 1$ parameter vectors and \odot denotes the Hadamard (element-by-element) product. Σ_η is a positive-definite covariance matrix and P_u is a positive-definite correlation matrix. The model implies that $\Sigma_t = Q_t^{1/2} P_u Q_t^{1/2}$ and mitigates the curse of dimensionality by assuming absence of volatility transmission effects across assets and constant correlations similar to the CCC-GARCH model of Bollerslev (1990). Both assumptions are clearly violated in practice (see e.g. Yu and Meyer, 2006). In order to allow for volatility spillovers the basic model can be extended by a vector autoregressive moving average (VARMA) structure for h_t .

Factor SV models are conceptually similar to factor GARCH models and further reduce the curse of dimensionality in multivariate volatility modeling while allowing for dynamic correlation structures. The factor SV approach was originally proposed by Harvey et al. (1994), and extended by Shephard (1996), Pitt and Shephard (1999b), Jacquier et al. (1999) and Doz and Renault (2006). The models feature a natural link to the arbitrage pricing theory (APT) of Ross (1976) and received great attention in the literature. Nevertheless, in practice model identification issues arise and the models still impose rather heavy restrictions since the same set of parameters determining time-variation in variances also determines the covariance and correlation dynamics.

A recent strand of MSV modeling tries to relax this intrinsic tension between variance

and covariance/correlation dynamics: Yu and Meyer (2006) propose a bivariate SV model which assumes AR(1) dynamics for the Fisher (1915)-transformed correlation coefficient. The model features the obvious drawback that it cannot be generalized to dimensions $k > 2$. Tsay (2005) proposes an MSV model based on latent AR(1) processes for the distinct elements of the Cholesky factor of the covariance matrix. Since the model builds on the Cholesky factorization, inference on the volatility dynamics depends on the sorting of the assets in the return vector. Recent approaches of Philipov and Glickman (2006) and Asai and McAleer (2009) introduce a new class of MSV models, which is based on the inverse-Wishart distribution for covariance matrices. This distributional assumption appears natural since the Wishart distribution is defined on the domain of positive definite matrices. Since the Wishart distribution is a multivariate generalization of the gamma distribution the Wishart multivariate SV (WMSV) approach can be interpreted as a natural generalization of a univariate inverse-gamma stochastic volatility model e.g. analyzed by Gander and Stephens (2007). Wishart SV models promise particularly flexible (co)variance dynamics including flexible dynamic correlation patterns and volatility transmission effects across assets.

Chapter 2 of the present thesis analyzes the properties of the basic WMSV model. It is found that the model has series problems in accommodating the strong persistence in daily asset return volatilities and the excess kurtosis of the return distribution. The existing literature on volatility modeling suggests that persistence in asset return volatilities may be caused by the presence of distinct volatility regimes (see e.g. Diebold, 1986, and Lamoureux and Lastrapes, 1990) induced by economic forces like business-cycle effects and periods of financial crisis. A volatility process featuring sudden shifts between various volatility levels is known to generate long-memory like persistence patterns which are typical for high-frequency return volatilities. Lamoureux and Lastrapes (1990) suggest to apply Markov switching (MS) processes as a way to model persistence within and switches between regimes. So et al. (1998) apply Markov switching volatility regimes to univariate SV models while Lopes and Carvalho (2007) extend the univariate framework to multivariate Markov switching factor SV models. Against this background Chapter 2 seeks to improve the empirical and theoretical properties of the WMSV approach and generalizes the basic Wishart multivariate stochastic volatility model to encompass regime switching behavior. The proposed MS WMSV model allows for state-dependent (co)variance and correlation levels and state-dependent volatility transmission across assets. A strengthening of volatility spillovers and return correlations in periods of high market volatility indicates contagion, i.e. crisis-related increases in return- and volatility dependencies (see e.g. Forbes and Rigobon, 2002, and Chiang and Wang, 2011), which tends to reinforce financial crises. The MS WMSV model allows for an assessment of contagion effects in returns and volatilities, which is important in order to understand the international propagation of financial distress. An empirical application to daily return data of five European stock indices shows that the proposed regime-switching specification substantially improves the model fit relative to the basic WMSV model. The

model's out-of-sample performance is evaluated in a Value-at-Risk (VaR) forecasting experiment. The VaR framework is of particular importance for financial managers since, for example, regulatory capital requirements for the market risk exposure of commercial banks are now explicitly based on VaR estimates and include a penalty for model inaccuracy (see e.g. Lopez and Walter, 2001). The MS WMSV model outperforms a range of competing volatility models from the literature with respect to unconditional coverage of the 5% VaR level. Chapter 2 is currently submitted to "Computational Statistics and Data Analysis" for possible publication.

1.1.3 The Realized Volatility Approach

The work of Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004) provides the basis for a direct modeling of consistent estimates of variances and covariances of daily asset returns, so-called realized (co)variances. The framework builds on the increasing availability of intra-day asset return data and effectively exploits the respective information content without having to explicitly model the high-frequency returns. The difference to GARCH and SV models is twofold: The former approaches focus on the conditional distribution of daily asset returns in order to forecast the daily return covariance matrix. The realized volatility approach avoids specifying a conditional distribution for asset returns and directly models and forecasts the realized covariance matrix. Furthermore, GARCH and SV models condition on low-frequency, i.e. daily, asset return data, while the realized volatility approach conditions on high-frequency, i.e. intra-day, return data comprising additional information on the variability of the return process.

Referring to Eq. (1.3) of Section 1.1.2 above and without loss of generality assuming zero drift components for the log-price diffusion, the daily log-return vector is given by

$$\xi_t = S_t - S_{t-1} = \int_{t-1}^t \Sigma^{1/2}(s) dW(s). \quad (1.11)$$

The quantity of interest is the integrated covariance matrix $\Sigma_t = \int_{t-1}^t \Sigma(s) ds$ of the period- t asset return vector ξ_t . For notational convenience we normalize the length of the trading day to 1. Exploiting results from stochastic process theory (see e.g. Protter, 2004) and essentially assuming a continuous semi-martingale log-price process $S(t)$, the daily integrated covariance matrix is equal to its quadratic variation over the same interval,

$$[S] = \text{plim}_{n \rightarrow \infty} \sum_{j=1}^n \{S(t_j) - S(t_{j-1})\} \{S(t_j) - S(t_{j-1})\}' \quad (1.12)$$

for any sequence of partitions $0 = t_0 < t_1 < \dots < t_n = 1$ with $\sup_j \{t_{j+1} - t_j\} \rightarrow 0$ for $n \rightarrow \infty$. In period t we observe, say, $M + 1$ intra-day log-price vectors $S_{t,i}$, $i = 0, 1, \dots, M$.

According to Eq. (1.12) the period- t realized covariance matrix

$$R_t = \sum_{i=1}^M \xi_{t,i} \xi_{t,i}' \quad (1.13)$$

computed using M intra-day log-returns $\xi_{t,i} = S_{t,i} - S_{t,i-1}$, $i = 1, \dots, M$, represents a consistent non-parametric estimate for the latent integrated covariance matrix Σ_t .⁴

Consistency indicates that the realized covariance matrix should be computed using the highest-frequency data available. However, if intra-day prices are recorded at very short intervals, say every second, the realized covariance matrix would be seriously contaminated by market microstructure noise induced by, but not limited to, the bid-ask bounce, discreteness of prices and non-synchronous trading (see e.g. Barndorff-Nielsen and Shephard, 2004). To put it another way, the observed log-price process is actually not a semi-martingale. If ignored, the presence of microstructure effects results in potentially large biases of the realized volatility measure (see e.g. Brown, 1990). In order to cope with microstructure noise the empirical finance literature suggests to sample at lower frequencies, e.g. every 5 minutes, or to apply intra-day-sampling at various sub-grids to exploit intra-day information more efficiently as proposed by Zhang et al. (2005).

Having computed a sequence of realized covariance matrices $\{R_t = (r_{ij,t})\}$ for $t = 1, \dots, T$, the realized variances and covariances can be modeled directly. Pioneering multivariate approaches to model the dynamics of realized (co)variances are found in Gouriéroux et al. (2009), Jin und Maheu (2011), Chiriac and Voev (2011), and Bauer and Vorkink (2011). The specification proposed by Gouriéroux et al. (2009) extends the Wishart distribution of the sample covariance for i.i.d. multivariate Gaussian random variables by allowing the multivariate Gaussian random variables to be serially correlated. Under the resulting Wishart autoregressive (WAR) process the realized covariance has a transition distribution which is noncentral Wishart with a non-centrality parameter depending on lagged covariances and a fixed scale matrix. As such the WAR model naturally accommodates positive definiteness of predicted covariance matrices without any parametric restrictions. The density of the noncentral Wishart distribution features the hypergeometric function, which has to be approximated numerically. Since this approximation is time-consuming, Maximum Likelihood estimation is practically infeasible and Method of Moment estimation has to be applied. The approach followed by Jin und Maheu (2011) is similar to the MSV approach of Philipov and Glickman (2006) and also relies on a Wishart transition distribution. The model assumes a central rather than a noncentral Wishart distribution, and decomposes its scale matrix into multiplicative components, which are driven by sample averages of lagged realized covariance

⁴A strand of literature analyzes the presence of discontinuous jump components in the semi-martingale log-price process dS_t . If price-jumps are present, the realized volatility approach measures the sum of both continuous and discontinuous volatility components (see e.g. Barndorff-Nielsen and Shephard, 2004b, and Huang and Tauchen, 2005). The analysis of discontinuous volatility components is not subject of this thesis and is left for future research.

matrices. The model is estimated by simulation based Bayesian inference techniques. In order to account for positive definiteness, the approaches of Chiriac and Voev (2011) and Bauer and Vorkink (2011) use appropriate transformations of the covariance matrix. The former approach is based upon a Cholesky decomposition of the covariance matrix and assumes fractionally integrated VARMA processes in order to model the elements of the Cholesky factor. The Cholesky factorization implies that inference on volatility dynamics depends on the sorting of the assets in the return vector. The approach of Bauer and Vorkink (2011) transforms the covariance matrix by using the matrix logarithm function and specifies the individual elements of the transformation as functions of factors driven by lagged volatilities and lagged returns. The nonlinearity of the matrix logarithm induces problems in deriving the marginal effects of the various forecasting variables driving the factor process and results in biased volatility forecasts. Hence bias correction methods have to be applied.

Chapter 3 of the present thesis proposes a conditional autoregressive Wishart (CAW) model for the analysis of realized covariance matrices of asset returns. The model assumes a simple autoregressive moving average structure for the scale matrix of the central Wishart distribution, which allows for complex serial and cross-sectional dependencies in variances and covariances. Under the CAW model the predicted covariance matrix depends on lagged covariance matrices as well as on their lagged predictions. The model therefore represents a dynamic generalization of the models proposed by Gouriéroux et al. (2009) and Jin und Maheu (2011), where the predicted covariance matrix is specified as a function of lagged covariances only. The model furthermore accounts for symmetry and positive definiteness of the predicted covariance matrices without imposing parametric restrictions and can easily be estimated by Maximum Likelihood (ML). In addition, the model allows to derive in a straightforward manner conditions for stationarity and other important time series properties. A further advantage of the CAW approach is that its baseline specification can easily be generalized. Chapter 3 illustrates two CAW model extensions which are specifically designed to capture long-memory like dependence patterns in the variances and covariances. For this purpose, the CAW specification is combined with the mixed data sampling (MIDAS) approach of Ghysels et al. (2005, 2006) and, alternatively, with a heterogeneous autoregressive (HAR) component as used by Corsi (2009) and Bonato et al. (2009). Chapter 3 is joint work with Vasyl Golosnoy and Roman Liesenfeld and represents a slightly extended version of the article “The conditional autoregressive Wishart model for multivariate stock market volatility” by Vasyl Golosnoy, Bastian Gribisch, and Roman Liesenfeld, which is published in the “Journal of Econometrics” 167, 2012, p. 211-223.

The particularly flexible mean dynamics of the CAW approach makes the model an ideal tool for analyzing volatility transmission effects on international financial markets. Chapter 4 of the present thesis extends the CAW model in order to investigate the short-term interdependence of the realized variances and covariance of the US Dow Jones and the German stock index DAX. Against the background of an apparently increasing integration of

international financial markets it is interesting to see to what extent a volatility shock generated by news in one market spills over onto the volatility observed in the next market to trade (see e.g. Engle et al., 1990, Hamao et al., 1990, and Melvin and Melvin, 2003). Similarly, it is of interest whether those spillover effects are more pronounced during periods of very high volatility associated with severe financial crises like the subprime crisis of 2008-2009 (see e.g. Chiang and Wang, 2011). The direct modeling of realized variances and covariances offers the advantage that these measures are typically more informative about the true volatility than corresponding conditional (co)variances obtained from MGARCH or MSV models. The use of high-frequency data is therefore expected to result in improved inference on volatility transmission across markets relative to the aforementioned models. In order to analyze intra-day volatility transmission patterns a novel sequential phase model is proposed, which accounts for the three distinct geographical intra-day trading periods of the US and German stock market: (1) the Germany-US trading overlap period, (2) the US-only trading period, and (3) the Germany-only trading period. The resulting model facilitates a detailed analysis of the short-term causal effects of news generating intra-day volatility in one market onto subsequent trading on this and the other market. In addition, the framework is used in order to investigate whether the short-term volatility transmission mechanism is significantly different during the recent subprime crisis than before and after the crisis which would indicate volatility contagion effects, i.e. intensifying cross-market volatility dependencies during periods of financial crisis (see e.g. Chiang and Wang, 2011). The extended CAW framework of Chapter 4 offers a new approach to modeling short term volatility transmission effects using high-frequency data. The approach differs from existing empirical studies on volatility transmissions across stock markets with overlapping trading hours like those of Engle, et al. (2012), Dimpfl and Jung (2012) and Chiang and Wang (2011) by accounting for the interdependence between the variances as well as the covariance. The approach therefore allows to account for two potential channels of volatility spillovers, namely, via a direct volatility transmission from one market to the other through its variance and via an indirect transmission through its covariance. Furthermore, the extended CAW approach explicitly accounts for the contemporaneous interdependence between the variances during the overlapping trading periods. Chapter 4 is joint work with Vasyl Golosnoy and Roman Liesenfeld and is currently submitted to the “Journal of Financial Econometrics” for possible publication.

Besides the requirement of positivity for covariance matrix forecasts and the ability to account for complex serial and cross-sectional (co)variance dependencies, the curse of dimensionality remains a main challenge in multivariate volatility modeling. Compared to MGARCH and MSV models the realized volatility approach facilitates a precise measurement of volatility in higher dimensions. Models for the realized covariance matrix are directly fitted to the time-series of $k(k+1)/2$ realized (co)variances, while MGARCH and MSV models are estimated based on the returns of k assets only. Hence, the number of observations

per parameter is significantly larger for realized volatility models than for similarly parameterized MGARCH and MSV specifications. Nevertheless, realized volatility models still suffer from the curse of dimensionality: The Cholesky approach of Chiriac and Voev (2011) and the Wishart models of Gouriéroux et al. (2009), Jin and Maheu (2011) as well as the CAW approach of Chapter 3 are heavily parameterized in order to allow for sufficiently flexible dynamic (cross-)correlation structures and the authors restrict their empirical applications to at most six assets. Bauer and Vorkink (2011) limit the application of their factor model to a five-dimensional asset portfolio, since the authors are mainly interested in assessing the predictive power of the various forecasting variables. Recently Bauwens and Storti (2011) proposed a CAW model featuring DCC dynamics, which allows for the forecasting of high-dimensional covariance matrices. Their empirical application comprises realized (co)variances for 50 NYSE stocks. Although the CAW-DCC model tackles the curse of dimensionality, this achievement does not come without a cost: the model imposes heavy restrictions on the correlation dynamics.

Chapter 5 introduces a new flexible latent dynamic factor model which alleviates the curse of dimensionality in multivariate volatility modeling and can be readily applied to the forecasting of potentially high-dimensional realized covariance matrices. The factor specification is motivated by persistent common dynamics of realized (co)variance series. Similar to the approach of Bauer and Vorkink (2011) the model is based on the matrix logarithm function which enables the modeling of log-(co)variances in Euclidean space, preserving positive definiteness and symmetry of covariance matrix forecasts without the necessity of imposing restrictions on the parameter space. By modeling the dynamics of the common factors as heterogeneous autoregressive processes (HAR, see Corsi, 2009) and assuming AR(1) processes for idiosyncratic dynamics the model mitigates the curse of dimensionality while allowing for rich (co)variance dynamics. A simulated Bayesian estimation approach using Markov Chain Monte Carlo (MCMC) techniques enables straightforward estimation of the model parameters and forecasting of covariance matrices. Since the elements of the matrix logarithm of a covariance matrix can be interpreted as approximations to correlations and logarithmic variances, the factor model allows for analyzing the presence of joint risk-factors related to market-risk and diversification risk (see e.g. Krishnan et al., 2009, and Driessen et al., 2009). The model is applied to two data sets of five- and 30-dimensional covariance matrices of NYSE traded stocks and the model's forecasting performance is assessed via a comprehensive out-of-sample experiment including a set of prominent forecasting models from the relevant literature. Besides conducting a statistical evaluation based on the mean squared error criterion Chapter 5 also addresses the practitioners point of view by investigating the performance of mean-variance optimal portfolios selected using the various volatility models.

The following section presents the outline of the present thesis.

1.2 Outline

Chapter 2 considers the Wishart multivariate SV (WMSV) approach proposed by Philipov and Glickman (2006) and Asai and McAleer (2009). The model assumes conditionally normal distributed asset returns and an inverse-Wishart distributed covariance matrix. The WMSV model promises particularly flexible (co)variance dynamics including time-varying correlation structures. The chapter discusses the stochastic properties of the basic Wishart MSV model and illustrates a new flexible Markov Switching (MS) WMSV model. The MS WMSV model allows for state-dependent shifts in the unconditional mean of (co)variances and correlations as well as state-dependent volatility spillover effects. The MS approach thereby captures sudden changes in the volatility level and induces long-memory like persistence patterns which are typical for high-frequency return volatilities. The basic WMSV and the MS WMSV model are applied to daily returns of five European stock indices. Parameter estimates are obtained using Bayesian Monte Carlo Markov Chain (MCMC) methods. Model diagnostic tests are conducted in order to check the model's ability in capturing (co)variance dynamics and the distributional characteristics of the underlying return data. The models' out-of-sample performance is evaluated in a VaR forecasting application.

Chapter 3 applies the conditional Wishart approach to the direct modeling of realized covariance matrices. A novel conditional autoregressive Wishart (CAW) model is proposed as a flexible modeling tool for the realized covariance matrix of asset returns. Its baseline specification assumes a simple autoregressive moving average structure for the scale matrix of the Wishart distribution allowing to account for complex serial dependencies in the variances and covariances. Under the CAW model the predicted covariance matrix depends on lagged covariance matrices as well as on their lagged predictions. The CAW model can easily be estimated by means of Maximum Likelihood (ML). To explicitly capture the long-run fluctuations in the variances and covariances the CAW specification is combined with the mixed data sampling (MIDAS) approach of Ghysels et al. (2005, 2006) and, alternatively, with a heterogeneous autoregressive (HAR) component as used by Corsi (2009) and Bonato et al. (2009). The CAW models are applied to daily realized covariance matrices for the returns of five NYSE traded stocks.

Chapter 4 extends the CAW approach in order to investigate the short-term interdependence of the realized variances and covariance of the US Dow Jones and the German stock index DAX. A novel sequential phase model is proposed, which accounts for the three distinct geographical intra-day trading periods of the US and German stock market. The model is applied to a detailed analysis of the short-term causal effects of news generating intra-day volatility in one market onto subsequent trading on this and the other market. An impulse response analysis provides information not only about the direct but also the indirect effects of volatility shocks. The framework is finally used in order to investigate whether the short-term volatility transmission mechanism is significantly different during the recent subprime

crisis than before and after the crisis which would indicate volatility contagion effects.

Chapter 5 introduces a new flexible latent dynamic factor model which tackles the curse of dimensionality in multivariate volatility modeling and can be readily applied to the forecasting of potentially high-dimensional realized covariance matrices. The model is based on the matrix logarithm function which enables the modeling of log-(co)variances in Euclidean space. The approach combines latent heterogeneous autoregressive processes for the common factor structure with idiosyncratic AR(1) factors for series-specific dynamics and mitigates the curse of dimensionality while allowing for rich (co)variance dynamics. Parameter estimates are obtained using simulation based Bayesian inference techniques. Joint factors can be interpreted as risk-factors related to market-risk and diversification risk. The model is applied to two data sets of five- and 30-dimensional covariance matrices of NYSE traded stocks. The model's forecasting performance is assessed via a comprehensive out-of-sample experiment including a range of prominent forecasting models from the relevant literature.

Chapter 6 summarizes the main results of the thesis.

Chapter 2

Multivariate Wishart Stochastic Volatility and Changes in Regime

2.1 Introduction

In contrast to the GARCH approach where volatility is modeled as a deterministic function of past return innovations, the stochastic volatility (SV) model introduced by Taylor (1982, 1986) assumes volatility to have its own stochastic process. Kim et al. (1998) find that simple SV models typically fit the daily asset return data as well as more heavily parameterized GARCH models. Basic SV models are furthermore natural discrete-time versions of continuous-time models which build the foundation of modern financial theory including generalizations of the Black-Scholes option pricing formula. However, developing flexible multivariate SV specifications proved to be complicated. Proposed multivariate SV (MSV) models, e.g. employed by Harvey et al. (1994), Danielsson (1998) and Smith and Pitts (2006), typically feature vectors of log-volatilities interacting through a constant correlation structure. The assumption of constant correlation is generally rejected by the data. Yu and Meyer (2006) applied nine alternative MSV models to a bivariate exchange rate series and found strong evidence for dynamic correlations. Factor SV models accommodate time-varying correlation patterns where the covariance and correlation dynamics are driven by time-variation in factor volatilities. This imposes restrictions since the covariances are not allowed to move independently from the variances. Recently Philipov and Glickman (2006) and Asai and McAleer (2009) introduced a new class of MSV models which assumes a conditionally inverse Wishart distributed covariance matrix. The Wishart distribution is a multivariate generalization of the gamma distribution and is defined on the domain of positive-definite matrices. The proposed model therefore naturally generalizes stochastic scalar variances to covariance matrices rather than vectors of log-variances. Wishart SV models promise particularly flexible (co)variance dynamics including flexible dynamic correlation patterns and volatility transmission effects across assets.

The present chapter analyzes the stochastic properties of the basic Wishart MSV (WMSV) model and proposes a new flexible Markov Switching (MS) WMSV model. The MS WMSV model allows for state-dependent shifts in the unconditional means of (co)variances and corre-

lations and state-dependent volatility transmission across assets, so-called volatility spillover effects. It has long been argued that strong persistence in asset return volatilities may be due to shifts in the unconditional mean of the volatility process (see e.g. Diebold, 1986, and Lamoureux and Lastrapes, 1990). A volatility process featuring sudden shifts between various volatility levels is known to generate long-memory like persistence patterns which are typical for high-frequency return volatilities. Lamoureux and Lastrapes (1990) suggest to apply Markov switching models as a way to model persistence within and switches between regimes. The MS approach allows to capture changes in the volatility level which are due to economic forces like business cycle downturns (see Hamilton and Susmel, 1994) as well as sudden changes which are due to unusual market events like the Lehman Brothers bust in 2008 or the 1987 stock market crash (see So et al., 1998). The idea of changes in volatility regimes is supported by various tests indicating multiple structural breaks for the conditional variance of asset return series spanning long time periods (see Andreou and Ghysels, 2002, for an overview). States of panic-like mood induce a higher volatility level compared to “calm” periods. Lamoureux and Lastrapes (1990) argue that sudden shifts in the variance, if unaccounted for, may bias upward persistence estimates. This has a clear practical implication: biased persistence estimates negatively affect volatility forecasts (see Haas et al, 2004). Fast tracking of structural changes in the (co)variance structure helps to avoid this bias. Hamilton and Susmel (1994), Gray (1996) and Haas et al. (2004) proposed univariate ARCH and GARCH models with regime switching. So et al. (1998) suggest to apply Markov switching volatility regimes to univariate SV models while Lopes and Carvalho (2007) extend the univariate framework to multivariate MS SV modeling and propose a factor SV model featuring univariate MS processes for the common factors’ variance dynamics. Limiting the MS process to a few common factors imposes restrictions in multivariate volatility modeling. The proposed MS WMSV model contributes to the literature by allowing for sudden shifts in the (co)variance level affecting all elements of the covariance matrix independently from one another. The model thereby offers particularly flexible volatility and correlation dynamics including long-memory type of persistence patterns, state-dependent (co)variance and correlation levels and volatility transmission effects across assets. Crisis-related strengthening of volatility spillovers and return correlations indicates contagion effects (see e.g. Forbes and Rigobon, 2002, and Chiang and Wang, 2011), which are known to reinforce financial crisis events (see e.g. Diebold and Yilmaz, 2009). The MS WMSV model allows to assess the presence of contagion effects in returns and volatilities, which is important in order to understand the international propagation of financial distress.

The proposed MS WMSV model is applied to daily returns of five European stock indices. Model diagnostic tests are conducted in order to check the model’s ability in capturing (co)variance dynamics and the distributional characteristics of the observed return data. The results show that the MS extension substantially improves the model fit of the basic WMSV approach. The estimates furthermore indicate intensifying return correlation and volatil-

ity transmission in periods of financial turmoil. The models' out-of-sample performance is evaluated in a VaR forecasting application. The MS WMSV model outperforms a range of competing volatility models from the literature with respect to unconditional coverage of the 5% VaR level.

The outline of the chapter is as follows: Section 2.2 illustrates the basic WMSV model and the MS WMSV model, the Bayesian simulation based estimation scheme and model diagnostics based on standardized returns. Section 2.3 presents estimation- and model diagnostic results and the VaR forecasting application. Section 2.4 summarizes the findings.

2.2 Model Specification, Inference and Diagnostics

2.2.1 The Basic WMSV Model

Consider the stochastic k -dimensional return vector ξ_t and its stochastic $k \times k$ covariance matrix $\Sigma_t = (\sigma_{ij,t})$ at time period t ($t = 1, \dots, T$). The basic WMSV model is given by

$$\xi_t | \Sigma_t \sim N(0, \Sigma_t), \quad (2.1)$$

$$\Sigma_t^{-1} | \Sigma_{t-1}^{-1} \sim \mathcal{W}_k(\nu, S_t / \nu), \quad (2.2)$$

where the return vector ξ_t is assumed to be mean-corrected. \mathcal{W}_k denotes the law of a k -dimensional central Wishart distribution with $\nu > k$ degrees of freedom and a $k \times k$ symmetric and positive definite scale matrix S_t / ν , where $S_t = (s_{ij,t})$. By specifying a conditional Wishart distribution for the precision matrix Σ_t^{-1} instead of the covariance matrix Σ_t the WMSV framework generalizes the univariate inverse gamma SV model which is e.g. discussed in Gander and Stephens (2007).

Using the properties of the Wishart and inverse Wishart distribution of Σ_t^{-1} and Σ_t , respectively, we obtain (see Muirhead, 1982)

$$E[\Sigma_t^{-1} | \Sigma_{t-1}^{-1}] = S_t, \quad (2.3)$$

$$E[\Sigma_t | \Sigma_{t-1}] = \frac{1}{\nu - k - 1} S_t^{-1}. \quad (2.4)$$

In order to allow for serial and cross-correlations across the variances and covariances the scale matrix in period t is assumed to depend on lagged (co)variances:

$$S_t = \Sigma_{t-1}^{-d/2} A \Sigma_{t-1}^{-d/2}, \quad (2.5)$$

where A is a positive definite $k \times k$ parameter matrix and d is a scalar persistence parameter.¹

¹The assumed functional form of the scale matrix S_t corresponds to the Wishart Inverse Covariance (WIC) model of Asai and McAleer (2009). Philipov and Glickman (2006) assume a similar specification: $S_t = A^{1/2} \Sigma_{t-1}^{-d} A^{1/2}$.

Based on the spectral decomposition $\Sigma_t^{-1} = V_t \Lambda_t V_t'$ we obtain

$$\Sigma_t^{-d/2} = V_t \Lambda_t^{\frac{d}{2}} V_t', \quad (2.6)$$

where V_t denotes the matrix of orthogonal eigenvectors of Σ_t^{-1} and Λ_t denotes the corresponding diagonal matrix of eigenvalues. The power operator is defined to work element-wise. Note that $\Sigma_t^{-d/2} \Sigma_t^{-d/2} = \Sigma_t^{-d}$. The quadratic form of S_t in Eq. (2.5) ensures a positive definite scale matrix.

(Co)variance dynamics are governed by the parameter matrix A and the scalar d , which directs the persistence of the (co)variance process. This can be seen by rewriting the specification of Σ_t^{-1} using the properties of the Wishart distribution: Denoting the $k \times k$ identity matrix by I_k and the lower triangular Cholesky factor of A by L , i.e. $A = LL'$, we obtain

$$\Sigma_t^{-1} = \frac{1}{\nu} \Sigma_{t-1}^{-d/2} L \mathcal{W}_k(\nu, I_k) L' \Sigma_{t-1}^{-d/2}, \quad (2.7)$$

which yields an autoregressive representation for the logarithmic determinant of Σ_t^{-1}

$$\ln |\Sigma_t^{-1}| = -k \ln(\nu) + \ln |A| + d \ln |\Sigma_{t-1}^{-1}| + \ln |\mathcal{W}_k(\nu, I_k)|. \quad (2.8)$$

The condition for weak stationarity of the logarithmic determinant of the Wishart process is therefore given by $|d| < 1$. Philipov and Glickman (2006) acknowledge that deriving analytical conditions for weak stationarity of the (co)variances themselves may not be possible. In practice, d should be additionally restricted to positivity to rule out stochastic processes for Σ_t^{-1} which alternate between powers of inverses. While d determines the strength of inter-temporal relationships, A can be interpreted as a measure of “inter-temporal sensitivity” (see Philipov and Glickman, 2006): Without restrictions on this matrix, all elements of Σ_t are allowed to depend on their own lag and the lags of all remaining (co)variances. Restricting A to a diagonal matrix completely excludes volatility spillover effects. Eqs. (2.4) and (2.5) show that inter-temporal (co)variance transmission is actually measured by A^{-1} .

Since no closed form analytical expression can be derived, I simulate unconditional (co)-variance moments based on a two-dimensional WMSV model and a variety of parameter constellations in order to further analyze the influence of the model parameters A^{-1} , ν and d on distributional and dynamic characteristics. For each structural model parameter five parameter values are considered: The parameter sets for d and ν are $d \in \{0.2, 0.4, 0.6, 0.8, 0.9\}$ and $\nu \in \{20, 40, 60, 80, 90\}$. The matrices A_i^{-1} , $i = 1, \dots, 5$, are characterized by overall increasing matrix entries in i on each single position in A_i^{-1} . Let $\text{vech}(\cdot)$ denote the operator that stacks the lower triangular portion, including the diagonal of a matrix into a vector. In order to reflect realistic (co)variance dynamics $\text{vech}(A_1^{-1})$ is set to its point estimate $\text{vech}(A_1^{-1}) = (0.96, 0.02, 0.96)'$ obtained by fitting the basic WMSV model to a bivariate series of daily stock index returns for France and Germany (see the data description in Sec-

tion 2.3 below). For $i = 2, \dots, 5$ we obtain $\text{vech}(A_i^{-1}) = (1.2, 2, 1.2)' \odot \text{vech}(A_{i-1}^{-1})$, where \odot denotes element-wise multiplication². Figure 2.1 and Figure 2.2 show that increasing the elements of A^{-1} has a significant positive effect on the overall (co)variance and correlation level. The effects of ν and d in contrast appear comparatively minor. Figure 2.3 depicts simulated autocorrelation functions for the first asset's variance. The persistence appears to be solely driven by d . Corresponding plots for the second variance and the covariance are not presented here but confirm that d drives serial correlation for the whole (co)variance process. Figure 2.4 depicts simulated cross-correlation functions for the variances of the first and second asset return. The functions show that cross-asset volatility persistence is solely captured by A^{-1} . Spillover effects increase with increasing matrix entries in A^{-1} . Summarizing the results, while d drives the overall (co)variance persistence, the volatility and correlation level as well as the strength of cross-asset volatility transmission effects are captured by A^{-1} . The role of the d.o.f. parameter ν becomes apparent by considering (co)variances of the Σ_t^{-1} elements based on the properties of the Wishart distribution (see Muirhead, 1982):

$$\text{Cov}(\sigma_{ij,t}^{-1}, \sigma_{lm,t}^{-1} | \Sigma_{t-1}^{-1}) = \frac{1}{\nu} (s_{il,t} \cdot s_{jm,t} + s_{im,t} \cdot s_{jl,t}), \quad (2.9)$$

for $i, j, l, m = 1, \dots, k$, where $\sigma_{ij,t}^{-1}$ denotes the ij 'th element of Σ_t^{-1} . Hence ν directly effects the dependence structure within the (co)variance process.

2.2.2 The Markov Switching WMSV Model

This section describes a new Markov switching (MS) WMSV model, which induces state-dependent covariance and correlation levels and state-dependent volatility spillover effects. This is accomplished by allowing the parameter matrix A of the Basic WMSV model to switch between different realizations. Suppose that $s_t \in \{1, 2\}$ is an unobserved two-state Markov process with transition probability matrix

$$\begin{bmatrix} \Pr(s_t = 1 | s_{t-1} = 1) & \Pr(s_t = 2 | s_{t-1} = 1) \\ \Pr(s_t = 1 | s_{t-1} = 2) & \Pr(s_t = 2 | s_{t-1} = 2) \end{bmatrix} = \begin{bmatrix} (1 - e_1) & e_1 \\ e_2 & (1 - e_2) \end{bmatrix}, \quad (2.10)$$

where e_1 denotes the probability of switching from state 1 in period $t - 1$ to state 2 in period t and e_2 the probability of switching from state 2 in period $t - 1$ to state 1 in period t . The latent state variable s_t defines a particular regime characterized by a regime-specific parameter matrix A_{s_t} . The 2-regime MS model is then given by

$$\xi_t | \Sigma_t \sim N(0, \Sigma_t), \quad (2.11)$$

$$\Sigma_t^{-1} | \Sigma_{t-1}^{-1} \sim \mathcal{W}_k(\nu, S_t / \nu), \quad S_t = \Sigma_{t-1}^{-d/2} A_{s_t} \Sigma_{t-1}^{-d/2}, \quad (2.12)$$

²The simulation results are found to be robust to variations in the parameter values.

together with Eq. (2.10). According to the simulation results of Section 2.2.1 above the MS WMSV model allows for structural changes in the (co)variance/correlation level and volatility transmission intensity, where the timing of the shifts is captured by the latent Markov process.

The MS WMSV model as specified is unidentified. A sufficient condition for identification is restricting the first diagonal element of the matrix difference $\tilde{A} = A_2 - A_1$ to be positive. Note that it is straight forward to also allow the parameters ν and d to change according to the same Markov process. The goal is, however, to capture clusters of low and high risk in the market as captured by small and large values in A . Also note that the model can easily be generalized to more than two volatility states. This would however significantly increase the dimension of the parameter space since the number of parameters in A is proportional to the square of the number of assets. The results of Lopes and Carvalho (2007) and Carvalho and Lopes (2007) indicate the empirical sufficiency of a 2-regime model, which preserves parsimony in multivariate volatility modeling. Two states imply two (co)variance and correlation levels, which correspond to times of high and low risk in the market.

2.2.3 Estimation and Diagnostics

Following Philipov and Glickman (2006) a Bayesian estimation approach is applied for inference on the (MS) WMSV model's parameter vector $\theta^{\text{WMSV}} = (\text{vech}(A)', \nu, d)'$ or $\theta^{\text{MS WMSV}} = (\text{vech}(A_1)', \text{vech}(A_2)', \nu, d, e_1, e_2)'$, respectively. Bayesian estimation is particularly attractive for complex multivariate models including a large number of parameters. High-dimensionality of the parameter vector involves practical problems of the classical estimation scheme due to the numerical maximization of the likelihood function. These complications can be avoided by making use of tractable Bayesian estimation techniques. The objective of primary interest is the joint posterior distribution of the model parameters, whose moments can be used to generate point estimates and to assess the according parameter uncertainty. The posterior distribution is proportional to the product of the likelihood function and the parameters' joint prior distribution. The likelihood function of the basic WMSV model is a high-dimensional integral

$$L(\{\xi_t\}_{t=1}^T | \theta^{\text{WMSV}}) = \int \dots \int_{\Sigma_1, \dots, \Sigma_T} \prod_{t=1}^T P(\xi_t | \Sigma_t) \times P(\Sigma_t | \Sigma_{t-1}, \theta^{\text{WMSV}}) d\Sigma_1, \dots, d\Sigma_T. \quad (2.13)$$

This integral is analytically intractable and its evaluation requires simulation-based estimation techniques. The Monte Carlo Markov Chain (MCMC) approach became increasingly popular in the last decades and can be readily applied for Bayesian inference within the WMSV framework. The MCMC scheme generates draws from the joint posterior distribution of the model parameters via simulating an irreducible and aperiodic Markov chain. Under some mild regularity conditions the latter converges to the parameters' joint poste-

rior distribution. The Markov chain is generated by the Gibbs sampling algorithm, which involves iterative drawing from the full conditional distributions of the model parameters, where the parameter vector is augmented by the set of latent variables³. Bayesian point estimates are obtained by averaging the Gibbs draws after convergence of the Markov chain⁴. Estimation uncertainty is captured by the sample standard deviation of the Gibbs draws. Following Lopes and Carvalho (2007) full conditional sampling of the state sequence $\{s_t\}_{t=1}^T$ is achieved by Forward Filtering Backward Sampling (FFBS) using the Hamilton filter (see Hamilton and Susmel, 1994). All derivations of full conditional distributions are given in the Appendix. If specific distributions are not available in closed form, but known up to an integrating constant, the Metropolis-Hastings algorithm is applied for simulation purposes.

Having estimated the models, diagnostic tests are applied in order to check the model's ability to account for the observed (co)variance dynamics and distributional characteristics of the return series. Diagnostic tests on (co)variance dynamics are conducted from the vector of standardized Pearson residuals

$$\begin{aligned}
 e_t^* &= \text{Var}[\xi_t | \mathcal{F}_{t-1}]^{-\frac{1}{2}} \xi_t \\
 &= E[\xi_t \xi_t' | \mathcal{F}_{t-1}]^{-\frac{1}{2}} \xi_t \\
 &= E[E[\xi_t \xi_t' | \Sigma_t] | \mathcal{F}_{t-1}]^{-\frac{1}{2}} \xi_t \\
 &= E[\Sigma_t | \mathcal{F}_{t-1}]^{-\frac{1}{2}} \xi_t,
 \end{aligned} \tag{2.14}$$

where $\mathcal{F}_{t-1} = \{\xi_t\}_{t=1}^{t-1}$ and $E[\Sigma_t | \mathcal{F}_{t-1}]^{-\frac{1}{2}}$ denotes the inverse Cholesky factor of $E[\Sigma_t | \mathcal{F}_{t-1}]$. The filtered covariance estimate $E[\Sigma_t | \mathcal{F}_{t-1}]$ constitutes a high-dimensional integral, which can be approximated by the sample mean over draws from the respective conditional distribution:

$$e_t^* = E[\Sigma_t | \mathcal{F}_{t-1}]^{-\frac{1}{2}} \xi_t \cong \left(\frac{1}{M} \sum_{j=1}^M \Sigma_t^{(j)} \right)^{-\frac{1}{2}} \xi_t, \tag{2.15}$$

where $\Sigma_t^{(j)}$ denotes a draw from $f(\Sigma_t | \mathcal{F}_{t-1})$, which is obtained by applying the standard particle filter algorithm illustrated by Pitt and Shephard (1999) and given in the Appendix, and M is the simulation sample size. For a correctly specified model, the standardized residuals $e_{i,t}^*$ in the vector e_t^* are serially uncorrelated in levels, squares and cross-products. The series can therefore be used for diagnostic checking of the assumed dynamic structure, e.g. using the Ljung-Box test on serial correlation.

The model's ability in reflecting the distributional characteristics of the underlying return data is checked following Kim et al. (1998) and Liesenfeld and Richard (2003). The approach requires the computation of the conditional probability that the i 'th return $\xi_{i,t}$ is less than

³For details on the Gibbs sampling algorithm and Monte Carlo Markov Chain methods see e.g. Bauwens et al. (1999) and the Appendix of the current chapter.

⁴I.e. after a certain number of burn-in iterations of the Gibbs sampler.

the actually observed return $\xi_{i,t}^o$, i.e. $Pr(\xi_{i,t} \leq \xi_{i,t}^o | \mathcal{F}_{t-1})$. Again applying standard particle filtering this probability can be approximated by

$$Pr(\xi_{i,t} \leq \xi_{i,t}^o | \mathcal{F}_{t-1}) \cong u_{i,t}^M = \frac{1}{M} \sum_{j=1}^M Pr(\xi_{i,t} \leq \xi_{i,t}^o | \sigma_{ii,t}^{(j)}), \quad (2.16)$$

where $\sigma_{ii,t}^{(j)}$ denotes the i 'th diagonal element of $\Sigma_t^{(j)}$, drawn from $f(\Sigma_t | \mathcal{F}_{t-1})$, $j = 1, \dots, M$. Under the Null of a correctly specified model the $\{u_{i,t}^M\}_{t=1}^T$ sequence is iid uniform distributed on $[0, 1]$ for all $i = 1, \dots, k$ and can be mapped into the standard normal distribution via the inverse of the according cdf: $e_{i,t}^M = F_N^{-1}(u_{i,t}^M)$. Statistical tests for normality of $e_{i,t}^M$ can be based on the Jarque-Bera test statistic.

2.3 Empirical Application

2.3.1 Data

The (MS) WMSV models are applied to daily AR(p) pre-filtered stock index log-returns⁵ for France, Germany, Italy, Switzerland and the UK from January 2, 2003, to December 31, 2008, leaving a sample of 1565 observations⁶. The return series are illustrated in Figure 2.5 and descriptive statistics are given in Table 2.1. All series feature excess kurtosis, insignificant autocorrelation in returns and significant autocorrelation in squared returns. The reported sample correlations indicate a huge degree of co-movement for all five stock indices.

2.3.2 Estimation Results

Basic WMSV Model

Table 2.2 presents the estimation results for the basic WMSV model. The chosen prior distributions are overall uninformative and also reported in Table 2.2. The estimation is based on 50,000 Gibbs iterations and a burn-in of 15,000 iterations. The convergence of the generated Markov chains is checked using convergence diagrams (not presented here) as e.g. applied by Liesenfeld and Richard (2008) and Ross (2002). All parameter estimates are significant at the 5% level. MC standard errors addressing the numerical accuracy of the simulation based estimation scheme are calculated using a correlation consistent Parzen window based spectral estimator for the variance of the sample mean (see Kim et al., 1998). The ratio of MC standard error to posterior standard deviation addresses the proportion of variation in the estimates due to simulation relative to the variation induced by the data. The mean ratio is about 10% which exceeds the ratio of 1% obtained by Kim et al. (1998) for the

⁵Datastream DS market indices.

⁶The daily prices p_t are transformed into continuously compounded rates $r_t = 100 \times \ln(p_t/p_{t-1})$ which are then filtered for AR(p) processes according to the Akaike information criterium.

basic univariate log-normal SV model and a comparable number of Gibbs iterations. This reduction in numerical accuracy reflects the relative complexity of the WMSV framework, as illustrated by the integration problem in Eq. (2.13). Within the WMSV framework posterior inference implies $k(k+1)/2 \times T$ -dimensional integration as opposed to T -dimensional integration for the basic SV model.

The estimated persistence parameter $d = 0.95$ implies strong persistence of the (co)-variance process and the significant off-diagonal elements in A^{-1} indicate the presence of volatility spillover effects. Figures 2.6 and 2.7 depict smoothed estimates of dynamic standard deviations and correlations. The results imply strong volatility clustering and accentuated volatility peaks at the beginning of 2003 and in 2006, and a large volatility cluster slowly building up from the middle of 2007. The latter is caused by the financial crisis originating in the US subprime market. Figure 2.7 shows strong co-movement and significant dynamics in the correlation series.

Table 2.3 shows Ljung-Box diagnostic test results for the series of Pearson residual cross-products. 14 out of 15 series do not pass the test of the Null of no serial correlation at the 1% significance level. This implies considerable problems of the baseline WMSV model in accommodating the strong serial and cross-sectional correlation of daily asset return (co)variances. Figure 2.8 shows sample autocorrelation functions for the squared residual series which support the Ljung-Box results. The plots show significant serial correlation for up to 50 lags. Yet the model successfully accounts for a major portion of the highly persistent (co)variance dynamics. Table 2.4 shows diagnostic results on distributional characteristics. The Jarque-Bera test indicates significant deviations from normality for all residual series. This finding is mainly due to unexplained excess kurtosis of the return distribution. The basic WMSV model has problems in capturing the fat tails of daily asset return data. The residuals are furthermore skewed to the left, which suggests the presence of asymmetric effects, e.g. the leverage effect of Black (1976) and Christie (1982).⁷ The previous findings are supported by qq-plots depicted in Figure 2.9. The plots show severe deviations from normality in the tails of the residual distribution.

MS WMSV Model

Table 2.2 shows estimation results and prior distributions for the two-state Markov switching WMSV model. Allowing for Markov switching regimes enables the WMSV framework to accommodate structural changes in asset-return volatilities. Yet the model extension increases the parameter space by $k(k+1)/2 + 2$ additional parameters and a latent state variable. Compared to the basic WMSV model and given a comparable number of Gibbs sequences, this increase in model complexity is reflected by increasing numerical standard errors: The mean ratio of MC standard error to posterior standard deviation amounts to 20%, which is

⁷The modeling of asymmetric effects is not subject of this thesis and is left for future research.

twice as high compared to the basic WMSV approach.

14 out of 15 estimates in A_2^{-1} significantly exceed their corresponding estimates in A_1^{-1} . This suggests an overall higher volatility and correlation level in the second state, which is supported by numerical approximations of unconditional means of volatility and correlation presented in Table 2.5. A higher correlation level under turbulent market conditions is a commonly observed phenomenon (see e.g. Solnik et al., 1996) and can be interpreted as contagion in the lines of Forbes and Rigobon (2002), i.e. crisis-related increases in return dependencies. Increasing asset correlation in periods of turmoil indicates that diversification opportunities tend to vanish when they are needed most. Figure 2.10 depicts smoothed state and volatility estimates for France and Germany obtained under the basic WMSV and the MS WMSV model. The figure shows that MS WMSV implied volatility significantly exceeds basic WMSV implied volatility in periods with high probability for the second state ($s_t = 2$). In particular, assuming that a volatility state has been realized if the corresponding smoothed state probability exceeds 0.5, the second volatility state covers two pronounced clusters of exceedingly high market volatility: the period of Iraq war in March 2003 and preceding oil price fluctuations as well as the subprime crisis period slowly building up from the mid of 2007 and finally culminating in a huge volatility cluster initiated by the Lehman Brothers bust on September 15, 2008. The high-volatility state additionally covers particular events like the terrorist attacks in Madrid and London on March 11, 2004, and July 7, 2005, respectively, which had pronounced effects on international stock markets. State-dependent regime switching allows for a fast adaption to structural changes like crisis-related increases in volatility levels. This helps to avoid an overestimation of the model-implied volatility persistence which is likely to occur if structural changes in the volatility process are not taken into account: The estimate of the persistence parameter d obtained under the MS WMSV model is significantly lower compared to the corresponding estimate obtained under the basic WMSV model (see Table 2.2). The estimated diagonal elements of the transition probability matrix $\Pr(s_t = 1|s_{t-1} = 1) = 0.92$ and $\Pr(s_t = 2|s_{t-1} = 2) = 0.60$ imply long duration in each regime with a predominance of the low volatility regime. The estimated unconditional probability for state 2 is 0.17.⁸ The estimates of A_1^{-1} and A_2^{-1} suggest intensifying volatility transmission effects in periods of high market volatility. Table 2.5 shows that model-implied one-period ahead volatility cross-correlations increase significantly by switching from state 1 to state 2. This indicates intensified volatility spillovers in uncertain periods and implies contagion in volatilities (see e.g. Chiang and Wang, 2011, and Diebold and Yilmaz, 2009). The presence of contagion stimulates international propagation of crisis effects as e.g. observed for the U.S. subprime crisis, which spread out around the world through various economic and financial links. A potential source of such changes in market dependencies could be the boost of intensity at which news hits international financial markets when entering a turbulent crisis period. This prompts investors to strengthen their monitoring of financial

⁸See Hamilton, 1994, p. 683, for the computation of unconditional state probabilities.

market transactions in order to gather new critical information about their investments and fundamentally reassess the vulnerability of other financial markets (see e.g. Bekaert et al., 2011).

Table 2.3 shows Ljung-Box diagnostic test results for the series of Pearson residual cross-products. 13 out of 15 series pass the test of the Null of no serial correlation at the 1% significance level. This finding is supported by sample ACFs of squared residual series depicted in Figure 2.11. Compared to the basic WMSV approach the MS framework captures the strong persistence of asset return (co)variances by combining structural shifts in the mean of the volatility process with volatility persistence in each regime. Table 2.4 shows diagnostics on distributional characteristics. Compared to the basic WMSV model the results show remarkable improvements in capturing the excess kurtosis of the return distribution. According to the Jarque-Bera test results we cannot reject the Null of normality for two out of five series at the 1% significance level. The significant reduction of residual kurtosis is caused by the mixture effects of the Markov switching process (see e.g. Haas et al., 2004). Figure 2.12 depicts QQ-plots which confirm the Jarque-Bera test results (compare to Figure 2.9). Since it is a widely accepted fact that conditional normality in standard SV and GARCH models does not capture the excess kurtosis of financial return series, fat-tailed conditional return distributions, like the multivariate Student- t distribution, represent an alternative popular way of accounting for excess kurtosis. For an initial investigation I fitted a WMSV model with conditionally multivariate Student- t distributed returns to the European asset return data. In contrast to the MS WMSV model the respective residual series still implied considerable problems in capturing the excess kurtosis of the return data.

2.3.3 Forecasting Results

This section assesses the out-of-sample performance of the WMSV model in a Value-at-Risk (VaR) forecasting experiment. VaR measures indicate the portfolio value that could be lost over a given time-period with a specified confidence level α . Given a k -dimensional vector of portfolio weights w the level α VaR forecast of a portfolio return $\xi_{p,t}$ at time t given return information up to period $t - 1$ is computed as

$$\text{VaR}_{p,t|t-1}(\alpha) = \sqrt{\hat{\sigma}_{p,t|t-1}} F^{-1}(\alpha), \quad (2.17)$$

where $F^{-1}(\alpha)$ denotes the α -percentile of the cumulative one-step-ahead distribution assumed for portfolio returns and $\hat{\sigma}_{p,t|t-1}$ denotes the model-based portfolio variance forecast using return information up to period $t - 1$. The VaR framework is of particular importance for financial managers since, for example, regulatory capital requirements for the market risk exposure of commercial banks are explicitly based on VaR estimates and include a penalty for model inaccuracy (see Lopez and Walter, 2001).

According to common practice (see e.g. Lopez and Walter, 2001, and Chib et al., 2006)

I conduct 5% VaR forecasts for an equally weighted portfolio of the considered five European stock indices. The out-of-sample window covers 262 trading days from January 2, 2008 through December 31, 2008. All models are re-estimated daily and new forecasts are generated based on the updated parameter estimates. I consider a range of prominent competing forecasting models, where the choices are motivated by the popularity of the models in the academic literature. The following specifications are used:

1. The BEKK-GARCH(p,q) model of Engle and Kroner (1995) assumes $\xi_t = H_t^{1/2}v_t$, where $v_t \sim \mathcal{N}(0, I_k)$ and $H_t^{1/2}$ is the lower triangular Cholesky factor of the conditional covariance matrix H_t , which is specified as

$$H_t = D_0 D_0' + \sum_{i=1}^p D_i H_{t-i} D_i' + \sum_{j=1}^q G_j [\xi_{t-j} \xi_{t-j}'] G_j', \quad (2.18)$$

where D_0 is a lower triangular $k \times k$ matrix. D_i, G_j are $k \times k$ matrices which may be restricted to diagonality to reduce the dimension of the parameter space (Diagonal BEKK-GARCH(p,q) model).

2. The Dynamic Conditional Correlation (DCC)-GARCH(p,q) model of Engle (2002) assumes conditional normality for the return vector ξ_t and scalar GARCH(p,q) dynamics for the conditional variances $\{h_{ii,t}\}_{i=1}^k$. The modeling of dynamic conditional correlations is based on the decomposition

$$H_t = D_t P_t D_t, \quad (2.19)$$

where $D_t = \text{diag}(\sqrt{h_{11,t}}, \dots, \sqrt{h_{kk,t}})$ and P_t is a $k \times k$ conditional correlation matrix. The latter is expressed as

$$P_t = (\text{diag}(Q_t))^{-\frac{1}{2}} Q_t (\text{diag}(Q_t))^{-\frac{1}{2}}, \quad (2.20)$$

with Q_t being a $k \times k$ symmetric, positive definite matrix given by

$$Q_t = (1 - \alpha - \beta) \bar{Q} + \alpha u_{t-1} u_{t-1}' + \beta Q_{t-1}, \quad (2.21)$$

where α and β are positive scalar parameters and u_t is the k -dimensional vector of standardized residuals with elements

$$u_{i,t} = \frac{\xi_{i,t}}{\sqrt{h_{ii,t}}}, \quad i = 1, \dots, k. \quad (2.22)$$

\bar{Q} is the unconditional covariance matrix of u_t which is consistently estimated by the according sample covariance matrix.

3. The Constant Conditional Correlation (CCC)-GARCH(p,q) model of Bollerslev (1990) is obtained by restricting the DCC-GARCH(p,q) model setting $P_t = P$, where P is the

sample correlation matrix of returns.

4. The Exponentially Weighted Moving Average (EWMA) approach is a simple forecasting model, which is commonly used for risk management purposes (see RiskMetrics, J.P. Morgan, 1996). The model assumes conditional normality for returns and a conditional covariance matrix

$$H_t = (1 - \lambda)\xi_{t-1}\xi'_{t-1} + \lambda H_{t-1}. \quad (2.23)$$

For the empirical application λ is set to its typical value for daily asset return data given by 0.94.

Details on obtaining forecasts given the multivariate GARCH and EWMA models are e.g. provided by Chib et al. (2006). For standard MGARCH models the portfolio return's cumulative one-step ahead distribution is normal. VaR forecasts are then obtained as the α -percentile of the corresponding normal distribution for portfolio returns. Chib et al. (2006) illustrate how to obtain VaR forecasts within the simulation based MCMC scheme: The Gibbs sampling algorithm allows for a direct simulation from the predictive densities of the individual asset returns. The VaR forecast is then obtained by the (left-tail) quantile of interest.

The accuracy of obtained VaR estimates is evaluated using the unconditional and conditional coverage tests illustrated by Lopez and Walter (2001) and e.g. applied by Chib et al. (2006). The test of unconditional coverage is explicitly incorporated into the Basel bank capital requirements. Defining an indicator variable

$$I_t = \begin{cases} 1 & \text{if } \xi_{p,t} < \text{VaR}_{p,t|t-1}, \\ 0 & \text{if } \xi_{p,t} \geq \text{VaR}_{p,t|t-1}, \end{cases} \quad (2.24)$$

and denoting the number of out-of-sample observations by T^* , the “hit-rate” is obtained as $\hat{\alpha} = \gamma/T^*$, where $\gamma = \sum_{t=1}^{T^*} I_t$. Accurate VaR forecasts should feature an unconditional coverage $\hat{\alpha}$ close to α . The hypothesis $E[\hat{\alpha}] = \alpha$ can be tested using the statistic

$$LR_{uc} = 2\{ \ln[\hat{\alpha}^\gamma (1 - \hat{\alpha})^{T^* - \gamma}] - \ln[\alpha^\gamma (1 - \alpha)^{T^* - \gamma}] \}, \quad (2.25)$$

which is under the Null asymptotically $\chi^2(1)$ distributed.

Since financial asset returns are heteroscedastic, volatility models that ignore (co)variance dynamics will provide VaR estimates that may have unconditional coverage, but will have incorrect conditional coverage at any point in time (see Lopez and Walter, 2001). Christoffersen (1998) proposes a test of conditional coverage by jointly testing for correct unconditional coverage and independence in the hit-rate series, where the independence hypothesis is tested against the hypothesis of first-order Markov dependence. Define T_{ij} as the number of observations in state j ($I_t = j$) after having been in state i in the previous pe-

riod ($I_{t-1} = i$), where $i, j \in \{0, 1\}$ (see Eq. 2.24). Also denote $\pi_{01} = T_{01}/(T_{00} + T_{01})$ and $\pi_{11} = T_{11}/(T_{10} + T_{11})$. Under the alternative hypothesis the likelihood function is $L_A = (1 - \pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_{10}} \pi_{11}^{T_{11}}$. Under the null hypothesis of independence, the likelihood is instead $L_0 = (1 - \pi)^{T_{00} + T_{10}} \pi^{T_{01} + T_{11}}$, where $\pi = (T_{01} + T_{11})/T$ and $\pi_{01} = \pi_{11} = \pi$. The test statistic for independence is then given by

$$LR_{\text{ind}} = 2\{\ln L_A - \ln L_0\}, \quad (2.26)$$

which is asymptotically $\chi^2(1)$ distributed. To jointly test for correct unconditional coverage and independence Christoffersen (1998) apply the test statistic

$$L_{cc} = LR_{\text{uc}} + LR_{\text{ind}}, \quad (2.27)$$

which is under the Null asymptotically $\chi^2(2)$ distributed.

Table 2.6 presents the forecasting results. The basic WMSV model shows the overall worst VaR forecasting performance within the range of considered volatility models - the hit-rate amounts to 15% and the LR test statistics on unconditional and conditional coverage are highly significant. The overestimation of coverage may be attributed to the model's deficiency in capturing the leptokurtic distribution of daily asset returns (see the diagnostic test results in Section 2.3.2). Extending the basic WMSV model by Markov switching regimes significantly improves the VaR forecasting results: The hit-rate amounts to 8.78%, which is closest to the 5% level across all considered volatility models. Based on the test results we cannot reject the Null of correct unconditional coverage at the 1% significance level. Compared to the basic WMSV model this improvement of unconditional coverage can be attributed to the mixture effects of the Markov switching process, which induce additional probability mass in the tails of the return distribution. Nevertheless, the joint Null of unconditional coverage and independence in the hit-rate series is rejected for both WMSV specifications. This indicates potential improvements regarding the flexibility of the dynamic specification of the scale matrix S_t (see Eq. 2.5), e.g. by introducing multiplicative volatility components as in Jin and Maheu (2011). All competing volatility models show violations of unconditional coverage, which may be explained by the overall high volatility level in 2008 inducing strong excess kurtosis in the return series. The Null of conditional coverage is rejected for all models, except for the DCC-GARCH(2,1) and the BEKK-GARCH(1,1) model.

Summarizing the results, the MS extension of the basic WMSV model significantly improves the model's in-sample and out-of-sample properties.

2.4 Summary

This chapter proposes a new Markov switching (MS) extension to the basic Wishart MSV (WMSV) model of Philipov and Glickman (2006) and Asai and McAleer (2009). The proposed model allows for particularly flexible (co)variance dynamics including state-dependent shifts in the unconditional mean of (co)variances and correlations as well as state-dependent volatility transmission effects across assets. The MS approach captures sudden changes in the volatility level related to particular events like increasing market uncertainty induced by the 2005 terrorist attacks in London as well as lasting structural changes due to financial crisis e.g. induced by the collapse of the US subprime mortgage market in 2007. Markov switching volatility regimes generate long-memory like persistence patterns which are typical for high-frequency return volatilities.

The WMSV model is applied to daily returns of five European stock indices. Parameter estimates are obtained using Bayesian Monte Carlo Markov Chain (MCMC) methods. The estimation results indicate the presence of a high-volatility and a low-volatility regime where states of high market volatility correspond to increasing market correlations. This indicates the presence of contagion effects in asset returns in the lines of Forbes and Rigobon (2002) as well as vanishing diversification benefits in periods of turmoil. The high-volatility states are accompanied by increasing volatility transmission effects across assets. This indicates volatility contagion, i.e. crisis-related increases in inter-asset volatility dependencies (see e.g. Chiang and Wang, 2011, and Diebold and Yilmaz, 2009). Contagion effects stimulate the international propagation of crisis as e.g. observed for the U.S. subprime crisis, which spread out around the world through various transmission channels.

Model diagnostics show that the MS WMSV model alleviates the shortcoming of the basic WMSV model in accommodating the strong persistence of daily asset return (co)variances. The model prevents the underestimation of (co)variances in periods of high market volatility resulting in an improved model fit to the leptokurtic return distribution. A Value-at-Risk (VaR) forecasting experiment shows that the MS WMSV model outperforms a range of competing volatility models from the literature with respect to unconditional coverage of the 5% VaR level.

2.5 Technical Details

2.5.1 Simulation Based Bayesian Inference

Denote the sample data by $\Xi = \{\xi_1, \dots, \xi_T\}$, the likelihood function by $L(\Xi|\theta)$ and the parameters' joint prior distribution by $\pi(\theta)$. The object of primary interest is the posterior distribution

$$P(\theta|\Xi) \propto L(\Xi|\theta) \pi(\theta) \quad (2.28)$$

and corresponding moments. The Gibbs sampling algorithm generates samples from the posterior distribution, which can then be used to approximate respective moments and to produce density estimates. In order to implement Gibbs sampling the parameter vector θ is partitioned into K blocks $\theta = \{\theta_1, \dots, \theta_K\}$, which are conveniently chosen in order to enable sampling from the respective full conditional distributions. The latter are proportional to the product of the likelihood function and the prior distribution. Hence the full conditional distribution of a single block θ_k

$$p(\theta_k|\theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_K, \Xi) \quad (2.29)$$

can be deduced via isolating the density kernel of θ_k conditional on all remaining blocks and the data Ξ . The Gibbs sampling algorithm now proceeds as follows: Given an initialization $\theta^{(0)}$ the algorithm simulates iteratively for $r = 1, \dots, R$ trajectories $\theta^{(r)}$ from the full conditional distributions

$$\begin{aligned} & p(\theta_1|\theta_2^{(r-1)}, \dots, \theta_K^{(r-1)}, \Xi), \\ & p(\theta_2|\theta_1^{(r)}, \theta_3^{(r-1)}, \dots, \theta_K^{(r-1)}, \Xi), \\ & \vdots \\ & p(\theta_K|\theta_1^{(r)}, \dots, \theta_{K-1}^{(r)}, \Xi). \end{aligned} \quad (2.30)$$

The algorithm thereby generates an irreducible and aperiodic Markov chain which under some mild regularity conditions e.g. given in Chib (2001) converges to the parameters' joint posterior distribution. The Gibbs draws conducted until convergence (so-called burn-in phase) are discarded and only the remaining draws are used for estimation purposes. The prior distribution $\pi(\theta)$ is typically factorized into the product of marginal priors $\pi(\theta) = \pi_{\theta_1} \cdot \pi_{\theta_2} \dots \pi_{\theta_K}$, which are conveniently chosen in order to allow for direct sampling from the full conditional distributions.

Tanner and Wong (1987) introduced data augmentation in order to include latent variables into the parameter vector. For the basic WMSV model the latent variables are given by the covariance matrices $\Sigma_1, \dots, \Sigma_T$. The MS WMSV model further extends the set of latent

variables by including the Markov states s_1, \dots, s_T . The Gibbs sampling algorithm is then applied to generate samples from the augmented posterior distribution $P(\theta, \Sigma_1, \dots, \Sigma_T | \Xi)$ or $P(\theta, \Sigma_1, \dots, \Sigma_T, s_1, \dots, s_T | \Xi)$, respectively.

In many practical situations it is not possible to obtain a closed form solution for the full conditional distribution. For the WMSV model the full conditional distributions of the covariance matrices Σ_t , for $t = 1, \dots, T$, and the parameters ν and d are known up to a multiplicative constant. In such cases the Metropolis-Hastings algorithm can be applied for simulation purposes (see Metropolis et al., 1953, and Hastings, 1970). The algorithm simulates random draws from the target distribution $p(\theta_k | \cdot)$ using an approximate distribution, the so-called proposal distribution $q(\theta_k | \cdot)$, for which random draws are easily available. The algorithm proceeds as follows

1. initialize $\theta_k^{(0)}$.
2. For $z = 1, \dots, Z$
 - a) draw a candidate $\theta_k^{(*)}$ from the proposal density $q(\theta_k | \cdot)$.
 - b) Calculate the ratio

$$\alpha = \frac{p(\theta_k^{(*)} | \cdot) q(\theta_k^{(z-1)} | \cdot)}{q(\theta_k^{(*)} | \cdot) p(\theta_k^{(z-1)} | \cdot)}. \quad (2.31)$$

- c) Set

$$\theta_k^{(z)} = \begin{cases} \theta_k^{(*)}, & \text{with probability } \min(\alpha, 1), \\ \theta_k^{(z-1)}, & \text{otherwise.} \end{cases} \quad (2.32)$$

Under some regularity conditions the sequence $\{\theta_k^{(z)}\}$ converges in distribution to $p(\theta_k | \cdot)$ (see Gelman et al., 2003, and Tsay, 2005). A more general version of the Metropolis-Hastings algorithm allows the proposal density in iteration z to depend on $\theta_k^{(z-1)}$.

The functional forms of the full conditional distributions are illustrated in Sections 2.5.2 and 2.5.3 for the basic WMSV model and the MS WMSV model, respectively. Section 2.5.4 illustrates the particle filter algorithm which is applied in order to obtain filtered volatility and Markov state estimates.

2.5.2 Full Conditional Distributions: Basic WMSV Model

The basic WMSV model is outlined in Eqs. (2.1), (2.2) and (2.5). The joint prior distribution is assumed to factor into the product of marginal prior distributions given by

1. a Wishart prior $\pi_{A^{-1}}(Q_0, \gamma_0)$ for A^{-1} with scale matrix Q_0 and d.o.f. parameter γ_0 ;
2. a uniform prior $\pi_d(0, 1)$ on $[0, 1]$ for d ;
3. a gamma prior $\pi_\nu(\alpha_0, \beta_0)$ for $\nu - k$ with shape parameter α_0 and scale parameter β_0 .

Denoting the augmented parameter vector by θ^{aug} we obtain

$$P(\theta^{\text{aug}}|\Xi) \propto \prod_{t=1}^T f(\xi_t|\Sigma_t^{-1}) \times f(\Sigma_t^{-1}|\Sigma_{t-1}^{-1}, \theta) \\ \times \pi_{A-1}(Q_0, \gamma_0) \times \pi_d(0, 1) \times \pi_\nu(\alpha_0, \beta_0). \quad (2.33)$$

In order to simplify notation, the vector of remaining model parameters for each parameter block is denoted by θ_-^{aug} . The full conditional distributions are obtained as follows:

Full conditional distribution of Σ_t^{-1} :

For notational convenience suppressing dependence on model parameters, the kernel of the full conditional distribution of Σ_t^{-1} is obtained as

$$p(\Sigma_t^{-1}|\theta_-^{\text{aug}}) \propto f(\xi_t|\Sigma_t^{-1}) \times f(\Sigma_t^{-1}|\Sigma_{t-1}^{-1}) \times f(\Sigma_{t+1}^{-1}|\Sigma_t^{-1}) \\ \propto |\Sigma_t^{-1}|^{(\nu-k-d\nu)/2} \times \exp\{-0.5 \text{tr}[(S_t^{-1} + \xi_t\xi_t')\Sigma_t^{-1}]\} \\ \times \exp\{-0.5 \text{tr}[S_{t+1}^{-1}\Sigma_{t+1}^{-1}]\} \\ \propto \mathcal{W}_k^\kappa(\Sigma_t^{-1}|\tilde{\nu}, \tilde{S}_t) \times f(\Sigma_t^{-1}), \quad (2.34)$$

where $\mathcal{W}_k^\kappa(\Sigma_t^{-1}|\cdot)$ denotes a Wishart kernel in Σ_t^{-1} and

$$\tilde{\nu} = \nu(1-d) + 1, \quad (2.35)$$

$$\tilde{S}_t = (S_t^{-1} + \xi_t\xi_t')^{-1}, \quad (2.36)$$

$$f(\Sigma_t^{-1}) = \exp\{-0.5 \text{tr}[S_{t+1}^{-1}\Sigma_{t+1}^{-1}]\}, \quad (2.37)$$

$$S_t = \Sigma_t^{-d/2} A \Sigma_t^{-d/2}. \quad (2.38)$$

The full conditional distribution of Σ_t^{-1} is known up to an integrating constant and the Metropolis-Hastings (MH) algorithm is applied in order to obtain samples from $p(\Sigma_t^{-1}|\theta_-^{\text{aug}})$. The proposal density is given by $\mathcal{W}_k(\nu, \tilde{S}_t)$. For the empirical application the number of Metropolis-Hastings iterations is set to $Z = 1$, i.e. the Metropolis-Hastings chain converges with the Gibbs sampler.

Full conditional distribution of A^{-1} :

The full conditional distribution of A^{-1} is Wishart since

$$\begin{aligned}
 p(A^{-1}|\theta_-^{\text{aug}}) &\propto \pi_{A^{-1}}(Q_0, \gamma_0) \prod_{t=1}^T f(\Sigma_t^{-1}|\Sigma_{t-1}^{-1}) \\
 &\propto \pi_{A^{-1}}(Q_0, \gamma_0) |A^{-1}|^{(T\nu)/2} \\
 &\quad \times \exp \left\{ -0.5 \text{tr} \left[\nu \sum_{t=1}^T \Sigma_{t-1}^{d/2} \Sigma_t^{-1} \Sigma_{t-1}^{d/2} A^{-1} \right] \right\} \\
 &\propto \pi_{A^{-1}}(Q_0, \gamma_0) \times \mathcal{W}_k^{\nu}(A^{-1}|\gamma, U),
 \end{aligned} \tag{2.39}$$

where

$$U^{-1} = \nu \sum_{t=1}^T \Sigma_{t-1}^{d/2} \Sigma_t^{-1} \Sigma_{t-1}^{d/2}, \tag{2.40}$$

$$\gamma = T\nu + k + 1, \tag{2.41}$$

and hence

$$\begin{aligned}
 p(A^{-1}|\theta_-^{\text{aug}}) &\propto \pi_{A^{-1}}(Q_0, \gamma_0) \times \mathcal{W}_k^{\nu}(A^{-1}|\gamma, U) \\
 &\propto |A^{-1}|^{(\gamma_0 + \gamma - 2k - 2)/2} \exp\{-0.5 \text{tr}[(Q_0^{-1} + U^{-1})A^{-1}]\}.
 \end{aligned} \tag{2.42}$$

Therefore

$$A^{-1}|\theta_-^{\text{aug}} \sim \mathcal{W}_k(\tilde{\gamma}, \tilde{U}), \tag{2.43}$$

where

$$\tilde{U}^{-1} = Q_0^{-1} + U^{-1}, \tag{2.44}$$

$$\tilde{\gamma} = \gamma_0 + \gamma - k - 1. \tag{2.45}$$

Full conditional distribution of ν and d :

The full conditional distributions of the parameters ν and d are not obtained in closed form and the Metropolis-Hastings algorithm is used for simulation issues. Since $\nu > k$ and $d \in (0, 1)$, truncated normal proposal densities are applied where mean and variance are given by the optimum and the corresponding Hessian obtained after numerically optimizing the posterior distribution's density kernel.

The kernel of the full conditional distribution of d is obtained as

$$\begin{aligned}
 p(d|\theta_-^{\text{aug}}) &\propto \pi_d(0, 1) \prod_{t=1}^T |\Sigma_{t-1}^{-1}|^{-d\nu/2} \\
 &\quad \times \exp \left\{ -0.5 \operatorname{tr} \left[\left((1/\nu) \Sigma_{t-1}^{-d/2} A \Sigma_{t-1}^{-d/2} \right)^{-1} \Sigma_t^{-1} \right] \right\} \\
 &\propto \exp \left\{ d\psi - 0.5 \operatorname{tr} [Q(d)A^{-1}] \right\},
 \end{aligned} \tag{2.46}$$

where

$$\psi = -\frac{\nu}{2} \sum_{t=1}^T \ln(|\Sigma_{t-1}^{-1}|), \tag{2.47}$$

$$Q(d) = \sum_{t=1}^T \nu \Sigma_{t-1}^{d/2} \Sigma_t^{-1} \Sigma_{t-1}^{d/2}. \tag{2.48}$$

The kernel of the full conditional distribution of ν is obtained as

$$\begin{aligned}
 p(\nu|\theta^{\text{aug}}) &\propto \pi_\nu(\alpha_0, \beta_0) \times \prod_{t=1}^T f(\Sigma_t^{-1}|\Sigma_{t-1}^{-1}) \\
 &\propto \exp\{(\alpha - 1) \ln(\nu - k) - \beta(\nu - k)\} \\
 &\quad \times \left(\frac{|\nu A^{-1}|^{\nu/2}}{2^{\nu k/2} \prod_{j=1}^k \Gamma((\nu - j + 1)/2)} \right)^T \\
 &\quad \times \prod_{t=1}^T |Q_t^{-1}|^{\nu/2} \exp\{-0.5 \operatorname{tr} [Q^{-1}A^{-1}]\},
 \end{aligned} \tag{2.49}$$

where

$$Q_t^{-1} = \Sigma_{t-1}^{d/2} \Sigma_t^{-1} \Sigma_{t-1}^{d/2}, \tag{2.50}$$

$$Q^{-1} = \nu \sum_{t=1}^T \Sigma_{t-1}^{d/2} \Sigma_t^{-1} \Sigma_{t-1}^{d/2}. \tag{2.51}$$

2.5.3 Full Conditional Distributions: Markov Switching MWSV Model

The MS WMSV model is outlined in Eqs. (2.10), (2.11) and (2.12). The joint prior distribution is again assumed to factor into the product of marginal prior distributions. Given the state sequence $s = (s_1, s_2, \dots, s_T)'$, the derivation of the full conditional distributions for Σ_t^{-1} , A_1 , A_2 , ν and d is analogous to the illustrations of the previous section, except that we have to condition on $A_{s_t} \forall t = 1, \dots, T$ instead of A .

Full conditional distribution of $s = (s_1, s_2, \dots, s_T)'$:

Denoting $\underline{\Sigma}_t^{-1} = \{\Sigma_1^{-1}, \dots, \Sigma_t^{-1}\}$ and exploiting the Markov property of s_t , the full conditional density of the state vector s can be factorized as

$$\begin{aligned} p(s|\theta_-^{\text{aug}}) &= P(s|\underline{\Sigma}_T^{-1}, \theta) \\ &= P(s_T|\underline{\Sigma}_T^{-1}, \theta) \times P(s_{T-1}|s_T, \underline{\Sigma}_T^{-1}, \theta) \times \dots \times P(s_1|s_2, \underline{\Sigma}_T^{-1}, \theta) \\ &= P(s_T|\underline{\Sigma}_T^{-1}, \theta) \times P(s_{T-1}|s_T, \underline{\Sigma}_{T-1}^{-1}, \theta) \times \dots \times P(s_1|s_2, \underline{\Sigma}_1^{-1}, \theta). \end{aligned} \quad (2.52)$$

The conditional probabilities

$$P(s_t|s_{t+1}, \underline{\Sigma}_t^{-1}, \theta) = \frac{P(s_{t+1}|s_t) \times P(s_t|\underline{\Sigma}_t^{-1}, \theta)}{P(s_{t+1}|\underline{\Sigma}_t^{-1}, \theta)} \quad (2.53)$$

are obtained by the ‘‘Hamilton filter’’ which - given a starting value for $P(s_0|\underline{\Sigma}_0^{-1}, \theta)$ (e.g. stationary probabilities, see Hamilton, 1994, p. 683) - proceeds recursively in five steps $\forall t \in \{1, \dots, T\}$:

$$I \quad P(s_t, s_{t-1}|\underline{\Sigma}_{t-1}^{-1}, \theta) = P(s_t|s_{t-1}) \times P(s_{t-1}|\underline{\Sigma}_{t-1}^{-1}, \theta) \quad (2.54)$$

$$II \quad P(s_t|\underline{\Sigma}_{t-1}^{-1}, \theta) = \sum_{s_{t-1}} P(s_t, s_{t-1}|\underline{\Sigma}_{t-1}^{-1}, \theta) \quad (2.55)$$

$$III \quad f(\Sigma_t^{-1}, s_t|\underline{\Sigma}_{t-1}^{-1}, \theta) = f(\Sigma_t^{-1}|s_t, \Sigma_{t-1}^{-1}, \theta) \times P(s_t|\underline{\Sigma}_{t-1}^{-1}, \theta) \quad (2.56)$$

$$IV \quad f(\Sigma_t^{-1}|\underline{\Sigma}_{t-1}^{-1}, \theta) = \sum_{s_t} f(\Sigma_t^{-1}, s_t|\underline{\Sigma}_{t-1}^{-1}, \theta) \quad (2.57)$$

$$V \quad P(s_t|\underline{\Sigma}_t^{-1}, \theta) = \frac{f(\Sigma_t^{-1}, s_t|\underline{\Sigma}_{t-1}^{-1}, \theta)}{f(\Sigma_t^{-1}|\underline{\Sigma}_{t-1}^{-1}, \theta)}. \quad (2.58)$$

The whole state sequence $s = (s_1, s_2, \dots, s_T)'$ can then be sampled backward recursively based on Eq. (2.52).

Full conditional distributions of e_1 and e_2 :

Using beta prior distributions $\pi_{e_i}(\alpha_{i,0}, \beta_{i,0})$, $i \in \{1, 2\}$, the kernel of the full conditional

distribution of e_i is obtained as

$$\begin{aligned} p(e_i|\text{rest}) &\propto \pi_{e_i}(\alpha_{i,0}, \beta_{i,0}) \times \prod_{j=1}^{g_i} e_i \prod_{j=1}^{h_i} (1 - e_i) \\ &\propto e_i^{\alpha_{i,0}-1} (1 - e_i)^{\beta_{i,0}-1} \times e_i^{g_i} (1 - e_i)^{h_i}, \end{aligned} \quad (2.59)$$

where g_i denotes the number of switches from state i to state $i-$ (not state i) and h_i denotes the number of periods where the state does not change. The full conditional distribution of e_i is therefore beta with parameters $\alpha_i = \alpha_{i,0} + g_i$ and $\beta_i = \beta_{i,0} + h_i$, $i \in \{1, 2\}$.

2.5.4 A Particle Filter Algorithm for the Basic WMSV and the MS WMSV Model

This section illustrates the basic particle filter algorithm of Pitt and Shephard (1999). Considering the basic WMSV model, the return vectors ξ_t , $t = 1, \dots, T$, are conditionally independent given the unobserved Markovian states $\{\sigma_t = \text{vech}(\Sigma_t)\}_{t=1}^T$. A particle filtering algorithm uses simulation to carry out on-line filtering. The filter requires known parametric forms of the measurement density $f(\xi_t|\sigma_t)$ and the transition density $f(\sigma_t|\sigma_{t-1})$. For the basic WMSV model these densities are given by Eqs. (2.1) and (2.2). Furthermore, it must be possible to simulate from the transition density.

In general, filtering proceeds in two basic steps: The first step consists in propagating the period- t filtering density $f(\sigma_t|\mathcal{F}_t)$ into the future via the transition density $f(\sigma_{t+1}|\sigma_t)$. The resulting prediction density is then given by

$$f(\sigma_{t+1}|\mathcal{F}_t) = \int f(\sigma_{t+1}|\sigma_t) f(\sigma_t|\mathcal{F}_t) d\sigma_t. \quad (2.60)$$

The filtering density in period $t + 1$ is then obtained as

$$f(\sigma_{t+1}|\mathcal{F}_{t+1}) = \frac{f(\xi_{t+1}|\sigma_{t+1}) f(\sigma_{t+1}|\mathcal{F}_t)}{f(\xi_{t+1}|\mathcal{F}_t)}, \quad (2.61)$$

where $f(\xi_{t+1}|\mathcal{F}_t) = \int f(\xi_{t+1}|\sigma_{t+1}) f(\sigma_{t+1}|\mathcal{F}_t) d\sigma_{t+1}$.

For the WMSV model the involved integrals are analytically intractable and their evaluation requires simulation-based particle filtering techniques. The standard particle filter proceeds in 3 steps:

1. Initialize M draws $\sigma_1^{(1)}, \dots, \sigma_1^{(M)}$ from $f(\sigma_1|\mathcal{F}_0)$ e.g. via sampling from a Wishart distribution centered at the sample covariance matrix of the data.
2. For $t = 1, \dots, T$
 - a) Simulate from $f(\sigma_t|\mathcal{F}_t)$ according to Eq. (2.61) using the sampling/importance re-sampling method of Rubin (1987): Re-sample the M draws from $f(\sigma_t|\mathcal{F}_{t-1})$

using weights

$$\pi_j = \frac{\omega_j}{\sum_{i=1}^M \omega_i}, \quad \omega_j = f(\xi_t | \sigma_t^{(j)}), \quad j = 1, \dots, M. \quad (2.62)$$

- b) Iterate the draws from $f(\sigma_t | \mathcal{F}_t)$ obtained in the previous step forward using the transition density $f(\sigma_{t+1} | \sigma_t)$. The resulting draws represent a sample from $f(\sigma_{t+1} | \mathcal{F}_t)$ according to Eq. (2.60).

The standard particle filter algorithm illustrated above is easily extended to filtering within the MS WMSV framework. The state vector is then augmented to include the additional state s_t , where the according transition density is given by Eq. (2.10).

Table 2.1: Descriptive Statistics for the Daily Index Log Returns

Statistic	France	Germany	Italy	Switzerland	UK
Sample correlation	1.00	0.72	0.91	0.89	0.90
	.	1.00	0.68	0.66	0.69
	.	.	1.00	0.83	0.89
	.	.	.	1.00	0.84
	1.00
Mean	0.00	0.00	0.00	0.00	0.00
Std. dev.	1.26	1.34	1.13	1.11	1.18
Kurtosis	12.48	24.67	14.47	11.84	12.86
Skewness	-0.03	0.80	-0.23	-0.09	-0.47
Minimum	-8.35	-8.64	-9.01	-7.50	-8.54
Maximum	9.60	16.24	9.19	9.68	8.34
$LB_r(10)$	4.98	1.34	2.91	7.91	1.56
$LB_{r,2}(30)$	2293.70*	1137.94*	2409.56*	2341.55*	2583.31*

$LB_r(10)$: Ljung-Box test statistic for the return series at 10 lags. $LB_{r,2}(30)$: Ljung-Box test statistic for the squared return series at 30 lags. The number of observations for each series is 1,565.

*: Significant at the 1% level.

Table 2.2: Estimation Results

		Basic WMSV Model																
		A^{-1}															ν	d
Estimate		a_{11}	a_{21}	a_{31}	a_{41}	a_{51}	a_{22}	a_{32}	a_{42}	a_{52}	a_{33}	a_{43}	a_{53}	a_{44}	a_{54}	a_{55}		
MC Std. dev. $\times 10^2$		0.88	0.02	0.04	0.04	0.04	0.92	0.02	0.01	0.02	0.88	0.02	0.03	0.88	0.03	0.88	69.05	0.95
Post. Std. Dev. $\times 10^2$		0.06	0.02	0.02	0.03	0.03	0.04	0.02	0.03	0.02	0.06	0.02	0.03	0.05	0.02	0.05	3.34	0.03
$q_{0.025}$		0.57	0.32	0.34	0.37	0.35	0.47	0.32	0.31	0.31	0.57	0.31	0.33	0.54	0.33	0.56	48.09	0.27
$q_{0.975}$		0.87	0.02	0.03	0.03	0.03	0.91	0.01	0.01	0.01	0.86	0.02	0.02	0.87	0.02	0.87	67.88	0.94
		0.89	0.03	0.05	0.05	0.04	0.93	0.03	0.02	0.02	0.89	0.03	0.04	0.89	0.03	0.89	69.91	0.95
		MS WMSV Model															ν	d
		A_1^{-1}															ν	d
Estimate		a_{11}	a_{21}	a_{31}	a_{41}	a_{51}	a_{22}	a_{32}	a_{42}	a_{52}	a_{33}	a_{43}	a_{53}	a_{44}	a_{54}	a_{55}		
MC Std. dev. $\times 10^2$		0.75	0.05	0.07	0.07	0.07	0.84	0.04	0.03	0.04	0.75	0.04	0.05	0.76	0.05	0.75	80.12	0.89
Post. Std. Dev. $\times 10^2$		0.18	0.08	0.10	0.07	0.08	0.21	0.06	0.08	0.11	0.25	0.08	0.14	0.19	0.09	0.19	1.17	0.17
$q_{0.025}$		0.90	0.46	0.51	0.44	0.46	0.92	0.43	0.43	0.50	1.10	0.46	0.60	0.96	0.49	0.94	49.41	0.61
$q_{0.975}$		0.74	0.04	0.06	0.06	0.06	0.82	0.03	0.02	0.03	0.73	0.04	0.04	0.74	0.04	0.73	79.16	0.88
		0.77	0.06	0.08	0.08	0.08	0.86	0.05	0.03	0.05	0.77	0.05	0.06	0.78	0.06	0.76	81.09	0.90
		A_2^{-1}															e_1	e_2
Estimate		a_{11}	a_{21}	a_{31}	a_{41}	a_{51}	a_{22}	a_{32}	a_{42}	a_{52}	a_{33}	a_{43}	a_{53}	a_{44}	a_{54}	a_{55}		
MC Std. dev. $\times 10^2$		1.21	0.11	0.20	0.19	0.19	1.30	0.11	0.11	0.10	1.17	0.10	0.19	1.24	0.14	1.27	0.08	0.40
Post. Std. Dev. $\times 10^2$		0.49	0.46	0.22	0.56	0.67	0.66	0.45	0.49	0.57	0.40	0.51	0.31	0.77	0.82	0.46	0.08	0.04
$q_{0.025}$		3.03	2.46	2.01	2.72	3.20	6.31	2.48	2.57	2.77	2.77	2.63	2.07	3.74	3.65	3.03	0.70	0.98
$q_{0.975}$		1.15	0.07	0.16	0.14	0.13	1.23	0.06	0.06	0.05	1.11	0.04	0.15	1.16	0.07	1.21	0.07	0.38
		1.26	0.17	0.24	0.24	0.25	1.37	0.16	0.16	0.16	1.22	0.15	0.23	1.31	0.20	1.33	0.10	0.42

95% a posteriori high density region: $[q_{0.025}; q_{0.975}]$. Basic WMSV Model: Burn-in: 15,000; Gibbs sequences: 50,000; Gamma prior for ν implies $E[\nu] = 70$, $\sqrt{\text{Var}[\nu]} = 10$; Wishart prior for A^{-1} ; scale matrix $Q_0 = I_5$, d.o.f. $\gamma_0 = 6$. MS WMSV Model: Burn-in: 20,000; Gibbs sequences: 50,000; Gamma prior for ν implies $E[\nu] = 80$, $\sqrt{\text{Var}[\nu]} = 10$; Wishart prior for A_1^{-1} and A_2^{-1} ; scale matrix $Q_0 = I_5$, d.o.f. $\gamma_0 = 6$. Beta prior for e_1 implies $E[e_1] = 0.09$, $\sqrt{\text{Var}[e_1]} = 0.1$. Beta prior for e_2 implies $E[e_2] = 0.4$, $\sqrt{\text{Var}[e_2]} = 0.1$.

Table 2.3: Model Diagnostics Results: Pearson Residuals

Ljung-Box test statistics for residual cross-products, 50 lags														
$e_1^* \times e_1^*$	$e_1^* \times e_2^*$	$e_1^* \times e_3^*$	$e_1^* \times e_4^*$	$e_1^* \times e_5^*$	$e_2^* \times e_2^*$	$e_2^* \times e_3^*$	$e_2^* \times e_4^*$	$e_2^* \times e_5^*$	$e_3^* \times e_3^*$	$e_3^* \times e_4^*$	$e_3^* \times e_5^*$	$e_4^* \times e_4^*$	$e_4^* \times e_5^*$	$e_5^* \times e_5^*$
2653.21*	1398.42*	2594.82*	2571.13*	2858.81*	1220.47*	1229.21*	1353.14*	1804.84*	2721.31*	2543.02*	2903.62*	2555.77*	3031.24*	3016.18*
Data														
Basic WMSV model														
315.85*	86.33*	84.84*	137.37*	213.15*	262.70*	149.52*	146.56*	203.59*	71.32	140.14*	76.58*	281.41*	115.75*	308.12*
MS WMSV model														
75.19	37.91	56.84	55.53	88.44*	44.21	64.66	58.00	48.49	61.59	70.53	63.49	41.05	59.94	82.56*

The Pearson residual series are generated by particle filter techniques (see the Appendix). The particle filtering is based on 100,000 particles. *: Significant at the 1% level.

Table 2.4: Distributional Diagnostics

	France	Germany	Italy	Switzerland	UK
Basic WMSV model					
Mean	0.01	0.02	0.03	0.02	0.02
Std. Dev.	0.99	0.99	0.99	0.99	0.99
Kurtosis	3.72	3.94	3.81	3.73	3.62
Skewness	-0.27	-0.23	-0.45	-0.27	-0.33
JB-Test	53.89*	72.14*	96.20*	54.46*	54.31*
MS WMSV model					
Mean	0.04	0.04	0.06	0.04	0.04
Std. Dev.	1.06	1.04	1.06	1.06	1.06
Kurtosis	2.95	3.05	3.02	2.92	2.95
Skewness	-0.17	-0.15	-0.29	-0.19	-0.21
JB-Test	8.11	6.29	22.25*	9.83*	12.18*

Std. Dev.: Standard Deviation. JB-Test: Jarque-Bera test. The residual series are generated by particle filter techniques (see the Appendix). The particle filtering is based on 100,000 particles. *: Significant at the 1% level.

Table 2.5: Simulation Results for the Fitted MS WMSV Model

	France	Germany	Italy	Switzerland	UK	France	Germany	Italy	Switzerland	UK
	$E[\Sigma_t], \mathbf{A}_1$					$E[\Sigma_t], \mathbf{A}_2$				
France	0.39	0.29	0.28	0.28	0.28	84.62	63.07	71.57	71.67	81.99
Germany	0.29	0.51	0.24	0.21	0.24	63.07	58.40	54.34	54.83	61.77
Italy	0.28	0.24	0.31	0.23	0.24	71.57	54.34	63.69	60.11	71.30
Switzerland	0.28	0.21	0.23	0.32	0.24	71.67	54.83	60.11	65.63	69.77
UK	0.28	0.24	0.24	0.24	0.31	81.99	61.77	71.30	69.77	85.22
	$E[\text{Corr}_t], \mathbf{A}_1$					$E[\text{Corr}_t], \mathbf{A}_2$				
France	1.00	0.64	0.82	0.80	0.82	1.00	0.89	0.98	0.96	0.97
Germany	0.64	1.00	0.59	0.52	0.60	0.89	1.00	0.89	0.88	0.87
Italy	0.82	0.59	1.00	0.72	0.77	0.98	0.89	1.00	0.93	0.97
Switzerland	0.80	0.52	0.72	1.00	0.75	0.96	0.88	0.93	1.00	0.93
UK	0.82	0.60	0.77	0.75	1.00	0.97	0.87	0.97	0.93	1.00
	$\text{Corr}[\sigma_{ii,t-1}, \sigma_{jj,t}], \mathbf{A}_1$					$\text{Corr}[\sigma_{ii,t-1}, \sigma_{jj,t}], \mathbf{A}_2$				
France	0.89	0.34	0.53	0.51	0.50	0.89	0.66	0.80	0.78	0.78
Germany	0.34	0.89	0.28	0.23	0.27	0.66	0.89	0.65	0.62	0.63
Italy	0.53	0.28	0.89	0.41	0.47	0.80	0.65	0.89	0.70	0.78
Switzerland	0.51	0.22	0.41	0.89	0.43	0.77	0.62	0.70	0.89	0.71
UK	0.50	0.27	0.47	0.43	0.89	0.77	0.63	0.78	0.71	0.89

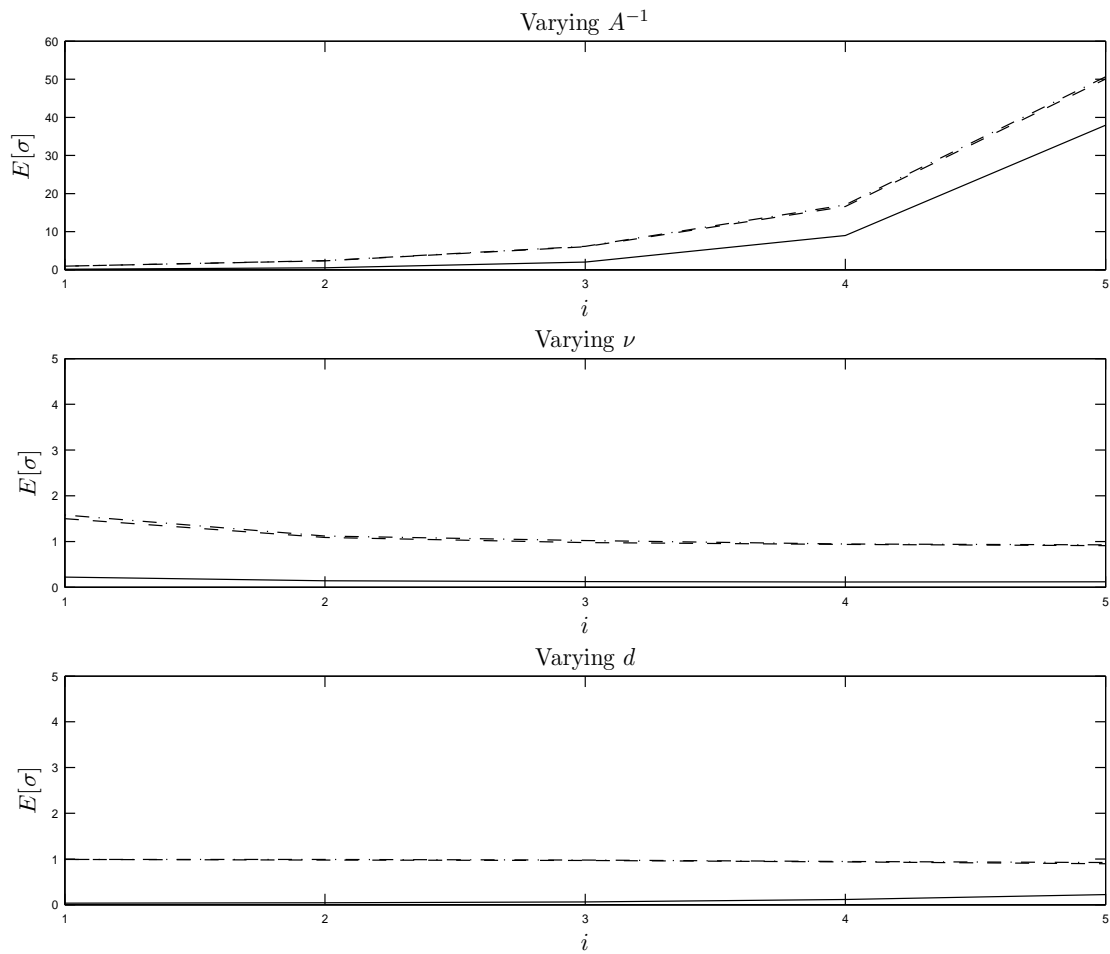
Simulation sample size: $T = 20,000$. Corr_t denotes the correlation matrix implied by Σ_t . $\text{Corr}[\sigma_{ii,t-1}, \sigma_{jj,t}]$: i is the row-index and j is the column-index of the respective panel. All parameters are set to their point estimates under the MS framework.

Table 2.6: VaR Forecasting Results

Model	(p, q)	5% VaR			
		Hit-Rate	LR_{uc}	LR_{ind}	LR_{cc}
DCC-GARCH	(2, 1)	0.0916	0.0054	0.3277	0.0129
CCC-GARCH	(1, 1)	0.0992	0.0012	0.2229	0.0025
BEKK-GARCH	(1, 1)	0.0916	0.0054	0.8804	0.0206
D-BEKK-GARCH	(1, 1)	0.1107	< 0.0001	0.8929	< 0.0001
EWMA		0.0954	0.0026	0.2717	0.0058
WMSV		0.1527	< 0.0001	0.2921	< 0.0001
MS WMSV		0.0878	0.0108	0.0352	0.0042

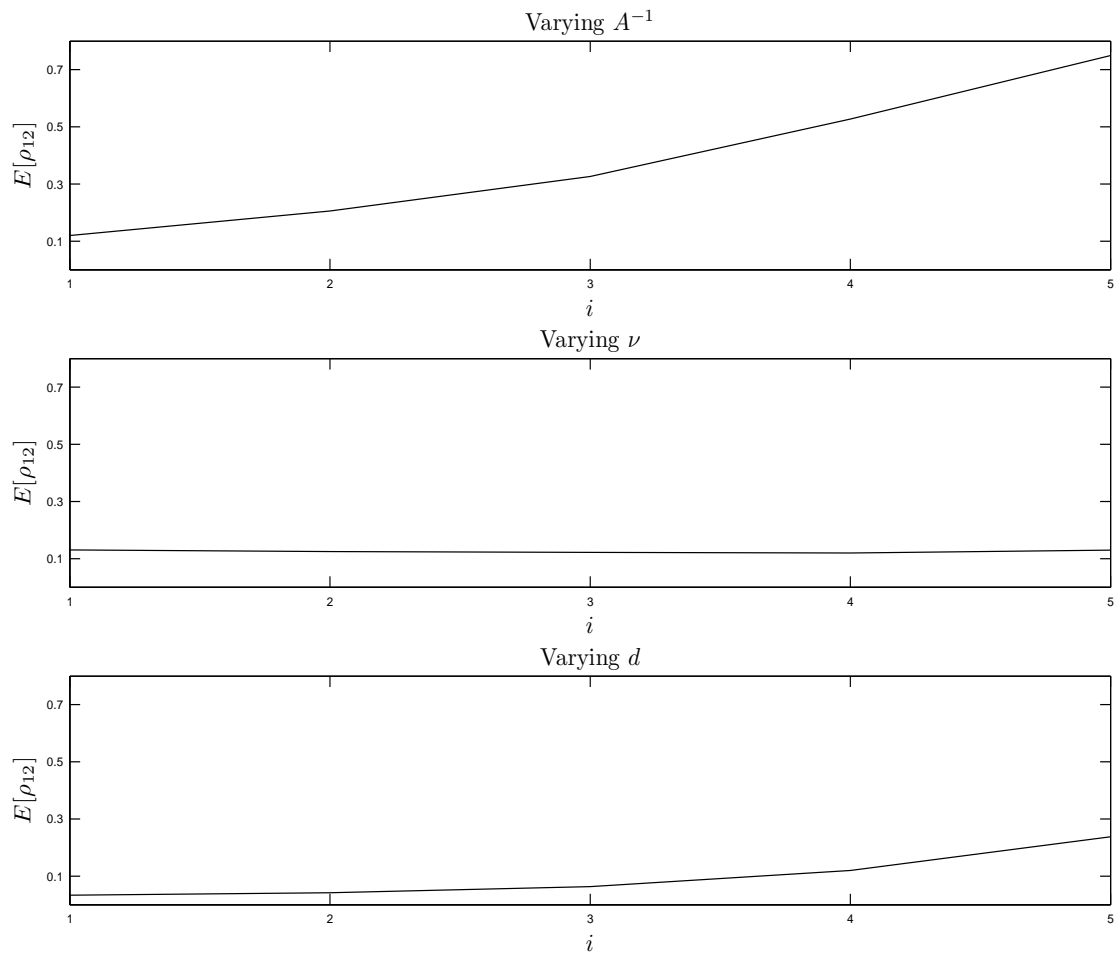
The table reports hit-rates and p -values for the likelihood ratio tests of unconditional coverage, independence, and conditional coverage of the 5% VaR level. If model orders are quoted, models up to order (3, 3) have been estimated and the presentation is limited to the best performing models according to the hit-rate criterion. D-BEKK-GARCH: Diagonal BEKK-GARCH.

Figure 2.1: Simulated Means of Covariances



i is the index on the respective parameter set. Simulation sample size: $T = 20,000$. Dashed line: σ_{11} ; solid line: σ_{12} ; dashdotted line: σ_{22} . All remaining model parameters are kept constant at $\text{vech}(A^{-1}) = (0.96, 0.02, 0.96)'$, $\nu = 80$ and $d = 0.8$, respectively.

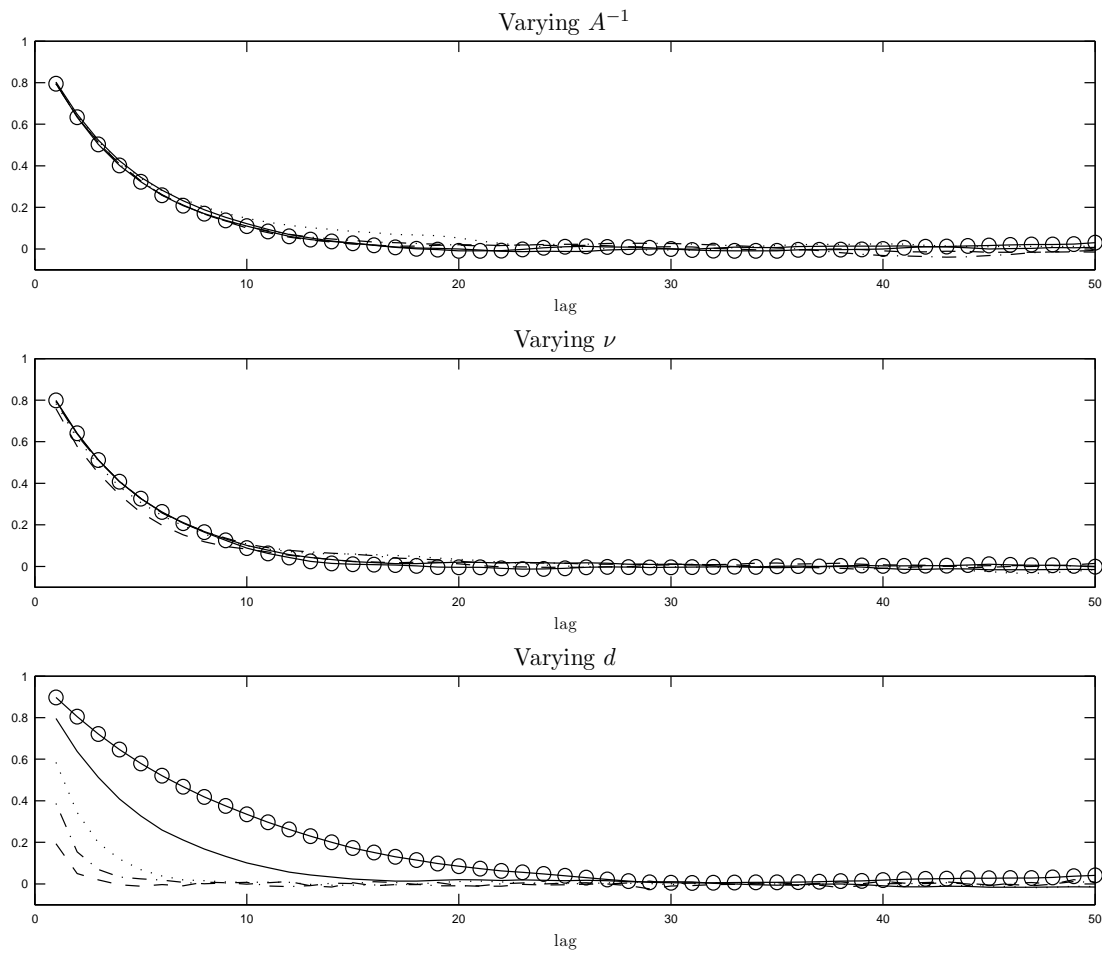
Figure 2.2: Simulated Means of Correlation



i is the index on the respective parameter set. Simulation sample size: $T = 20,000$. All remaining model parameters are kept constant at $\text{vech}(A^{-1}) = (0.96, 0.02, 0.96)'$, $\nu = 80$ and $d = 0.8$, respectively.

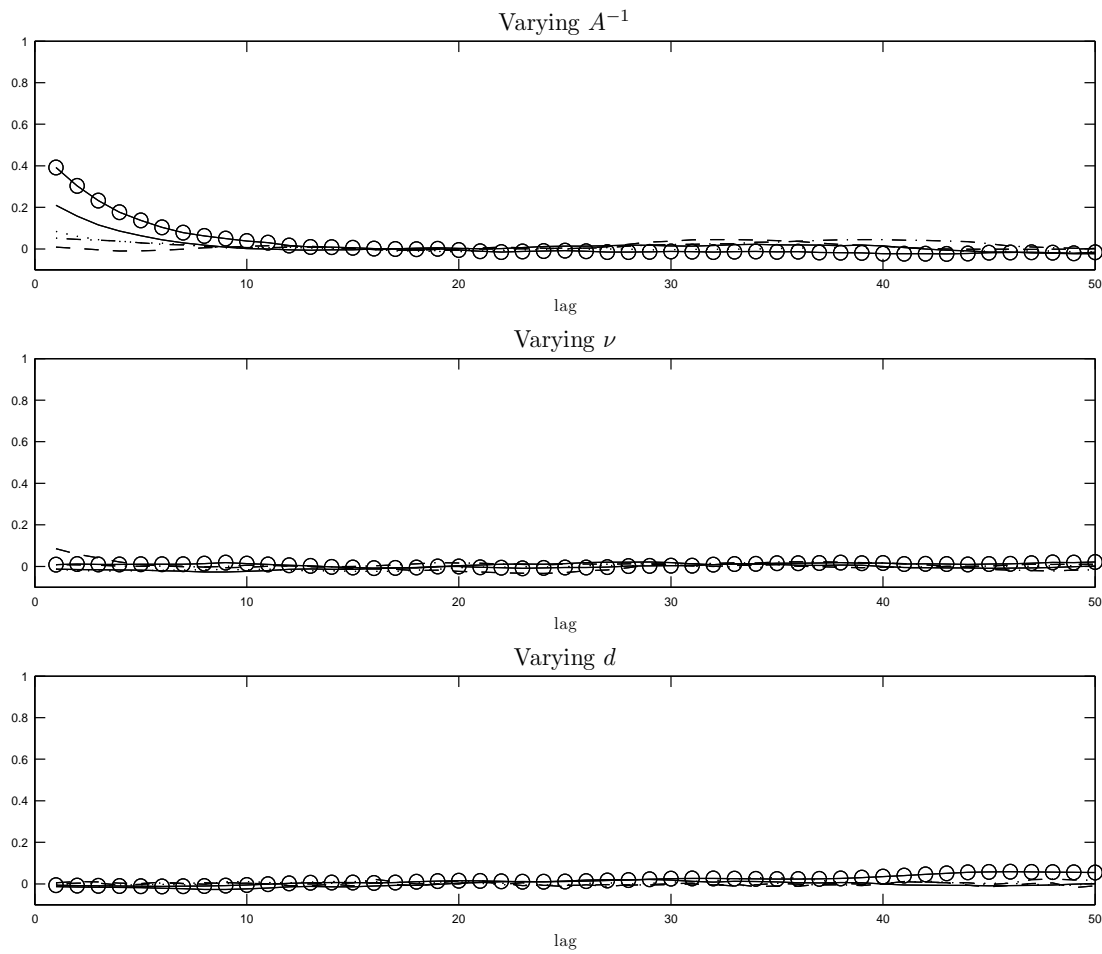
$$\rho_{12} = \sigma_{12} / \sqrt{\sigma_{11}\sigma_{22}}.$$

Figure 2.3: Simulated Autocorrelation Functions for σ_{11}



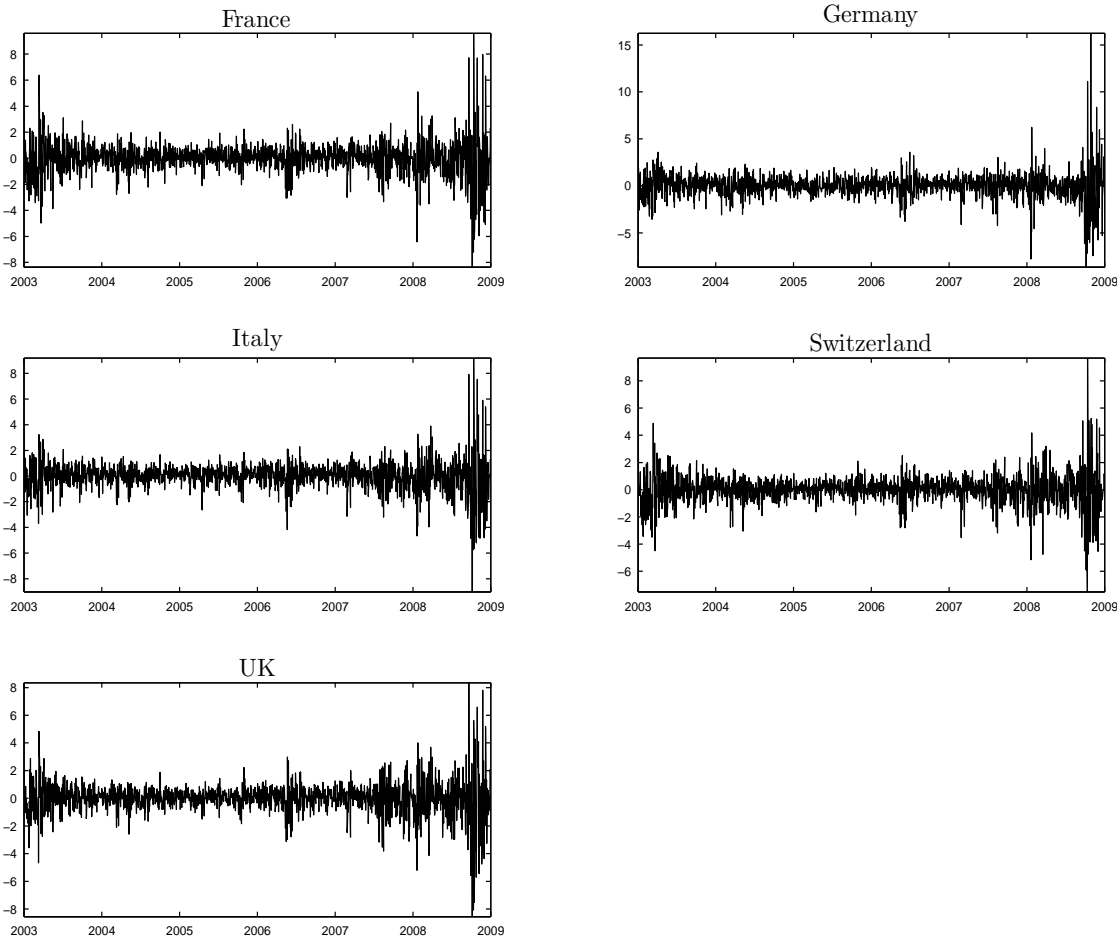
Simulation sample size: $T = 20,000$. i is the index on the respective parameter set. Dashed line: $i = 1$; dashdotted line: $i = 2$; dotted line: $i = 3$; solid line: $i = 4$; \circ : $i = 5$. All remaining model parameters are kept constant at $\text{vech}(A^{-1}) = (0.96, 0.02, 0.96)'$, $\nu = 80$ and $d = 0.8$, respectively.

Figure 2.4: Simulated Cross-Correlation Functions



Depicted are simulated cross-correlations $\text{Corr}[\sigma_{11,t}, \sigma_{22,t-q}]$ for $q = 1, \dots, 50$. Simulation sample size: $T = 20,000$. i is the index on the respective parameter set. Dashed line: $i = 1$; dashdotted line: $i = 2$; dotted line: $i = 3$; solid line: $i = 4$; \circ : $i = 5$. All remaining model parameters are kept constant at $\text{vech}(A^{-1}) = (0.96, 0.02, 0.96)'$, $\nu = 80$ and $d = 0.8$, respectively.

Figure 2.5: Return Series



Log-returns of Datastream DS market indices. The number of observations for each series is 1565.

Figure 2.6: Smoothed Volatility Estimates and Corresponding Return Series: Basic WMSV Model

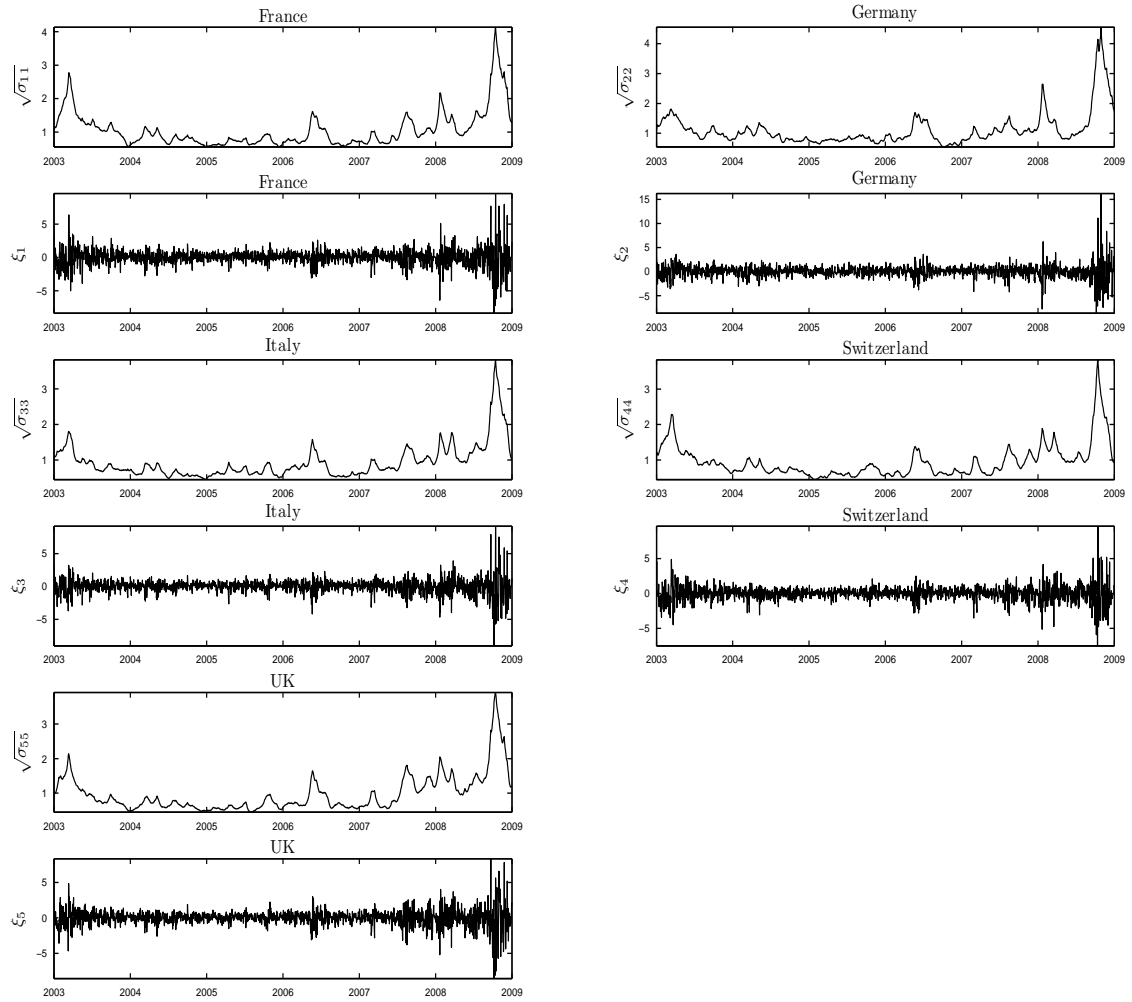


Figure 2.7: Smoothed Correlation Estimates: Basic WMSV Model

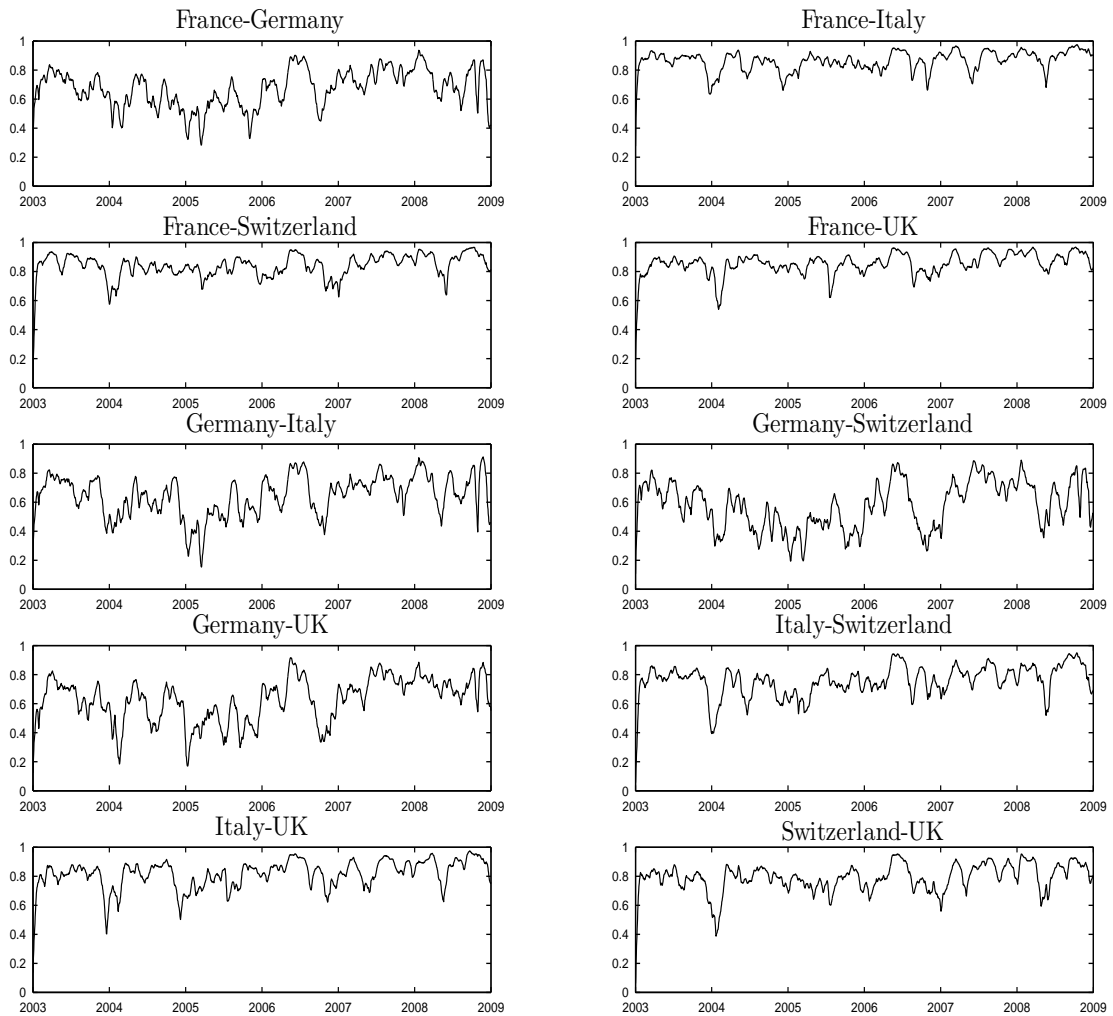
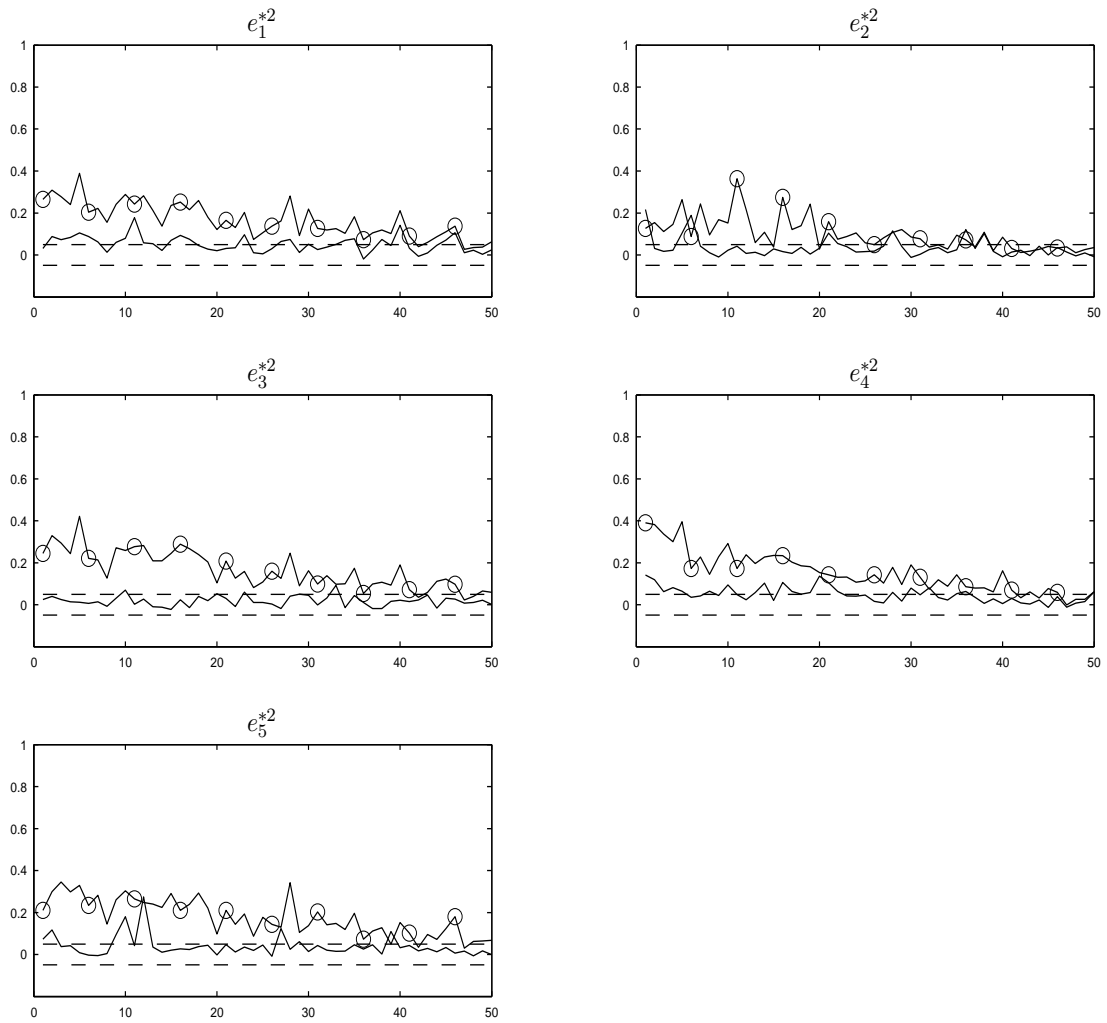
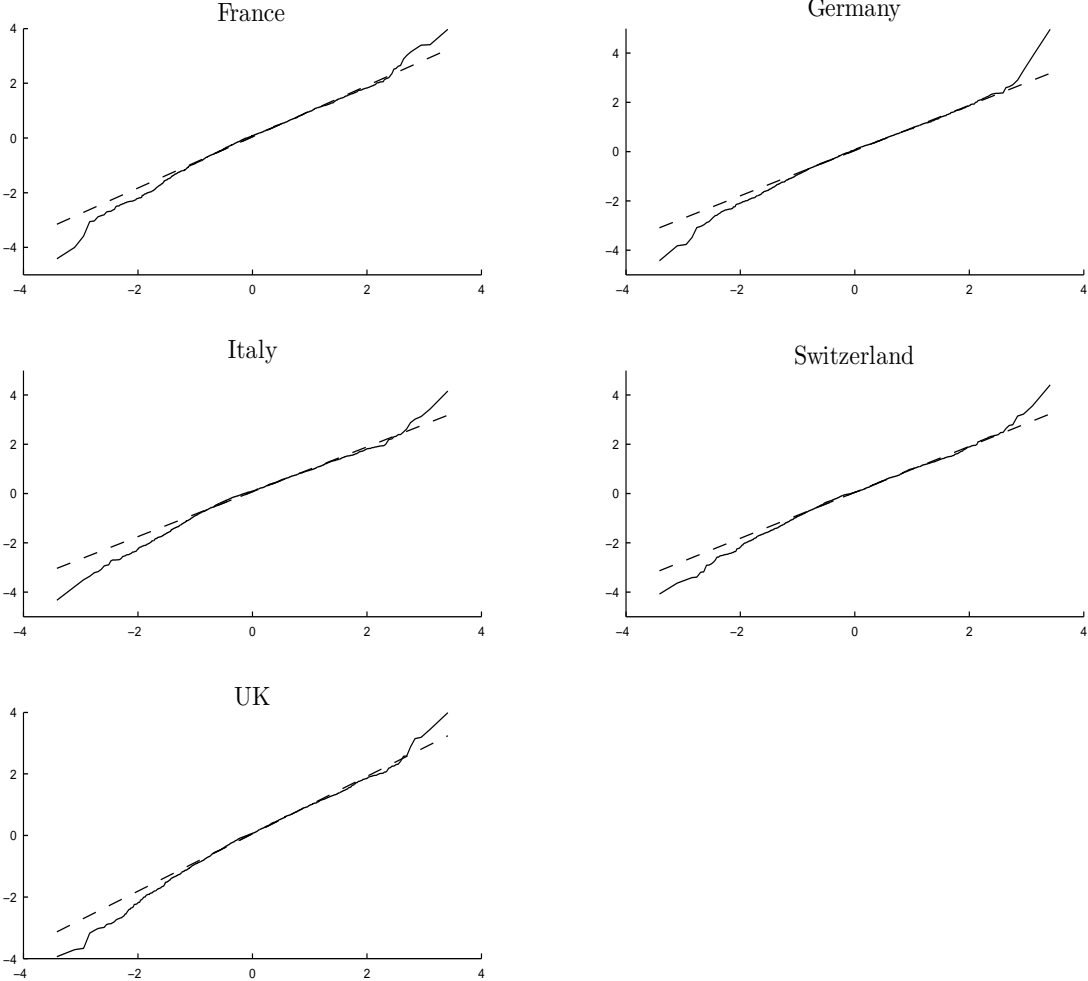


Figure 2.8: Sample Autocorrelation Functions of Squared Residual Series: Basic WMSV Model



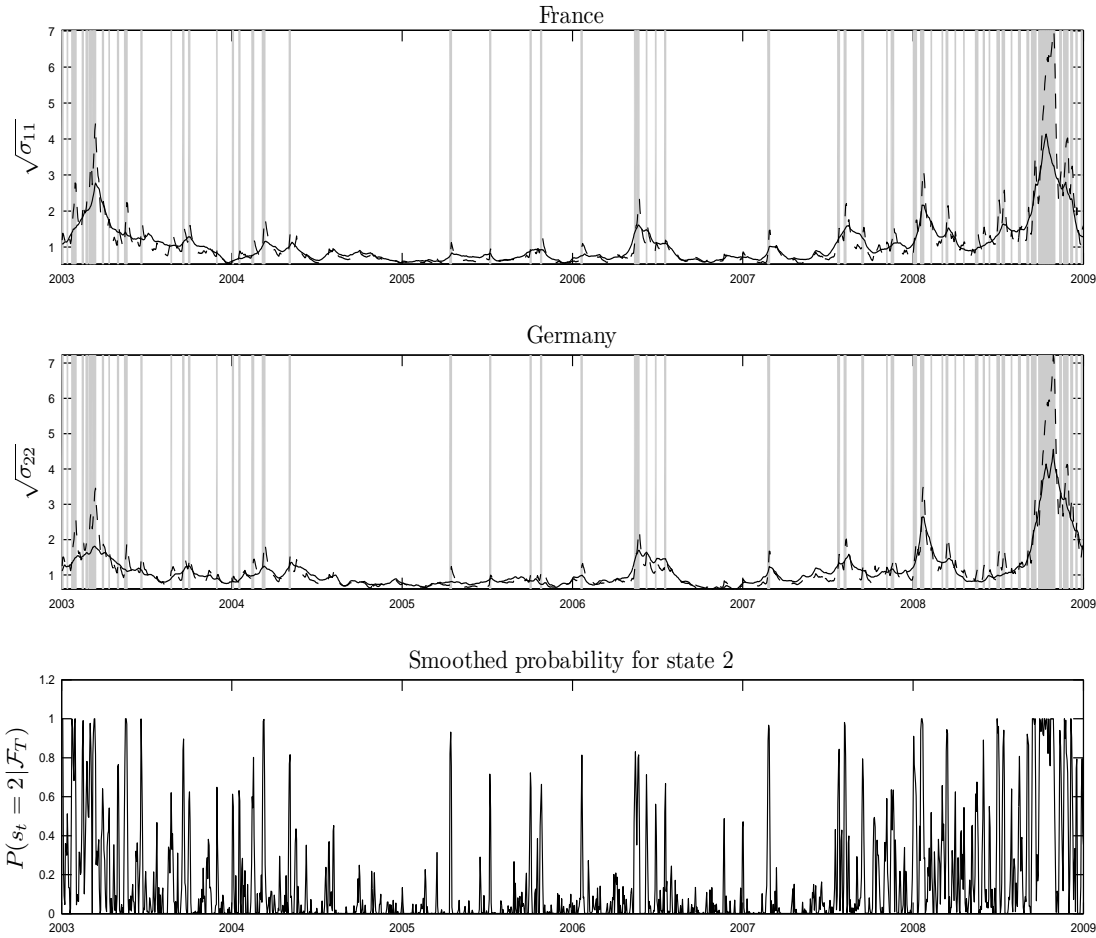
Solid line: ACF of squared residual series, basic WMSV model; \circ : ACF of squared return series; dashed line: 95% Bartlett confidence bands for no serial dependence.

Figure 2.9: Residual QQ-Plots: Basic WMSV Model



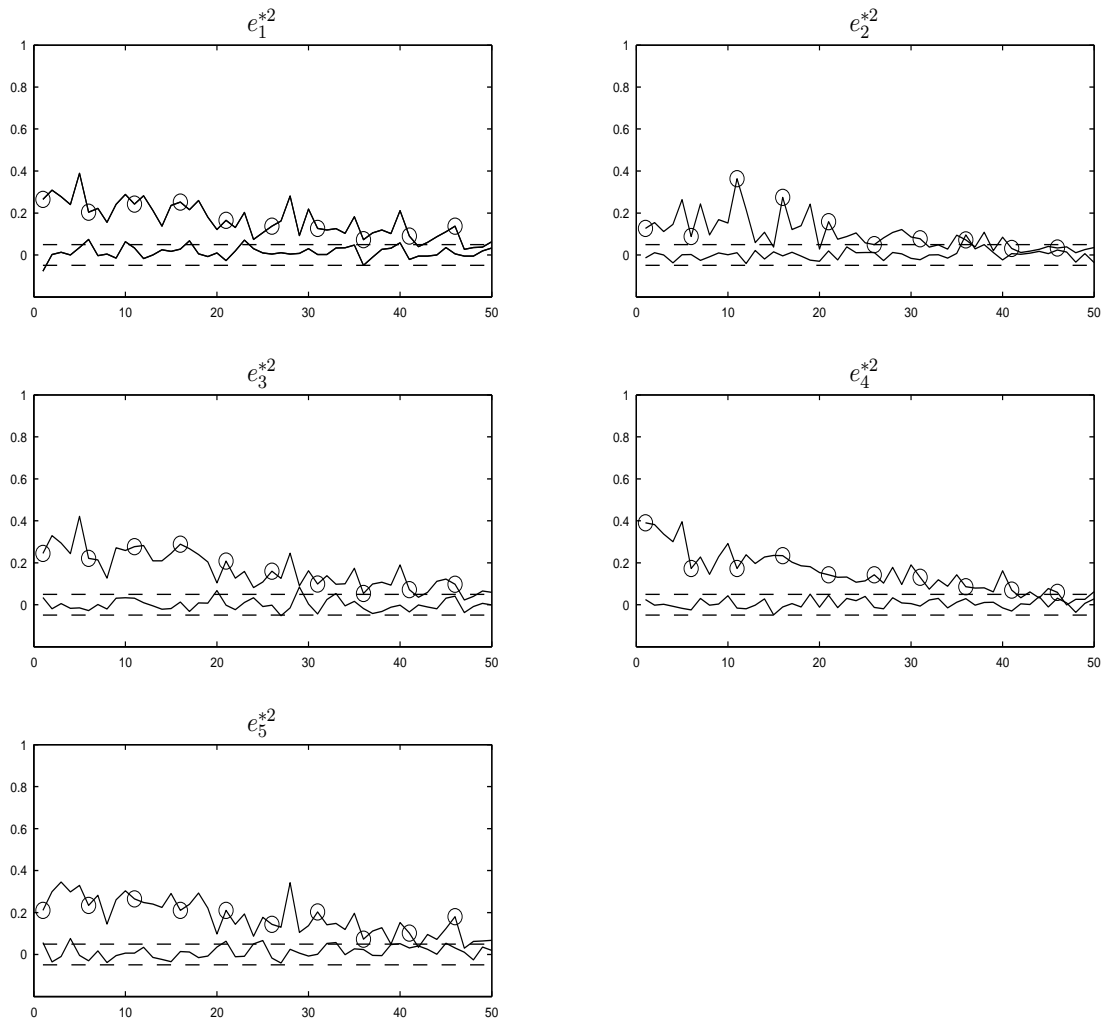
The qq-plots refer to the residual series e_t on distributional diagnostics.

Figure 2.10: Smoothed Volatility and Markov State Estimates



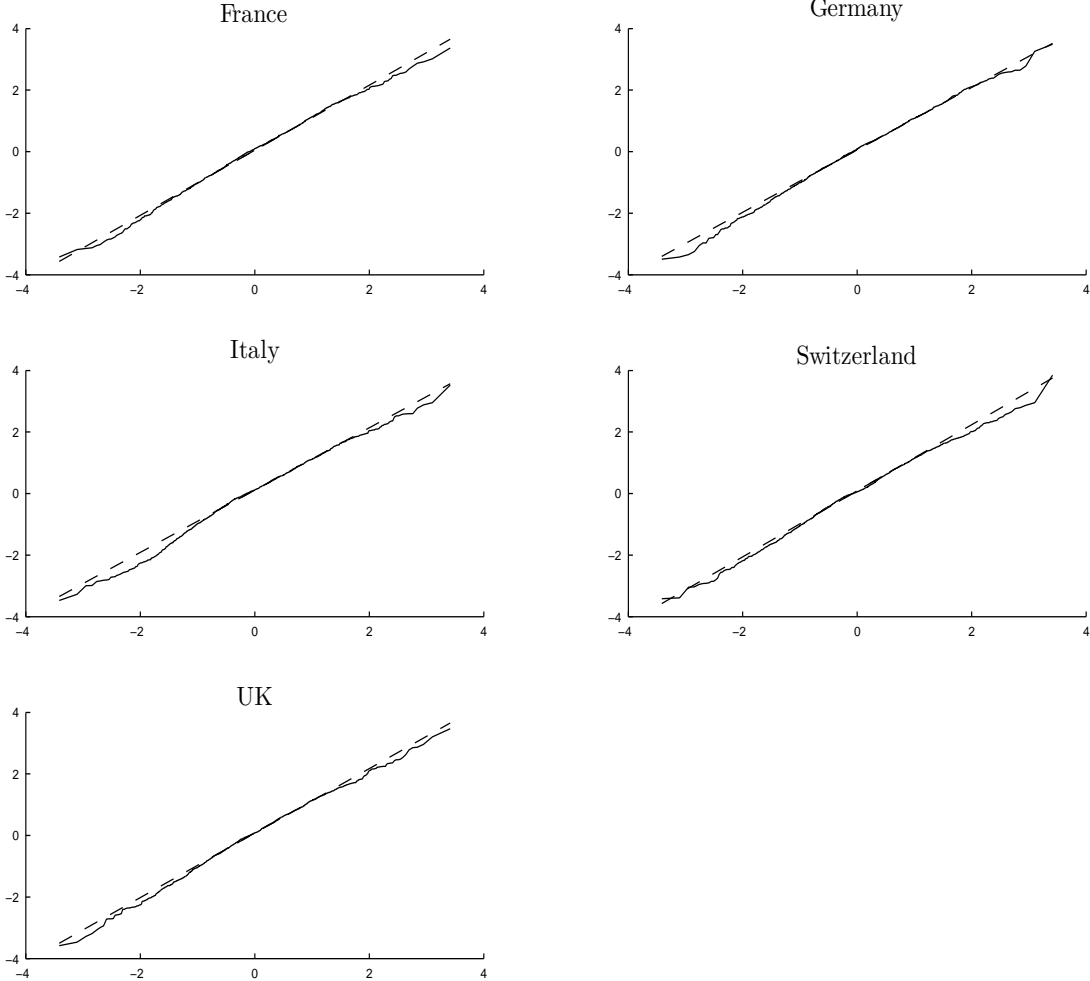
Solid line: Basic WMSV model; dashed line: MS WMSV model. The gray shaded areas mark periods where the smoothed state probability exceeds 0.5.

Figure 2.11: Sample Autocorrelation Functions of Squared Residual Series: MS WMSV Model



Solid line: ACF of squared residual series, MS WMSV model; \circ : ACF of squared return series; dashed line: 95% Bartlett confidence bands for no serial dependence.

Figure 2.12: Residual QQ-Plots: MS WMSV Model



The qq-plots refer to the residual series e_t on distributional diagnostics.

Chapter 3

The Conditional Autoregressive Wishart Model

3.1 Introduction

Multivariate GARCH and SV models are popular approaches for the estimation and modeling of volatility on financial markets. Yet an alternative approach attracted substantial interest in recent years: The increasing availability of high-frequency return data allows to construct realized variances and covariances as precise estimates for the variances and covariances of low-frequency returns (see e.g., Andersen et al. 2003 and Barndorff-Nielsen and Shephard, 2004). Compared to MGARCH or MSV models the realized volatility approach facilitates a more precise measurement and forecasting of daily asset return volatilities. Realized variances and covariances can be modeled directly as advocated, for example, by Andersen et al. (2003). The direct modeling of realized variances and covariances avoids specifying a conditional distribution for asset returns - this, however, precludes applications, which focus on predicting the return density or respective quantiles, like VaR forecasting as applied in the previous chapter. Multivariate models for the realized covariance matrix should satisfy two important requirements, namely that the predicted covariance matrices remain positive definite, and, second, that the specification is parsimoniously parameterized yet empirically realistic with the ability to account for the strong serial dependence typically observed for realized variances and covariances.

Pioneering multivariate approaches to model the dynamics in the realized covariance matrix are found in [Gourieroux et al. \(2009\)](#), [Jin und Maheu \(2011\)](#), [Chiriac and Voev \(2011\)](#), and [Bauer and Vorkink \(2011\)](#). The specification proposed by [Gourieroux et al. \(2009\)](#) extends the Wishart distribution of the sample covariance for i.i.d. multivariate Gaussian random variables by allowing the multivariate Gaussian random variables to be serially correlated. Under the resulting Wishart autoregressive (WAR) process the realized covariance has a transition distribution which is noncentral Wishart with a non-centrality parameter depending on lagged covariances and a fixed scale matrix. As such the WAR model naturally accommodates the positive definiteness of predicted covariance matrices without any parametric restrictions. The density of the noncentral Wishart distribution features the hy-

pergeometric function, which has to be approximated numerically. Since this approximation is time-consuming, Maximum Likelihood estimation is practically infeasible and Method of Moment estimation has to be applied. The approach followed by Jin und Maheu (2011) also relies on a Wishart transition distribution, but assumes a central rather than a noncentral Wishart distribution, and decomposes its scale matrix into multiplicative components, which are driven by sample averages of lagged realized covariance matrices. The model is estimated by simulation based Bayesian inference techniques. In order to account for positive definiteness, the approaches of Chiriac and Voev (2011) and Bauer and Vorkink (2011) use appropriate transformations of the covariance matrix. The former approach is based upon a Cholesky decomposition of the covariance matrix and assumes fractionally integrated VARMA processes for the individual elements of the Cholesky factor. Since the model builds on the Cholesky factorization, inference on the volatility dynamics depends on the sorting of the assets in the return vector. The latter approach transforms the covariance matrix by using the matrix logarithm function and specifies the individual elements of the transformation as functions of latent factors driven by lagged volatilities and lagged returns. The nonlinearity of the matrix logarithm induces problems in deriving the marginal effects of the various forecasting variables driving the factor process and results in biased volatility forecasts. Hence bias correction methods have to be applied.

The present chapter adopts a novel conditional autoregressive Wishart (CAW) approach and proposes a new flexible dynamic model for the realized covariance matrix of asset returns. Its baseline specification assumes a simple autoregressive moving average structure for the scale matrix of the Wishart distribution allowing to account for complex serial dependencies in the variances and covariances. In particular, under the model the predicted covariance matrix depends on lagged covariance matrices as well as on their lagged predictions. As such it presents a dynamic generalization of the models proposed by Gourioux et al. (2009) and Jin und Maheu (2011), where the predicted covariance matrix is specified as a function of lagged covariances only. The model also accounts for symmetry and positive definiteness of the predicted covariance matrices without imposing parametric restrictions and can easily be estimated by Maximum Likelihood (ML). In addition, it allows to derive in a straightforward manner conditions for stationarity and other important time series properties. A further advantage of the CAW approach is that its baseline specification can easily be generalized. Since daily asset return (co)variances are known to feature long-memory like persistence patterns, the present chapter explores two extensions of the baseline CAW model, which are specifically designed to capture the long-run fluctuations in the variances and covariances: The CAW specification is combined with the mixed data sampling (MIDAS) approach of Ghysels et al. (2005, 2006) and, alternatively, with a heterogeneous autoregressive (HAR) component as used by Corsi (2009) and Bonato et al. (2009).

The outline of the chapter is as follows: Section 3.2 introduces the baseline CAW model and discusses its stochastic properties, estimation and model diagnostics as well as extensions

of the baseline model. The empirical application to NYSE data is presented in Section 3.3. Section 3.4 summarizes the findings. The proofs are provided in the Technical Appendix.

3.2 Model Specification, Inference and Diagnostics

3.2.1 The CAW(p, q) Model

Consider the stochastic, symmetric positive definite matrix $R_t = (r_{ij,t})$ of realized covariances with dimension $k \times k$ recorded at time t ($t = 1, \dots, T$). The matrix R_t given the past history $\mathcal{F}_{t-1} = \{R_{t-1}, R_{t-2}, \dots\}$ is assumed to follow a central Wishart distribution

$$R_t | \mathcal{F}_{t-1} \sim \mathcal{W}_k(\nu, S_t/\nu), \quad (3.1)$$

where $\nu > k$ is the scalar degree of freedom, and S_t/ν is the $k \times k$ symmetric, positive definite scale matrix with $S_t = (s_{ij,t})$, such that the conditional mean and covariances are (see Muirhead, 1982)

$$E(R_t | \mathcal{F}_{t-1}) = S_t, \quad \text{Cov}(r_{ij,t}, r_{lm,t} | \mathcal{F}_{t-1}) = \frac{1}{\nu} (s_{il,t} \cdot s_{jm,t} + s_{im,t} \cdot s_{jl,t}) \quad (3.2)$$

for $i, j, l, m = 1, \dots, k$. The density function for $R_t | \mathcal{F}_{t-1}$ has the form

$$f(R_t | \mathcal{F}_{t-1}) = \frac{|S_t/\nu|^{-\nu/2} |R_t|^{(\nu-k-1)/2}}{2^{\nu k/2} \pi^{k(k-1)/4} \prod_{i=1}^k \Gamma([\nu+1-i]/2)} \exp \left\{ -\frac{1}{2} \text{tr}(\nu S_t^{-1} R_t) \right\}, \quad (3.3)$$

where $\Gamma(\cdot)$ denotes the Gamma function. In order to account for serial- and cross-correlation across the elements in R_t the matrix-variate process S_t is assumed to follow the linear recursion of order (p, q)

$$S_t = CC' + \sum_{i=1}^p B_i S_{t-i} B_i' + \sum_{j=1}^q A_j R_{t-j} A_j', \quad (3.4)$$

where C is a $k \times k$ lower-triangular matrix and A_j, B_i are $k \times k$ parameter matrices. This recursion of order (p, q) resembles the BEKK-GARCH(p, q) specification of Engle and Kroner (1995) for the conditional covariance in models for multivariate returns, and has the appealing property to guarantee the symmetry and positive-definiteness of the conditional mean S_t essentially without imposing parametric restrictions on (C, A_j, B_i) as long as the initial matrices $S_0, S_{-1}, \dots, S_{-p+1}$ are symmetric and positive definite. In fact, S_t is positive definite if the null spaces of C', A_1', \dots, A_q' and B_1', \dots, B_p' all intersect only at the origin (see Engle and Kroner, 1995, Proposition 2.5).

In the CAW(p, q) model defined by Equations (3.1) and (3.4), the degree of freedom of the Wishart distribution ν is associated with the overall variability of the covariances in R_t , which is decreasing in ν . The joint dynamic behavior of the covariances is directed by the

parameters in the A_j and B_i matrices. However, note that those parameters do not have direct interpretations in terms of their impact on the conditional mean S_t , as they enter the model in a quadratic form. In the bivariate case with $k = 2$ and $(p = 0, q = 1)$, for example, the elements in S_t obtain as

$$\begin{aligned} s_{11,t} &= c_{11}^2 + a_{11}^2 r_{11,t-1} + 2a_{11}a_{12}r_{12,t-1} + a_{12}^2 r_{22,t-1}, \\ s_{12,t} &= c_{11}c_{21} + a_{11}a_{21}r_{11,t-1} + (a_{11}a_{22} + a_{12}a_{21})r_{12,t-1} + a_{12}a_{22}r_{22,t-1}, \\ s_{22,t} &= c_{22}^2 + c_{21}^2 + a_{22}^2 r_{22,t-1} + 2a_{22}a_{21}r_{12,t-1} + a_{21}^2 r_{11,t-1}. \end{aligned} \quad (3.5)$$

This representation additionally shows that the 9 autoregressive coefficients representing the marginal effects of the lagged $r_{ij,t}$ s are parameterized using 4 parameters only. Hence, the CAW(p, q) model imposes over-identifying restrictions on the autoregressive coefficients.

Moreover, note that under the CAW model not only the conditional means of the elements in R_t but also their conditional variances and covariances given by Equation (3.2) are driven by lagged observations via the recursion (3.4), which generates nonlinear serial dependence in the time series behavior of R_t .

The CAW(p, q) specification can be interpreted as a state-space model with S_t as a state variable measured by the observable matrix R_t and with measurement density given by Equation (3.3). The corresponding measurement equation obtains as (see Muirhead, 1982, p. 95, Theorem 3.2.11)

$$R_t = \frac{1}{\nu} S_t^{1/2} U_t (S_t^{1/2})', \quad U_t \sim \mathcal{W}_k(\nu, I_k), \quad (3.6)$$

where $S_t^{1/2}$ denotes the lower-triangular Cholesky factor of S_t such that $S_t = S_t^{1/2} (S_t^{1/2})'$ and U_t represents the measurement error following a standardized Wishart distribution with ν degree of freedom and a scale matrix given by the identity matrix I_k . This allows to interpret S_t as the ‘true’ integrated covariance for a broad class of multivariate continuous-time stochastic volatility processes which is, under fairly general conditions, consistently estimated by the realized covariance R_t (see Barndorff-Nielsen and Shephard, 2004, and Chapter 1, Section 1.1.3). Within this context, the matrix U_t in Equation (3.6) plays the role of the corresponding estimation error.

The CAW(p, q) model as specified is unidentified. Sufficient conditions for identification are that the main diagonal elements of C , denoted by c_{ii} , and the first diagonal element for each of the matrices A_j, B_i denoted by $a_{11,j}$ and $b_{11,i}$ are restricted to be positive (see Engle and Kroner, 1995).

The CAW(p, q) model is designed to capture complex dynamic interactions across $k(k + 1)/2$ elements of the realized covariance matrix for the returns of k assets. It involves $k(k + 1)/2 + (p + q)k^2 + 1$ parameters. For the multivariate GARCH class of models with such a highly parameterized covariance process, the estimation can be computationally very

demanding when the number of assets increases. However, note that the CAW model is directly fitted to $k(k+1)/2$ realized (co-)variances, while the corresponding GARCH models are estimated based on the returns of k assets only. Hence, the number of observations per parameter is significantly larger for the CAW model than that for the corresponding GARCH specification, such that the curse-of-dimensionality problem appears to be less acute for the CAW model. Furthermore, the number of CAW-parameters can be reduced by imposing restrictions on the matrices (A_j, B_i) . A natural restriction is to impose a diagonal structure on the dynamics of S_t by assuming that A_j and B_j are diagonal matrices. This reduces the number of parameters to $k(k+1)/2 + (p+q)k + 1$.

The CAW model is related to the Wishart autoregressive (WAR) model introduced by Gouriéroux et al. (2009), which is based upon a conditional non-central Wishart distribution for R_t . Under the WAR model, it is the matrix of non-centrality parameters of the Wishart distribution which is assumed to depend on lagged R_t s, rather than the scale matrix as under the CAW model. In particular, the WAR(p) process is characterized by ν degrees of freedom, a fixed scale matrix S , and a matrix of non-centrality parameters given by $S^{-1}(\sum_{i=1}^p A_i R_{t-i} A_i')$ such that

$$E(R_t | \mathcal{F}_{t-1}) = \nu \cdot S + \sum_{i=1}^p A_i R_{t-i} A_i'. \quad (3.7)$$

Hence, the WAR(p) and CAW($0, q$) model with $q = p$ have conditional expectations for the covariance matrix of the same form and with the same number of parameters. This allows to interpret the CAW(p, q) model as a dynamic generalization of the WAR(p) specification. Note, however, that the two model specifications are nonnested, except for the trivial cases, that the WAR(0) obtains as a restricted CAW(p, q) and the CAW($0, 0$) represents a restricted WAR(p) model.

3.2.2 Stochastic Properties of the CAW(p, q) Model

For the discussion of the stochastic properties of the CAW model, it proves convenient to use its VARMA representation which obtains from the recursion (3.4).

Let $\text{vech}(\cdot)$ denote the operator that stacks the lower triangular portion, including the diagonal of a matrix into a vector, and let $\text{vec}(\cdot)$ denote the operator that stacks all columns of a matrix into a vector. Then defining $r_t = \text{vech}(R_t)$, $s_t = \text{vech}(S_t)$ and $c = \text{vech}(CC')$, the vector representation of recursion (3.4) is

$$s_t = c + \sum_{i=1}^p \mathcal{B}_i s_{t-i} + \sum_{j=1}^q \mathcal{A}_j r_{t-j}, \quad (3.8)$$

where $(\mathcal{A}_j, \mathcal{B}_i)$ are $n \times n$ matrices with $n = k(k+1)/2$. They obtain as

$$\mathcal{A}_j = L_k(A_j \otimes A_j)D_k, \quad \mathcal{B}_i = L_k(B_i \otimes B_i)D_k, \quad (3.9)$$

where L_k and D_k denote the elimination and duplication matrix, respectively, defined so that $\text{vec}(X) = D_k \text{vech}(X)$ and $\text{vech}(X) = L_k \text{vec}(X)$ for any symmetric $k \times k$ matrix X (see Lütkepohl, 1996, p. 9-10).

Notice further that r_t can be written as

$$r_t = E(r_t | \mathcal{F}_{t-1}) + v_t = s_t + v_t, \quad \text{with} \quad E(v_t) = 0, \quad E(v_t v_s') = 0 \quad \forall s \neq t, \quad (3.10)$$

where v_t is a martingale difference. By plugging $s_{t-i} = r_{t-i} - v_{t-i}$ ($i = 1, \dots, p$) into Equation (3.8), the CAW(p, q) can be represented as a VARMA($\max(p, q), p$) model:

$$r_t = c + \sum_{i=1}^{\max(p,q)} (\mathcal{B}_i + \mathcal{A}_i) r_{t-i} - \sum_{j=1}^p \mathcal{B}_j v_{t-j} + v_t, \quad (3.11)$$

with $\mathcal{A}_{q+1} = \dots = \mathcal{A}_p = 0$ if $q < p$ and $\mathcal{B}_{p+1} = \dots = \mathcal{B}_q = 0$ if $p < q$. From the VARMA representation (3.11) we immediately obtain the conditions for the existence of the unconditional mean for the CAW(p, q) model, which are given in the following proposition.

Proposition 1. *The unconditional mean of the CAW(p, q) model (3.1) - (3.4) is finite if and only if all eigenvalues of the matrix $\Psi_1 = \sum_{i=1}^{\max(p,q)} (\mathcal{B}_i + \mathcal{A}_i)$ are less than 1 in modulus. In that case the unconditional mean is given by*

$$E(r_t) = \bar{r} = \left(I_n - \sum_{i=1}^{\max(p,q)} (\mathcal{B}_i + \mathcal{A}_i) \right)^{-1} c. \quad (3.12)$$

The following discussion of the second moments of r_t , which represent the fourth moments of the asset returns, is based on the VMA(∞) representation of the CAW(p, q) model and resembles that of Hafner (2003) who derives the existence conditions and the analytic expressions for the fourth moments of multivariate GARCH processes.

The VMA(∞) representation which obtains from the VARMA($\max(p, q), p$) specification (3.11) is given by (see Lütkepohl, 2005, p. 424)

$$r_t = \bar{r} + \sum_{i=0}^{\infty} \Phi_i v_{t-i}, \quad \text{with} \quad \Phi_i = -\mathcal{B}_i + \sum_{j=1}^i (\mathcal{A}_j + \mathcal{B}_j) \Phi_{i-j}, \quad i = 1, 2, \dots, \quad \Phi_0 = I_n. \quad (3.13)$$

Then the autocovariance and variance of r_t , provided that they exist, have the form

$$\Gamma(\tau) = E[(r_t - \bar{r})(r_{t-\tau} - \bar{r})'] = \sum_{i=0}^{\infty} \Phi_{\tau+i} E(v_t v_t') \Phi_i', \quad \tau = 1, 2, \dots, \quad (3.14)$$

and

$$\Gamma(0) = E(r_t r_t') - \bar{r} \bar{r}' = \sum_{i=0}^{\infty} \Phi_i E(v_t v_t') \Phi_i'. \quad (3.15)$$

The following lemma establishes the particular relationship between the second moment of r_t and the second moment of the state process s_t obtained under the conditional Wishart distribution in Equation (3.1).

Lemma 1. *Under the CAW(p, q) model (3.1) - (3.4) and the assumption that $E(r_t r_t')$ exists,*

$$\text{vec}[E(r_t r_t')] = (\Omega + I_{n^2}) \text{vec}[E(s_t s_t')], \quad (3.16)$$

with

$$\Omega = \frac{1}{\nu} (L_k \otimes L_k) [I_{k^2} \otimes (I_{k^2} + K_{kk})] (I_k \otimes K_{kk} \otimes I_k) (D_k \otimes D_k), \quad (3.17)$$

where K_{kk} denotes the commutation matrix (as given in Lütkepohl, 1996 p. 115).

Based upon this result it becomes possible to derive the necessary and sufficient conditions for the existence of $E(r_t r_t')$ and its explicit form, which are given in Proposition 2.

Proposition 2. *The unconditional second moment for the CAW(p, q) model (3.1) - (3.4) is finite if and only if all eigenvalues of the matrix $\Psi_2 = \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) \Omega$ are less than 1 in modulus. In that case the second moment is given by*

$$\text{vec}[E(r_t r_t')] = (\Omega + I_{n^2}) \left(I_{n^2} - \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) \Omega \right)^{-1} \text{vec}(\bar{r} \bar{r}'). \quad (3.18)$$

Proposition 2 implies that under the CAW(p, q) model the process $\{R_t\}$ is covariance stationary if and only if the eigenvalues of the matrix $\sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) \Omega$ are less than 1 in modulus. The explicit form of the unconditional variance obtains by inserting \bar{r} and $E(r_t r_t')$ as given by Equations (3.12) and (3.18), respectively, into $\Gamma(0) = E(r_t r_t') - \bar{r} \bar{r}'$. The unconditional variance of the martingale difference v_t , which is required for the computation of the autocovariance for r_t in Equation (3.14), is given by

$$\text{vec}[E(v_t v_t')] = (I_{n^2} - (\Omega + I_{n^2})^{-1}) \text{vec}[E(r_t r_t')], \quad (3.19)$$

which follows from Equation (3.16) and the fact that $E(v_t v_t') = E(r_t r_t') - E(s_t s_t')$ (see Equation 3.10).

If the order (p, q) of the CAW model is small, it is possible to obtain a more convenient expression for the second moment than that in Equation (3.18). For the CAW(1,1) specification, for example, the following results are derived:

Corollary 1. *The unconditional second moment for the CAW(1,1) model is finite if and*

only if all eigenvalues of the matrix

$$\Delta = (\mathcal{A}_1 \otimes \mathcal{A}_1)(\Omega + I_{n^2}) + (\mathcal{B}_1 \otimes \mathcal{A}_1) + (\mathcal{A}_1 \otimes \mathcal{B}_1) + (\mathcal{B}_1 \otimes \mathcal{B}_1) \quad (3.20)$$

are less than 1 in modulus. In that case the mean and the second moment are given by

$$E(r_t) = \bar{r} = (I_k - (\mathcal{A}_1 + \mathcal{B}_1))^{-1} c, \quad (3.21)$$

$$\text{vec}[E(r_t r_t')] = (\Omega + I_{n^2}) (I_{n^2} - \Delta)^{-1} \text{vec}(cc' + c\bar{r}'(\mathcal{A}_1 + \mathcal{B}_1)' + (\mathcal{A}_1 + \mathcal{B}_1)\bar{r}c'). \quad (3.22)$$

3.2.3 Estimation and Diagnostics

Estimation of the parameters $\psi = (\nu, \text{vech}(C)', \text{vec}(B_1)', \dots, \text{vec}(B_p)', \text{vec}(A_1)', \dots, \text{vec}(A_q)')$ of the CAW(p, q) model can be carried out by maximizing the log-likelihood function using numerical techniques routinely available in standard software packages. The log-likelihood function obtains as

$$\begin{aligned} \mathcal{L}(\psi) = \sum_{t=1}^T \left\{ -\frac{\nu k}{2} \ln(2) - \frac{k(k-1)}{4} \ln(\pi) - \sum_{i=1}^k \ln \Gamma\left(\frac{\nu+1-i}{2}\right) \right. \\ \left. - \frac{\nu}{2} \ln \left| \frac{S_t}{\nu} \right| + \left(\frac{\nu-k-1}{2}\right) \ln |R_t| - \frac{1}{2} \text{tr}(\nu S_t^{-1} R_t) \right\}. \end{aligned} \quad (3.23)$$

The ML-estimates presented below are obtained by using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization procedure. In order to obtain good starting values for the likelihood optimization a ‘bottom-up’ model estimating strategy is used. The proceedings start with estimating a CAW(0,1) model. The order (p, q) is then increased successively where the estimates for one order are used as starting values for the next. At each step the likelihood optimization is repeated by using a set of different starting values in order to check for a local optimum. Positivity of the diagonal elements c_{ll} , $a_{11,i}$, $b_{11,j}$ and the degree of freedom ν are enforced by estimating $\sqrt{c_{ll}}$, $\sqrt{a_{11,i}}$, $\sqrt{b_{11,j}}$, and $\sqrt{\nu}$.

For identification of the order (p, q) the Schwarz’s (1978) information criterion is used. Asymptotically, the Schwarz model selection is equivalent to a Bayesian model selection based on the posterior probability. For linear ARMA models the Schwarz criterion is consistent, selecting asymptotically the correct specification (see Geweke and Meese, 1981). However, there seems to be no published result establishing the consistency of the Schwarz procedure for nonlinear time series models, which would formally justify its use for the CAW model (see Leeb and Pötscher, 2009). Hence, the order identification is supplemented by diagnostic tests for the fitted models. They are conducted from the vector of standardized residuals

$$e_t^* = \text{Var}(r_t | \mathcal{F}_{t-1})^{-1/2} [r_t - E(r_t | \mathcal{F}_{t-1})], \quad (3.24)$$

where $\text{Var}(r_t | \mathcal{F}_{t-1})^{-1/2}$ denotes the inverse Cholesky factor of the conditional covariance

$\text{Var}(r_t|\mathcal{F}_{t-1})$. The elements of the conditional covariance are given in Equation (3.2). For a correctly specified model, the standardized residuals $e_{ij,t}^*$ in the vector e_t^* are serially uncorrelated. Hence, the series $e_{ij,t}^*$ can be used for diagnostic checking of the assumed dynamic structure.

3.2.4 Extensions of the Baseline CAW Model

Due to the nonlinear serial dependence and the large number of different elements in R_t , given by $n = k(k+1)/2$, a low order CAW(p, q) model can be expected to accommodate a large variety of dynamic patterns in the variances and covariances of asset returns, including a long-memory type of persistence. The ability to accommodate long-memory type dependence patterns can be expected due to the well-known fact that low-order multivariate VARMA models typically imply univariate ARMA specifications of a very high order (see, e.g., Cubadda et al., 2009). Nevertheless, it might be useful to consider extensions of the basic CAW(p, q) model, introduced in Section 3.2.1, which are specifically designed to capture strong persistence in asset return volatilities. Chapter 2 illustrated regime switching volatility models in order to capture long-memory like persistence patterns. While Markov switching volatility regimes constitute an interesting extension of the CAW approach, the current chapter focusses on an alternative strand of literature initiated by the seminal work of Engle and Lee (1999), who proposed component GARCH models with short- and long-run components as a way to capture complex dependence structures in the volatility. In the following two of such component model extensions are introduced, where the CAW model is combined with the MIDAS (mixed data sampling) approach of Engle et al. (2009) and Colacito et al. (2011), and with a HAR (heterogenous autoregressive) component as used by Corsi (2009) and Bonato et al. (2009).

MIDAS-CAW Model

Component GARCH models with short- and long-run components have been proven to be useful representations of complex dependence structures in the volatility. Under the GARCH-MIDAS component model, recently proposed by Engle et al. (2009), the short-run component is specified as a mean-reverting GARCH process based on daily returns that moves around a long-run component driven by realized volatilities computed over a monthly, quarterly or semi-annual basis. Following this idea the scale matrix S_t for the daily covariance matrix R_t in Equations (3.1)-(3.3) is decomposed into a secular component M_t and a mean-reverting short-run component S_t^* :

$$E(R_t|\mathcal{F}_{t-1}) = S_t = C_t S_t^* C_t', \quad \text{with} \quad M_t = C_t C_t', \quad (3.25)$$

where C_t is the lower-triangular Cholesky factor of the secular component M_t . The short-run component S_t^* is assumed to follow a covariance-stationary CAW(p, q) process with

$E(S_t^*) = I_k$, namely,

$$S_t^* = \left(I_k - \sum_{j=1}^q A_j A_j' - \sum_{i=1}^p B_i B_i' \right) + \sum_{i=1}^p B_i S_{t-i}^* B_i' + \sum_{j=1}^q A_j \left[C_{t-j}^{-1} R_{t-j} (C_{t-j}')^{-1} \right] A_j'. \quad (3.26)$$

The long-run component M_t is specified as a parsimonious multivariate extension of the univariate MIDAS polynomial proposed by Engle et al. (2009). This multivariate extension is applied to realized covariance matrices ($\bar{R}_t^{(m)}$) computed over a horizon of m trading days using rolling samples that change from day to day. In particular, the MIDAS component is given by the following weighted sum of L lags of m -period realized covariances:

$$M_t = \bar{C} \bar{C}' + \theta \cdot \sum_{\ell=1}^L \varphi_\ell(\omega) \cdot \bar{R}_{t,\ell}^{(m)}, \quad (3.27)$$

$$\bar{R}_{t,\ell}^{(m)} = \sum_{\tau=t-m \cdot \ell}^{t-m \cdot (\ell-1)-1} R_\tau, \quad \ell = 1, \dots, L, \quad (3.28)$$

where \bar{C} denotes a lower triangular matrix, θ is a slope parameter restricted to be non-negative, and $\varphi_\ell(\cdot)$ represents a scalar-valued function of weights. Following Engle et al. (2009), the function $\varphi_\ell(\cdot)$ is specified using so-called Beta weights, defined as

$$\varphi_\ell(\omega) = \frac{\left(1 - \frac{\ell}{L}\right)^{\omega-1}}{\sum_{j=1}^L \left(1 - \frac{j}{L}\right)^{\omega-1}}, \quad (3.29)$$

where the parameter ω controls the weights' decay pattern. In the empirical application discussed below, for the m -period realized covariance in Equation (3.28) a window length m of one month (20 trading days) is used and a lag order L of 12 is taken such that the MIDAS filter aggregates daily covariances of one year.

Note that the MIDAS filter in Equations (3.27)-(3.29) with the same weighting scheme and the same slope across all series imposes a common pattern in the long-run dynamics for all elements in the covariance matrix, thereby preserving parsimony for the specification of the secular component. This restriction is justified by the finding that the long-run movements of the individual realized (co)variances (see Figure 3.1) appear to be very similar.

An interesting alternative to the parametric MIDAS filter for the secular component would be to use for M_t a non-parametric function which smoothes realized covariances in the spirit of the multivariate component GARCH approach recently proposed by Hafner and Linton (2010)¹. In order to make such an approach amenable to out-of-sample forecasting one could substitute the two-sided kernel used by Hafner and Linton by a one-sided kernel involving past covariances only.

¹See also Chapter 4, Sections 4.2 and 4.3 for non-parametric inference on the long-run volatility component.

HAR-CAW Model

A related alternative to the MIDAS component specification of Engle et al. (2009) for highly persistent volatility processes is the HAR model proposed by Corsi (2009), which is a special case of a MIDAS regression (see Ghysels et al. 2007). It accommodates the long-memory type of dependence patterns in daily volatility by a hierarchical autoregressive specification including lagged daily as well as weekly and monthly volatilities. Bonato et al. (2009) extend this univariate approach by combining the multivariate WAR process with HAR dynamics. Following this idea, consider the following specification for the scale matrix S_t in Equations (3.1)-(3.3):

$$S_t = CC' + AR_{t-1}A' + A^{(w)}\bar{R}_{t-1}^{(w)}A^{(w)'} + A^{(bw)}\bar{R}_{t-1}^{(bw)}A^{(bw)'} + A^{(m)}\bar{R}_{t-1}^{(m)}A^{(m)'}, \quad (3.30)$$

with $\bar{R}_{t-1}^{(x)}$ denoting the realized covariance computed over a time window $x = \{w, bw, m\}$, where w stands for the weekly (5 days), bw for the biweekly (10 days), and m for the monthly (20 days) horizon. A and $A^{(x)}$ are $k \times k$ parameter matrices.

Using the vector representation of specification (3.30), it can be written as a restricted CAW(0,20) model:

$$\begin{aligned} s_t = & c + [\mathcal{A} + \mathcal{A}^{(w)} + \mathcal{A}^{(bw)} + \mathcal{A}^{(m)}] \cdot r_{t-1} + \dots + [\mathcal{A}^{(w)} + \mathcal{A}^{(bw)} + \mathcal{A}^{(m)}] \cdot r_{t-5} \\ & + [\mathcal{A}^{(bw)} + \mathcal{A}^{(m)}] \cdot r_{t-6} + \dots + [\mathcal{A}^{(bw)} + \mathcal{A}^{(m)}] \cdot r_{t-10} \\ & + [\mathcal{A}^{(m)}] \cdot r_{t-11} + \dots + [\mathcal{A}^{(m)}] \cdot r_{t-20}, \end{aligned} \quad (3.31)$$

where the matrices \mathcal{A} , $\mathcal{A}^{(x)}$ are obtained as described in Equation (3.9). Hence, the restrictions imposed by the HAR approach take the form of a step function for the autoregressive weight matrices. Also note that this representation of the HAR-CAW model allows to use directly the results discussed in Section 3.2.2 for the derivation of its stochastic properties.

3.3 Empirical Application

3.3.1 Data

The CAW model introduced in Section 3.2 is used to analyze the dynamics of daily realized covariance matrices for five stocks traded at the New York Stock Exchange: American Express (AXP), Citigroup (C), General Electric (GE), Home Depot (HD), and International Business Machines (IBM). The data set represents an updated version of the data set evaluated by Chiriac and Voev (2011). The daily realized covariance matrix can be computed as $R_t = \sum_{j=1}^M \xi_{t,j} \xi_{t,j}'$, where $\xi_{t,j}$ is the vector of returns for the $k = 5$ stocks computed for the j th 5-minute interval of trading day t between 9:30 a.m. and 4:00 p.m. (see Chapter 1, Section 1.1.3). This realized covariance measure is further refined, as in Chiriac and Voev,

by averaging over 30 subsampling subgrids per day in order to exploit the data richness more efficiently and to cope with market microstructure noise. The sample period starts at January 1, 2000, and ends on December 31, 2009, covering 2514 trading days. The first 240 covariance matrices are reserved as starting values for the initialization of the MIDAS filter discussed in Section 3.2.4. This leaves a sample of $T = 2274$ observations. The top row of Figure 3.1 shows the time series plots of five representative elements of the realized covariance matrix: the variances r_{11} (AXP), r_{22} (C), and the covariances r_{21} (C-AXP), r_{31} (GE-AXP), r_{32} (GE-C). The plots reveal a U-shaped pattern in the variance and covariance time series. During the early 2000s, in the aftermath of the dot-com bubble, and during the subprime crisis starting in 2008 both the level as well as the volatility of the variances and covariances are significantly higher than in the middle part of the sample. Descriptive statistics are provided in Table 3.1. The empirical distribution of the variances and covariances is heavily skewed to the right and is highly leptokurtic. The autocorrelation functions of the variances and covariances plotted in the middle row of Figure 3.1 die out at a very slow rate indicating very strong serial correlation.

3.3.2 Estimation Results

Baseline CAW Model

The first attempt to describe the full sample data uses the baseline $CAW(p,q)$ model as given by Equations (3.1) and (3.4), with lag orders (p,q) ranging from (0,1) to (3,3). The upper panel of Table 3.2 reports the values of the maximized log-likelihood function, Schwarz's information criterion, the largest eigenvalues of the estimated matrices $\Psi_1 = \sum_{i=1}^{\max(p,q)} (\mathcal{B}_i + \mathcal{A}_i)$ and $\Psi_2 = \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) \Omega$, and the results of diagnostic checks on the standardized residuals e_t^* , obtained for the fitted unrestricted CAW specifications.

Of the unrestricted models, the Schwarz-preferred specification is the $CAW(2,2)$ with 116 parameters. The largest eigenvalues of the estimated Ψ_1 and Ψ_2 matrices imply that the unrestricted $CAW(2,2)$ has finite unconditional first-order moments, but is not covariance stationary due to an explosive behavior in the second-order moments. Furthermore, note that the largest eigenvalue of Ψ_1 is very close to unity indicating an extremely high persistence in the conditional mean. The result of the Ljung-Box test for serial correlation in the standardized residuals $e_{ij,t}^*$ using 100 lags reveals that the unrestricted $CAW(2,2)$ model successfully accounts for the dynamics in the variance of three stocks and in five covariances, but points towards residual autocorrelation of the variance for two stocks and five covariances. The movement away from the Schwarz-preferred $CAW(2,2)$ specification to the $CAW(2,1)$ and $CAW(2,3)$ improves slightly the approximation of the serial correlation insofar as it increases the number of residual series which pass the Ljung-Box test at the 1% level from 8 to 10, out of 15. Hence, the class of unrestricted CAW models can account for the serial correlation in most, though not all, elements of the realized covariance matrix. This is

corroborated by the autocorrelation functions of the standardized residuals of the CAW(2,2) model (see bottom row of Figure 3.1), indicating that it dramatically reduces the serial correlation for the raw data in the middle row of Figure 3.1. The ML parameter estimates for the unrestricted CAW(2,2) model are provided in the upper panel of Table 3.3. They indicate that for the leading autocorrelation matrices A_1 and B_1 all diagonal elements are statistically significant at the 1% level, while many of the off-diagonal entries are not significantly different from zero. This seems to suggest the use of either a diagonal CAW specification obtained by restricting the matrices B_i and A_j in Equation (3.4) to be diagonal or a partially restricted CAW model obtained by setting the insignificant parameters equal to zero.

The estimation results for the diagonal CAW models are reported in the middle panel of Table 3.2 and the lower panel of Table 3.3. The Schwarz criterion typically favors the parsimonious diagonal CAW(p,q) specifications over their unrestricted counterparts and selects the diagonal CAW(3,2) with 41 parameters as the best diagonal model. However, the diagonal CAW specifications are clearly dominated by their unrestricted counterparts in terms of accounting for serial correlation in the covariance matrix. In particular, under all unrestricted specifications the number of residual series which pass the Ljung-Box test for autocorrelation is larger than under the corresponding diagonal model, except for the model with lag order (0, 1). Finally note that the largest eigenvalues of the estimated matrices characterizing the unconditional moments indicate that all fitted diagonal CAW specifications are characterized by an unbounded unconditional mean.

The last row of Table 3.2 reports the goodness of fit measures for the model obtained by setting in the Schwarz-preferred unrestricted specification, the CAW(2,2), the insignificant parameters to zero. In terms of the Schwarz criterion, this partially restricted CAW(2,2) model represents the overall best CAW formulation but does not improve the approximation of the serial correlation relative to the unrestricted CAW(2,1) and CAW(2,3) specification.

MIDAS-CAW and HAR-CAW Model

In order to account explicitly for long-run fluctuations the baseline CAW models are generalized by including a MIDAS component as described in Equations (3.25)-(3.29). The upper panel of Table 3.4 summarizes goodness-of-fit measures for those CAW extensions. Here again, we consider unrestricted and diagonal specifications as well as the partially constrained version of the Schwarz-preferred unrestricted model.

For each lag order (p, q), the inclusion of the MIDAS component significantly increases the value of the maximized log-likelihood function, indicative for a much better fit. The Schwarz-preferred MIDAS specification is the partially restricted MIDAS-CAW(2,2)-model. The largest eigenvalues of Ψ_1 and Ψ_2 , obtained for all MIDAS specifications, are all less than one and are noticeably smaller than under the baseline CAW models. This suggests that the inclusion of the MIDAS component has significant effects on the dynamic struc-

ture of the model. The same finding is reported by Engle et al. (2009) for a univariate GARCH model, where the inclusion of a MIDAS filter substantially reduces the persistence in the GARCH component. The results of the Ljung-Box test for autocorrelation in the standardized residuals show that the MIDAS specifications better account for the dynamics in the realized covariance matrix than their baseline counterparts in Table 3.2. However, none of the fitted MIDAS-CAW specifications can completely capture the serial correlation in all the variances and covariances. The parameter estimates of the unrestricted and diagonal MIDAS-CAW(2,2) model provided in Table 3.5 show that under both specifications the slope parameter for the MIDAS component θ is significantly larger than zero, which underscores the importance of allowing for a long-run component. Figure 3.2 shows plots of the predicted variance, covariance and correlation for the AXP and C stock together with their respective long-run MIDAS components obtained under the MIDAS-CAW(2,2) model. As expected, the MIDAS component explains a significant part of the variation in the predicted conditional (co)variances and correlations.

An alternative to the MIDAS approach is given by the HAR-CAW model as specified by Equations (3.1) and (3.30). The goodness-of-fit measures provided in the lower part of Table 3.4 reveal that the Schwarz criterion strongly favors the unrestricted and diagonal HAR-CAW models over the unrestricted and diagonal baseline CAW in Table 3.2. However, the Schwarz criterion for both HAR specifications remains substantially larger than that for the best MIDAS counterparts. Furthermore, the Ljung-Box test for residual autocorrelation shows that the HAR specifications also have difficulties capturing the full dynamics in the realized covariance matrix.

All in all, the Schwarz criterion together with the Ljung-Box test for residual autocorrelation indicate that the partially restricted MIDAS-CAW(2,2) model represents the preferred CAW specification for the data. However, none of the considered CAW specifications can fully account for the serial correlation. In particular, for all specifications the Ljung-Box test rejects the null for the residuals of the covariances r_{32} (GE-C), r_{42} (HD-C), r_{52} (IBM-GE), r_{54} (IBM-HD) and the variance r_{55} (IBM). Figure 3.3 displays the ACF for the first 100 lags of those residuals obtained from the unrestricted MIDAS-CAW(2,2) model. It reveals that for the covariance residuals those rejections are mainly due to a significant negative first-order autocorrelation, while the variance residuals exhibit significant serial correlation at various lag orders.

A possible explanation for this lack of fit is that the implicit restrictions imposed by the CAW model on the autoregressive coefficients within and across the S_t equations (see Equation 3.5) are too restrictive. Hence, in order to more completely capture the dynamics in the realized covariance one might consider to eliminate those restrictions and to generalize the recursions (3.4) and (3.26) by including further positive definite terms as in the BEKK-

GARCH models for returns. For the baseline CAW model such an extension has the form

$$S_t = CC' + \sum_{m=1}^z \sum_{i=1}^p B_{i,m} S_{t-i} B'_{i,m} + \sum_{m=1}^z \sum_{j=1}^q A_{j,m} R_{t-j} A'_{j,m}. \quad (3.32)$$

For an initial investigation a MIDAS-CAW model with lag order $(p, q) = (1, 1)$ and $z = 2$ is fitted to the data. The Ljung-Box test for the standardized residuals obtained from this extended MIDAS-CAW(1,1) model (not presented here) indicates that it accounts for the serial correlation of 12 elements in the covariance matrix including that of r_{32} (GE-C) and r_{42} (HD-C), but still cannot fully capture the dynamics of r_{52} (IBM-GE), r_{54} (IBM-HD) and r_{55} (IBM). A more complete exploration of this highly parameterized generalization of the CAW model would go beyond the scope of the present chapter and is left for future research.

3.3.3 Forecasting Results

This section compares the out-of-sample-forecast performance of the CAW-specifications and alternative forecasting models, focusing on forecast horizons of $h = \{1, 5, 10\}$ days. The forecast of the h -period-ahead realized covariance, denoted by $\hat{R}_{t+h} = E(R_{t+h}|\mathcal{F}_t)$, can be compared with the ex-post realization of the realized covariance R_{t+h} . Every model is re-estimated daily and new forecasts are generated based upon the updated parameter estimates. For the forecast experiment two out-of-sample windows are used. The first is selected to be prior to the recent subprime crisis and covers the period from July 2, 2007 through June 30, 2008, where the volatility is comparably low (see the dark-gray shaded area in the top row of Figure 3.1). The second window covers a time period during the crisis with a very high volatility. It starts at July 1, 2008 and ends June 30, 2009 (see the light-gray shaded area in top row of Figure 3.1). So together, the two scenarios represent a balanced assessment of the forecast performance of the CAW model.

For the baseline CAW and the HAR-CAW models, given the parameter estimates, the h -step-ahead forecasts are easily obtained by recursion. In particular, it can be shown for the baseline CAW(p, q) model that

$$\begin{aligned} E(r_{t+h}|\mathcal{F}_t) = E(s_{t+h}|\mathcal{F}_t) &= c + \mathcal{B}_1 E(s_{t+h-1}|\mathcal{F}_t) + \cdots + \mathcal{B}_p E(s_{t+h-p}|\mathcal{F}_t) \\ &\quad + \mathcal{A}_1 E(r_{t+h-1}|\mathcal{F}_t) + \cdots + \mathcal{A}_q E(r_{t+h-q}|\mathcal{F}_t), \end{aligned} \quad (3.33)$$

where

$$E(s_{t+h-\tau}|\mathcal{F}_t) = \begin{cases} s_{t+h-\tau}, & \text{if } \tau \geq h-1 \\ E(r_{t+h-\tau}|\mathcal{F}_t), & \text{if } \tau < h-1. \end{cases} \quad (3.34)$$

The forecasts under the HAR-CAW model are obtained by exploiting its restricted CAW(0,20) representation as given by Equation (3.31).

Under the MIDAS-CAW model the functional relationship between $E(r_{t+1}|\mathcal{F}_t) = s_{t+1}$ and

$\mathcal{F}_t = \{r_t, r_{t-1}, \dots\}$ as specified in Equations (3.25)-(3.29) is non-linear. This implies that the forecast $E(r_{t+h}|\mathcal{F}_t)$ for $h > 1$ depends upon the entire h -step-ahead forecast distribution $f(r_{t+h}|\mathcal{F}_t)$, which is not available in a closed form. Hence, this distribution is approximated by Monte-Carlo (MC) simulation from the convolution of the h one-step-ahead forecast distributions $\{f(r_{t+\tau}|\mathcal{F}_{t+\tau-1})\}_{\tau=1}^h$ as specified by the model and forecasts $E(r_{t+h}|\mathcal{F}_t)$ are evaluated by MC integration. The MC integration is implemented using 10,000 artificial trajectories simulated from the convolution of the one-step-ahead forecast distributions.

As alternative forecasting models for the daily realized covariance, a simple Exponentially Weighted Moving Average (EWMA) approach applied to the realized covariance matrices, a BEKK-GARCH(p,q) model and a DCC-GARCH(p,q) model fitted to the daily stock returns are considered. The EWMA model, which is often used in risk management systems like RiskMetrics (see J.P. Morgan, 1996) to forecast variances and covariances, is given by

$$E(r_t|\mathcal{F}_{t-1}) = (1 - \lambda)r_{t-1} + \lambda E(r_{t-1}|\mathcal{F}_{t-2}), \quad (3.35)$$

where λ is set to its typical value given by 0.94. For this approach the h -step-ahead forecast obtains as $E(r_{t+h}|\mathcal{F}_t) = E(r_{t+h-1}|\mathcal{F}_t)$.

The BEKK-GARCH(p,q) model and the DCC-GARCH(p,q) model are illustrated in Chapter 2, Section 2.3.3. Since the diagonal BEKK-GARCH models with the best out-of-sample performance dominate the unrestricted BEKK-GARCH specifications under all scenarios, only the results for the diagonal versions are reported.

Following Ledoit et al. (2003) the root-mean-square error (RMSE) based on the Frobenius norm of the forecast error is used in order to assess the predictive accuracy for a given model. The measure is given by

$$FN_h = \frac{1}{T_h} \sum_t \|R_{t+h} - \hat{R}_{t+h}\| = \frac{1}{T_h} \sum_t \left[\sum_{i,j} (r_{ij,t+h} - \hat{r}_{ij,t+h})^2 \right]^{1/2}, \quad (3.36)$$

where T_h is number of forecast periods. Alternative measures which can be used to evaluate the forecast performance of multivariate volatility models, including the mean-absolute error (MAE) based on the Frobenius norm and the RMSE and MAE based on the Euclidian norm, are discussed in Laurent et al. (2009).

Table 3.6 summarizes the results on the forecast accuracy of the different models for the out-of-sample window prior to the subprime crisis, and Table 3.7 reports those during the crisis. For the time period before the crisis, the diagonal MIDAS-CAW(3,2) model outperforms the other CAW specifications as well as the GARCH models at the 1-day horizon, whereas at the 5-days horizon the diagonal MIDAS-CAW(2,2) yields the most accurate forecasts. The best CAW model for 10-day forecasts is the diagonal MIDAS-CAW(2,1). It is, however, somewhat outperformed by the diagonal BEKK-GARCH(3,3) specification. Also note that

the differences in the forecast accuracy between the best CAW model, the EWMA and the best GARCH specification becomes smaller if we move from shorter to longer horizons.

As expected, during the crisis the forecasting accuracy of all models substantially deteriorates relative to their accuracy before the crisis, which is indicated by the large increase of the RMSE across all models and time horizons (see Table 3.7). However, the best CAW models for the 1-day and 5-days horizon, the unrestricted CAW(3,2) and the diagonal MIDAS-CAW(0,1), respectively, still outperform the EWMA and the GARCH specifications and again the BEKK-GARCH(3,3) yields slightly more accurate 10-day forecasts than the preferred CAW model.

Overall, the out-of-sample performance of the CAW approach in a moderately volatile as well as in a highly volatile time period appears to be good relative to the competing models especially at the shorter horizons. Among the CAW specifications, a diagonal MIDAS model typically yields the most accurate forecasts, which indicates the importance of allowing for a long-run component.

3.4 Summary

The realized volatility approach facilitates a more precise measurement and forecasting of daily asset return volatilities compared to MGARCH or MSV models. The current chapter contributes to the growing literature on realized volatility modeling by proposing a conditional autoregressive Wishart (CAW) model for the analysis of realized covariance matrices of asset returns. The model is designed to represent complex temporal interdependencies across variances and covariances and is based upon an autoregressive moving average structure for the scale matrix of the central Wishart distribution. Under the CAW model the predicted covariance matrix depends on lagged covariance matrices as well as on their lagged predictions. The model therefore represents a dynamic generalization of the models proposed by *Gourieroux et al. (2009)* and *Jin und Maheu (2011)*, where the predicted covariance matrix is specified as a function of lagged covariances only. A further advantage offered by the CAW approach is that its baseline specification is easily generalizable. In order to explicitly account for long-memory type dependence patterns of realized (co)variances the CAW specification is combined with the mixed data sampling (MIDAS) approach of *Ghysels et al. (2005, 2006)* and, alternatively, with a heterogeneous autoregressive (HAR) component as used by *Corsi (2009)* and *Bonato et al. (2009)*.

The empirical application to daily realized covariance matrices for the returns of five stocks shows that the CAW model typically outperforms GARCH-type volatility models in 1-period ahead volatility forecasting. This finding reflects the gains of the direct modeling of realized (co)variances opposed to daily return data based GARCH models: The direct modeling of realized volatility measures effectively exploits intra-day return information on the variability of the return process resulting in improved (short-term) volatility forecasting

performance. In terms of accounting for the observed dynamic behavior of the realized covariances as well as in terms of out-of-sample covariance predictions the MIDAS-CAW specification shows overall better results compared to the baseline CAW and the HAR-CAW alternatives. Furthermore, the MIDAS-CAW model is found to remove most, though not all, of the observed serial dependence in the variances and covariances. This indicates that in order to more completely capture the highly complex dynamics in the realized covariance matrix it might be useful to consider further alternative extensions of the baseline CAW model, e.g. extended autoregressive recursions of the form (3.32), allowing for a more flexible approximation of the intricate serial correlation observed for the covariance matrix of asset returns. Model diagnostics for the multivariate Wishart SV model illustrated in Chapter 2 indicate the importance of accounting for asymmetric effects of positive and negative news on the covariance matrix. Within the CAW framework such effects could be accounted for using a model extension in line with that of Cappiello et al. (2006) for a multivariate GARCH model. Chapter 2 discussed Markov switching volatility regimes in order to account for the empirically observed strong persistence of daily asset return volatilities and structural changes induced by economic forces like business cycle downturns as well as sudden changes which are due to unusual market events. Extending the CAW approach to Markov switching regimes (i.e. by allowing the parameters of the VARMA specification (3.4) to differ across regimes) is left for future research.

From the computational point of view, the CAW model applied to the realized covariances of five stocks is found to be fairly easy to estimate despite its comparably large number of parameters. This, together with the fact that the similarly parameterized BEKK-GARCH model has been successfully estimated for ten stocks (see, e.g., Chib et al. 2006), suggests that the CAW model may be applicable to about ten stocks.

3.5 Technical Details

Proof for Lemma 1. First, the functional form of the conditional variance of $r_t | \mathcal{F}_{t-1}$ obtained under the conditional Wishart distribution is derived. Since $r_t = \text{vech}(R_t) = L_k \text{vec}(R_t)$, we can write

$$\begin{aligned} \text{Var}(r_t | \mathcal{F}_{t-1}) &= \text{Var}(\text{vech}(R_t) | \mathcal{F}_{t-1}) = \text{Var}(L_k \text{vec}(R_t) | \mathcal{F}_{t-1}) \\ &= L_k \text{Var}(\text{vec}(R_t) | \mathcal{F}_{t-1}) L_k'. \end{aligned} \quad (3.37)$$

Under the Wishart distribution in Equation (3.1) the conditional variance $\text{Var}(\text{vec}(R_t) | \mathcal{F}_{t-1})$ is (see Muirhead, 1982, p. 90)

$$\text{Var}(\text{vec}(R_t) | \mathcal{F}_{t-1}) = \frac{1}{\nu} (I_{k^2} + K_{kk}) (S_t \otimes S_t), \quad (3.38)$$

where K_{kk} is the commutation matrix defined so that $K_{mn} \text{vec}(X) = \text{vec}(X')$ for any $m \times n$ matrix X . Due to $\text{vec}(ABC) = (C' \otimes A) \text{vec}(B)$ we obtain from Equations (3.37) and (3.38)

$$\text{vec}[\text{Var}(r_t | \mathcal{F}_{t-1})] = \frac{1}{\nu} (L_k \otimes L_k) \text{vec}[(I_{k^2} + K_{kk}) (S_t \otimes S_t)]. \quad (3.39)$$

Since $\text{vec}(AB) = (I_p \otimes A) \text{vec}(B)$ for $A (m \times n)$, $B (n \times p)$ and $\text{vec}(A \otimes B) = (I_n \otimes K_{sm} \otimes I_r) [\text{vec}(A) \otimes \text{vec}(B)]$ for $A (m \times n)$, $B (r \times s)$ (see Lütkepohl, 1996, p. 97), we can write

$$\begin{aligned} \text{vec}[\text{Var}(r_t | \mathcal{F}_{t-1})] &= \frac{1}{\nu} (L_k \otimes L_k) \cdot [I_{k^2} \otimes (I_{k^2} + K_{kk})] \text{vec}(S_t \otimes S_t) \\ &= \frac{1}{\nu} (L_k \otimes L_k) \cdot [I_{k^2} \otimes (I_{k^2} + K_{kk})] (I_k \otimes K_{kk} \otimes I_k) [\text{vec}(S_t) \otimes \text{vec}(S_t)], \end{aligned} \quad (3.40)$$

where $\text{vec}(S_t) \otimes \text{vec}(S_t) = (D_k \otimes D_k) \text{vec}(s_t s_t')$. Thus

$$\text{vec}[\text{Var}(r_t | \mathcal{F}_{t-1})] = \Omega \text{vec}(s_t s_t'), \quad (3.41)$$

with

$$\Omega = \frac{1}{\nu} (L_k \otimes L_k) [I_{k^2} \otimes (I_{k^2} + K_{kk})] (I_k \otimes K_{kk} \otimes I_k) (D_k \otimes D_k). \quad (3.42)$$

The law of iterated expectations applied to $\text{Var}(r_t | \mathcal{F}_{t-1}) = \text{E}(r_t r_t' | \mathcal{F}_{t-1}) - s_t s_t'$ leads to

$$\text{E}[\text{Var}(r_t | \mathcal{F}_{t-1})] = \text{E}(r_t r_t') - \text{E}(s_t s_t'), \quad (3.43)$$

such that $\text{E}(r_t r_t') = \text{E}[\text{Var}(r_t | \mathcal{F}_{t-1})] + \text{E}(s_t s_t')$. Taking vecs and accounting for Equation

(3.41) we obtain

$$\begin{aligned}
 \text{vec}[\mathbb{E}(r_t r_t')] &= \text{vec}(\mathbb{E}[\text{Var}(r_t | \mathcal{F}_{t-1})]) + \text{vec}[\mathbb{E}(s_t s_t')] & (3.44) \\
 &= \mathbb{E}(\text{vec}[\text{Var}(r_t | \mathcal{F}_{t-1})]) + \text{vec}[\mathbb{E}(s_t s_t')] \\
 &= \Omega \text{vec}[\mathbb{E}(s_t s_t')] + \text{vec}[\mathbb{E}(s_t s_t')] = (\Omega + I_{n^2}) \text{vec}[\mathbb{E}(s_t s_t')],
 \end{aligned}$$

which completes the proof. \square

Proof for Proposition 1. The VARMA representation in Equation (3.11) allows us to write $\mathbb{E}(r_t) = c + \sum_{i=1}^{\max(p,q)} (\mathcal{B}_i + \mathcal{A}_i) \mathbb{E}(r_t)$, which can be solved for $\mathbb{E}(r_t) = \bar{r}$ to obtain Equation (3.12) if and only if all eigenvalues of the matrix $\sum_{i=1}^{\max(p,q)} (\mathcal{B}_i + \mathcal{A}_i)$ are less than 1 in modulus. \square

Proof for Proposition 2. Since $\text{Var}(r_t) = \Gamma(0) = \mathbb{E}(r_t r_t') - \bar{r} \bar{r}'$ and $\mathbb{E}(v_t v_t') = \mathbb{E}(r_t r_t') - \mathbb{E}(s_t s_t')$ (see Equation 3.10), we obtain from covariance Equation (3.15)

$$\mathbb{E}(r_t r_t') = \sum_{i=0}^{\infty} \Phi_i [\mathbb{E}(r_t r_t') - \mathbb{E}(s_t s_t')] \Phi_i' + \bar{r} \bar{r}', \quad (3.45)$$

$$\text{vec}[\mathbb{E}(r_t r_t')] = \sum_{i=0}^{\infty} (\Phi_i \otimes \Phi_i) \text{vec}[\mathbb{E}(r_t r_t') - \mathbb{E}(s_t s_t')] + \text{vec}(\bar{r} \bar{r}'). \quad (3.46)$$

Applying the result of Lemma 1 that $\text{vec}[\mathbb{E}(r_t r_t')] = (\Omega + I_{n^2}) \text{vec}[\mathbb{E}(s_t s_t')]$, we obtain

$$\text{vec}[\mathbb{E}(r_t r_t')] = \sum_{i=0}^{\infty} (\Phi_i \otimes \Phi_i) \Omega \text{vec}[\mathbb{E}(s_t s_t')] + \text{vec}(\bar{r} \bar{r}'). \quad (3.47)$$

Since $\Phi_0 = I_n$, Equation (3.47) can be rewritten as

$$\text{vec}(\bar{r} \bar{r}') = \left(I_{n^2} - \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) \Omega \right) \text{vec}[\mathbb{E}(s_t s_t')]. \quad (3.48)$$

If and only if all eigenvalues of the matrix $\Psi_2 = \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) \Omega$ are less than 1 in modulus, Equation (3.48) can be solved for $\text{vec}[\mathbb{E}(s_t s_t')]$ to obtain

$$\text{vec}[\mathbb{E}(s_t s_t')] = \left(I_{n^2} - \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) \Omega \right)^{-1} \text{vec}(\bar{r} \bar{r}'). \quad (3.49)$$

Inserting Equation (3.49) into $\text{vec}[\mathbb{E}(r_t r_t')] = (\Omega + I_{n^2}) \text{vec}[\mathbb{E}(s_t s_t')]$ completes the proof. \square

Proof for Corollary 1. The mean $\mathbb{E}(r_t)$ is obtained directly from Proposition 1. Furthermore, note that the CAW(1,1) model can be written as a VARMA(1,1) with a VMA(∞)

representation characterized by the parameters (see Equation 3.13)

$$\Phi_0 = I_n, \quad \Phi_1 = \mathcal{A}_1, \quad \Phi_2 = (\mathcal{A}_1 + \mathcal{B}_1)\mathcal{A}_1, \dots, \Phi_i = (\mathcal{A}_1 + \mathcal{B}_1)^{i-1}\mathcal{A}_1.$$

Then using the result that $AC \otimes BD = (A \otimes B)(C \otimes D)$ (see Lütkepohl, 1996, p. 19), we can write under the assumption that the second moment exists

$$\begin{aligned} \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) &= \sum_{i=1}^{\infty} [(\mathcal{A}_1 + \mathcal{B}_1)^{i-1}\mathcal{A}_1] \otimes [(\mathcal{A}_1 + \mathcal{B}_1)^{i-1}\mathcal{A}_1] \\ &= \sum_{i=0}^{\infty} [(\mathcal{A}_1 + \mathcal{B}_1) \otimes (\mathcal{A}_1 + \mathcal{B}_1)]^i (\mathcal{A}_1 \otimes \mathcal{A}_1) \\ &= [I_{n^2} - (\mathcal{A}_1 + \mathcal{B}_1) \otimes (\mathcal{A}_1 + \mathcal{B}_1)]^{-1} (\mathcal{A}_1 \otimes \mathcal{A}_1) \\ &= Q^{-1}(\mathcal{A}_1 \otimes \mathcal{A}_1) \quad (\text{say}). \end{aligned} \tag{3.50}$$

Plugging Equation (3.50) into Equation (3.18) we obtain

$$\begin{aligned} \text{vec}[\mathbf{E}(r_t r_t')] &= (\Omega + I_{n^2})(I_{n^2} - Q^{-1}(\mathcal{A}_1 \otimes \mathcal{A}_1)\Omega)^{-1} \text{vec}(\bar{r}\bar{r}') \\ &= (\Omega + I_{n^2})(Q^{-1}[Q - (\mathcal{A}_1 \otimes \mathcal{A}_1)\Omega])^{-1} \text{vec}(\bar{r}\bar{r}') \\ &= (\Omega + I_{n^2})(I_{n^2} - [\mathcal{A}_1 + \mathcal{B}_1] \otimes [\mathcal{A}_1 + \mathcal{B}_1] - (\mathcal{A}_1 \otimes \mathcal{A}_1)\Omega)^{-1} Q \text{vec}(\bar{r}\bar{r}') \\ &= (\Omega + I_{n^2})(I_{n^2} - \Delta)^{-1} Q \text{vec}(\bar{r}\bar{r}'), \end{aligned} \tag{3.51}$$

where $Q \text{vec}(\bar{r}\bar{r}') = \text{vec}(cc' + c\bar{r}'(\mathcal{A}_1 + \mathcal{B}_1)' + (\mathcal{A}_1 + \mathcal{B}_1)\bar{r}c')$, which completes the proof. \square

Table 3.1: Descriptive Statistics for the Realized Variances and Covariances

Stock	Mean	Max.	Min.	Std. dev.	Skewness	Kurtosis
Realized Variance						
AXP (r_{11})	5.37	263.93	.07	10.61	9.29	169.27
C (r_{22})	9.03	887.96	.11	33.58	13.16	252.87
GE (r_{33})	3.84	164.31	.10	8.15	8.23	108.32
HD (r_{44})	4.07	167.63	.16	6.05	10.34	226.59
IBM (r_{55})	2.61	74.34	.11	4.09	6.30	70.54
Realized Covariance						
C-AXP (r_{21})	3.06	149.53	-.55	8.46	8.57	107.46
GE-AXP (r_{31})	1.98	102.29	-1.47	4.78	8.27	115.41
HD-AXP (r_{41})	1.81	87.18	-2.45	3.98	8.20	117.53
IBM-AXP (r_{51})	1.38	45.89	-1.97	3.01	7.05	72.14
GE-C (r_{32})	2.38	132.78	-.58	6.85	9.63	135.58
HD-C (r_{42})	2.07	106.88	-2.35	5.19	9.13	133.13
IBM-C (r_{52})	1.58	85.90	-3.27	3.95	9.47	140.91
HD-GE (r_{43})	1.57	86.39	-1.14	3.49	9.71	170.72
IBM-GE (r_{53})	1.28	50.76	-.33	2.74	7.62	90.66
IBM-HD (r_{54})	1.22	79.42	-1.20	2.88	12.10	151.14

Note: The number of observations for each (co)variance series is 2274.

Table 3.2: Optimized Log-likelihood and Diagnostics for the Baseline CAW Model

(p,q)	dim(ψ)	Log-lik.	Schwarz	meig1	meig2	p -value for the Ljung-Box test for the residuals e_{ijt}^* , 100 lags													
						e_{11}^*	e_{21}^*	e_{31}^*	e_{41}^*	e_{51}^*	e_{22}^*	e_{32}^*	e_{42}^*	e_{52}^*	e_{33}^*	e_{43}^*	e_{53}^*	e_{44}^*	e_{54}^*
Unrestricted CAW(p,q)																			
(0,1)	41	-20592	41501	1.068		.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
(1,1)	66	-17258	35027	1.022		.04	.00	.01	.00	.55	.01	.00	.00	.00	.00	.16	.01	.01	.00
(1,2)	91	-17027	34758	1.016		.07	.02	.10	.00	.29	.00	.00	.00	.00	.02	.46	.02	.01	.00
(2,1)	91	-16971	34645	1.009		.07	.02	.02	.01	.12	.05	.00	.00	.00	.01	.54	.01	.09	.00
(2,2)	116	-16835	34566	.999	5.681	.11	.01	.06	.00	.09	.00	.00	.00	.00	.03	.17	.03	.06	.00
(2,3)	141	-16813	34715	.998	3.363	.10	.01	.04	.01	.09	.01	.00	.00	.00	.02	.17	.03	.06	.00
(3,2)	141	-16740	34571	.995	1.068	.29	.00	.06	.00	.02	.01	.00	.00	.00	.06	.27	.00	.21	.00
(3,3)	166	-16718	34718	.995	1.184	.20	.01	.04	.00	.01	.00	.00	.00	.00	.08	.45	.00	.21	.00
Diagonal CAW(p,q)																			
(0,1)	21	-20623	41408	1.067		.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
(1,1)	26	-17313	34827	1.013		.02	.00	.01	.00	.11	.00	.00	.00	.00	.00	.10	.00	.01	.00
(1,2)	31	-17297	34833	1.011		.02	.01	.01	.00	.10	.00	.00	.00	.00	.00	.13	.00	.01	.00
(2,1)	31	-17179	34598	1.011		.04	.01	.00	.00	.05	.01	.00	.00	.00	.01	.07	.00	.08	.00
(2,2)	36	-17165	34609	1.008		.03	.01	.00	.00	.03	.01	.00	.00	.00	.01	.10	.00	.07	.00
(2,3)	41	-17164	34646	1.008		.03	.01	.00	.00	.03	.01	.00	.00	.00	.01	.10	.00	.08	.00
(3,2)	41	-17074	34465	1.006		.09	.00	.01	.00	.01	.00	.00	.00	.00	.02	.18	.00	.25	.00
(3,3)	46	-17073	34501	1.005		.09	.00	.01	.00	.01	.00	.00	.00	.00	.02	.18	.00	.25	.00
Partially restricted CAW(2,2)																			
(2,2)	67	-16861	34241	.999	3.969	.10	.01	.09	.01	.08	.00	.00	.00	.04	.08	.04	.05	.00	.00

Note: dim(ψ) is the number of parameters; meig1 and meig2 are the maximum eigenvalues of the matrices $\Psi_1 = \sum_{i=1}^{\max(p,q)} (\mathcal{B}_i + \mathcal{A}_i)$ and $\Psi_2 = \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) \Omega$, respectively, see Equations (3.12) and (3.18); bold p -values indicate significance at the 1% level.

Table 3.3: ML-Parameter Estimates for the Baseline CAW(2,2) model

		Unrestricted CAW(2,2)			
Param.	Estimate	Param.	Estimate	Param.	Estimate
A_1	.798*	-.024	.148	.073*	.004
	.031*	-.011	-.083*	.011	.179*
	.736*	-.012	-.068*	.134*	-.017
	.024*	-.013	-.038*	.159*	-.074*
B_1	.024	.574*	.029	.024	-.142*
	.022*	.008	-.032	.029	-.059*
	.011	-.005	.589*	.019*	.011
	-.039*	-.044	-.081*	-.121*	-.053*
B_2	.849*	.075*	.004	-.196*	.006
	.014	-.166*	-.085*	.024	-.080*
	.630*	-.007	.104*	.077*	-.185*
	-.003	-.024	.104*	.000	.002
ν	.035*	.530*	.290*	-.006	-.014
	.055*	.001	.356*	.074*	-.593*
	-.025*	.006	-.023	-.024*	.215*
	.015	-.010	.597*	-.011	-.004
	21.689*				-.526*

		Diagonal CAW(2,2)	
Param.	Estimate	Param.	Estimate
A_1	.813*	A_2	.328*
B_1	.718*	B_2	.668*
	.535*		.452*
ν	.593*		.465*
	21.343*		.421*
			.086
			.496*

Note: The estimated model is described in Equations (3.1) and (3.4). The first diagonal element for A_1 , A_2 , B_1 , B_2 represents the estimates of $\sqrt{a_{1,1,1}}$, $\sqrt{a_{1,1,2}}$, $\sqrt{b_{1,1,1}}$, $\sqrt{b_{1,1,2}}$. * Significant at the 1% level.

Table 3.5: ML-Parameter Estimates for the MIDAS-CAW(2,2) Model

Unrestricted MIDAS-CAW(2,2)					
Param.	Estimate	Param.	Estimate	Param.	Estimate
A_1	.779*	.016	.071*	.000	.156*
	.059*	-.014	.000	.045	-.123*
	.679*	-.018	.045	.110*	.066*
	.028	-.041*	.102	.110*	-.113*
B_1	.011	.013	-.003	-.037	.068
	-.001	.556*	-.212*	-.187*	-.266*
	.007	.004	.108*	-.077*	.009
	-.010	.591*	.132*	.065	.022
B_2	.003	.089*	.063	.394*	-.099*
	-.002	.033	-.097	.232*	-.233*
	.709*	-.081*	.283*	-.080	.036
	-.007	.575*	.031	-.045	.065
$\sqrt{\theta}$.925*	.007	.157*	-.047	.010
	.040	.206*	.146*	.020*	.212*
	.127*	.312*	.135*	.135*	.020*
	-.017	.177*	-.080	.271*	-.001
ν	.063*	.177*	-.080	-.045	-.087
	-.072*	.249*	.010	-.047	.107*
	21.991*	ω	9.385*		

Diagonal MIDAS-CAW(2,2)		
Param.	Estimate	Estimate
A_1	.647*	.291*
	.698*	.410*
B_1	.467*	.584*
	.441*	.413*
$\sqrt{\theta}$.916*	.053
	21.666*	9.693*

Note: The estimated model is described in Equations (3.1) and (3.25)-(3.29). The first diagonal element for A_1 , A_2 , B_1 , B_2 represents the estimates of $\sqrt{a_{1,1,1}}$, $\sqrt{a_{1,1,2}}$, $\sqrt{b_{1,1,1}}$, $\sqrt{b_{1,1,2}}$. * Significant at the 1% level.

Table 3.6: Evaluation of Forecasting Accuracy Prior to the Subprime Crisis

Model	(p, q)	Frobenius norm of forecast error		
		$h = 1$	$h = 5$	$h = 10$
Unrestricted CAW	(0,1)	6.961	10.298	14.371
	(1,1)	6.288	8.030	9.543
	(1,2)	6.263	7.922	9.306
	(2,1)	6.150	7.889	9.261
	(2,2)	6.122	7.852	9.243
	(2,3)	6.150	7.896	9.290
	(3,2)	6.074	7.834	9.193
	(3,3)	6.090	7.848	9.223
Diagonal CAW	(0,1)	6.693	8.396	9.831
	(1,1)	6.228	7.878	9.220
	(1,2)	6.214	7.841	9.157
	(2,1)	6.142	7.871	9.181
	(2,2)	6.118	7.814	9.080
	(2,3)	6.112	7.796	9.051
	(3,2)	6.070	7.797	9.028
	(3,3)	6.062	7.768	8.982
Partially restricted CAW(2,2)		6.127	7.855	9.177
Unrestricted MIDAS-CAW	(0,1)	6.257	8.277	9.192
	(1,1)	6.241	8.025	8.977
	(1,2)	6.233	8.008	8.995
	(2,1)	6.156	8.022	9.137
	(2,2)	6.148	7.951	9.074
	(2,3)	6.173	7.927	9.038
	(3,2)	6.132	7.938	9.123
	(3,3)	6.121	7.923	9.094
Diagonal MIDAS-CAW	(0,1)	6.236	8.279	9.187
	(1,1)	6.113	7.738	8.745
	(1,2)	6.119	7.704	8.746
	(2,1)	6.073	7.666	8.702
	(2,2)	6.052	7.647	8.719
	(2,3)	6.056	7.662	8.725
	(3,2)	6.029	7.655	8.737
	(3,3)	6.031	7.650	8.735
Partially restricted MIDAS-CAW(2,2)		6.191	7.909	9.021
Unrestricted HAR-CAW		6.233	8.278	9.679
Diagonal HAR-CAW		6.108	7.960	9.165
EWMA		7.380	8.310	9.135
Diagonal BEKK-GARCH	(0,1)	8.562	8.953	9.056
	(1,1)	8.435	9.056	9.479
	(1,2)	7.955	8.594	9.032
	(2,1)	8.059	8.705	9.127
	(2,2)	7.603	8.265	8.733
	(2,3)	7.499	8.164	8.644
	(3,2)	7.496	8.161	8.638
	(3,3)	7.441	8.100	8.572
DCC-GARCH	(0,1)	8.565	8.844	8.986
	(1,1)	8.332	9.570	12.21
	(1,2)	8.093	8.717	9.521
	(2,1)	8.552	8.696	9.254
	(2,2)	8.259	8.575	9.171
	(2,3)	8.178	8.663	9.212
	(3,2)	8.318	9.020	9.595
	(3,3)	8.099	8.998	9.611

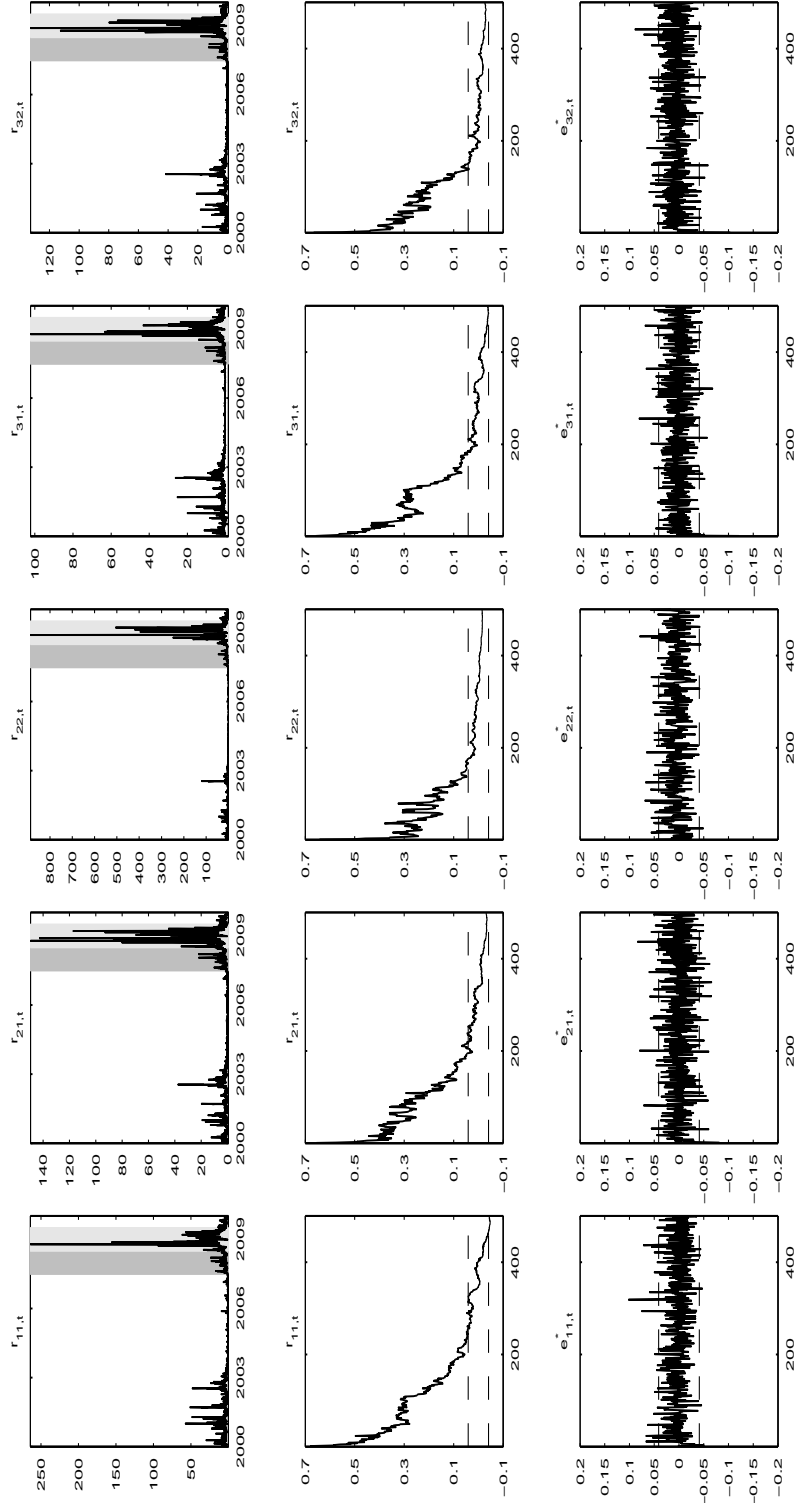
Note: Reported are the average Frobenius norm of the forecast error as given by Equation (5.18). Bold numbers indicate the smallest value of the average Frobenius norm.

Table 3.7: Evaluation of Forecasting Accuracy During the Subprime Crisis

Model	(p, q)	Frobenius norm of forecast error		
		$h = 1$	$h = 5$	$h = 10$
Unrestricted CAW	(0,1)	59.559	99.496	150.587
	(1,1)	53.892	78.403	91.232
	(1,2)	53.627	77.627	89.941
	(2,1)	53.659	76.947	87.489
	(2,2)	53.163	75.478	85.534
	(2,3)	52.879	74.359	83.848
	(3,2)	52.474	72.505	80.319
	(3,3)	52.656	73.009	80.927
Diagonal CAW	(0,1)	58.188	85.663	110.700
	(1,1)	53.725	76.516	86.481
	(1,2)	53.519	75.671	84.816
	(2,1)	53.376	75.256	83.791
	(2,2)	53.100	74.273	81.982
	(2,3)	53.051	74.075	81.621
	(3,2)	52.924	73.435	80.357
	(3,3)	52.799	73.131	79.858
Partially restricted CAW(2,2)		52.882	74.126	82.990
Unrestricted MIDAS-CAW	(0,1)	55.614	70.658	75.279
	(1,1)	54.945	75.803	81.936
	(1,2)	54.502	74.858	80.585
	(2,1)	54.275	74.480	80.372
	(2,2)	54.129	73.350	79.758
	(2,3)	54.006	72.557	78.107
	(3,2)	53.977	74.169	81.876
	(3,3)	54.007	73.780	81.386
Diagonal MIDAS-CAW	(0,1)	55.380	69.868	74.799
	(1,1)	54.532	75.586	83.365
	(1,2)	54.480	75.554	83.925
	(2,1)	54.382	75.421	83.435
	(2,2)	54.346	75.761	84.871
	(2,3)	54.364	75.778	84.902
	(3,2)	54.276	75.732	84.616
	(3,3)	54.314	75.880	84.734
Partially restricted MIDAS-CAW(2,2)		53.846	73.009	79.115
Unrestricted HAR-CAW		52.974	73.222	81.130
Diagonal HAR-CAW		53.113	72.681	79.176
EWMA		62.361	70.068	73.955
Diagonal BEKK-GARCH	(0,1)	68.419	74.626	75.745
	(1,1)	76.163	83.243	87.509
	(1,2)	68.877	76.106	78.978
	(2,1)	70.016	77.285	80.620
	(2,2)	66.641	73.600	76.187
	(2,3)	64.875	71.877	74.325
	(3,2)	64.238	71.553	74.058
	(3,3)	63.550	71.028	73.424
DCC-GARCH	(0,1)	72.821	81.084	79.512
	(1,1)	78.542	89.918	93.638
	(1,2)	74.321	87.940	90.696
	(2,1)	74.162	85.360	90.447
	(2,2)	73.338	84.770	90.452
	(2,3)	73.167	84.189	89.774
	(3,2)	73.042	84.535	90.060
	(3,3)	72.121	83.859	89.246

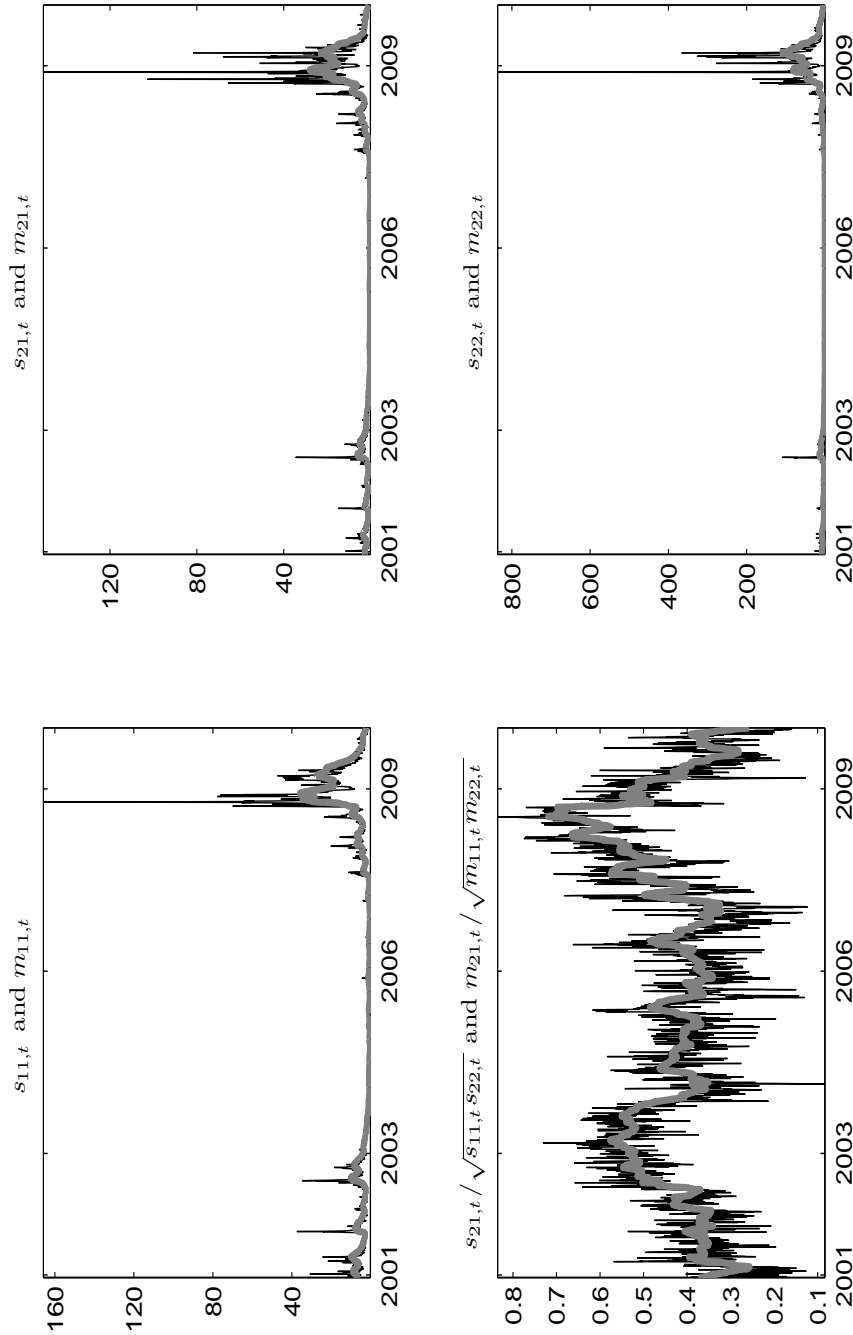
Note: Reported are the average Frobenius norm of the forecast error as given by Equation (5.18). Bold numbers indicate the smallest value of the average Frobenius norm.

Figure 3.1: Time Series and ACFs of Realized Variances, Covariances and Residual Series



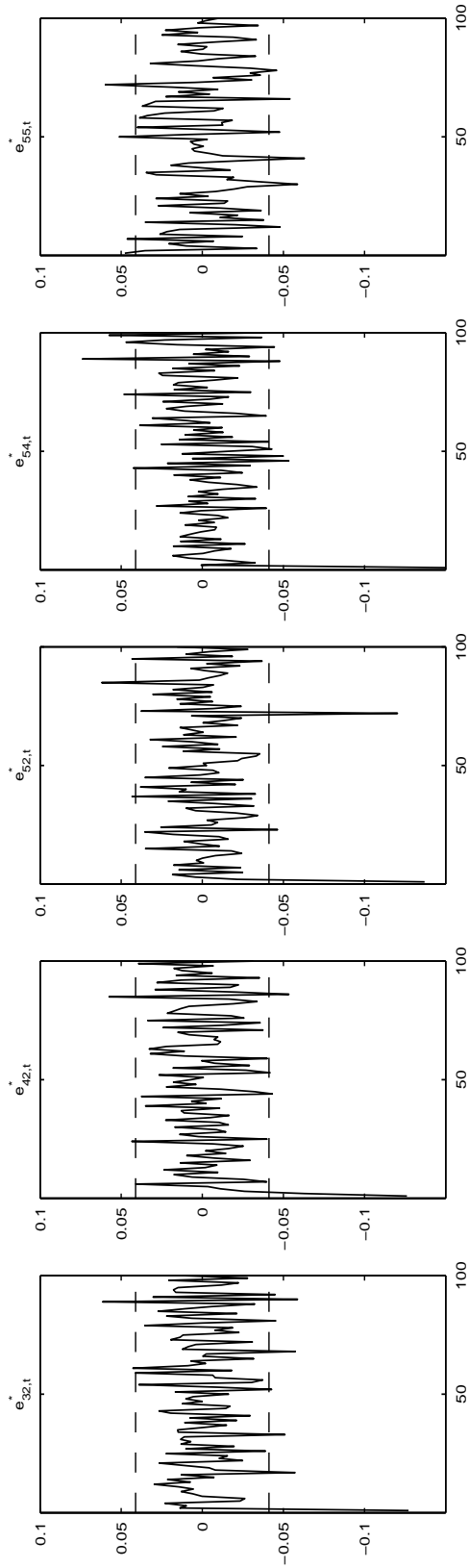
Time series of realized variances and covariances $r_{ij,t}$ for AXP ($i = 1$), C ($i = 2$), and GE ($i = 3$) stock (top row); sample autocorrelation function of realized variances and covariances $r_{ij,t}$ (middle row); sample autocorrelation function of standardized residuals $e_{ij,t}^*$ from the unrestricted baseline CAW(2,2) model (bottom row); the gray shaded areas mark the two out-of-sample windows used in the forecast experiment and the dashed lines indicate the 95% Bartlett confidence bands for no serial dependence.

Figure 3.2: Predicted (Co)Variances, Correlations and Secular Components obtained under the MIDAS-CAW(2,2) Model



Predicted (co)variances $s_{i,j,t}$ and correlations $s_{i,j,t}/\sqrt{s_{i,i,t}s_{j,j,t}}$ for AXP ($i = 1$) and C ($j = 2$) stock, together with their predicted secular components $m_{i,j,t}$ and $m_{i,j,t}/\sqrt{m_{i,i,t}m_{j,j,t}}$ obtained under the MIDAS-CAW(2,2) model. The thick grey line represents the secular component.

Figure 3.3: Sample Autocorrelation Function of Standardized Residuals from Unrestricted MIDAS-CAW(2,2) Model



Sample autocorrelation function of the standardized residuals $e_{j,t}^*$ for the covariances r_{32} (GE-C), r_{42} (HD-C), r_{52} (IBM-GE), r_{54} (IBM-HD) and the variance r_{55} (IBM) from the unrestricted MIDAS-CAW(2,2) model; the dashed lines indicate the 95% Bartlett confidence bands for no serial dependence.

Chapter 4

Intra-Daily Volatility Spillovers between the US and German Stock Markets

4.1 Introduction

The particularly flexible mean dynamics of the conditional autoregressive Wishart (CAW) approach for realized covariance matrices, illustrated in the previous chapter, makes the model an ideal tool for analyzing volatility transmission effects. In order to investigate the short-term volatility transmission mechanism between the US and the German stock market the current chapter develops a novel three-phase model based upon the CAW approach. The direct modeling of realized variances and covariances offers the advantage that these measures are typically more informative about the true volatility than corresponding conditional (co)variances obtained from MGARCH or MSV models. The use of high-frequency data is therefore expected to result in improved inference on volatility transmission across markets.

A common finding of empirical studies devoted to asset-return variances and covariances across international financial markets is their high degree of contemporaneous and temporal interdependence. This interdependence, which plays an important role for international portfolio allocation and financial risk management, is often attributed to information transmissions across financial markets. This view is based upon the hypothesis that the arrival process of economic news and the trading dynamics in response to news are key determinants of the short-run dynamics of asset-return volatility (see, e.g., Kyle, 1985). Against the background of an apparently increasing integration of international financial markets it is interesting to see to what extent a volatility shock generated by news in one market spills over onto the volatility observed in the next market to trade. As noted by Hamao et al. (1990) and Wongswan (2006), such spillovers could represent a causal phenomenon across markets that trade sequentially; alternatively they could reflect shocks which are generated by news relevant to the global economy and impinging concurrently on the volatility across international markets. Similarly, it is of interest whether those spillover effects are more pronounced during periods of very high volatility associated with severe financial crises like that of 2007-2009. While this crisis had its origin in the US sub-prime mortgage market, it spread out increasing the volatility across international financial markets above and be-

yond a level which can be explained by a ‘regular’ fluctuation. A potential channel of such a volatility contagion is that an initially local crisis in one country generates news that prompt investors to fundamentally reassess the general vulnerability of other national markets (see, e.g., Bekaert, et al., 2011).

The strand of empirical literature concerned with volatility spillovers on international financial markets goes back to the early papers of Engle, et al. (1990) and Hamao, et al. (1990), in which GARCH models fitted to intra-day returns are used to measure the volatility transmissions from one period to the next within markets (‘heat waves’) and across markets (‘meteor showers’). The former study uses four intra-day returns per day of the yen-US dollar exchange rate associated with four distinct geographic market segments with non-synchronous trading hours (Tokyo, Europe, New York, Pacific), and reports significant spillovers between the different market segments, indicating that volatility in international markets behaves like a meteor shower. Hamao, et al. (1990) rely on close-to-open and open-to-close returns and find spillovers from the US to the Japanese stock market but not conversely.

More recent studies examining volatility transmissions between international markets use high-frequency return data in order to construct realized variances or ranges between the largest and smallest log prices as precise estimates for the volatility of low-frequency returns and model those estimates directly. Compared to conditional variances obtained from MGARCH or MSV models this offers the advantage that high-frequency based volatility measures provide additional information on the variability of the return process. Such approaches are found, e.g., in Engle, et al. (2012), Bubák, et al. (2011) for markets with synchronous trading hours and in Melvin and Melvin (2003), Dimpfl and Jung (2012), Chiang and Wang (2011), for markets with nonsynchronous business hours. The study of Engle, et al. (2012) uses a multivariate multiplicative error model (MEM) for the vector of daily volatilities approximated by the daily ranges and applies this approach to measure the volatility transmissions across eight East Asian stock markets and to examine changes in the transmission mechanism during the 1997-1998 East Asian crisis. In order to analyze the short-term interdependence of the realized variances for the exchange rates of four European currencies against the US dollar, Bubák, et al. (2011) propose a multivariate version of the heterogeneous autoregressive (HAR) model of Corsi (2009). Melvin and Melvin (2003) investigate volatility spillovers of the Deutsche mark-US dollar and yen-US dollar exchange rate across geographical market segments, while Dimpfl and Jung (2012) examine spillovers across the stock markets in Europe, the US and Japan. Both studies rely on structural vector autoregressive (VAR) models for the realized volatilities accounting for the time differences in trading hours of the markets under consideration. Using a range-based conditional autoregressive volatility model for the stock markets of the G7 countries, Chiang and Wang (2011) examine changes in the volatility transmission mechanism due to the subprime mortgage crises.

The present chapter investigates the short-term interdependence of the realized variances and covariance of the US Dow Jones and the German stock index DAX. For this purpose a novel sequential phase model is developed, which accounts for the three distinct geographical intra-day trading periods of the US and German stock market: (1) the Germany-US trading overlap period, (2) the US-only trading period, and (3) the Germany-only trading period. The approach consists of three separate reduced-form time-series specifications, one for each intra-day period. For the covariance matrix of the Germany-US trading overlap period the conditional autoregressive Wishart (CAW) model of Chapter 3 is extended to include the lagged variances of the other two intra-day periods as additional covariates. The two variances of the US-only and the Germany-only trading periods are assumed to follow a corresponding conditional autoregressive Gamma distribution, which obtains from the CAW model for the covariance matrix as a natural marginal specification for the variances. The resulting sequential three-phase model facilitates a detailed analysis of the short-term causal effects of news generating intra-day volatility in one market onto subsequent trading on this and the other market, i.e. both meteor-shower and heat-wave effects. The analysis of the direct causal effects is supplemented by an impulse-response analysis, which provides information not only about the direct but also the indirect effects of volatility shocks. As such, the impulse-response analysis also accounts, e.g., for the indirect effect of a volatility shock during the afternoon trading on the German market on its volatility at the next morning via the US trading which has taken place in the meantime. In addition, the framework is used in order to investigate whether the short-term volatility transmission mechanism is significantly different during the recent subprime crisis than before and after the crisis which would indicate volatility contagion effects.

By accounting for the interdependence between the variances as well as the covariance the illustrated approach differs from existing empirical studies on volatility transmissions across stock markets with overlapping trading hours like those of Engle, et al. (2012), Dimpfl and Jung (2012) and Chiang and Wang (2011), which solely investigate the dynamic interdependence of variance measures. As such, the approach allows to account for two potential channels of volatility spillovers, namely, via a direct volatility transmission from one market to the other through its variance and via an indirect transmission through its covariance. A reason to expect an indirect transmission channel via the covariance is the empirical evidence that variance shocks tend to increase international market correlation, as documented e.g. in Solnik, et al. (1996) and Forbes and Rigobon (2002). These correlation increases might then be transmitted to variances in subsequent periods as a reaction on the corresponding changes in diversification risks (see, e.g., Driessen, et al., 2009). Furthermore, the approach explicitly accounts for the contemporaneous interdependence between the variances during the overlapping trading periods which is ignored in the studies of Engle, et al. (2012) and Dimpfl and Jung (2012). In order to properly identify the direct causal effects of news on subsequent volatility on the domestic and foreign markets, it is critical to explicitly account for

the indirect effects transmitted via the covariance and for the contemporaneous dependence among the variances.

The outline of the chapter is as follows: Section 4.2 describes the schedule of trading hours of the German and US stock market and the adjustments made to remove long-run trend effects from the realized variance and covariance series. Section 4.3 introduces the sequential three-phase model and discusses its properties. Section 4.4 presents the design of the impulse response analysis for the three phase model. The empirical results are presented in Section 4.5, while Section 4.6 summarizes the findings.

4.2 Trading Times, Data and Adjustments

In order to investigate short-term volatility spillovers for the German and US stock market on an intra-daily basis the analysis has to account for their non-synchronous opening hours, leading to three distinct intra-day trading periods. The different trading intervals in Central European Time (CET) associated with different trading regimes are illustrated in Figure 4.1. Assume that a global business day t starts with the opening of the New York Stock Exchange at 3:30 pm CET. From 3:30 pm to 5:30 pm the US and German stock market are simultaneously open. This joint trading period of two hours length is referred to as period 1 of a trading day. In period 2 which lasts from 5:30 pm to 10:00 pm, only the US market is open. The last interval of a trading day referred to as period 3, starts at 9:30 am when the German market opens and ends with the re-opening of the US market at 3:30 pm.

To model the dynamic process of intra-daily volatilities accounting for this chronological ordering of overlapping and non-overlapping trading periods a sequential three-phase model is proposed, which treats the volatility for the three intra-day periods separately by specifying three dynamic reduced-form models, one for each intra-day period. In order to obtain volatility measures for the three intra-day periods, high-frequency data is used to construct realized variances and covariances as direct estimates of the corresponding variances and covariances of returns. The data consists of synchronized 1-minute prices sampled with previous-tick interpolation for the German stock index DAX and the US Dow Jones industrial index (DJ). The sample period begins at January 2, 1996 and ends on December 29, 2010 covering $T = 3645$ trading days. For intra-day period 1 the realized covariance matrix for the DAX and DJ can be computed as $V_{t,1}^{(us,g)} = \sum_{i=1}^{n_1} \xi_{t,i} \xi'_{t,i}$, where $\xi_{t,i} = (\xi_{t,i}^{(us)}, \xi_{t,i}^{(g)})$ is the vector of the DJ and DAX log returns computed for the 5-minute interval i in period 1 of trading day t . The number of 5-minute intervals in this period is $n_1 = 24$. In the sequel, the diagonal (variance) elements of the period-1 realized covariance matrix $V_{t,1}^{(us,g)}$ are denoted by $v_{t,1}^{(us)}$ and $v_{t,1}^{(g)}$ and the off-diagonal (covariance) element by $v_{t,1}^{(us,g)}$. The realized variance of the DJ in period 2 and that of the DAX in period 3 can be computed analogously as $v_{t,2}^{(us)} = \sum_{i=n_1+1}^{n_2} [\xi_{t,i}^{(us)}]^2$ and $v_{t,3}^{(g)} = \sum_{i=n_2+1}^{n_3} [\xi_{t,i}^{(g)}]^2$, respectively, where the number of 5-minute intervals for the second period is $n_2 - n_1 = 54$ and that for the third

period $n_3 - n_2 = 78$. These realized variance and covariance measures are further refined by averaging over subsampling subgrids per intra-day period in order to cope with market microstructure noise (see, Zhang, et al., 2005). The resulting measures are consistent estimators of the daily quadratic (co)variation, which is the sum of both continuous and jump components (see, e.g. Huang and Tauchen, 2005). However, as two broad stock indices are considered, price jumps in individual assets tend to be averaged out, such that the importance of the jump components for the indices can be expected to be limited. Finally, the realized variances and covariances of the three intra-day periods are normalized by the length of the respective intra-day period.

Figure 4.2 shows the time series plots of the resulting realized (co)variances for the three intra-day periods. These plots reveal a common cyclical long-term behavior across the five variance and covariance time series with its largest peak during the subprime crisis starting in 2008. A number of authors attribute those long-term shifts in the volatility to changes in the global macroeconomic and financial environment and interpret them as evidence against global stationarity – see, e.g., Engle, et al. (2009) and the literature cited therein¹. In order to capture those long-run movements Engle and Rangel (2008), Engle, et al. (2009) and Hafner and Linton (2010) use component volatility models with a long-run and a short-run component, where the former is associated with the state of the economy while the latter is related to day-to-day liquidity concerns and the arrival of news process triggering trading activities in response to news. Here it is not of interest to explain or model the long-term volatility rather than the short-term volatility transmission between the two stock markets from one intra-day period to the next. Therefore, the common long-run shifts are removed from the realized (co)variances prior to the analysis of the short-term patterns. Following Hafner and Linton (2010) a nonparametric two-sided kernel procedure is used to estimate the long-run components of the covariance matrix $V_{t,1}^{(us,g)}$ and the variances $v_{t,2}^{(us)}$ and $v_{t,3}^{(g)}$.² The corresponding estimates for the long-run components of the three intra-day periods are obtained as

$$M_{t,1} = \frac{\sum_{s=1}^T \mathcal{K}\left(\frac{t-s}{hT}\right) \cdot V_{t,1}^{(us,g)}}{\sum_{s=1}^T \mathcal{K}\left(\frac{t-s}{hT}\right)}, \quad m_{t,2} = \frac{\sum_{s=1}^T \mathcal{K}\left(\frac{t-s}{hT}\right) \cdot v_{t,2}^{(us)}}{\sum_{s=1}^T \mathcal{K}\left(\frac{t-s}{hT}\right)}, \quad m_{t,3} = \frac{\sum_{s=1}^T \mathcal{K}\left(\frac{t-s}{hT}\right) \cdot v_{t,3}^{(g)}}{\sum_{s=1}^T \mathcal{K}\left(\frac{t-s}{hT}\right)}, \quad (4.1)$$

respectively, where h denotes the bandwidth and $\mathcal{K}(\cdot)$ is a scalar-valued kernel function. Here a two-sided quartic kernel function is used and the bandwidth is set to $h = 0.05$, such that about 10% of the data are used for local averaging³. Note that the same weighting

¹For an initial investigation, the model proposed in Section 4.3 below has been fitted to the raw realized (co)variances $V_{t,1}^{(us,g)}$, $v_{t,2}^{(us)}$ and $v_{t,3}^{(g)}$ and it is found that they can not be represented by a covariance stationary specification due to an explosive behavior in their conditional means.

²In order to conduct out-of-sample forecasting, one could replace the two-sided by a one-sided kernel, as proposed by Hafner and Linton (2010, p. 65). However, the aim of the chapter is to identify short-term causal effects of news on subsequent foreign and domestic market volatilities.

³Experiments with other values for the bandwidth ranging from 0.025 to 0.075 showed that the qualitative results of the analysis of the short-term volatility patterns reported below remain essentially unchanged

scheme and bandwidth is used for all five variance and covariance time series, which imposes implicitly a common pattern in the long-run dynamics for all of them. This restriction could be justified by the finding that the long-run movements of the five time series appear to be very similar (see Figure 4.2).

To detrend the realized (co)variances, they are normalized by their estimated long-run components given by Equation (4.1) and plotted in Figure 4.2. In particular, the detrended realized covariance matrix for intra-day period 1, denoted by $R_{t,1}^{(us,g)}$, obtains as

$$R_{t,1}^{(us,g)} = C_{t,1}^{-1} V_{t,1}^{(us,g)} (C_{t,1}^{-1})', \quad \text{with} \quad M_{t,1} = C_{t,1} C_{t,1}', \quad (4.2)$$

where $C_{t,1}$ is the lower-triangular Cholesky factor of the period-1 long-run component $M_{t,1}$. The detrended realized variances of the intra-day periods 2 and 3 are constructed analogously by

$$r_{t,2}^{(us)} = v_{t,2}^{(us)} / m_{t,2}, \quad r_{t,3}^{(g)} = v_{t,3}^{(g)} / m_{t,3}, \quad (4.3)$$

respectively.

Figure 4.3 shows the plots of the detrended realized variance and covariance time series and reveals that the normalization makes them more homogeneous. This allows to focus on the short-term dynamic structure under the assumption of global stability. Descriptive statistics for the detrended realized (co)variance series are provided in Table 4.1. The mean of the variances is close to unity and that of the covariance close to zero, which is to be expected given the normalization rule given by Equations (4.2) and (4.3). The empirical distribution of the variances is leptokurtic and slightly skewed to the right while that of the covariance is skewed to the left. The Ljung-Box statistics including 50 lags indicate strong serial correlation.

4.3 Model Specification, Inference and Diagnostics

For the detrended volatility of the three intra-day periods given by $\{R_{t,1}^{(us,g)}, r_{t,2}^{(us)}, r_{t,3}^{(g)}\}_{t=1}^T$ a sequential three-phase approach is used to model the short-term dynamic structure of the volatility of the US and German stock market related to the news arrival process and to analyze the information transmission effects between and within markets. Since the sequentially ordered intra-day periods are non-overlapping, the volatility originating from previous intra-day periods is a pre-determined variable for the current period. This suggests the following sequential factorization of the conditional joint density of $(R_{t,1}^{(us,g)}, r_{t,2}^{(us)}, r_{t,3}^{(g)})$ given the information set \mathcal{F}_{t-1} available at the end of day $t-1$:

$$f(R_{t,1}^{(us,g)}, r_{t,2}^{(us)}, r_{t,3}^{(g)} | \mathcal{F}_{t-1}) = f(R_{t,1}^{(us,g)} | \mathcal{F}_{t-1}) \cdot f(r_{t,2}^{(us)} | R_{t,1}^{(us,g)}, \mathcal{F}_{t-1}) \cdot f(r_{t,3}^{(g)} | r_{t,2}^{(us)}, R_{t,1}^{(us,g)}, \mathcal{F}_{t-1}). \quad (4.4)$$

for those alternative values of the bandwidth.

Based upon this natural decomposition of the daily joint distribution, three separate reduced form models are specified, one for each of the three intra-day periods designed to measure the volatility transmission effects for the two stock markets from one intra-day period to the coming ones.

4.3.1 Period 1: Germany-US Trading Overlap

We start with detailing the specification for the covariance matrix of period 1, $R_{t,1}^{(us,g)}$. A particular convenient and flexible dynamic specification for this symmetric positive definite matrix of dimension 2×2 is provided by the CAW model illustrated in Chapter 3, which assumes a central Wishart distribution for $R_{t,1}^{(us,g)} | \mathcal{F}_{t-1}$. The specific CAW model for $R_{t,1}^{(us,g)}$ adopted here includes preceding period-2 and period-3 variances as additional explanatory variables and takes the form

$$R_{t,1}^{(us,g)} | \mathcal{F}_{t-1} \sim \mathcal{W}_2(\nu_1, S_{t,1}/\nu_1), \quad (4.5)$$

$$S_{t,1} = G_1 G_1' + \sum_{i=1}^{q_1} \sum_{\ell=1}^{\bar{q}_1} A_{i\ell,1} R_{t-i,1}^{(us,g)} A_{i\ell,1}' + \sum_{i=1}^{p_1} \sum_{\ell=1}^{\bar{p}_1} B_{i\ell,1} S_{t-i,1} B_{i\ell,1}' + \sum_{i=1}^{z_1} \sum_{\ell=1}^{\bar{z}_1} D_{i\ell,1} \bar{R}_{t-i}^{(us,g)} D_{i\ell,1}', \quad (4.6)$$

where \mathcal{W}_2 denotes the law of a central Wishart distribution for a 2×2 matrix, $\nu_1 > 2$ is the scalar degree of freedom, and $S_{t,1}/\nu_1$ represents the 2×2 positive definite scale matrix, such that the conditional mean is $E(R_{t,1}^{(us,g)} | \mathcal{F}_{t-1}) = S_{t,1}$. In the linear autoregressive recursion for the conditional mean (4.6), which resembles the BEKK-GARCH specification of Engle and Kroner (1995), $\bar{R}_t^{(us,g)} = \text{diag}(r_{t,2}^{(us)}, r_{t,3}^{(g)})$ is a diagonal matrix containing the variances of period 2 and 3, and G_1 , $A_{i\ell,1}$, $B_{i\ell,1}$ and $D_{i\ell,1}$ are 2×2 parameter matrices, where G_1 has a lower-triangular form. While the summation limits (q_1, p_1, z_1) determine the number of lagged terms, the limits $(\bar{q}_1, \bar{p}_1, \bar{z}_1)$ control the generality of the process. The most general process ensures that the number of parameters in the matrices $A_{i\ell,1}$, $B_{i\ell,1}$ and $D_{i\ell,1}$ is equal to the number of marginal effects of the different elements in the lagged $R_{t,1}^{(us,g)}$, $S_{t,1}$, and $\bar{R}_t^{(us,g)}$ matrices on the distinct elements in $S_{t,1}$. However, the model as specified is unidentified. Sufficient conditions for identification are given by Engle and Kroner (1995, Proposition 2.3). For a model with $(\bar{q}_1, \bar{p}_1, \bar{z}_1) = (1, 1, 1)$, for example, these conditions are that the main diagonal elements of G_1 and the first diagonal element for each of the matrices $A_{i1,1}$, $B_{i1,1}$ and $D_{i1,1}$ are restricted to be positive.

The contemporaneous dependence of the volatility for the two markets implied by the CAW model becomes manifest in the behavior of the conditional covariance matrix of the realized (co)variance, denoted by $\text{Var}[\text{vec}(R_{t,1}^{(us,g)}) | \mathcal{F}_{t-1}]$. Under the conditional Wishart distribution in Equation (4.5) this conditional covariance matrix is (see Muirhead, 1982)

$$\text{Var}[\text{vec}(R_{t,1}^{(us,g)}) | \mathcal{F}_{t-1}] = \frac{1}{\nu_1} (I + K_{44})(S_{t,1} \otimes S_{t,1}), \quad (4.7)$$

where I is the identity matrix, $\text{vec}(\cdot)$ denotes the operator that stacks all columns of a matrix into a vector, and K_{44} is the commutation matrix defined so that $K_{44}\text{vec}(W) = \text{vec}(W')$ for any 4×4 matrix W .

As discussed in Chapter 3, the CAW model (4.5)-(4.6) can be interpreted as a state-space model with $S_{t,1}$ as a state variable measured by the observable matrix $R_{t,1}^{(\text{us,g})}$ so that $S_{t,1}$ can be regarded as the ‘true’ integrated covariance matrix of period 1 for a broad class of continuous-time stochastic volatility processes approximated by $R_{t,1}^{(\text{us,g})}$ (see Barndorff-Nielsen and Shephard, 2004). The dynamic specification assumed for $S_{t,1}$ is designed to capture complex dynamic interactions across the covariance and variances for the returns of the US and German stock market of period 1 as well as their dependencies from the preceding variances for the corresponding returns of periods 2 and 3. It is easy to see that the direct volatility spillover effects from one market to future trading periods of the other market are directed by the non-diagonal elements in $A_{i\ell,1}$, $B_{i\ell,1}$ and $D_{i\ell,1}$ parameter matrices. For a specification with $q_1 = \bar{q}_1 = z_1 = \bar{z}_1 = 1$ and $p_1 = 0$ with parameter matrices $G_1 = (g_{.jk})$, $A_{11,1} = (a_{.jk})$ and $D_{11,1} = (d_{.jk})$, for example, the conditional mean of the period-1 DJ variance $s_{t,1}^{(\text{us})}$ obtains from Equation (4.6) as

$$s_{t,1}^{(\text{us})} = g_{.11}^2 + a_{.11}^2 r_{t-1,1}^{(\text{us})} + 2a_{.11}a_{.12}r_{t-1,1}^{(\text{us,g})} + a_{.12}^2 r_{t-1,1}^{(\text{g})} + d_{.11}^2 r_{t-1,2}^{(\text{us})} + d_{.12}^2 r_{t-1,3}^{(\text{g})}. \quad (4.8)$$

Hence, the effect of a shock in the period-3 DAX variance $r_{t-1,3}^{(\text{g})}$ to the next period-1 DJ variance is directed by $d_{.12}$ and that of a shock in the period-1 DAX variance $r_{t-1,1}^{(\text{g})}$ by $a_{.12}$, respectively. However, note that the contemporaneous correlation among the period-1 (co)variances $r_{t-1,1}^{(\text{g})}$, $r_{t-1,1}^{(\text{us})}$ and $r_{t-1,1}^{(\text{us,g})}$ (see Equation 4.7) implies that the shock in the period-1 DAX variance $r_{t-1,1}^{(\text{g})}$ can spill over onto the next day period-1 DJ variance also indirectly via the variance $r_{t-1,1}^{(\text{us})}$ and the covariance $r_{t-1,1}^{(\text{us,g})}$.

4.3.2 Periods 2 and 3: US-Only and Germany-Only Trading

Since the conditional Wishart distribution assumed for the period-1 realized covariance matrix $R_{t,1}^{(\text{us,g})}$ implies that its diagonal variance elements follow a conditional Gamma distribution it is natural to assume such a conditional Gamma distribution also for the realized variances of period 2 and 3. The particular reduced form model used for the period-2 DJ variance $r_{t,2}^{(\text{us})}$ including preceding period-1 covariance matrices and period-3 variances as additional explanatory variables takes the form

$$r_{t,2}^{(\text{us})} | R_{t,1}^{(\text{us,g})}, \mathcal{F}_{t-1} \sim \mathcal{G}(\nu_2/2, 2s_{t,2}/\nu_2), \quad (4.9)$$

$$s_{t,2} = g_2 + d'_{0,2} R_{t,1}^{(\text{us,g})} d_{0,2} + \sum_{i=1}^{q_2} a_{i,2} r_{t-i,2}^{(\text{us})} + \sum_{i=1}^{p_2} b_{i,2} s_{t-i,2} + \sum_{i=1}^{z_2} d'_{i,2} R_{t-i,1}^{(\text{us,g})} d_{i,2} + \sum_{i=1}^{w_2} c_{i,2} r_{t-i,3}^{(\text{g})}, \quad (4.10)$$

where \mathcal{G} denotes the law of a Gamma distribution, $\nu_2/2$ is the shape parameter of the Gamma distribution, and $2s_{t,2}/\nu_2$ represents its scale parameter such that $s_{t,2}$ is the conditional mean, i.e., $s_{t,2} = \mathbb{E}(r_{t,2}^{(us)} | R_{t,1}^{(us,g)}, \mathcal{F}_{t-1})$. The linear autoregressive recursion for $s_{t,2}$ given by Equation (4.10) is characterized by the scalar parameters $g_2, a_{i,2}, b_{i,2}, c_{i,2}$, which are restricted to be positive, and the two-dimensional parameter vectors $d_{i,2}$. As mentioned in the context of the CAW model above, $s_{t,2}$ can be interpreted as the true integrated variance of period 2 measured by $r_{t,2}^{(us)}$. The direct volatility transmission effects from the German stock market to the period-2 DJ volatility are driven by the second elements of the vectors $d_{i,2}$ and the parameters $c_{i,2}$. Additionally, we have an indirect transmission of an impulse in the period-1 DAX variance $r_{t-i,1}^{(g)}$ through its contemporaneous correlation with period-1 DJ variance $r_{t-i,1}^{(us)}$ and the corresponding covariance $r_{t-i,1}^{(us,g)}$.

The final component of the sequential three-phase model for the volatility of the US and German stock market consists of a reduced form specification for the period-3 variance of the DAX $r_{t,3}^{(g)}$, which takes a similar form as that for the period-2 DJ variance, namely

$$r_{t,3}^{(g)} | r_{t,2}^{(us)}, R_{t,1}^{(us,g)}, \mathcal{F}_{t-1} \sim \mathcal{G}(\nu_3/2, 2s_{t,3}/\nu_3), \quad (4.11)$$

$$s_{t,3} = g_3 + d'_{0,3} R_{t,1}^{(us,g)} d_{0,3} + c_{0,3} r_{t,2}^{(us)} + \sum_{i=1}^{q_3} a_{i,3} r_{t-i,3}^{(g)} + \sum_{i=1}^{p_3} b_{i,3} s_{t-i,3} + \sum_{i=1}^{z_3} d'_{i,3} R_{t-i,1}^{(us,g)} d_{i,3} + \sum_{i=1}^{w_3} c_{i,3} r_{t-i,2}^{(us)}. \quad (4.12)$$

Here the spillover parameters are given by the first elements of the two-dimensional vectors $d_{i,3}$ and the coefficients $c_{i,3}$ driving the direct transmission effects.

4.3.3 Model Properties, Estimation and Diagnostics

The three-phase model introduced in Equations (4.5)-(4.12) is expected to accommodate a large variety of dynamic patterns in the process of intra-day realized variances and covariances. In order to obtain the stability conditions of this process ensuring the existence of the stationary mean, its VARMA representation is used. Let $r_t = (\text{vech}(R_{t,1}^{(us,g)})', r_{t,2}^{(us)}, r_{t,3}^{(g)})'$ and $s_t = (\text{vech}(S_{t,1})', s_{t,2}, s_{t,3})'$, where $\text{vech}(\cdot)$ denotes the operator that stacks the lower triangular portion, including the diagonal of a matrix, into a vector. Then the system of interdependent recursions (4.6), (4.10), and (4.12) from one day to the next can be written as (see Chapter 3)

$$s_t = g + \Delta_0 r_t + \sum_{i=1}^q \mathcal{A}_i r_{t-i} + \sum_{i=1}^p \mathcal{B}_i s_{t-i}, \quad (4.13)$$

where g is a 5-dimensional vector and $\Delta_0, \mathcal{A}_i, \mathcal{B}_i$ are 5×5 matrices which are straightforward functions of the parameters characterizing the three recursions. The lag orders are $q = \max\{q_1, z_1, q_2, z_2, w_2, q_3, z_3, w_3\}$ and $p = \max\{p_1, p_2, p_3\}$. r_t can be written as $r_t = s_t + v_t$, where v_t is a martingale difference with $\mathbb{E}(v_t) = 0$ and $\mathbb{E}(v_t v_s') = 0$ for all $s \neq t$ so that the

VARMA representation of r_t obtains as

$$r_t = g^* + \sum_{i=1}^{\max\{p,q\}} (\mathcal{A}_i^* + \mathcal{B}_i^*) r_{t-i} - \sum_{i=1}^p \mathcal{C}_i^* v_{t-i}^* + v_t^*, \quad (4.14)$$

where $g^* = (I - \Delta_0)^{-1}g$, $\mathcal{A}_i^* = (I - \Delta_0)^{-1}\mathcal{A}_i$, $\mathcal{B}_i^* = (I - \Delta_0)^{-1}\mathcal{B}_i$, $\mathcal{C}_i^* = \mathcal{B}_i^*(I - \Delta_0)$ and $v_t^* = (I - \Delta_0)^{-1}v_t$. It follows that the three-phase model is stable with a stationary mean $E(r_t) = [I - \sum_{i=1}^{\max\{p,q\}} (\mathcal{A}_i^* + \mathcal{B}_i^*)]^{-1}g^*$ if and only if all eigenvalues of the matrix $\sum_{i=1}^{\max\{p,q\}} (\mathcal{A}_i^* + \mathcal{B}_i^*)$ are less than 1 in modulus (see, e.g. Lütkepohl, 2005).

The parameter vector θ of the three-phase model (4.5)-(4.12) is made up of the coefficients in the autoregressive specifications for the conditional means $S_{t,1}$, $s_{t,2}$, and $s_{t,3}$ plus the degree of freedom ν_1 and shape parameters ν_2 and ν_3 . They can be estimated by maximizing the log-likelihood function

$$\mathcal{L}(\theta) = \sum_{t=1}^T \ln f(R_{t,1}^{(us,g)} | \mathcal{F}_{t-1}) + \sum_{t=1}^T \ln f(r_{t,2}^{(us)} | R_{t,1}^{(us,g)}, \mathcal{F}_{t-1}) + \sum_{t=1}^T \ln f(r_{t,3}^{(g)} | r_{t,2}^{(us)}, R_{t,1}^{(us,g)}, \mathcal{F}_{t-1}). \quad (4.15)$$

Since there are no parametric restrictions across the log-likelihood components for the three intra-day periods, the complete log-likelihood can be maximized by separately maximizing the three components provided in Equation (4.15). Thus, the maximum likelihood estimation is conducted without imposing the stationarity constraints given above, which involve the parameters of all three components. Stationarity is then checked using the resulting estimates.

For identification of the orders for each of the three model components, that is $(q_1, p_1, z_1, \bar{q}_1, \bar{p}_1, \bar{z}_1)$ for the period-1 CAW specification, (q_2, p_2, z_2, w_2) for the period-2 Gamma model, and (q_3, p_3, z_3, w_3) for the period-3 Gamma model, Schwarz's (1978) information criterion is used. The order identification is supplemented by diagnostic checks based upon the standardized Pearson residuals. For the period-1 CAW model they obtain as

$$(u_{t,1}^{(us)}, u_{t,1}^{(us,g)}, u_{t,1}^{(g)})' = \text{Var}[\text{vech}(R_{t,1}^{(us,g)}) | \mathcal{F}_{t-1}]^{-1/2} \left\{ \text{vech}(R_{t,1}^{(us,g)}) - E[\text{vech}(R_{t,1}^{(us,g)}) | \mathcal{F}_{t-1}] \right\}, \quad (4.16)$$

where $\text{Var}[\text{vech}(R_{t,1}^{(us,g)}) | \mathcal{F}_{t-1}]^{-1/2}$ is the inverse Cholesky factor of the conditional covariance matrix of $\text{vech}(R_{t,1}^{(us,g)})$, given in Equation (4.7). The Pearson residuals for the period-2 and period-3 Gamma specifications are constructed analogously by

$$u_{t,2}^{(us)} = \frac{r_{t,2}^{(us)} - E(r_{t,2}^{(us)} | R_{t,1}^{(us,g)}, \mathcal{F}_{t-1})}{\text{Var}(r_{t,2}^{(us)} | R_{t,1}^{(us,g)}, \mathcal{F}_{t-1})^{1/2}}, \quad u_{t,3}^{(g)} = \frac{r_{t,3}^{(g)} - E(r_{t,3}^{(g)} | r_{t,2}^{(us)}, R_{t,1}^{(us,g)}, \mathcal{F}_{t-1})}{\text{Var}(r_{t,3}^{(g)} | r_{t,2}^{(us)}, R_{t,1}^{(us,g)}, \mathcal{F}_{t-1})^{1/2}}, \quad (4.17)$$

respectively. The corresponding conditional variances are given as

$$\text{Var}(r_{t,2}^{(\text{us})} | R_{t,1}^{(\text{us,g})}, \mathcal{F}_{t-1}) = \frac{2s_{t,2}^2}{\nu_2}, \quad \text{Var}(r_{t,3}^{(\text{g})} | r_{t,2}^{(\text{us})}, R_{t,1}^{(\text{us,g})}, \mathcal{F}_{t-1}) = \frac{2s_{t,3}^2}{\nu_3}. \quad (4.18)$$

For a correctly specified model, these residuals are serially uncorrelated and not predictable by past realized (co)variances. In order to test this implication, each of the residual series $u_t^{(\cdot)}$ can be regressed on a constant and past realized (co)variances of all three intra-day periods and the joint hypothesis that all coefficients other than the constant are equal to zero can be tested by using the F -statistic.

4.4 Impulse-Response Analysis

The marginal effects given by the entries of Δ_0 , \mathcal{A}_i , and \mathcal{B}_i in Equation (4.13) measure the direct causal impact of volatility shocks on future variances. However, as noted earlier, there are also indirect volatility transmission channels. In order to examine the compound effect of volatility shocks in one market on subsequent volatility in both markets an impulse-response (IR) analysis is used. Since the three-phase volatility model is nonlinear the standard IR technique à la Sims (1980) developed for linear time series models is not applicable. Hence, the nonlinear IR strategy of Gallant, et al. (1993) is applied which involves a comparison of forecasts obtained when perturbing the vector of conditioning arguments in the conditional density (conditional mean profile) to baseline forecasts produced without such a perturbation (baseline profile). Using this approach the effects of shocks to the DJ and DAX variances appearing in the three different intra-day periods, i.e., shocks to $r_{t,1}^{(\text{us})}$, $r_{t,1}^{(\text{g})}$, $r_{t,2}^{(\text{us})}$, and $r_{t,3}^{(\text{g})}$, are analyzed by tracing them period by period through the system.

To simplify the notation for the following presentation the two time indices used for the (co)variances, i.e. the index for the trading day $t \in \{1, 2, \dots, T\}$ and that for the intra-day period for a given trading day $m \in \{1, 2, 3\}$, are replaced by a single time index for the sequence of consecutive intra-day periods, say, $\tau = \tau(t, m)$ such that $\tau \in \{1, 2, \dots, 3T\}$; next set

$$\mathbf{r}_\tau = \begin{cases} (r_\tau^{(\text{us})}, r_\tau^{(\text{us,g})}, r_\tau^{(\text{g})})', & \text{if } \tau \text{ is a Germany-US overlap period } \tau(t, 1) \\ r_\tau^{(\text{us})}, & \text{if } \tau \text{ is a US-only trading period } \tau(t, 2) \\ r_\tau^{(\text{g})}, & \text{if } \tau \text{ is a Germany-only trading period } \tau(t, 3) \end{cases}, \quad (4.19)$$

and denote the lags of $\mathbf{r}_{\tau+1}$ by $x_\tau = (\mathbf{r}'_\tau, \mathbf{r}'_{\tau-1}, \dots)'$. Then the j -step-ahead forecast, $j = 1, 2, \dots$, at time period τ of the (co)variance $\mathbf{r}_{\tau+j}$ for a given value of the conditioning arguments x is

$$\hat{\mathbf{t}}_j(x) = \text{E}(\mathbf{r}_{\tau+j} | x_\tau = x). \quad (4.20)$$

Under the three-phase model, those forecasts are easily obtained by recursion based upon

the conditional expectations (4.6), (4.10) and (4.12). Let δ denote a perturbation to the contemporaneous value of \mathbf{r}_τ and define $x^0 = (\hat{\mu}', \hat{\mu}', \hat{\mu}', \dots)'$ and $x^+ = (\hat{\mu}' + \delta', \hat{\mu}', \hat{\mu}', \dots)'$, where $\hat{\mu}$ denotes the sample means of \mathbf{r}_τ associated with the three periods. Note that δ and $\hat{\mu}$ represent 3-dimensional vectors for period 1 and scalars for periods 2 and 3. Then the impulse response is defined in terms of the relative net effect of a perturbation δ , i.e.,

$$\left\{ [\hat{\mathbf{t}}_j(x^+) - \hat{\mathbf{t}}_j(x^0)] ./ \hat{\mathbf{t}}_j(x^0) \right\}_{j=1}^{\infty}, \quad (4.21)$$

where $./$ denotes the element-wise division.

In the application below, δ for a period-2 DJ shock and a period-3 DAX shock is set to unity, which is roughly one sample standard deviation of $r_{t,2}^{(us)}$, and $r_{t,3}^{(g)}$. Following Gallant et al. (1993) conditional expectations are used in order to specify the vector δ for a typical shock of unity to one of the period-1 variances accounting for the contemporaneous correlation structure among the period-1 variables. In particular, when considering a perturbation of unity to the period-1 DJ variance $r_\tau^{(us)}$, then the remaining entries of the vector δ are specified such that the corresponding elements of $\hat{\mu} + \delta$ are equal to the conditional expectations of $r_\tau^{(us,g)}$ and $r_\tau^{(g)}$ given the value for $r_\tau^{(us)}$. These conditional expectations are approximated by the non-parametric Nadaraya-Watson kernel smoother with a rule-of-thumb bandwidth selection (see, e.g., Li and Racine 2007, p. 66ff). $(1 - \alpha)$ percent confidence bands around the IR function are constructed by drawing a sample of 10,000 simulated values for the parameter vector θ from the asymptotic distribution of the ML estimator for θ . For each simulated θ value the IR function is computed and an interval is put around the IR function obtained for the ML estimates, just wide enough to include $(1 - \alpha)$ percent of the simulated IR functions (see, Gallant, et al. 1993).

4.5 Empirical Application

4.5.1 Estimation Results

Table 4.2 reports the ML parameter estimates for the three-phase model given by Equations (4.5), (4.6) and (4.9)-(4.12) fitted to the full sample data described in Section 4.2 together with the results of diagnostic checks on the standardized residuals defined in Equations (4.16) and (4.17). The orders of the model components have been selected using the Schwarz-information criterion and are given by $(q_1, p_1, z_1, \bar{q}_1, \bar{p}_1, \bar{z}_1) = (2, 3, 1, 1, 1, 2)$ for the period-1 component, $(q_2, p_2, z_2, w_2) = (2, 3, 0, 1)$ for the period-2 component and $(q_3, p_3, z_3, w_3) = (2, 3, 0, 0)$ for the period-3 component.

The largest eigenvalue of the estimated characteristic matrix $\sum_{i=1}^3 (\mathcal{A}_i^* + \mathcal{B}_i^*)$ (see Equation 4.14) is given by 0.92 indicating that the short-term volatility process of the German and US stock market across the three intra-day periods is stable in mean, though with a fairly high persistence in the conditional mean. The results of the F -test for residual predictability using

50 lags reveals that the model successfully accounts for the joint dynamics of the DAX and DJ volatility. The standardized residuals for all three intra-day periods pass the F -test at the 1% significance level. Further increasing the model order beyond the Schwarz-preferred specification did not significantly improve the results of the diagnostic checks.

Table 4.3 reports the implied ML estimates of the marginal effects, which are given by the elements of the matrices Δ_0 , \mathcal{A}_i and \mathcal{B}_i in the vector representation of the three-phase model (see Equation 4.13). They reveal evidence for both heat-wave and meteor-shower effects. Across all intra-day periods the variances of both markets depend significantly on their own lags (heat waves) as well as on the lagged variances of the other market (meteor showers). Next, the covariance of the DAX and DJ returns in period 1 ($r_{t,1}^{(us,g)}$) has a significant effect on the subsequent period-3 and period-1 DAX variances. Since this covariance itself depends significantly on lagged DJ and DAX variances, it appears to be a further volatility transmission channel in addition to the transmission directly through the variances.

In order to analyze the immediate causal impact of variance shocks, Figure 4.4 provides a diagram of the marginal effects of each intra-day variance on the next-period variance of the respective domestic and foreign market together with the associated estimates taken from Table 4.3. The diagram reveals that variance shocks in each intra-day period have a significant effect on the next-period variance of the home market. Furthermore, it appears that the importance of those heat-wave effects depends on the currency of the transmitted information: more recent domestic news from the immediately preceding intra-day period is more important than older domestic news from a period separated by a non-trading period.

Next, the results also show for both stock markets significant causal effects of news which has generated volatility abroad. In general, those meteor-shower effects are somewhat smaller than the heat-wave effects. Similar to the heat waves, the importance of the meteor showers critically depends on the currency of the information and, additionally, on whether the news is from a period with or without a trading overlap. Specifically, news causing volatility on the German market during the trading-overlap period 1 does not have a significant direct causal effect on the DJ variance in the next trading period 2. This implies that the volatility impulse of global news hitting the markets when they are simultaneously trading is transmitted immediately during period 1 via the US variance to the subsequent US trading, which reflects the US economy's leading role for international stock markets. In sharp contrast to the DAX volatility of the joint-trading period 1, the DAX volatility in period 3 when the US market is closed has a relatively strong and significant causal effect on the next day period-1 DJ variance. Note that this effect of the DAX shock on the period-1 US volatility is even larger than the period-1 US response to a domestic shock from the previous trading period. This relatively strong impact of the period-3 DAX volatility is consistent with the result reported by Dimpfl and Jung (2012), that the European markets morning trading has significant impact on the US volatility. It can be explained by the fact that during period 3 the German market processes and aggregates global news generated after the closing of the US

market (including news from the Asian markets), which hit the US market when re-opening at the next day in period 1 as new information.

Turning to the causal importance of the US market for the German volatility, we find a significant marginal effect of the US volatility in period 2 when the German market is closed on the DAX variance in the next trading period. Interestingly, this effect from the US to the German market appears to be significantly smaller than the corresponding effect of the German to the US market, namely the impact of a period-3 DAX shock on the next period-1 DJ volatility. This seems to contradict the general assessment that it is the US stock market which is the leading market. However, this apparent contradiction can be resolved by accounting for the fact that news generating US volatility in period 2 hits the German market with a time delay, which is caused by the time gap between the closing of the US market and the opening of the German market, and after the news is processed by the trading on the Asian markets; in contrast, the news generating DAX volatility in period 3 arrives the US market immediately without such a time delay.

Taken all together, we can conclude that the short-run volatility dynamics of the German and US stock market are driven by both heat-wave and meteor-shower effects and that the importance of their immediate impacts on the next period to trade critically depends on how current the corresponding news is.

4.5.2 Impulse-Response Analysis

Although the marginal effects discussed above are suggestive about the impact of volatility shocks on the future volatility on the home and foreign market, they do not provide the complete picture. In particular, consider, e.g., a shock on the DAX volatility in period 1. Its direct impact on the subsequent volatility of the US market in period 2 and on the German market in period 3 is measured by the respective marginal effects. However, the period-1 DAX shock may also influence the period-2 US volatility indirectly, namely through a simultaneous change of the period-1 US variance and the covariance. Similarly, there is an indirect effect of the period-1 on the period-3 DAX volatility via an increase of the period-2 US volatility. Information about the compound impact consisting of the direct and indirect effects of shocks on subsequent volatility is provided by the IR function defined in Section 4.4. The values of the perturbation δ and of the conditioning argument for the base case $x_0 = (\hat{\mu}', \hat{\mu}', \hat{\mu}', \dots)$ which are used to compute the IR functions according to Equation (4.21) are summarized in Table 4.4.

Figure 4.5 displays the multi-period IR functions tracing trading period by trading period the response to volatility shocks occurring in the three different intra-day phases. Shocks in all periods lead to significant responses on the respective home and foreign market which die out approximately after two months. In general, the responses of both markets to domestic shocks are larger than to foreign shocks. Next note that the immediate responses

after one trading period essentially confirm the results gleaned from the immediate marginal effects discussed above. A notable exception is the significant spillover of a period-1 DAX shock onto the DJ volatility, while the corresponding direct marginal effect was found to be negligible (see Figure 4.4). This implies that the period-1 DAX volatility influences the subsequent US trading only indirectly via its contemporaneous correlation with the period-1 DJ variance. Hence, it appears that spillovers of shocks occurring on the German market during trading-overlap periods onto the subsequent US trading mainly reflect global news impinging concurrently on the volatility of both markets, rather than a causal phenomenon.

4.5.3 Impact of the Subprime Crisis

Recent studies, including those of Bubák, et al. (2011), Chiang and Wang (2011) and Engle, et al. (2012) report evidence for significant changes of the dynamic volatility transmission mechanism on international financial markets during financial crises. This finding and the enormous impact of the recent subprime crisis on the international financial markets suggest to treat the period during the subprime crisis differently from the periods before and after the crisis. For this purpose the current section allows for structural changes in the volatility spillover mechanism modeled by the three-phase specification. A potential source of such changes could be the boost of intensity at which news hits the international financial markets when entering a turbulent crisis period – especially for news from the country where the crisis originated. Additionally, those changes could also reflect an increase of the time of trading it takes to settle the differences in the traders' interpretation of news within and across regional markets, which can be expected due to a higher level of general uncertainty and the increased amount of information to be processed during crisis episodes.

In order to analyze the effects of the subprime crisis on the short-term volatility transmission mechanism, the model specification is extended by allowing the parameter values to be different during the crises than before and after the crises, while using the same orders for the three model components as selected in Section 4.5.1 above. This is implemented by means of a subprime crisis dummy, say CR_t , i.e. by defining the i th parameter of the model as $\theta_i = (1 - CR_t) \cdot \theta_{i0} + CR_t \cdot \theta_{i1}$ for all i , where θ_{i0} denotes the value of the parameter before and after the crisis and θ_{i1} is the value during the crisis. Following Bekaert, et al. (2011), the subprime crisis is defined to start at August 2007 and to end March 2009, shown as the dark-gray shaded area in Figures 4.2 and 4.3.

The ML estimation results for this extended three-phase model reveal that the inclusion of the subprime-crisis dummy substantially improves the fit of the model. The p -value of the likelihood-ratio test of the null hypothesis that the parameter values during the crisis period are equal to those before and after the crisis equals $8.4 \cdot 10^{-14}$, indicating a strong rejection of the null. Next, the parameter estimates (not presented here) imply a significant increase of the largest eigenvalue of the characteristic matrix (see Equation 14) from 0.90 during the non-crisis periods to 0.96 during the crisis. Hence, the crisis leads to a substantial rise of the

general persistence of volatility shocks in the German and US stock market.

Figure 4.6 provides the estimates of the immediate marginal effects of domestic and foreign volatility shocks on the next-period variances for both the non-crisis and the crisis period. The results indicate that before and after the crisis those effects remain typically very close to those obtained under the model specification without a crisis-dummy (see Figure 4.4). Next, the comparison between the marginal effects obtained for the non-crisis periods and those during the crisis reveals that the crisis had a major impact on the short-term volatility transmission mechanism. In particular, the crisis is associated with a substantial strengthening of the meteor-shower effects, e.g. the size of the marginal effect of the period-3 DAX variance on the subsequent period-1 DJ variance increased during the crisis by 10% and that of the period-2 DJ on the period-3 DAX variance by even 58%. Obviously, this particularly large increase of the immediate causal effect of the DJ on the DAX variance, dominating the increase of the causal effects of the DAX on the DJ variance, reflects the fact that the subprime crisis had its origin in the US subprime mortgage market and spread out across international stock markets via various economic and financial links, including mortgage-backed securities widely held by financial firms all over the world. Due to those links the investors' need to closely monitor the US market increased during the crisis in order to gather new critical information about investments in the German market. This in turn might have significantly intensified the causal effects in particular of US business news on the volatility in the German market. The results concerning the meteor-shower effects during the crisis and the non-crisis periods are in line with the findings of Diebold and Yilmaz (2009) and Engle, et al. (2012) who also report a substantial strengthening of the inter-market volatility linkages of regional markets induced by financial crises.

Figure 4.7 provides the IR functions of shocks occurring in the three intra-day periods obtained for the crisis and non-crisis periods. Here the same values of the perturbation and of the conditioning arguments for the base case are used as selected in Section 4.5.2 (see Table 4.4). Note that Figure 4.7 does not display confidence bands for the crisis period. The reason for this is that the volatility model is very close to non-stability during the crisis episode, as indicated by the largest eigenvalue of the characteristic matrix given by 0.96. Hence, using the asymptotic normal distribution of the ML-estimator in order to simulate artificial values of the model parameters for the construction of the confidence bands (as described in Section 4.4) often leads to simulated values violating the stability condition. The comparison of the IR functions during the crisis with those for non-crisis periods shows that the crisis leads to a substantial increase of the time it takes for volatility shocks to die out. This is fully in line with the finding that during the crisis episode the largest eigenvalue of the characteristic matrix is well above its value before and after the crisis. As mentioned above, this general rise of persistence in the volatility process could reflect an intensified information clustering and/or an increase of time needed by the investors to process and interpret new information. In particular, while the looming crisis initially appeared to be a

local phenomenon bound to the US market of subprime mortgages, it became progressively evident that this crisis has a global dimension. International financial markets experienced a clustering of news concerning the potential devaluation of various investments, which finally culminated in the Lehman Brothers Inc. bust of 2008. This intensified news clustering was accompanied by a large and long lasting uncertainty about the crisis implications on the real and financial sector of the global economy, which hindered a fast interpretation and pricing of news across international financial markets.

4.6 Summary

The current chapter extends the conditional autoregressive Wishart (CAW) approach of Chapter 3 in order to assess the short-term interdependence of the realized variances and covariance of the non-synchronously traded US Dow Jones and German DAX stock market indices. The proposed volatility model contributes to the literature by explicitly accounting for the chronological ordering of overlapping and non-overlapping trading periods while embodying the realized covariance as an additional indirect transmission channel for volatility shocks and accounting for the contemporaneous interdependence between the (co)variances during partially overlapping trading times.

Considering the 15-year period from 1996 to 2010, the common long-run trend is removed from the realized (co)variance series in order to focus on short-run volatility transmission patterns. Besides the analysis of direct marginal effects, an impulse response analysis serves for the quantification of compound (direct and indirect) effects of volatility shocks onto volatilities in subsequent periods. The empirical results show that both own market heat-wave and cross market meteor-shower effects are present across all intra-day trading periods. In general, the impact of heat-wave effects is larger than of meteor-shower effects. Furthermore, the importance of those effects is found to depend critically on the information currency, that means more recent news appear to be more important than older news. Finally, the results indicate considerable changes in the short-run volatility transmission mechanism during the recent subprime crisis period with a substantially stronger persistence of volatility shocks. The subprime crisis period shows up with much more pronounced meteor-shower effects of volatility shocks from the U.S. market, where the crisis originates, to the German market volatility.

Table 4.1: Descriptive Statistics for Detrended Realized Variances and Covariance

	Period 1			Period 2	Period 3
	$r_{t,1}^{(us)}$	$r_{t,1}^{(g)}$	$r_{t,1}^{(us,g)}$	$r_{t,2}^{(us)}$	$r_{t,3}^{(g)}$
Mean	0.97	0.95	0.00	0.95	0.96
Std. dev.	1.06	1.02	0.42	1.40	1.05
Skewness	8.54	8.38	-1.15	23.39	10.33
Kurtosis	132.03	142.46	117.88	912.44	191.98
Minimum	0.07	0.04	-8.29	0.06	0.02
Maximum	24.77	26.57	8.91	60.71	27.55
LB(50)	2003.2311	2092.11	138.80	849.12	1969.47

Table 4.2: ML-Parameter Estimates for the BIC Selected Sequential Three-Phase CAW Model

Period-1 component								
param.	estimate		param.	estimate		param.	estimate	
$A_{11,1}$.346*	.004	$A_{21,1}$.344*	-.032			
	.023	.355*		-.079*	.221*			
$B_{11,1}$.001	-.001	$B_{21,1}$.087	.066	$B_{31,1}$.274*	-.011
	-.001	.001		.030	.277*		-.018	.390*
$D_{11,1}$.504*	.598*	$D_{12,1}$	**	**			
	-.075*	.139*		.223*	.650*			
ν_1	9.864*							
Period-2 component								
param.	estimate		param.	estimate		param.	estimate	
$a_{1,2}$.136*		$a_{2,2}$.094*		$b_{3,2}$.119*	
$b_{1,2}$	<.001		$b_{2,2}$.122*				
$c_{1,2}$.141*							
$d'_{0,2}$.618*							
ν_2	6.370*							
		-.020						
Period-3 component								
param.	estimate		param.	estimate		param.	estimate	
$a_{1,3}$.217*		$a_{2,3}$.066*		$b_{3,3}$.129*	
$b_{1,3}$.102		$b_{2,3}$.035				
$c_{0,3}$.161*							
$d'_{0,3}$.186*							
ν_3	8.307*							
		.421*						
Log-lik.: -6217.12			BIC: 12861.22			Max. eigenvalue: 0.92		
p-values for F-test on residual predictability (50 lags)								
$u_{t,1}^{(us)}$	$u_{t,1}^{(g)}$	$u_{t,1}^{(us,g)}$	$u_{t,2}^{(us)}$	$u_{t,3}^{(g)}$				
0.037	0.015	0.105	1.000	0.825				

Note: selected model orders are $(q_1, p_1, z_1, \bar{q}_1, \bar{p}_1, \bar{z}_1) = (2, 3, 1, 1, 1, 2)$ in period 1; $(q_2, p_2, z_2, w_2) = (2, 3, 0, 1)$ in period 2; $(q_3, p_3, z_3, w_3) = (2, 3, 0, 0)$ in period 3; ** indicates identifying restrictions setting parameter values to zero. The max. eigenvalue refers to the estimated matrix $\sum_{i=1}^{\max(p,q)} (\mathcal{A}_i^* + \mathcal{B}_i^*)$ in Equation (4.14); * denotes significance at the 1% level.

Table 4.3: ML-Estimates of the Marginal Effects for the BIC Selected Sequential Three-Phase Model

Period-1 component								
Depend. Variable	$s_{t-1,1}^{(us)}$	$s_{t-1,1}^{(g)}$	$s_{t-1,1}^{(us,g)}$	$r_{t-1,1}^{(us)}$	$r_{t-1,1}^{(g)}$	$r_{t-1,1}^{(us,g)}$	$r_{t-1,2}^{(us)}$	$r_{t-1,3}^{(g)}$
$s_{t,1}^{(us)}$	<.001	<.001	<.001	.120*	<.001	.003	.254*	.357*
$s_{t,1}^{(g)}$	<.001	<.001	<.001	<.001	.126*	.016	.055*	.442*
$s_{t,1}^{(us,g)}$	<.001	<.001	<.001	.008	.001	.123*	-.038*	.083*
Depend. Variable	$s_{t-2,1}^{(us)}$	$s_{t-2,1}^{(g)}$	$s_{t-2,1}^{(us,g)}$	$r_{t-2,1}^{(us)}$	$r_{t-2,1}^{(g)}$	$r_{t-2,1}^{(us,g)}$		
$s_{t,1}^{(us)}$.008	.004	.012	.119*	.001	-.022		
$s_{t,1}^{(g)}$	<.001	.077*	.017	.006	.049*	-.035*		
$s_{t,1}^{(us,g)}$.003	.018	.026	-.027*	-.007	.079*		
Depend. Variable	$s_{t-3,1}^{(us)}$	$s_{t-3,1}^{(g)}$	$s_{t-3,1}^{(us,g)}$					
$s_{t,1}^{(us)}$.075*	<.001	-.006					
$s_{t,1}^{(g)}$	<.001	.152*	-.014					
$s_{t,1}^{(us,g)}$	-.005	-.004	.107*					

Period-2 component						
Depend. Variable	$s_{t-1,2}^{(us)}$	$r_{t-1,2}^{(us)}$	$r_{t-1,3}^{(g)}$	$r_{t,1}^{(us)}$	$r_{t,1}^{(g)}$	$r_{t,1}^{(us,g)}$
$s_{t,2}^{(us)}$	<.001	.136*	.141*	.382*	<.001	-.025
Depend. Variable	$s_{t-2,2}^{(us)}$	$r_{t-2,2}^{(us)}$	$r_{t-2,3}^{(g)}$	$s_{t-3,2}^{(us)}$		
$s_{t,2}^{(us)}$.122*	.094*	<.001	.119*		

Period-3 component						
Depend. Variable	$s_{t-1,3}^{(g)}$	$r_{t,2}^{(us)}$	$r_{t-1,3}^{(g)}$	$r_{t,1}^{(us)}$	$r_{t,1}^{(g)}$	$r_{t,1}^{(us,g)}$
$s_{t,3}^{(g)}$.102	.162*	.217*	.035*	.177*	.157*
Depend. Variable	$s_{t-2,3}^{(g)}$	$r_{t-1,2}^{(us)}$	$r_{t-2,3}^{(g)}$	$s_{t-3,3}^{(g)}$		
$s_{t,3}^{(g)}$.035	<.001	.066*	.129*		

Note: the marginal effects are given by the elements of the coefficient matrices Δ_0 , \mathcal{A}_i and \mathcal{B}_i characterizing the vector representation of the three-phase model – see Equation (4.13); * denotes significance at the 1% level.

Table 4.4: Shock Scenario of the Impulse Response Analysis

Shock	$\hat{\mu}'$			δ'		
	(us)	(us,g)	(g)	(us)	(us,g)	(g)
period-1 DAX	0.97	0.00	0.95	0.69	0.07	1
period-1 DJ	0.97	0.00	0.95	1	-0.02	0.63
period-2 DJ	0.95			1		
period-3 DAX			0.96			1

Figure 4.1: Daily Trading Hours for the US (DJ) and the German (DAX) Stock Markets

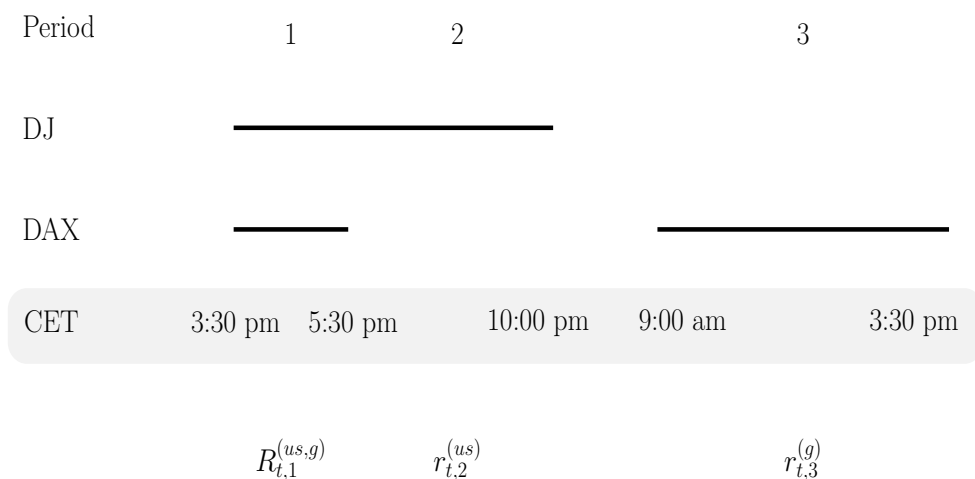
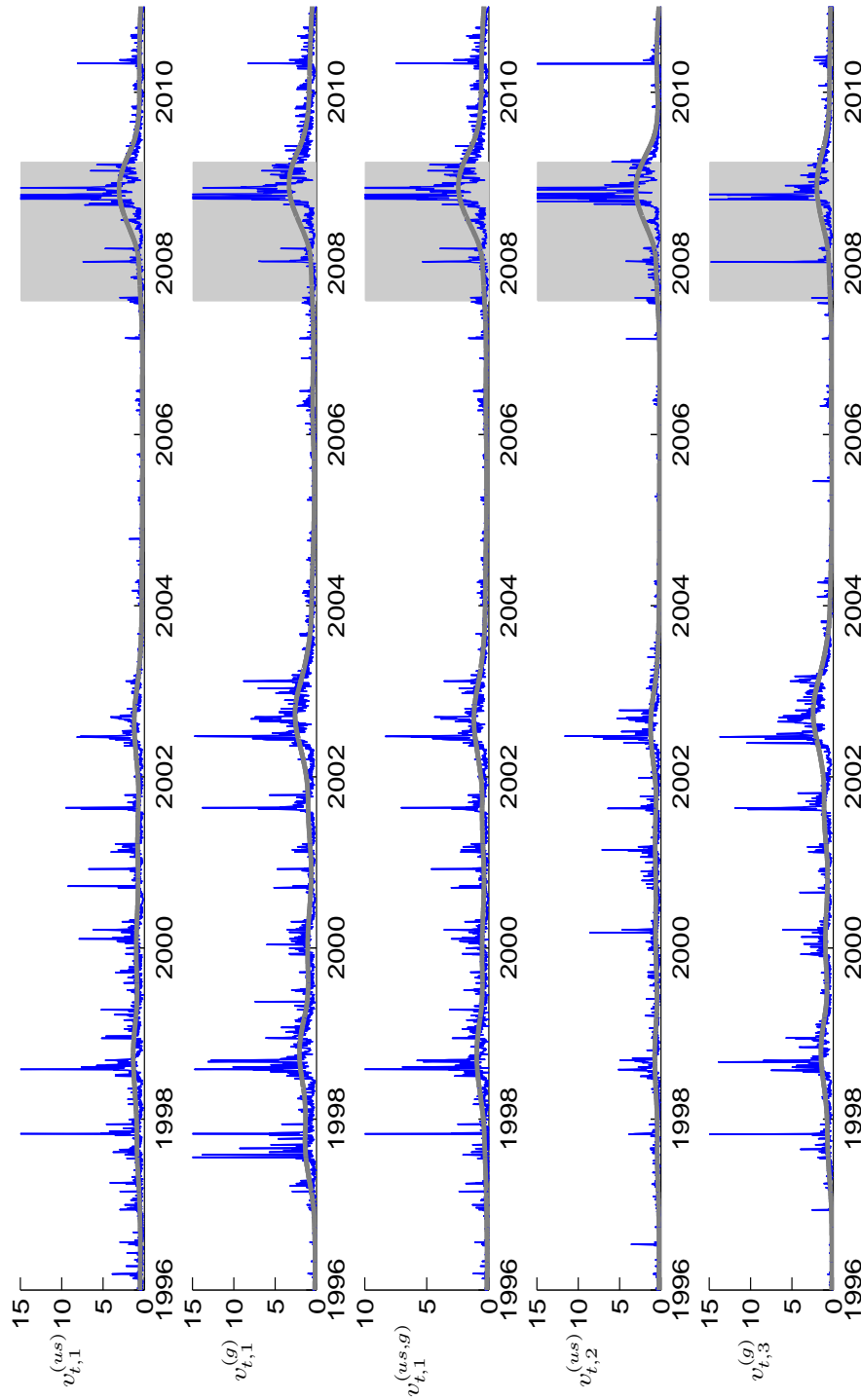
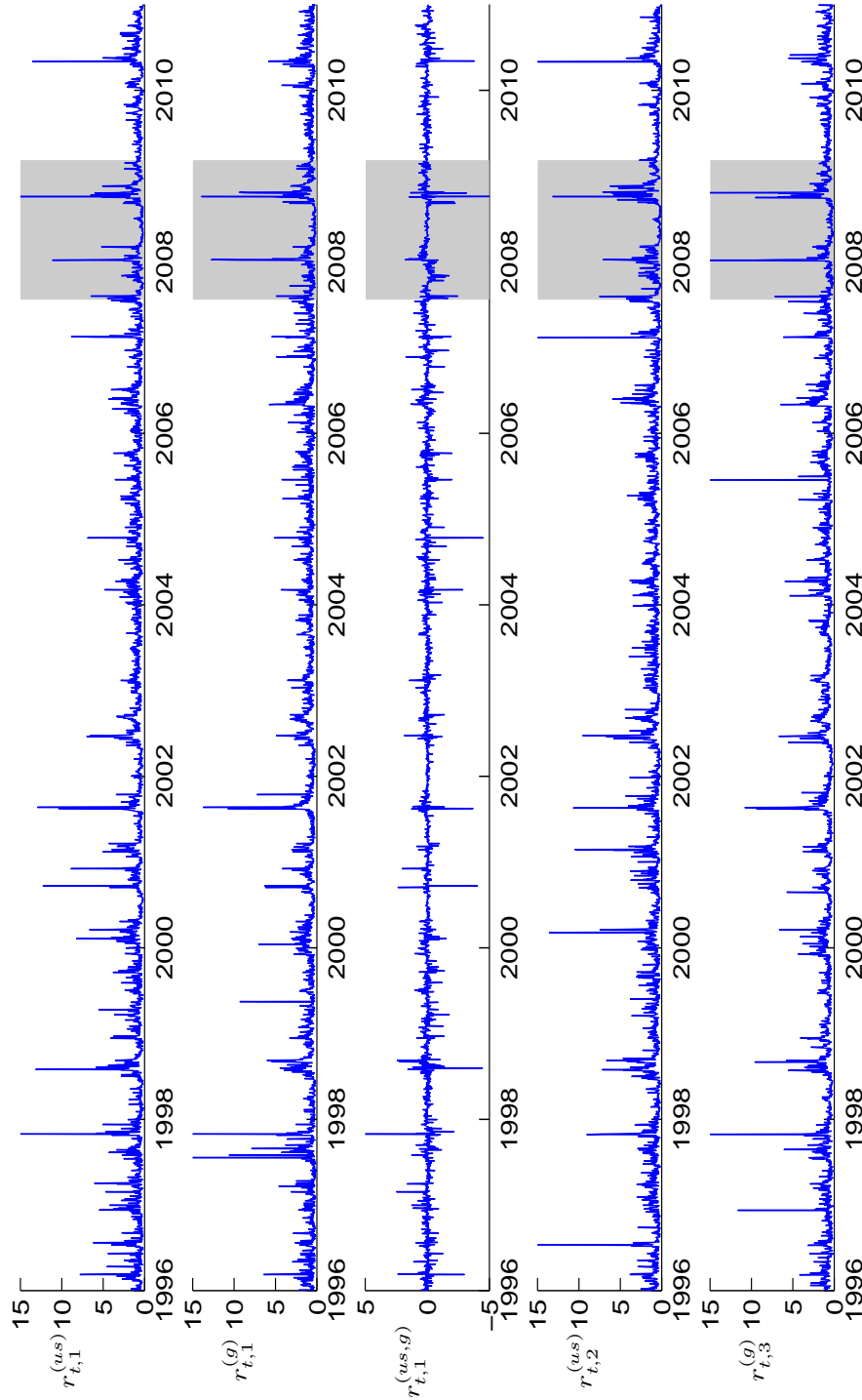


Figure 4.2: Time Series of Realized (Co)Variances and Long-Run Component for DAX and Dow Jones Stock Indices



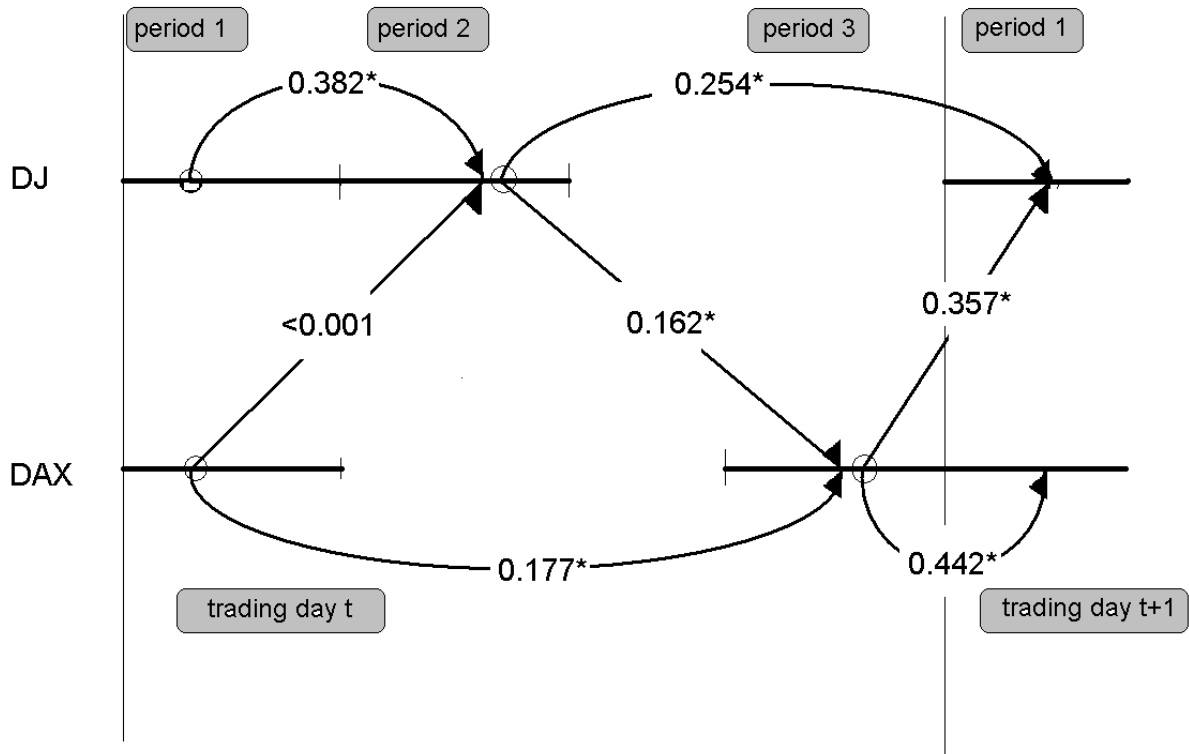
Time series of the realized (co)variances for the German DAX (g) and US Dow Jones (us) stock indices for the three intraday periods as in Figure 4.1 together with their estimated long-run component (thick gray line). The gray shaded area marks the subprime crisis period as defined in Section 4.5.

Figure 4.3: Time series of Detrended Realized (Co)Variances for DAX and Dow Jones Stock Indices



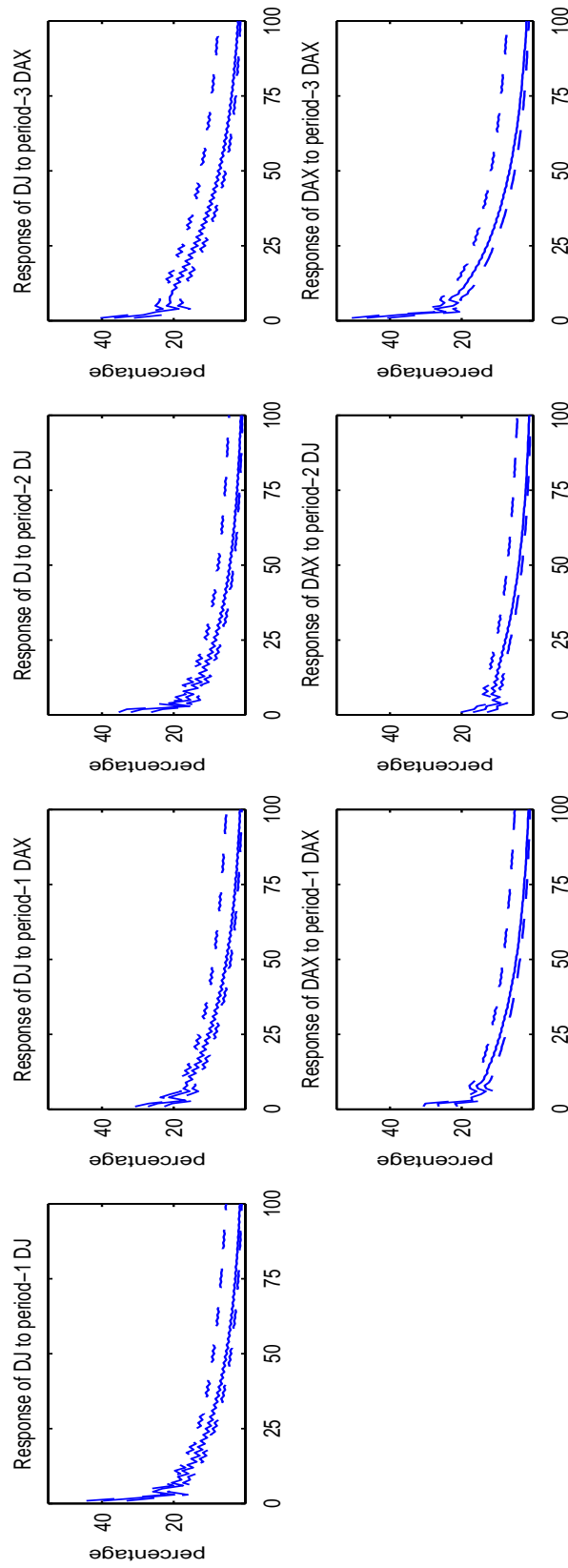
Time series of the detrended realized (co)variances for the German stock index DAX (g) and US Dow Jones stock index (us) for the three intra-day periods illustrated in Figure 4.1. The gray shaded area marks the subprime crisis period as defined in Section 4.5.

Figure 4.4: Estimates of the Direct Marginal Effects of Intra-Day Variances on the Next-Period Variance at Home and Abroad



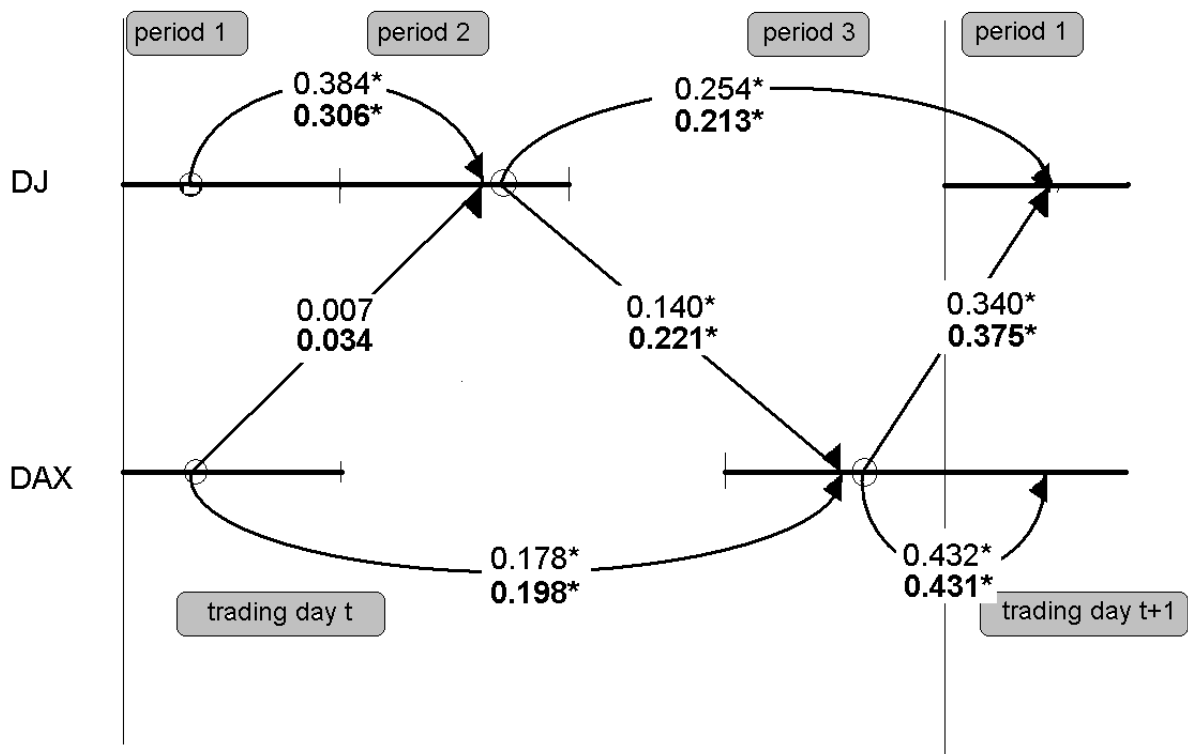
* denotes significance at the 1% level.

Figure 4.5: Estimated Impulse Response Functions



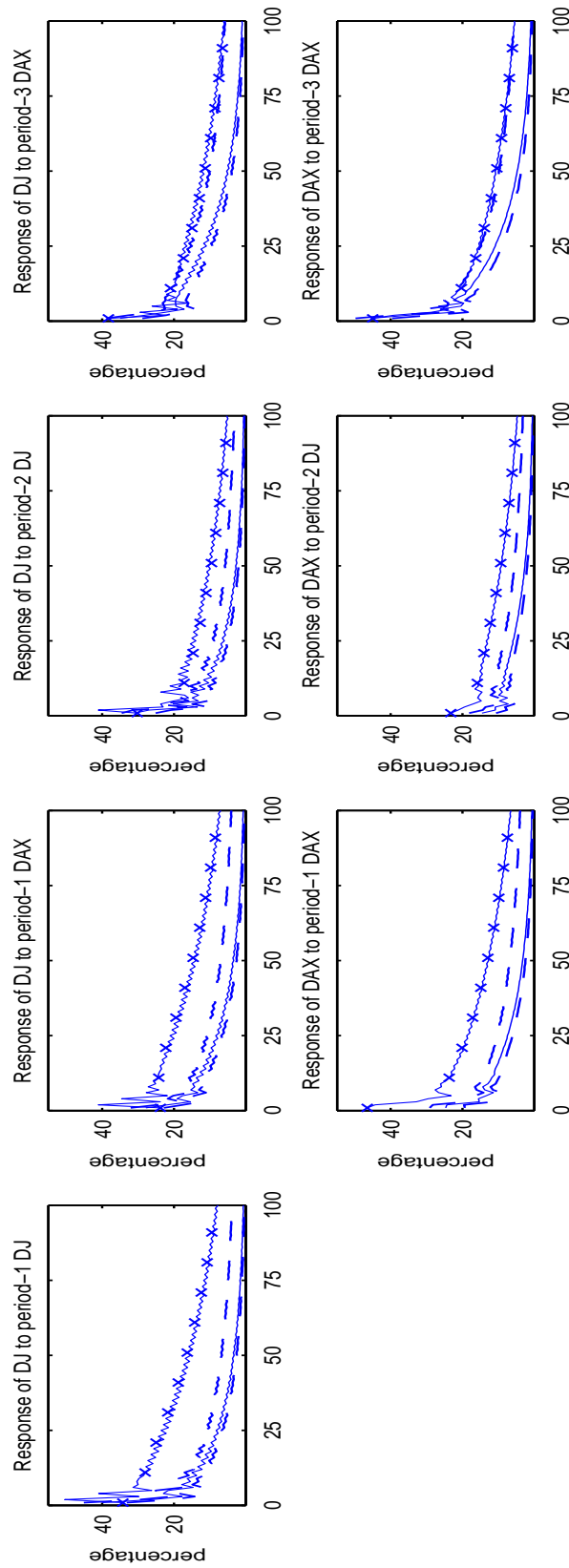
Estimated IR functions computed according to Equation (4.21). The dashed lines indicate 99% confidence bounds.

Figure 4.6: Estimates of Direct Marginal Effects during Non-Crisis Periods and during the Subprime Crisis



Estimates of the direct marginal effects of intra-day variance on the next-period variance at home and abroad during non-crisis periods and during the subprime crisis (bold numbers); * significant at the 1% level.

Figure 4.7: Estimated IR Functions during Non-Crisis Periods and during the Subprime Crisis



Estimated IR functions during non-crisis periods (solid line) and during the subprime crisis (crossed line) computed according to Equation (4.21). The dashed lines indicate 99% confidence bounds for non-crisis IR functions.

Chapter 5

A Latent Dynamic Factor Approach to Multivariate Stock Market Volatility

5.1 Introduction

The previous chapters discussed multivariate volatility models, which are specifically designed to capture complex serial and cross-sectional dependencies of variances and covariances of asset returns. Since the dimension of the object of interest is proportional to the square of the number of assets, these models tend to be heavily parameterized. This renders inference on the volatility dynamics complicated even for moderately sized portfolios. Yet empirical applications, e.g. the forecasting of optimal portfolio weight vectors in mean/variance portfolio optimization, typically require forecasts of high-dimensional covariance matrices. Models for the realized covariance matrix are fitted directly to time series of $k(k+1)/2$ realized (co)variances, while GARCH and MSV models are estimated based on the returns of k assets only. Hence, the number of observations per parameter is significantly larger for realized volatility models than for similarly parameterized GARCH and MSV specifications, such that the curse-of-dimensionality appears to be less acute. Nevertheless, the problem remains a main challenge. Chiriac and Voev (2011) use a fractionally integrated VARMA process to model the elements of the Cholesky factor of the realized covariance matrix. The model suffers from the curse of dimensionality and the authors restrict their empirical application to six assets. Several models proposed for realized covariance matrices are based on the conditional Wishart distribution (see Gouriéroux et al., 2009, Jin and Maheu, 2011, Noureldin et al., 2011, and the CAW model illustrated in Chapters 3 and 4). Empirical applications of the models are overall limited to portfolios of up to ten assets. Bauer and Vorkink (2011) propose a factor model for the distinct elements of the matrix logarithm of the covariance matrix. The factors are driven by lagged volatilities, lagged returns and other forecasting variables. Although the model could in general be applied to forecast high-dimensional covariance matrices, the authors restrict their application to a 5-dimensional case and focus on discussing the predictive power of the various forecasting variables. Since the matrix-log is a nonlinear function bias-correction methods have to be applied for the forecasting of volatilities. The non-linear nature of the model furthermore

complicates the analysis of the impact of the forecasting variables on volatility. Recently Bauwens and Storti (2011) proposed the CAW model featuring DCC dynamics, which allows for the forecasting of high-dimensional covariance matrices. Their empirical application comprises realized (co)variances for 50 NYSE stocks. Although the CAW-DCC model tackles the curse of dimensionality, this achievement does not come without a cost: the model imposes heavy restrictions on the correlation dynamics.

The present chapter proposes a novel flexible latent dynamic factor model for realized covariance matrices. The factor specification is motivated by persistent common dynamics of realized (co)variance series. The model is based on the matrix logarithm function which enables the modeling of log-(co)variances in Euclidean space, similar to the approach of Bauer and Vorkink (2011), preserving positive definiteness and symmetry of covariance matrix forecasts without the necessity of imposing restrictions on the parameter space. By modeling the dynamics of the common factors as heterogeneous autoregressive processes (HAR, see Corsi, 2009) and assuming AR(1) processes for the idiosyncratic dynamics the model mitigates (though not eliminates) the curse of dimensionality while allowing for rich (co)variance dynamics and can be readily applied to the forecasting of high-dimensional covariance matrices (say, ≤ 30 assets). In contrast to the observation driven approach of Bauer and Vorkink (2011) the proposed latent factor model allows for idiosyncratic (co)variance dynamics and offers enhanced flexibility in fitting the characteristics of the observed (co)variance series. The simulated Bayesian estimation approach using Markov Chain Monte Carlo (MCMC) techniques enables straightforward estimation of the model parameters and forecasting of covariance matrices without having to rely on bias correction methods as in Bauer and Vorkink (2011). Since the elements of the matrix logarithm of a covariance matrix can be interpreted as approximations to correlations and logarithmic variances, the factor model allows to investigate the presence of joint risk-factors related to market-risk and diversification risk (see e.g. Krishnan et al., 2009, and Driessen et al., 2009). In order to assess the model's forecasting performance a comprehensive out-of-sample experiment is conducted including a range of prominent forecasting models from the relevant literature. Besides a statistical evaluation based on the mean squared error criterion the chapter also addresses the practitioners point of view by investigating the performance of mean-variance optimal portfolios selected using the various forecasting models. An application to two data sets of 5- and 30-dimensional covariance matrices of NYSE traded stocks shows that the model outperforms proposed volatility models of the extant literature in out-of-sample forecasting for low as well as high-dimensional covariance matrices.

The outline of the chapter is as follows: Section 5.2 reviews the matrix logarithm and introduces the factor model, its estimation and model diagnostic tests. The empirical application to NYSE data is presented in Section 5.3. Section 5.4 concludes. Details on parameter estimation are provided in the appendix.

5.2 Model Specification, Inference and Diagnostics

5.2.1 The Matrix Logarithm

The current chapter is concerned with modeling the dynamics of the $n = k(k+1)/2$ distinct elements of the $k \times k$ matrix logarithm $Y_t = (y_{ij,t})$ of a time-varying $k \times k$ symmetric positive definite realized covariance matrix $R_t = (r_{ij,t})$ recorded at time t ($t = 1, \dots, T$).¹ The matrix logarithm is the inverse function of the matrix exponential, which is defined by the power series expansion

$$R_t = \text{expm}(Y_t) = \sum_{q=0}^{\infty} \frac{Y_t^q}{q!}, \quad (5.1)$$

where Y_t^0 is the identity matrix and Y_t^q denotes standard matrix multiplication of Y_t q times. From the spectral decomposition $R_t = L_t D_t L_t'$ we directly obtain $Y_t = \text{logm}(R_t) = L_t \ln(D_t) L_t'$, where $\ln(D_t)$ denotes a diagonal matrix of log-eigenvalues and L_t the corresponding matrix of eigenvectors. Taking the matrix logarithm of a real, positive definite matrix R_t results in a real, symmetric matrix Y_t and applying the matrix exponential function to a real symmetric matrix results in a real symmetric positive definite matrix (see Chiu et al., 1996, Lemma 1).

Denote the vector of the n distinct elements of the logarithmic covariance matrix Y_t by $y_t = \text{vech}(Y_t)$, where $\text{vech}(\cdot)$ is the operator that stacks the lower triangular portion including the diagonal of a matrix into a vector. The direct modeling of the $\{y_t\}_{t=1}^T$ series in Euclidean space proves convenient since the requirement of positive definiteness and symmetry of covariance matrices is readily fulfilled by the matrix exponential function.

As argued by Chiu et al. (1996) and Bauer and Vorkink (2011) there is no direct interpretation of the matrix logarithm in applications to covariance matrices. The elements of y_t can nevertheless be interpreted as approximations to correlations and logarithmic variances. Denoting the ij 'th element of the matrix Y_t^q by $y_{ij,t}^{[q]}$ and applying standard matrix multiplication to obtain $y_{ij,t}^{[q]} = \sum_{z=1}^k y_{iz,t} y_{zj,t}^{[q-1]}$ for $q \geq 2$ we can use Eq. (5.1) in order to write $\forall i = 1, \dots, k$

$$r_{ii,t} = \sum_{q=0}^{\infty} \frac{1}{q!} y_{ii,t}^{[q]} = 1 + y_{ii,t} + \sum_{q=2}^{\infty} \frac{1}{q!} \left[\sum_{z=1}^k y_{iz,t} y_{zi,t}^{[q-1]} \right] \quad (5.2)$$

$$= \sum_{q=0}^{\infty} \frac{1}{q!} y_{ii,t}^q + \nu_{ii,t} \quad (5.3)$$

$$= \exp(y_{ii,t}) + \nu_{ii,t}, \quad (5.4)$$

¹The realized covariance matrix R_t is computed by adding up the outer products of high-frequency (e.g. 5-minute) log-return vectors within a given day t (for details see Chapter 1, Section 1.1.3, and Section 5.3.1 below).

where repeated substitution reveals that $\nu_{ii,t} = \sum_{q=2}^{\infty} \frac{1}{q!} [\sum_{j=1}^{q-1} y_{ii,t}^{j-1} \sum_{z \neq i} y_{iz,t} y_{zi,t}^{[q-j]}]$. Hence $r_{ii,t} \cong \exp(y_{ii,t})$ and the diagonal elements of Y_t are approximations to logarithmic variances, where the approximation error $\nu_{ii,t}$ is a function of cross-products of the elements in Y_t . The properties of the approximation error, and hence the quality of the approximation itself, must be assessed using the specific data set at hand. Using $r_{ii,t} \cong \exp(y_{ii,t})$ and denoting the correlation coefficients by $(\rho_{ij,t})$, we obtain for $i \neq j$

$$r_{ij,t} = \rho_{ij,t} \sqrt{r_{ii,t} r_{jj,t}} \cong \rho_{ij,t} \exp\left(\frac{1}{2}(y_{ii,t} + y_{jj,t})\right). \quad (5.5)$$

Approximating $\exp\left(\frac{1}{2}(y_{ii,t} + y_{jj,t})\right)$ by a 1'st order TSE around $\frac{1}{2}(y_{ii,t} + y_{jj,t}) = 0$, we arrive at

$$r_{ij,t} \cong \rho_{ij,t} \left(1 + \frac{1}{2}(y_{ii,t} + y_{jj,t})\right) \quad \forall i \neq j. \quad (5.6)$$

Truncating the power series expansion of Eq. (5.1) at the second order, we obtain

$$\begin{aligned} r_{ij,t} &\cong y_{ij,t} + \frac{1}{2} \sum_{z=1}^k y_{iz,t} y_{zj,t} \\ &\cong y_{ij,t} + \frac{1}{2} y_{ij,t} (y_{ii,t} + y_{jj,t}) = y_{ij,t} \left(1 + \frac{1}{2}(y_{ii,t} + y_{jj,t})\right) \quad \forall i \neq j, \end{aligned} \quad (5.7)$$

where the second equation follows from the first by setting $y_{iz,t} y_{zj,t} = 0$ for $z \notin \{i, j\}$. Eqs. (5.6) and (5.7) imply $y_{ij,t} \cong \rho_{ij,t}$. The quality of this approximation depends on the quality of the log-variance approximation by the diagonal elements of the matrix logarithm and the overall variance level (see the 1'st order TSE around $\frac{1}{2}(y_{ii,t} + y_{jj,t}) = 0$ leading to Eq. 5.6). Section 5.3.1 below analyzes the quality of the approximation for a time-series of 5-dimensional covariance matrices. The results indicate that the elements of the logarithmic covariance matrices capture the dynamics of correlations and log-variances to a great extent.

5.2.2 The Dynamic Factor Model

The $k \times k$ realized covariance matrix R_t consistently estimates the latent integrated covariance matrix Σ_t of the k -dimensional period- t log-return vector ξ_t .² The literature on realized volatility modeling typically finds that realized variances and covariances as well as their logarithmic counterparts feature a common long-memory type of dependence pattern (see e.g. Bauer and Vorkink, 2011, Chiriac and Voev, 2011, Chapter 3, Section 3.2.4, and Figures 5.1 and 5.3 of the current chapter). Motivated by persistent common dynamics in logarithmic realized variances and covariances, I assume a persistent latent common factor structure for the n distinct elements of the matrix logarithm $x_t = \text{vech}(\log(\Sigma_t))$. Given the series of vectorized logarithmic realized covariance matrices $\{y_t\}_{t=1}^T$ the resulting state-space model

²See e.g. Chapter 1, Section 1.1.3, Andersen et al., 2003, Barndorff-Nielsen and Shephard, 2004, and the very general assumptions on the log-price process therein.

reads

$$x_t = a + B^c f_t^c + w_t \quad (5.8)$$

$$y_t = x_t + u_t, \quad u_t \stackrel{iid}{\sim} N(0, \Sigma_u), \quad (5.9)$$

where $a = (a_1, \dots, a_n)'$ is a vector of constants, B^c is a matrix of factor loadings

$$B^c = \begin{pmatrix} b_{1,1}^c & b_{1,2}^c & \dots & b_{1,p}^c \\ b_{2,1}^c & b_{2,2}^c & \dots & b_{2,p}^c \\ \vdots & & & \vdots \\ b_{n,1}^c & b_{n,2}^c & \dots & b_{n,p}^c \end{pmatrix}, \quad (5.10)$$

and $f_t^c = (f_{t,1}^c, \dots, f_{t,p}^c)'$ is a vector of p orthogonal dynamic latent factors driving the common dynamics of the variances and covariances in log-space. The n -dimensional vector w_t captures series specific random variation driven by an idiosyncratic factor structure $w_t = B^i f_t^i$, where $B^i = \text{diag}(b_1^i, b_2^i, \dots, b_n^i)'$ and $f_t^i = (f_{t,1}^i, \dots, f_{t,n}^i)'$. The measurement error u_t results from estimating the latent log-(co)variance process using realized (co)variances. In order to mitigate the curse of dimensionality the measurement error covariance matrix is assumed to be of diagonal type, $\Sigma_u = \text{diag}(\sigma_{u,1}^2, \dots, \sigma_{u,n}^2)'$.

In order to allow for common long-memory type of persistence patterns I adapt the heterogeneous autoregressive (HAR) model of Corsi (2009) to the modeling of common latent log-volatility factors. The HAR model forecasts volatility via a hierarchical autoregressive specification including lagged daily as well as weekly and monthly volatilities. The model amounts to a parsimonious and simple approach to modeling strong persistence in financial time series and represents an approximation to long-memory models. Assuming HAR structures for the common factors the respective dynamics are given by

$$f_{t,j}^c = \alpha_j^c + \phi_{j,1}^c f_{t-1,j}^c + \phi_{j,2}^c \sum_{i=1}^5 f_{t-i,j}^c + \phi_{j,3}^c \sum_{i=1}^{10} f_{t-i,j}^c + \phi_{j,4}^c \sum_{i=1}^{20} f_{t-i,j}^c + \eta_{t,j}^c, \quad (5.11)$$

where $\eta_{t,j}^c \sim N(0, \sigma_{c,j}^2)$ and $j = 1, \dots, p$. The HAR model results in a restricted AR(20) representation for the common factor dynamics (see Chapter 3, Section 3.2.4). Corsi (2009) and Audrino and Corsi (2010) find that HAR processes offer enhanced in-sample fit and out-of-sample forecasting performance in modeling log-volatilities and correlations, which are effectively approximated by the matrix logarithm. Due to the aggregation of flexible HAR dynamics for the p common factors, the factor structure is expected to accommodate a large variety of dependence patterns. Residual persistence is therefore expected to be short-lived and series-specific dynamics in w_t are assumed to be driven by AR(1) processes:

$$f_{t,j}^i = \alpha_j^i + \phi_j^i f_{t-1,j}^i + \eta_{t,j}^i, \quad \eta_{t,j}^i \sim N(0, \sigma_{i,j}^2), \quad (5.12)$$

where $j = 1, \dots, n$.

The model presented so far is unidentified. In order to identify the model the following restrictions are imposed: (i) B^c is restricted to a lower triangular matrix; (ii) the triangular elements of B^c and B^i are restricted to positivity; (iii) $\sigma_{c,j}^2 \stackrel{!}{=} 1$ and $\sigma_{i,j}^2 \stackrel{!}{=} 1 \forall j$ and (iv) $\alpha_j^c \stackrel{!}{=} 0$ and $\alpha_j^i \stackrel{!}{=} 0$. The identifying restrictions are proposed by Geweke and Zhou (1996) and are standard in the literature. An identified model comprises $4p + n(p + 4) - p(p - 1)/2$ parameters. Since the total number of parameters is a linear function in the number of time series n the model tackles the curse of dimensionality in multivariate volatility modeling. The property of weak stationarity of the underlying (co)variance process is easily checked via computing the characteristic roots of the factors' AR processes. The model then implies a stationary Gaussian distribution for the vector of logarithmic (co)variances y_t .

An important part of factor analysis is devoted to the interpretation of the common factors. From an asset pricing perspective we expect systematic variance dynamics reflected by the volatility of the latent market portfolio appearing e.g. in the CAPM asset pricing model (see Sharpe, 1964). A respective variance factor indicates un-diversifiable market risk. Krishnan et al. (2009) and Driessen et al. (2009) analyze the pricing of market-wide time-varying diversification benefits: So-called "correlation risk" is captured by a market-wide correlation factor. Investors would pay a premium for assets that perform well in states of high asset correlation, since increasing correlations imply lower diversification benefits and typically increasing market volatility. Driessen et al. (2009) assume a market-wide correlation factor and observe a significant pricing of correlation risk, which furthermore removes well-known biases in option pricing models. Krishnan et al. (2009) mention the importance of controlling for the market variance and asset-specific volatility when estimating correlation risk: If asset returns follow a one-factor model, the model-implied correlations are increasing in the asset betas and market variance and decreasing in idiosyncratic asset volatility, everything else held equal. The proposed factor model allows for investigating the presence of both systematic correlation and market risk.

5.2.3 Estimation and Diagnostics

Since the proposed factor model belongs to the class of linear Gaussian state space models, Maximum Likelihood estimation and Bayesian inference using Monte Carlo Markov Chain (MCMC) methods are straightforward to implement. In contrast to the ML approach, the Bayesian estimation scheme offers the advantage of avoiding high-dimensional numerical optimization of the log-likelihood function. In addition, Bayesian estimation easily accommodates nonlinear forecasting of covariance matrices within the MCMC sampling scheme. Standard Kalman filter based Maximum Likelihood estimation, in contrast, implies biased volatility forecasts due to the nonlinear matrix exponential function. I therefore apply Bayesian estimation with conjugate prior distributions for all model parameters. Forward

Filtering Backward Sampling (FFBS) serves for joint full conditional sampling of the latent factors (see Kim and Nelson, 1999). Details on the (overall uninformative) prior distributions, the implementation of the Gibbs sampling algorithm and the forecasting of covariance matrices are provided in the appendix.

The criterion of Onatski (2010) is applied in order to obtain an upper bound for the number of common factors. The criterion consistently estimates the number of factors in an approximate factor model while allowing for serially correlated idiosyncratic terms. The model selection is supplemented by model diagnostics. The diagnostic tests are based on Pearson residuals which are obtained as

$$e_t = \text{Var}[y_t|\mathcal{F}_{t-1}]^{-1/2} (y_t - E[y_t|\mathcal{F}_{t-1}]), \quad (5.13)$$

where \mathcal{F}_{t-1} is the information set including lagged observations up to period $t - 1$ and $\text{Var}[y_t|\mathcal{F}_{t-1}]^{-1/2}$ denotes the inverse Cholesky factor of $\text{Var}[y_t|\mathcal{F}_{t-1}]$. For a correctly specified model the standardized residuals $e_{ij,t}$ in e_t are serially and cross-sectionally uncorrelated. The modified Portmanteau test statistic represents a standard tool for detecting significant serial and cross-correlation in the residual series of multivariate econometric models. The modified Portmanteau statistic at l lags is

$$\bar{Q}_l = T^2 \sum_{i=1}^l (T - i)^{-1} \text{tr}(\hat{C}'_i \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}), \quad (5.14)$$

where $\hat{C}_i = 1/T \sum_{t=i+1}^T e_t e'_{t-i}$. Under general conditions $\bar{Q}_l \stackrel{a}{\sim} \chi^2(n^2l)$ (see Lütkepohl, 2005, p. 510). For low-dimensional cross-sections this test can be accompanied by F-test statistics for a regression of each single residual series on a constant and, say, 50 lags of the observed data y_t .³ This allows to detect single predictable residual series and offers a higher resolution in discovering violations of the null hypothesis.

5.3 Empirical Application

5.3.1 Data

The proposed dynamic factor model is applied to a 5-dimensional and a 30-dimensional data set of daily realized covariance matrices of equity market returns. The underlying stocks are traded at the New York Stock Exchange (NYSE) and listed in Table 5.1. The daily realized covariance matrix is computed as $R_t = \sum_{j=1}^M \xi_{t,j} \xi'_{t,j}$, where $\xi_{t,j}$ is the vector of returns for the $k = 5$ or $k = 30$ stocks computed for the j th 5-minute interval of trading day t between

³For $k = 5$ ($k = 30$) assets a regression of a single residual series on a constant and, say, 50 lags of the observed data y_t comprises $15 \times 50 + 1 = 751$ ($465 \times 50 + 1 = 23251$) regressors. Since the according data sets analyzed in Section 5.3 comprise 2514 (1564) observations, the regression becomes infeasible in the high-dimensional case.

9:30 a.m. and 4:00 p.m (see Chapter 1, Section 1.1.3). The ordering of the assets in the vector $\xi_{t,j}$ corresponds to the ordering of the assets in Table 5.1. Following Chiriac and Voev (2011) the realized (co)variance measure is further refined by averaging over 30 subsampling subgrids per day in order to exploit the data richness more efficiently and to cope with market microstructure noise. The sample period of the first data set (“Data Set 1”) starts at January 1, 2000, and ends on December 31, 2009, covering 2514 trading days including the sub-prime crisis. The data has already been studied in Chapter 3 and represents an updated version of the data set evaluated by Chiriac and Voev (2011). The application of the proposed factor model to 5-dimensional covariance matrices allows for an in-sample and out-of sample comparison to various volatility models from the relevant literature, where applications to higher dimensions, say larger than 10, are practically impossible. The second data set (“Data Set 2”) extends the first by 25 additional stocks randomly selected from the S&P 100. Since intra-day price data for the additional stocks is not freely available for the whole sample period of data set 1, I had to restrict the second data set to the period from February 2, 2002, to May 30, 2008, covering 1564 trading days⁴. The data set therefore (unfortunately) excludes the aftermath of the sub-prime crisis.

Figure 5.1 shows time series plots of the realized variances and covariances of the first data set. It reveals strong persistence and a common U-shaped pattern in the variance and covariance series. During the early 2000s, in the aftermath of the dot-com bubble, and during the recent sub-prime crisis starting in 2008 the level of the variances and covariances is significantly higher than in the middle part of the sample. Descriptive statistics are provided in Table 5.2. The empirical distribution of the variances and the covariances is highly skewed to the right and highly leptokurtic. The respective autocorrelation functions (ACFs) plotted in Figure 5.2 die out at a very slow rate indicating very strong serial correlation. Figures 5.3 and 5.4 illustrate the corresponding matrix-logarithmic time series and autocorrelation functions. Unsurprisingly, while the original series feature huge isolated volatility peaks, the logarithm greatly reduces the scale of these events letting the series appear much more homogeneous. The sample ACFs show persistent serial correlation, which is particularly pronounced for the diagonal elements of the matrix logarithm. Furthermore, the series feature distinct dynamic patterns for the diagonal and off-diagonal matrix-log elements. As discussed in Section 5.2.1 the diagonal elements can be interpreted as approximations to logarithmic variances while the off-diagonal elements can be interpreted as approximations to correlations. Figure 5.5 illustrates the quality of this approximation by comparing log-variance and correlation series with the matrix logarithm. The matrix-logarithmic series capture the dynamics of the original log-variance and correlation series to a great extent. The second data set extends the first and comprises 465 distinct series, where the according statistical properties are similar to the 15 time series of the first data set.

⁴I thank Giuseppe Storti for providing the data.

5.3.2 Estimation Results

Data Set 1: 5-Dimensional Covariance Matrix

Applying the Onatski criterion to data set 1 results in a maximum number of $p = 2$ common dynamic factors. I therefore estimate the factor model from Equations (5.8) to (5.12) including one and two common HAR factors. The MCMC scheme is based on 40,000 Gibbs iterations and a burn-in of 2,000 iterations. In order to assess the numerical accuracy of the estimates, MC standard errors are calculated using a correlation consistent Parzen window based spectral estimator for the variance of the sample mean (see Kim et al., 1998). The ratio of MC standard error to posterior standard deviation addresses the proportion of variation in the estimates due to simulation relative to the variation induced by the data. All numerical standard errors are within the range of 0.006-7% of the posterior standard deviations, indicating an acceptable balance of numerical and statistical uncertainty (see e.g. Kim et al., 1998). The chosen prior distributions are overall uninformative and given in the appendix. The parameter estimates indicate weak stationarity of the data generating processes for all considered models.

I now turn to the in-sample analysis. In order to enable a comparison to competing models, I also consider the MIDAS-CAW(3,3) model illustrated in Chapter 3, Section 3.2.4, and the DCC-CAW model of Bauwens and Storti (2011). In the empirical analysis of Chapter 3 the MIDAS-CAW(3,3) model was found to offer the overall best in-sample fit within the range of considered CAW specifications (see Table 3.4). The DCC CAW model of Bauwens and Storti (2011) assumes a conditional Wishart distribution for the realized covariance matrix R_t and decomposes the scale matrix S_t as

$$S_t = D_t P_t D_t, \quad (5.15)$$

where $D_t = \text{diag}(\sqrt{s_{11,t}}, \dots, \sqrt{s_{kk,t}})$ and P_t is a $k \times k$ conditional correlation matrix. The model assumes a dynamic equation for P_t similar to the DCC-GARCH model of Engle (2002):

$$P_t = (1 - \alpha - \beta) \bar{R} + \alpha Z_{t-1} + \beta P_{t-1}, \quad (5.16)$$

where Z_t denotes the realized correlation matrix at time t ,

$$Z_t = (\text{diag}(R_t))^{-\frac{1}{2}} R_t (\text{diag}(R_t))^{-\frac{1}{2}}, \quad (5.17)$$

where $\text{diag}(R_t)$ is the diagonal matrix whose i -th diagonal entry is the element $r_{ii,t}$, and α and β are positive scalar parameters. \bar{R} denotes the unconditional correlation matrix which is consistently estimated by the sample mean of realized correlation matrices. I follow Bauwens and Storti (2011) in assuming independent HAR dynamics for $\{s_{ii,t}\}_{i=1}^k$.

Table 5.3 shows Portmanteau diagnostic test results for the estimated Pearson residual

series. Note that the diagnostics for the factor models are based on residual series obtained for logarithmic (co)variance data, while the test results for the CAW models are based on residual series obtained for the original (co)variance data. Hence the results cannot be compared directly, but indicate how the models fit the dynamics of the respective original or logarithmic time series. Due to the rich serial and cross-sectional dynamics of the $n = 15$ (co)variance series the diagnostic tests indicate significant residual predictability for all considered models. Yet a comparison to the obtained test statistics for the raw data shows that all models successfully account for a major portion of the highly persistent (co)variance dynamics. The diagnostics for the 2-factor model indicate significant improvements compared to both competing models. This finding is particularly remarkable in relation to the flexible MIDAS-CAW approach. The dynamics implied by the MIDAS-CAW(3,3) model are driven by 168 (often insignificant) parameters as opposed to 82 parameters for the 2-factor model⁵. The previous findings are confirmed by additional F-test results for residual predictability presented in Table 5.4. Considering the 1% significance level and 50 lags of daily (logarithmic) realized (co)variances, the two-factor model successfully accounts for the predictability of twelve out of 15 logarithmic (co)variance series as opposed to five (co)variance series for the MIDAS-CAW and four (co)variance series for the DCC-CAW model. Figure 5.6 shows sample autocorrelation functions of the 15 residual series obtained for the fitted 2-factor model. The plots indicate that the model dramatically reduces the serial correlation in the raw data (compare to Figure 5.4). Summarizing the model diagnostic results, a parsimoniously parameterized factor model with two common factors and idiosyncratic dynamics offers a good fit to the complex serial and cross-sectional dynamics of the underlying logarithmic (co)variance data. Compared to the MIDAS-CAW approach the main source of parsimony is the reduction of cross-sectional dependence to loadings on a few persistent common factors.

Motivated by the model diagnostic results the subsequent analysis focuses on the 2-factor specification. Table 5.5 shows the parameter estimates. All estimated loadings are significantly different from zero. The estimates of the HAR parameters imply significance of the first, second and fourth HAR component. The estimated characteristic roots of the HAR-implied AR(20) processes are given by .9942 and .9893, respectively, and imply weak stationarity though strong persistence. The estimated AR(1) coefficients of the idiosyncratic factors reveal strong series specific serial correlation. Figure 5.7 shows bar plots of the fraction of total variance explained by the factors. Besides indicating the importance of idiosyncratic dynamics the figure shows that the first common factor is mainly associated with the diagonal elements of Y_t , while the second common factor appears to be almost exclusively driving off-diagonal dynamics. Referring to the properties of the matrix logarithm discussed in Section 5.2.1 this finding allows for the interpretation of common factors as market risk and correlation risk factor, in line with the asset pricing literature (see Section 5.2.2). Filtered

⁵The Schwarz-preferred MIDAS-CAW(2,2) model (see Chapter 3, Table 3.4) still comprises 118 parameters and offers slightly worse model diagnostic results compared to the MIDAS-CAW(3,3) model.

estimates are depicted in Figure 5.8 and confirm this interpretation. Comparing the second plot of Figure 5.8 with the time series in Figure 5.3 shows that the dynamics of the correlation factor mainly capture the persistently high correlation level in 2003 to 2007 when market volatility was comparably low. The second bar-plot in Figure 5.7 shows that the market risk factor has explanatory power for the correlation series, which are approximated by the off-diagonal elements in Y_t . This result corresponds to the common finding that high market volatility tends to be accompanied by strong correlation (see e.g. Solnik et al., 1996). Figure 5.9 depicts filtered estimates of the idiosyncratic factors. The time-series plots show that series-specific dynamics are mainly caused by the recent sub-prime crisis resulting in strong distinct reactions of a few logarithmic (co)variance series in 2007 to 2009 letting the dynamics of the corresponding idiosyncratic factors appear non-stationary. This finding indicates potential crisis related breaks in the idiosyncratic volatility and correlation structure which would motivate further research e.g. addressing Markov Switching regimes in volatility/correlation levels (see e.g. Lopes and Carvalho, 2007, and the analysis of Markov switching volatility models in Chapter 2), which are beyond the scope of this chapter.

Data Set 2: 30-Dimensional Covariance Matrix

The second data set comprises $n = 465$ distinct logarithmic (co)variance series of 30 asset returns covering the period from February 2, 2002, to May 30, 2008. The Onatski criterion suggests a maximum number of $p = 3$ common factors. The MCMC sampling scheme is based on 40,000 Gibbs iterations and a burn-in of 10,000 iterations. The numerical standard errors are within the range of 0.008-10% of posterior standard deviations. The parameter estimates imply weak stationarity of the data-generating processes.

Table 5.3 shows model diagnostic results. Since (MIDAS-)CAW models are generally not tractable for more than ten assets, the diagnostics are limited to the factor models and the DCC-CAW approach. The DCC-CAW results indicate significant residual predictability at any conventional significance level. The factor model residuals, in contrast, pass the test of the Null of no serial and cross-correlation at the 1% significance level for 75 as well as 100 lags. The large cross-sectional dimension precludes further testing for predictability of single residual series. Comparing the Portmanteau diagnostic test results for the factor model residuals obtained for data set 1 to the respective test results obtained for data set 2 reveals a better model fit to the observed logarithmic (co)variance dynamics in case of data set 2. Since the second data set does not cover the full extent of the sub-prime crisis this finding indicates a general problem of fitting the complex (co)variance dynamics in this particular period.

Motivated by the model diagnostic results the subsequent analysis focuses on the 3-factor model. Table 5.6 shows estimates of the HAR parameters, which imply significance of the first, second and fourth HAR component. The characteristic roots are given by .9906, .9723 and .9981 indicating weak stationarity and high persistence of the joint (co)variance dynam-

ics. Figures 5.10, 5.11 and 5.12 depict parameter estimates and 95% posterior confidence regions for the factor loadings, measurement error variances and the persistency of the idiosyncratic factors. Figure 5.12 shows that all idiosyncratic factor processes for the diagonal matrix-log elements (approximate log-variances) feature significant dynamics, where estimates imply overall strong persistence, though stationarity. Idiosyncratic dynamics for the off-diagonal elements, in contrast, are often insignificant. Figure 5.13 depicts the fraction of total variance of the $\{y_t\}_{t=1}^T$ series explained by the factors. The first row of plots in Figure 5.13 illustrates the importance of idiosyncratic variation for the series of (approximate) logarithmic variances with e.g. 30% explained variation for the American Express volatility and 26% explained variation for the volatility of the Dell stock. Significant explanatory power is also found for idiosyncratic factors of particular off-diagonal matrix-log elements (up to 22% explained variation), where the respective series mostly refer to approximate correlations involving Citigroup, JP Morgan Chase & Co., Intel, Dell and Microsoft, which have all been particularly affected by the sub-prime crisis inducing partly idiosyncratic volatility and correlation dynamics.

The particularly high fraction of variation of the diagonal matrix-log elements explained by the first common factor (30-80%, see Figure 5.13, second row of plots) motivates the interpretation as market risk factor. As already observed for data set 1, the market risk factor also captures significant variation of particular off-diagonal matrix-log series (up to 12%). The explanatory power of the second common factor is overall limited (up to 3% of total variation), but it appears to be mainly attributed to the off-diagonal elements in Y_t (compare the two plots in the third row of Figure 5.13). The factor is therefore interpreted as correlation risk factor. Figure 5.14 shows filtered factor estimates which confirm the previous factor interpretations (compare to Figure 5.8). In addition, the figure sheds light on the role of the third common factor, which appears to cover joint dynamics specifically linked to the sub-prime crisis inducing pronounced volatility and correlation peaks. Particularly high fractions of total variation explained by the third common factor are found for approximate log-variances of Citigroup (8%), JP Morgan Chase & Co. (7%), Intel (5%), and Microsoft (5%) and corresponding off-diagonal matrix-log elements (5-25% explained variation). As mentioned above, these stocks have been particularly affected by the sub-prime crisis.

5.3.3 Forecasting Results

I now compare the 1-day ahead forecasting performance of the dynamic factor model with alternative forecasting models from the relevant literature. Forecasts are denoted by $\hat{R}_{t+1} = E(R_{t+1}|\mathcal{F}_t)$. In addition to a statistical evaluation of the models' forecasting capabilities based on a root mean squared error (RMSE) criterion I follow Chiriac and Voev (2011) in also addressing potential economic benefits associated with accurate volatility forecasts. This is accomplished via evaluating the performance of portfolio optimization strategies based on volatility forecasts. For the first data set I follow the lines of Chapter 3, Section 3.3.3, and

select two out-of-sample windows: The first window is selected to be prior to the recent sub-prime crisis and covers the period from July 2, 2007 through June 30, 2008, with relatively low volatility (see the dark-gray shaded areas in Figures 5.1 and 5.3). The second window starts at July 1, 2008 and ends June 30, 2009 (see the light-gray shaded areas in Figures 5.1 and 5.3). The window covers the sub-prime crisis featuring a very high volatility level. The second data set ends at May 30, 2008, and allows for a slightly truncated version of the first forecasting window. All models are re-estimated daily and new forecasts are generated based on the updated parameter estimates. The set of competing models is given by the CAW, MIDAS-CAW, HAR-CAW, DCC-CAW, BEKK-GARCH, DCC-GARCH, and the EWMA approach. The CAW models are proposed in Chapter 3, Section 3.2.1, and the GARCH and EWMA models are illustrated in Chapter 2, Section 2.3.3.

Statistical Evaluation

In order to assess the predictive accuracy for a given forecasting model I follow Ledoit et al. (2003) in using the RMSE based on the Frobenius norm of the forecast error, given by

$$FN = \frac{1}{T^{\text{fore}}} \sum_t \|R_{t+1} - \hat{R}_{t+1}\| = \frac{1}{T^{\text{fore}}} \sum_t \left[\sum_{i,j} (r_{ij,t+1} - \hat{r}_{ij,t+1})^2 \right]^{1/2}, \quad (5.18)$$

where T^{fore} is the number of forecast periods.

Table 5.7 shows forecasting results for data set 1 and Table 5.8 presents those for data set 2. In forecasting the 5-dimensional covariance matrices of data set 1 the 2-factor model outperforms the competing models in terms of forecasting precision prior to the subprime crisis as well as in the crisis period. For data set 2 the set of competing models is substantially reduced since only the EWMA approach, the DCC-GARCH and the DCC-CAW model can be successfully applied to high-dimensional (co)variance forecasting. The lowest average Frobenius norm is obtained for the DCC-CAW approach closely followed by the 3-factor model. Note that the forecasting results for models based on daily asset return data are overall clearly inferior to the forecasting results obtained by volatility models using realized covariance matrices.

Economic Evaluation

In order to assess the economic value of the obtained volatility forecasts I follow Chiriac and Voev (2011) in constructing portfolios which maximize the utility of a risk-averse investor. Here I assume a second degree polynomial utility function and/or a conditional return distribution which is completely characterized by its first two moments (e.g. the normal distribution). The investor's portfolio optimization problem then reduces to the minimization of portfolio volatility via selecting the according asset weights while fixing a given expected

return (Markowitz, 1952).

I now assume an investor minimizing portfolio volatility subject to an expected portfolio return μ_p for the next trading day. The optimal portfolio is then given by the solution $\hat{\omega}_{t+1}$ to the quadratic problem

$$\hat{\omega}_{t+1} = \arg \min_{\omega_{t+1}} \omega'_{t+1} \hat{R}_{t+1} \omega_{t+1} \quad \text{s.t.} \quad \omega'_{t+1} E(\xi_{t+1}|\mathcal{F}_t) = \mu_p \quad \text{and} \quad \omega'_{t+1} \iota = 1, \quad (5.19)$$

where ω_{t+1} is the $k \times 1$ vector of portfolio weights chosen at t and held until $t + 1$, ι is a $k \times 1$ vector of ones, and μ_p is the target return. ξ_{t+1} denotes the 1-day ahead asset return vector. I assume serially uncorrelated daily asset returns which is typically met in practice and set $E(\xi_{t+1}|\mathcal{F}_t) \stackrel{!}{=} \mu$, where μ is approximated by the sample mean of returns. In order to assess the predictive accuracy of the considered models I compare the ex-post realization of the conditional portfolio mean and standard deviation. I therefore solve the minimization problem of Equation (5.19) resulting in an optimal weight vector $\hat{\omega}_{t+1}$ for each model and compute $\xi_{t+1}^p = \hat{\omega}'_{t+1} \xi_{t+1}$ and $\sigma_{t+1}^p = \sqrt{\hat{\omega}'_{t+1} R_{t+1} \hat{\omega}_{t+1}}$ for $t = T^*, T^* + 1, T^* + 2, \dots$, where T^* denotes the number of in-sample observations. Solving the optimal portfolio problem for various levels of the target portfolio return μ_p results in a predicted efficiency frontier, which characterizes the best mean-variance trade-off achievable by using a particular forecasting model. A suitable benchmark scenario is obtained by constructing the efficiency frontier using the ideal forecast $\hat{R}_{t+1} = R_{t+1}$.

Figures 5.15 to 5.17 show the obtained efficiency frontiers for the two data sets and forecasting windows averaged over the respective forecasting periods. As expected, the results show a by far lower achievable portfolio variance for a given portfolio mean in case of the ideal benchmark forecasts. Among the considered forecasting models the factor approach shows overall remarkably good ex-post mean-variance tradeoffs. The EWMA and the DCC-CAW models are nevertheless strong competitors. Selecting the global minimum variance portfolio as a natural reference point, the factor models stay unmatched in the rather calm forecasting phase I of data set 1 but are marginally outperformed in the turbulent phase II of data set 1 and the calm phase of data set 2. The results show considerable gains by the direct modeling of realized (co)variances opposed to daily return data based GARCH models where forecast-based ex post mean-variance tradeoffs are overall strictly inferior.

Summarizing the results, although the applied statistical and economic evaluation criteria are based on completely different objective functions, they overall result in the same models as the best performing ones, including the set of factor model specifications. This finding can be interpreted as evidence in favor of the factor model approach illustrated in this chapter.

5.4 Summary

In order to mitigate the curse of dimensionality in multivariate volatility modeling, this chapter illustrates a new flexible latent dynamic factor model for realized covariance matrices. The model is based on the matrix logarithm function which enables the modeling of log-(co)variances in Euclidean space, preserving positive definiteness and symmetry of covariance matrix forecasts without having to impose restrictions on the parameter space. By combining latent heterogeneous autoregressive processes (HAR, see Corsi, 2009) for the common factor structure with idiosyncratic AR(1) factors for series-specific dynamics the model effectively reduces the curse of dimensionality while allowing for rich (co)variance dynamics including a long-memory type of persistence. The simulated Bayesian estimation approach using Markov Chain Monte Carlo (MCMC) techniques enables straightforward estimation of the model parameters. An empirical application to realized (co)variances of up to 30 NYSE stocks shows that the model can be readily applied to the forecasting of high-dimensional covariance matrices. This enables practical applications like the forecasting of optimal portfolio weight vectors in mean/variance portfolio optimization, which typically requires (co)variance forecasts in high dimensions in order to effectively exploit diversification benefits. Since the elements of the matrix logarithm of a covariance matrix can be interpreted as approximations to logarithmic variances and correlations joint factors can be interpreted as risk-factors related to market-risk and diversification risk, offering a direct link to the recent asset pricing literature.

The empirical application to 5- and 30-dimensional realized covariance matrices of NYSE-traded stocks shows that the factor model successfully accounts for the observed dynamic behavior of up to 465 logarithmic (co)variance series, where two to three common factors appear overall sufficient in driving the cross-sectional dynamics. This finding implies significant dimension reduction in multivariate volatility modeling without substantially affecting in-sample fit. A comprehensive out-of-sample forecasting experiment based on the statistical root mean squared error criterium as well as an economic application to the forecasting of optimal portfolio weight vectors in mean/variance portfolio optimization shows that the factor model outperforms a range of prominent forecasting models from the relevant literature.

5.5 Technical Details

I now illustrate the Gibbs sampling algorithm for obtaining Bayesian point estimates of the parameters of the factor model presented in Section 5.2.2. According to Bayes' theorem the full conditional distribution of each sub-vector of the model's augmented parameter vector $\theta^{\text{aug}} = (a', \text{vec}(B^c)', \text{diag}(B^i)', \text{diag}(\Sigma_u)', \phi', f')$ is proportional to the product of the likelihood function and the sub-vector's joint prior distribution (see Chapter 2, Section 2.5.1). Let the vector ϕ embrace the factors' autoregressive coefficients and f summarize all $q = p+n$ factors for all time periods. The joint prior distribution is assumed to factorize into the product of marginal prior distributions. Details on the general Gibbs sampling algorithm are provided in Chapter 2, Section 2.5.

Full conditional sampling of a , B^c and B^i :

Given the factors f and the error variances $\sigma_{u,i}^2$ the model in Eqs. (5.8) and (5.9) reduces to n independent linear regressions $y_i = X_i\beta_i + u_i$, where $i = 1, \dots, n$ and y_i denotes the T -dimensional vector of observations $y_{t,i}$, X_i is a $T \times (q+1)$ regressor matrix including a constant and the factors $\{f_{t,j}^c\}_{j=1}^p$ and $f_{t,i}^i$, u_i is a T -dimensional vector of innovations $u_{t,i}$, and $\beta_i = (a_i, b_i)'$, where b_i denotes the vector of row parameters of the loadings matrix $B = (B^c, B^i)$. Assuming a joint normal prior distribution for β_i with mean $\mu_{0,i}$ and variance $\Sigma_{0,i}$, the according full conditional distribution is normal with mean and variance

$$\mu_i = \Sigma_i \left(\Sigma_{0,i}^{-1} \mu_{0,i} + \widehat{\Sigma}_{i,\text{ols}}^{-1} \widehat{\beta}_{i,\text{ols}} \right), \quad \Sigma_i = \left(\Sigma_{0,i}^{-1} + \widehat{\Sigma}_{i,\text{ols}}^{-1} \right)^{-1}, \quad (5.20)$$

where $\widehat{\beta}_{i,\text{ols}} = (X_i'X_i)^{-1}X_i'y_i$ and $\widehat{\Sigma}_{i,\text{ols}} = \sigma_{u,i}^2(X_i'X_i)^{-1}$ denote the ordinary least squares estimates of β_i and the residual variance. For the empirical application of Section 5.3.2 the following hyper-parameters are chosen: $\mu_{0,i} = 0.2$ and $\Sigma_{0,i} = 0.04 \times I_{q+1}$, implying a prior standard-deviation of 0.2.

Full conditional sampling of $\sigma_{u,i}^2$:

Given f , a and B and assuming inverse gamma prior distributions for the error variances $\sigma_{u,i}^2$ with parameters $\gamma_{0,i}$ and $\delta_{0,i}$ the full conditional distribution of $\sigma_{u,i}^2$ is inverse gamma with parameters

$$\gamma_i = \left(\frac{T}{2} + \gamma_{0,i} \right), \quad \delta_i = \left(\frac{u_i'u_i}{2} + \delta_{0,i}^{-1} \right)^{-1}. \quad (5.21)$$

For the empirical application of Section 5.3.2 the following hyper-parameters are chosen: $\gamma_{0,i} = 2.04$ and $\delta_{0,i} = 4.81$, implying a prior mean of 0.2 and a prior standard-deviation of 1.

Full conditional sampling of ϕ :

Given the factors f and assuming conjugate Gaussian priors, the factor persistencies are sampled analogously to β_i . For the empirical application of Section 5.3.2 the following hyper-parameters are chosen: $\mu_{0,i} = 0.5 \times \iota_{\tilde{q}}$ and $\Sigma_{0,i} = I_{\tilde{q}}$, where $\tilde{q} = 1$ for idiosyncratic factors

and $\tilde{q} = 4$ for common HAR factors; $i \in \{1, \dots, q\}$ and $\iota_{\tilde{q}}$ denotes a \tilde{q} -dimensional column vector of ones.

Full conditional distribution of the factors f :

All factors are drawn jointly by the ‘‘Forward Filtering Backward Sampling’’ (FFBS) scheme based on the Kalman filter (see Kim and Nelson, 1999). For the illustration of the FFBS method it proves convenient to write the (identified) factor model in standard state-space representation

$$y_t = a + Zs_t + u_t, \quad u_t \stackrel{iid}{\sim} N(0, \Sigma_u), \quad (5.22)$$

$$s_t = Hs_{t-1} + R\eta_t, \quad \eta_t \stackrel{iid}{\sim} N(0, I_q), \quad (5.23)$$

where s_t is the m -dimensional vector of latent state variables, Z and H are $n \times m$ and $m \times m$ matrices, and I_q is a q -dimensional identity matrix. For a q -factor model with idiosyncratic dynamics and p common HAR (AR(20)) factors we obtain $s_t = (f_t^i, f_t^c, f_{t-1}^c, \dots, f_{t-20}^c)'$, where f_t^i is the n -dimensional vector of idiosyncratic factors and f_t^c is a p -dimensional vector of common factors for period t . Hence $m = n + 20p$. Accordingly

$$Z = (B, 0_{n \times 19p}), \quad H = \begin{pmatrix} \text{diag}(\phi_1^i, \dots, \phi_n^i) & 0 & 0 & 0 & \dots & 0 & 0 \\ & \Phi_1 & \Phi_2 & \Phi_3 & \dots & \Phi_{19} & \Phi_{20} \\ & 0_{20p \times n} & I_p & 0 & 0 & \dots & 0 & 0 \\ & & 0 & I_p & 0 & \dots & 0 & 0 \\ & & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ & & 0 & 0 & 0 & \dots & I_p & 0 \end{pmatrix}, \quad (5.24)$$

where $0_{u \times v}$ denotes a $u \times v$ matrix of zeros and Φ_j , $j = 1, \dots, 20$, are p -dimensional diagonal matrices of HAR-model implied lag- j autoregressive parameters for the p common factors. Finally $R = [I_q, 0_{19p \times q}]$.

Denoting the set of states and data for $t = 1, \dots, T$ by $\underline{s}_T = \{s_t\}_{t=1}^T$ and $\underline{y}_T = \{y_t\}_{t=1}^T$, respectively, the joint full conditional density of the latent state variables for all time periods is obtained as

$$\begin{aligned} P(\underline{s}_T | \underline{y}_T) &= P(s_T | \underline{y}_T) \times P(s_{T-1} | s_T, \underline{y}_T) \times P(s_{T-2} | s_{T-1}, s_T, \underline{y}_T) \\ &\quad \times P(s_{T-3} | s_{T-2}, s_{T-1}, s_T, \underline{y}_T) \times \dots \times P(s_1 | s_2, s_3, \dots, s_T, \underline{y}_T) \\ &= P(s_T | \underline{y}_T) \times P(s_{T-1} | s_T, \underline{y}_{T-1}) \times P(s_{T-2} | s_{T-1}, \underline{y}_{T-2}) \\ &\quad \times P(s_{T-3} | s_{T-2}, \underline{y}_{T-3}) \times \dots \times P(s_1 | s_2, y_1) \\ &= P(s_T | \underline{y}_T) \times \prod_{t=1}^{T-1} P(s_t | s_{t+1}, \underline{y}_t), \end{aligned} \quad (5.25)$$

where I omit dependence on the parameter vector for the sake of readability. The second

step of the derivations in Eq. (5.25) is due to the state-space framework of Eqs (5.22) and (5.23) implying a Markov property for s_t and no additional information beyond s_{t+1} and \underline{y}_t relevant for predicting s_t . The decomposition of the joint full conditional density of \underline{s}_T in Eq. (5.25) implies that the whole \underline{s}_T sequence can be drawn jointly via recursive sampling from $P(s_t|s_{t+1}, \underline{y}_t)$. Note that the state-space model includes state-equations, which are identities. This results in a conditional variance of s_t given s_{t-1} being not positive definite. Therefore only the first q elements of s_t denoted by s_t^* can be conditioning factors in the full conditional sampling of \underline{s}_T and hence $P(\underline{s}_T|\underline{y}_T) = P(s_T|\underline{y}_T) \times \prod_{t=1}^{T-1} P(s_t|s_{t+1}^*, \underline{y}_t)$ (see Kim and Nelson, 1999, p. 194 ff.). Due to the model's linear Gaussian nature we directly obtain a conditional normal distribution for s_t given s_{t+1}^* and \underline{y}_t with mean and variance given by

$$s_{t|t, s_{t+1}^*} = s_{t|t} + P_{t|t} H^{*'} (H^* P_{t|t} H^{*'} + I_q)^{-1} (s_{t+1}^* - H^* s_{t|t}) \quad (5.26)$$

$$P_{t|t, s_{t+1}^*} = P_{t|t} - P_{t|t} H^{*'} (H^* P_{t|t} H^{*'} + I_q)^{-1} H^* P_{t|t}, \quad (5.27)$$

where H^* comprises the first q rows of H and I refer to standard Kalman filtering notation in denoting $s_{t|t} = E[s_t|\underline{y}_t]$ and $P_{t|t} = \text{Var}[s_t|\underline{y}_t]$. Both $s_{t|t}$ and $P_{t|t}$ are readily available from the Kalman filter algorithm. The prior derivations imply that the whole \underline{s}_T sequence can be drawn jointly full conditional via recursive sampling from conditional normal distributions with moments given in Eqs. (5.26) and (5.27).

Forecasting:

The forecast of the latent covariance matrix $V_{T+1} = \text{expm}(\text{vech}^{-1}(a + B f_{T+1}))$ is given by $E[V_{T+1}|\mathcal{F}_T]$, where vech^{-1} denotes the inverse function of the vech operator. Simulations from the forecast distribution can be obtained within the Gibbs sampling algorithm via running the Gibbs sampler based on all available data up to period T and simulating conditional on the respective Gibbs sweep $\theta^{\text{aug},(i)}$ the random matrix $V_{T+1}^{(i)} = \text{expm}(\text{vech}^{-1}(a^{(i)} + B^{(i)} f_{T+1}^{(i)}))$. A consistent simulation based estimate of $E[V_{T+1}|\mathcal{F}_T]$ is then obtained by computing the sample mean of the respective draws after convergence of the Gibbs sampler.

Table 5.1: NYSE Traded Stocks of Data Set 1 and Data Set 2

Name	Symbol	Name	Symbol
Alcoa	AA	Home Depot*	HD
Abbott Laboratories	ABT	Hewlett Packard	HPQ
Allstate	ALL	International Business Machines*	IBM
Amgen	AMGN	Intel	INTC
American Express*	AXP	Johnson & Johnson	JNJ
Bristol-Myers Squibb	BMY	JP Morgan Chase & Co.	JPM
Citigroup*	C	Kraft Foods	KFT
Colgate-Palmolive	CL	The Coca-Cola Company	KO
Cisco Systems	CSCO	McDonald's	MCD
DuPont	DD	Medtronic	MDT
Dell	DELL	Microsoft	MSFT
The Walt Disney Company	DIS	Pfizer	PFE
EMC Corporation	EMC	Wal-Mart	WMT
FedEx	FDX	Weyerhaeuser	WY
General Electric*	GE	Xerox	XRX

*: Stock belongs to data set 1.

Table 5.2: Descriptive Statistics

Stock	Mean	Max.	Min.	Std. dev.	Skewness	Kurtosis
Diagonal						
AXP (y_{11}) (r_{11})	0.59 (3.44)	5.29 (57.58)	-2.72 (0.07)	1.25 (4.68)	0.21 (4.23)	2.44 (32.78)
C (y_{22}) (r_{22})	0.71 (3.61)	6.68 (119.86)	-2.27 (0.11)	1.37 (5.91)	0.71 (7.65)	3.48 (108.49)
GE (y_{33}) (r_{33})	0.32 (2.43)	4.98 (51.40)	-2.3 (0.10)	1.09 (3.17)	0.51 (4.90)	3.07 (46.97)
HD (y_{44}) (r_{44})	0.74 (3.46)	4.64 (51.38)	-1.99 (0.16)	0.87 (3.97)	0.35 (3.92)	2.82 (28.01)
IBM (y_{55}) (r_{55})	0.16 (2.26)	3.77 (56.91)	-2.37 (0.12)	0.95 (3.05)	0.56 (5.98)	2.95 (67.60)
Off-Diagonal						
C-AXP (y_{21}) (r_{21})	0.35 (1.59)	1.14 (37.66)	-0.32 (-0.55)	0.19 (2.78)	0.54 (5.32)	3.64 (46.13)
GE-AXP (y_{31}) (r_{31})	0.28 (1.11)	0.88 (26.32)	-0.24 (-1.47)	0.15 (1.85)	-0.03 (5.90)	3.06 (58.08)
HD-AXP (y_{41}) (r_{41})	0.24 (1.16)	0.85 (27.66)	-0.30 (-2.46)	0.15 (1.97)	0.11 (5.33)	3.09 (47.60)
IBM-AXP (y_{51}) (r_{51})	0.24 (0.92)	0.66 (23.43)	-0.31 (-0.79)	0.14 (1.46)	-0.14 (5.65)	3.25 (55.89)
GE-C (y_{32}) (r_{32})	0.31 (1.24)	0.78 (41.69)	-0.23 (-0.58)	0.15 (2.12)	0.01 (7.02)	2.84 (91.59)
HD-C (y_{42}) (r_{42})	0.24 (1.27)	0.76 (27.34)	-0.25 (-0.93)	0.15 (2.17)	0.24 (5.02)	3.04 (39.51)
IBM-C (y_{52}) (r_{52})	0.25 (1.03)	0.81 (36.73)	-0.28 (-3.27)	0.14 (1.74)	0.03 (5.33)	3.01 (109.96)
HD-GE (y_{43}) (r_{43})	0.25 (1.04)	0.78 (26.85)	-0.29 (-1.14)	0.14 (1.70)	-0.29 (5.90)	3.14 (59.20)
IBM-GE (y_{53}) (r_{53})	0.28 (0.90)	0.73 (24.05)	-0.19 (-0.33)	0.14 (1.44)	-0.04 (0.76)	2.79 (57.77)
IBM-HD (y_{54}) (r_{54})	0.24 (0.87)	0.84 (18.32)	-0.31 (-1.20)	0.14 (1.34)	0.06 (5.21)	3.25 (44.18)

Descriptive statistics for the logarithmic and original covariance series of data set 1. Descriptive statistics for original (co)variance data in brackets.

Table 5.3: Portmanteau Diagnostic Test Results

				Lags:				
25	50	75	100		25	50	75	100
5-dim. covariance matrix					30-dim. covariance matrix			
Data: $\{R_t\}_{t=1}^T$								
121,838*	214,590*	300,239*	379,713*		5,507,128*	10,979,971*	16,437,898*	21,871,688*
1 common factor:								
6,889*	12,694*	18,540*	24,407*		5,430,351*	10,833,631*	16,221,666	21,586,103
2 common factors:								
6,556*	12,323*	18,096*	23,914*		5,428,924*	10,832,944*	16,222,752	21,589,451
3 common factors:								
					5,425,550*	10,827,337*	16,217,204	21,585,275
MIDAS-CAW(3,3):								
6,951*	12,878*	18,805*	24,584*					
DCC-CAW (HAR):								
6,820*	12,848*	18,798*	24,625*		5,501,739*	10,975,374*	16,423,168*	21,813,917*
1% critical values:								
5,874	11,602	17,305	22,996		5,413,277	10,822,071	16,230,127	21,637,801

* indicates significance at the 1% level.

Table 5.4: P-values of Residual F-tests

e_{11}	e_{21}	e_{31}	e_{41}	e_{51}	e_{22}	e_{32}	e_{42}	e_{52}	e_{33}	e_{43}	e_{53}	e_{44}	e_{54}	e_{55}
1 common factor:														
<.01*	<.01*	.03	.06	.17	.01	.19	.05	<.01*	<.01*	.11	<.01*	<.01*	.01	<.01*
2 common factors:														
.02	.02	.11	.11	.26	.19	.38	.06	.03	<.01*	.23	.02	<.01*	.02	<.01*
MIDAS-CAW(3,3):														
.89	.84	.01	<.01*	<.01*	<.01*	<.01*	<.01*	<.01*	<.01*	<.01*	<.01*	.91	<.01*	.04
DCC-CAW (HAR):														
.26	.03	.12	<.01*	<.01*	<.01*	<.01*	<.01*	<.01*	<.01*	<.01*	<.01*	.09	<.01*	<.01*

P-values of residual F-tests for a regression of each single residual series on a constant and 50 lags of the observed data. * indicates significance at the 1% level.

Table 5.5: Estimation Results: 2 Factor Model with Idiosyncratic AR(1) Dynamics, Data Set 1

Series	y_{11}	y_{21}	y_{31}	y_{41}	y_{51}	y_{22}	y_{32}	y_{42}	y_{52}	y_{33}	y_{43}	y_{53}	y_{44}	y_{54}	y_{55}
B^i	.2414	.0186	.0183	.0081	.0271	.1233	.0097	.0090	.0108	.0831	.0285	.0196	.1150	.0116	.1084
B_1^c	.2909	.0183	.0165	.0137	.0087	.3053	.0190	.0124	.0085	.2700	.0174	.0111	.2038	.0130	.2126
B_2^c	0	.0438	.0404	.0340	.0334	.0338	.0415	.0311	.0352	.0385	.0370	.0354	.0510	.0286	.0533
σ_u^2	.0605	.0142	.0128	.0142	.0122	.0733	.0125	.0133	.0118	.1015	.0116	.0127	.1058	.0139	.0959
f^i	ϕ_1^i	ϕ_2^i	ϕ_3^i	ϕ_4^i	ϕ_5^i	ϕ_6^i	ϕ_7^i	ϕ_8^i	ϕ_9^i	ϕ_{10}^i	ϕ_{11}^i	ϕ_{12}^i	ϕ_{13}^i	ϕ_{14}^i	ϕ_{15}^i
	.4399	.9879	.9062	.9876	.8009	.9585	.9888	.9916	.9885	.9740	.7628	.9218	.9576	.9496	.9793
f_1^c	$\phi_{1,1}^c$	$\phi_{1,2}^c$	$\phi_{1,3}^c$	$\phi_{1,4}^c$	ψ_1										
	.6677	.0410	.0045*	.0034*	.9942										
f_2^c	$\phi_{2,1}^c$	$\phi_{2,2}^c$	$\phi_{2,3}^c$	$\phi_{2,4}^c$	ψ_2										
	.6151	.0250	.0006*	.0108	.9893										

*: 95% posterior confidence region includes the null. Number of Gibbs sequences: 40,000; Burn-in: 2,000. ψ_i : characteristic root of the i 'th common factor's restricted AR(20) process (HAR).

Table 5.6: Estimation Results: HAR Parameter Estimates for 3 Factor Model with Idiosyncratic AR(1) Dynamics, Data Set 2

f_1^c	$\phi_{1,1}^c$	$\phi_{1,2}^c$	$\phi_{1,3}^c$	$\phi_{1,4}^c$	ψ_1	f_2^c	$\phi_{2,1}^c$	$\phi_{2,2}^c$	$\phi_{2,3}^c$	$\phi_{2,4}^c$	ψ_2
	.6152	.0433	.0080★	.0033★	.9906		.3043	.0575	.0189★	.0057★	.9723
f_3^c	$\phi_{3,1}^c$	$\phi_{3,2}^c$	$\phi_{3,3}^c$	$\phi_{3,4}^c$	ψ_3						
	.1967	.0228★	.0292★	.0195	.9981						

★: 95% posterior confidence region includes the null. Number of Gibbs sequences: 40,000; Burn-in: 10,000.

ψ_i : characteristic root of the i 'th common factor's restricted AR(20) process (HAR).

Table 5.7: Statistical Evaluation of Forecasting Accuracy: Data Set 1

Model	Phase 1		Phase 2		Model	Phase 1	Phase 2
	(p, q)		(p, q)				
Unrestricted CAW	(3,2)	6.074	(3,2)	52.474	EWMA	7.380	62.361
Diagonal CAW	(3,3)	6.062	(3,3)	52.799	DCC-CAW (HAR)	6.000	52.377
Unrestricted MIDAS-CAW	(3,3)	6.121	(3,2)	53.977	Unrestricted HAR-CAW	6.233	52.974
Diagonal MIDAS-CAW	(3,2)	6.029	(3,2)	54.276	Diagonal HAR-CAW	6.108	53.113
Unrestricted BEKK-GARCH	(2,3)	7.705	(2,3)	76.201			
Diagonal BEKK-GARCH	(3,3)	7.441	(3,3)	63.550			
DCC-GARCH	(1,2)	8.093	(3,3)	72.121			
Factor Models:							
1 common factor		5.979		48.771			
2 common factors		5.952		48.366			

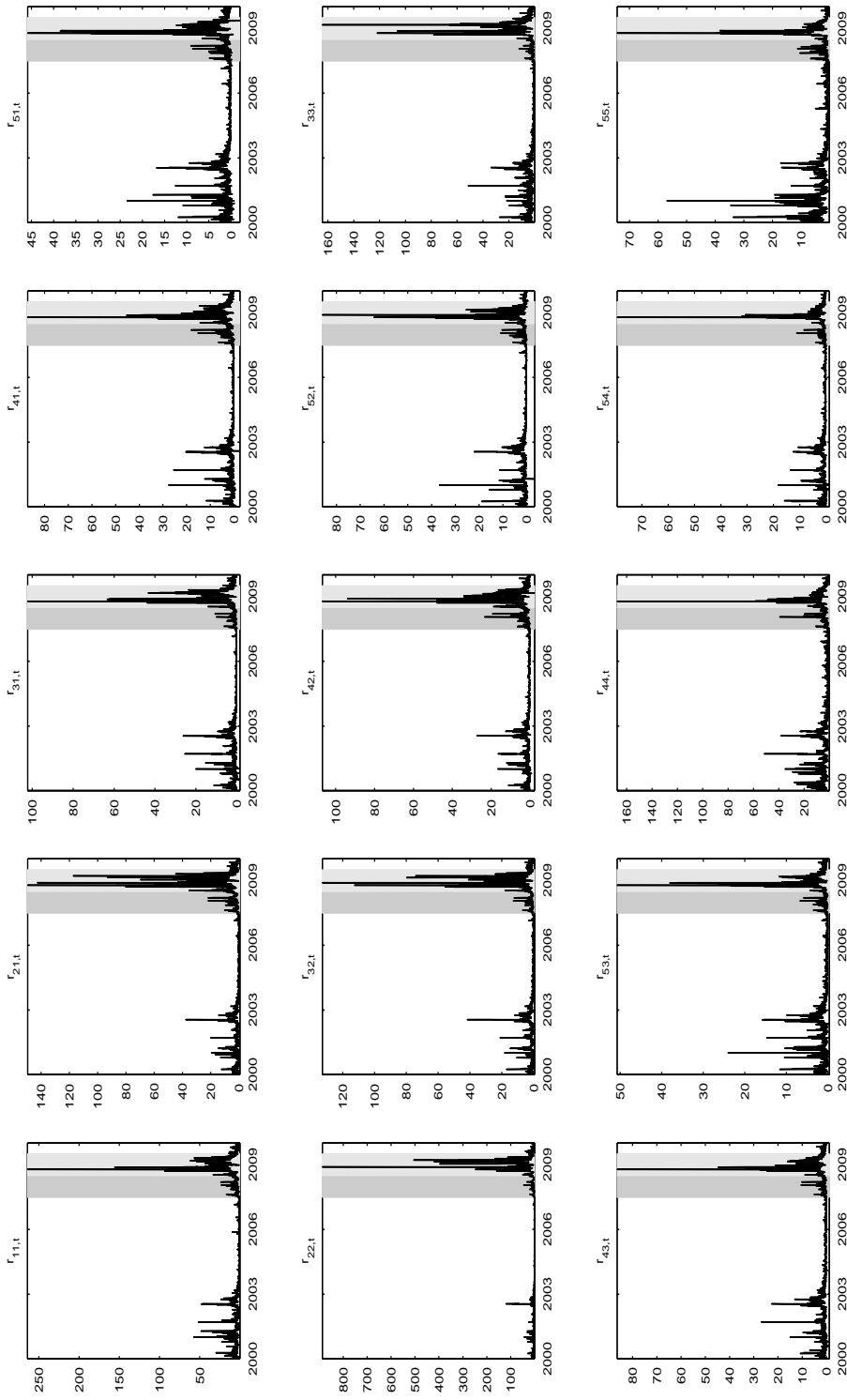
Reported is the average Frobenius norm of the forecast error. If model orders are quoted, models up to order (4, 4) have been estimated and the presentation is limited to the lowest obtained average Frobenius norm. Bold numbers indicate the smallest number of the average Frobenius norm.

Table 5.8: Statistical Evaluation of Forecasting Accuracy: Data Set 2, Phase 1

Model		Model	
DCC-GARCH (3,3)	25.861	1 common factor	21.473
DCC-CAW (HAR)	21.135	2 common factors	21.651
EWMA	24.800	3 common factors	21.257

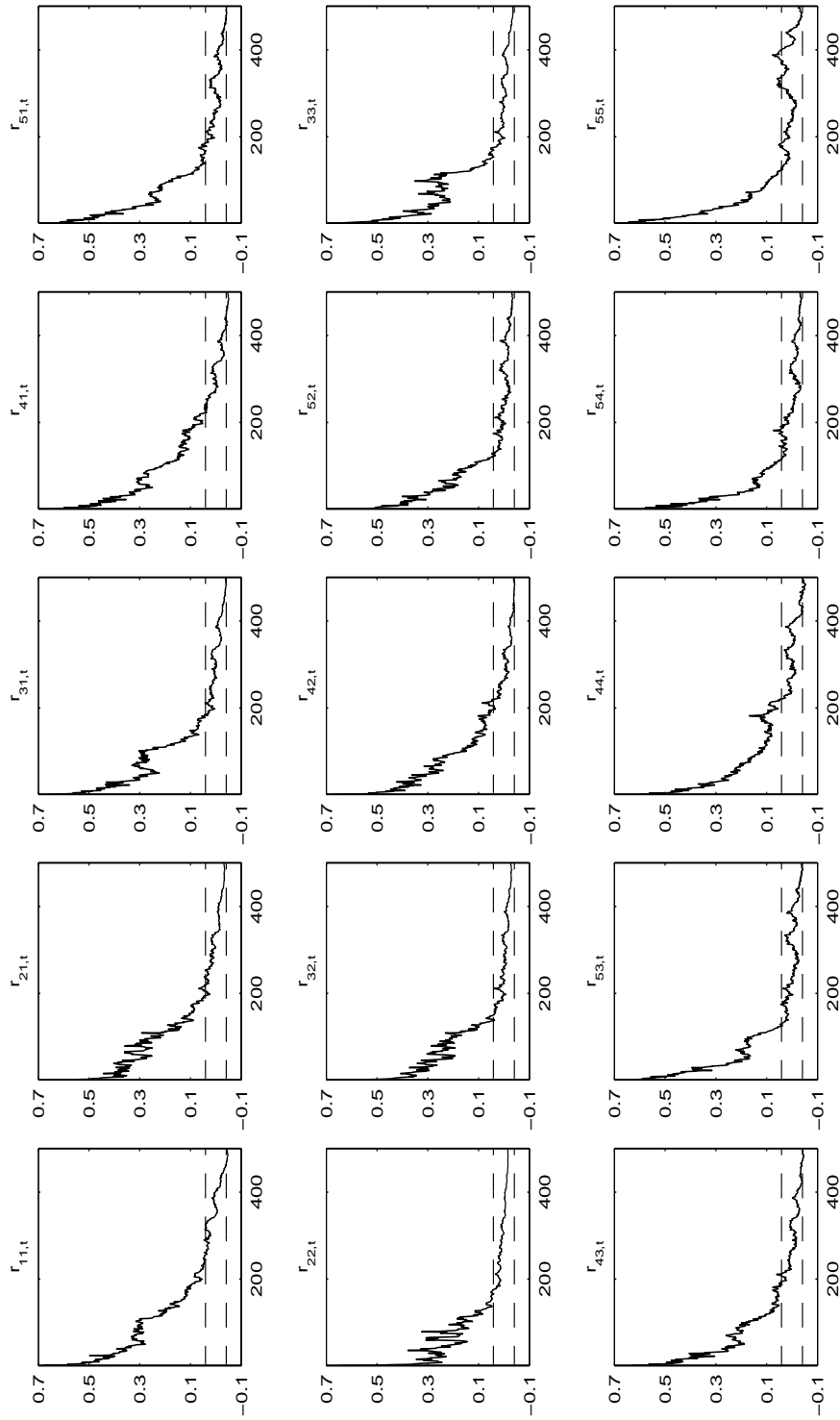
Reported is the average Frobenius norm of the forecast error. If model orders are quoted, models up to order (4,4) have been estimated and the presentation is limited to the lowest obtained average Frobenius norm. Bold numbers indicate the smallest number of the average Frobenius norm.

Figure 5.1: Time Series of Daily Realized Variances and Covariances: Data Set 1



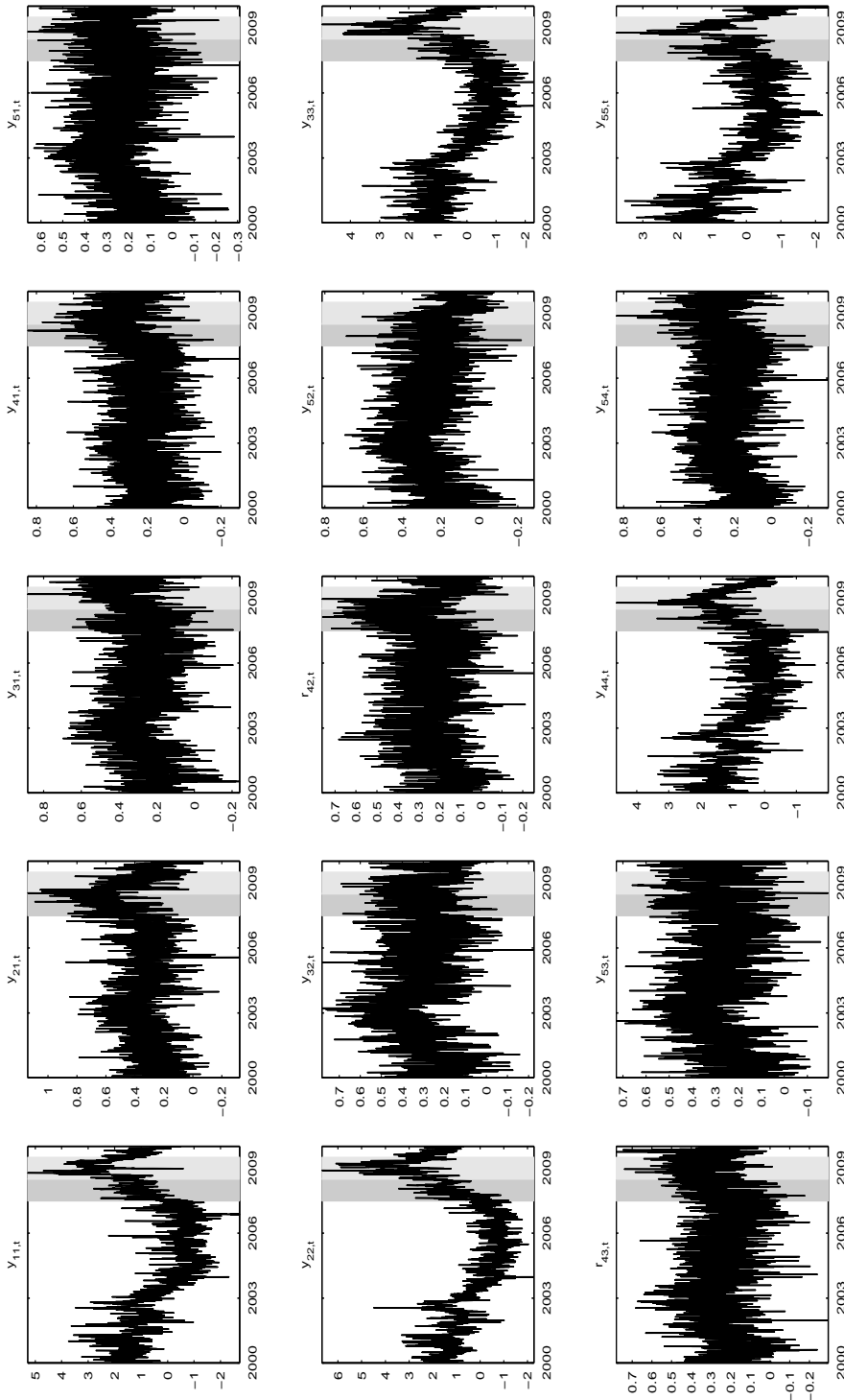
Time series of daily realized variances and covariances $r_{ij,t}$ for AXP ($i = 1$), C ($i = 2$), GE ($i = 3$), HD ($i = 4$), and IBM ($i = 5$) stock, data set 1; the gray shaded areas mark the two out-of-sample windows used in the forecast experiment.

Figure 5.2: Sample Autocorrelation Function of Daily Realized Variances and Covariances: Data Set 1



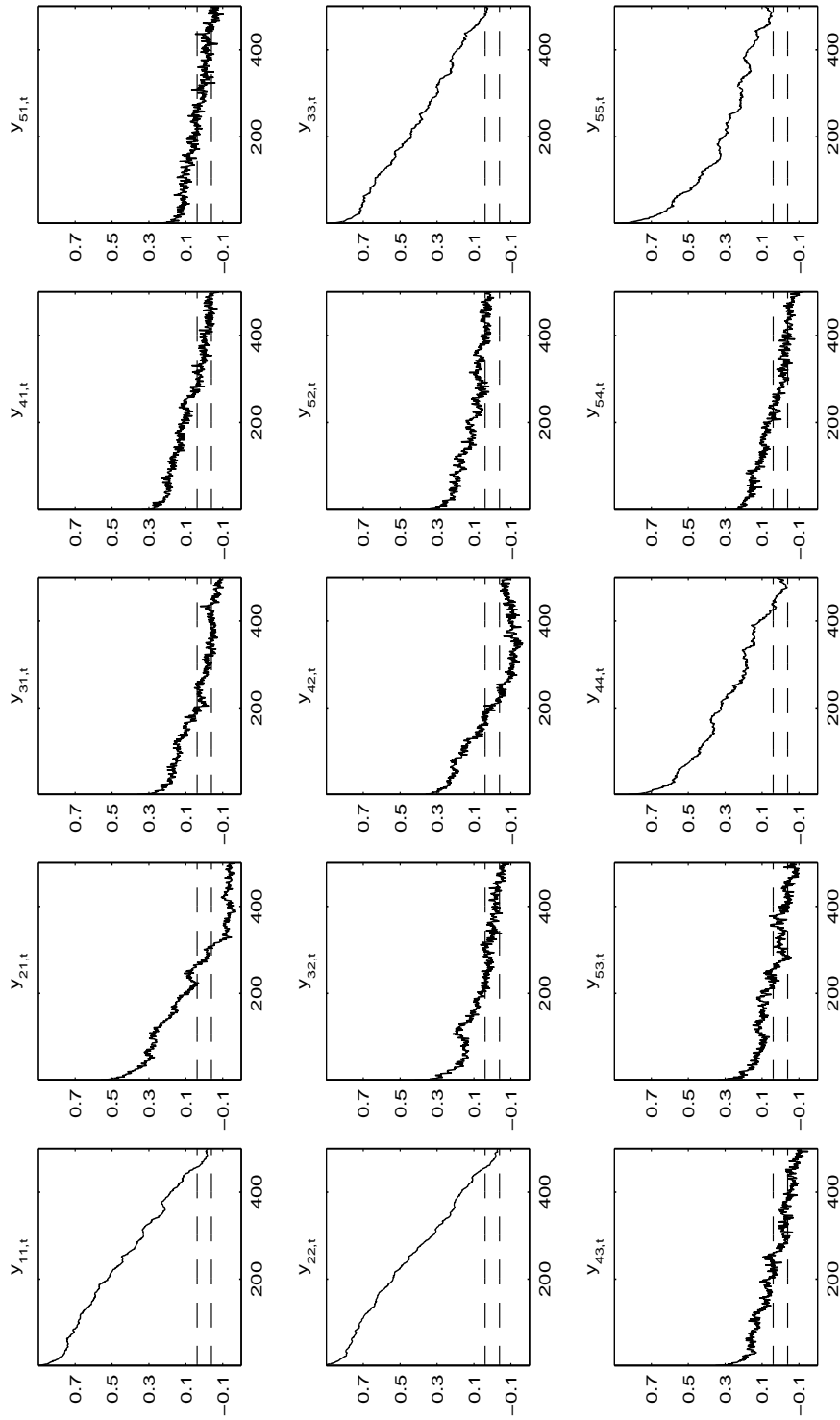
Sample autocorrelation function of daily realized variances and covariances $r_{ij,t}$ for AXP ($i = 1$), C ($i = 2$), GE ($i = 3$), HD ($i = 4$), and IBM ($i = 5$) stock, data set 1; the dashed lines indicate the 95% Bartlett confidence bands for no serial dependence.

Figure 5.3: Time Series of the Matrix Logarithm of Daily Realized Variances and Covariances: Data Set 1



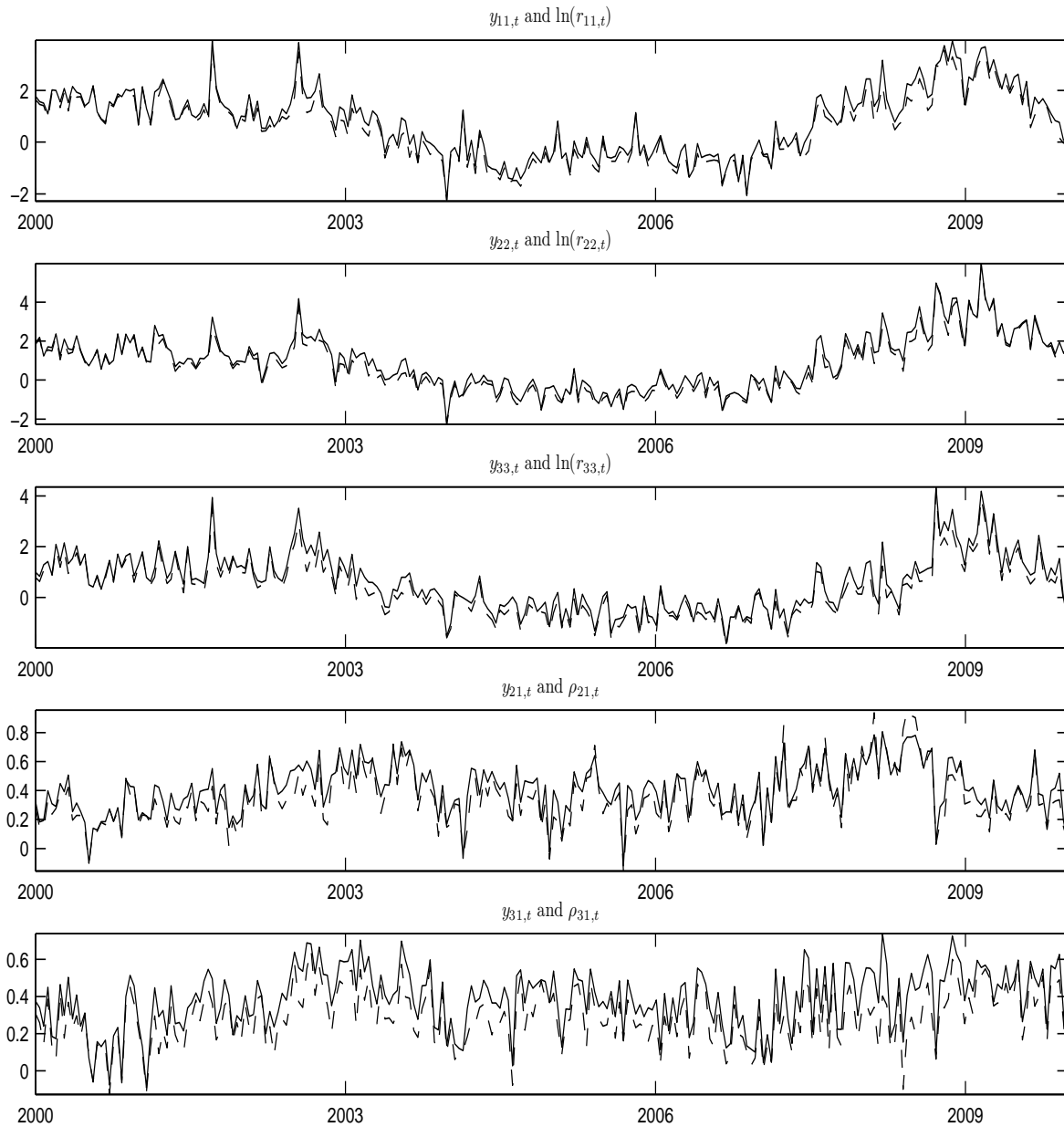
Time series of the matrix logarithm $y_{i,j,t}$ of daily realized variances and covariances $r_{i,j,t}$ for AXP ($i = 1$), C ($i = 2$), GE ($i = 3$), HD ($i = 4$), and IBM ($i = 5$) stock, data set 1; the gray shaded areas mark the two out-of-sample windows used in the forecast experiment.

Figure 5.4: Sample Autocorrelation Function of the Matrix Logarithm of Daily Realized Variances and Covariances: Data Set 1



Sample autocorrelation function of the matrix logarithm $y_{i,j,t}$ of daily realized variances and covariances $r_{i,j,t}$ for AXP ($i = 1$), C ($i = 2$), GE ($i = 3$), HD ($i = 4$), and IBM ($i = 5$) stock, data set 1; the dashed lines indicate the 95% Bartlett confidence bands for no serial dependence.

Figure 5.5: Logarithmic Variances and Correlations vs. Matrix-Logarithmic Approximations



Logarithmic variances and correlations and corresponding matrix-logarithmic approximations for AXP ($i = 1$), C ($i = 2$) and GE ($i = 3$), data set 1. Dashed lines: matrix-log. For reasons of legibility every 100'th observation is plotted.

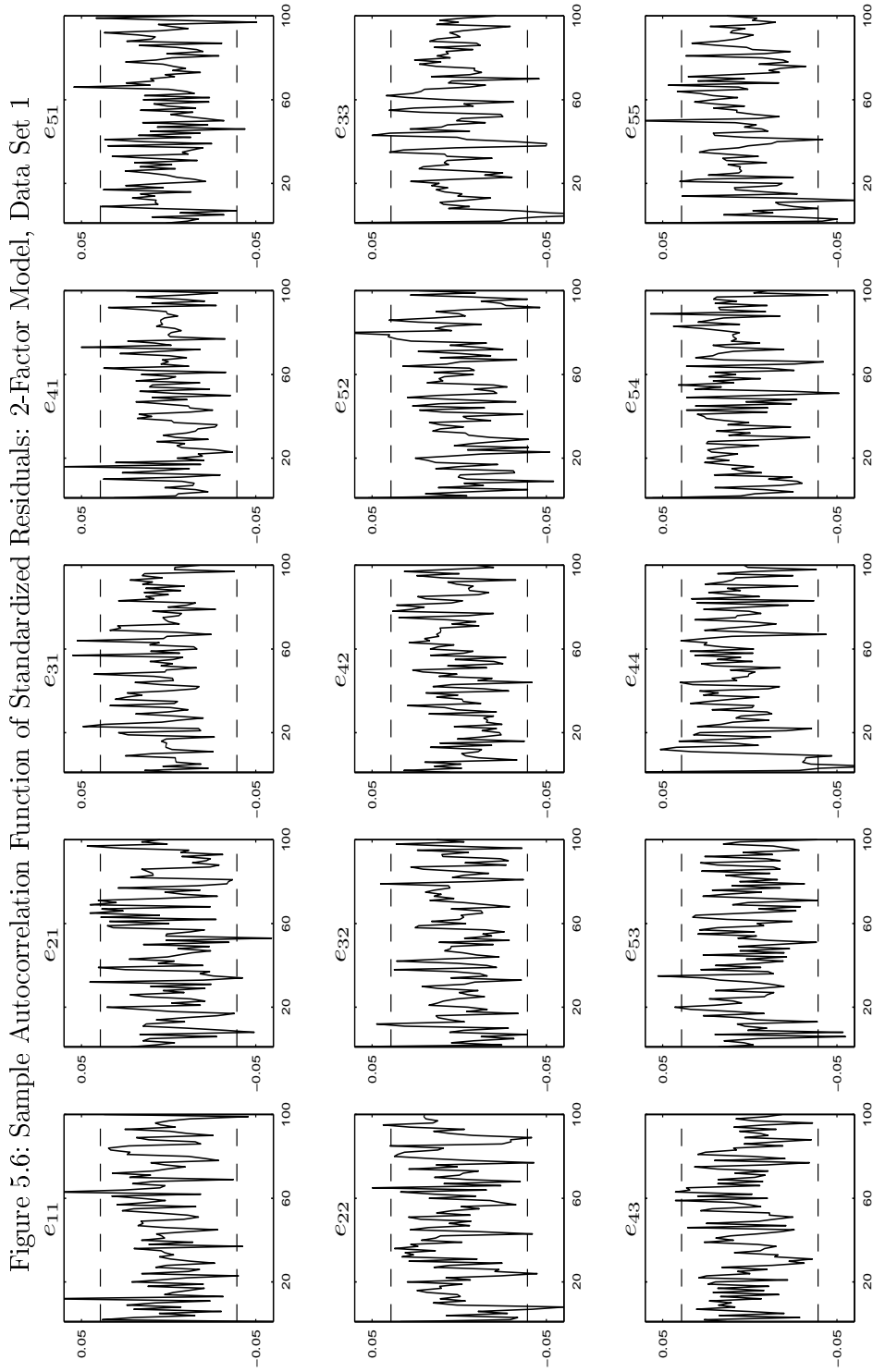


Figure 5.6: Sample Autocorrelation Function of Standardized Residuals: 2-Factor Model, Data Set 1

Sample autocorrelation function of the standardized residuals $e_{j,t}$ from the model with 2 common factors; data set 1. The dashed lines indicate 95% Bartlett confidence bands for no serial dependence.

Figure 5.7: Fraction of Total Variance Explained by Factors: 2-Factor Model, Data Set 1

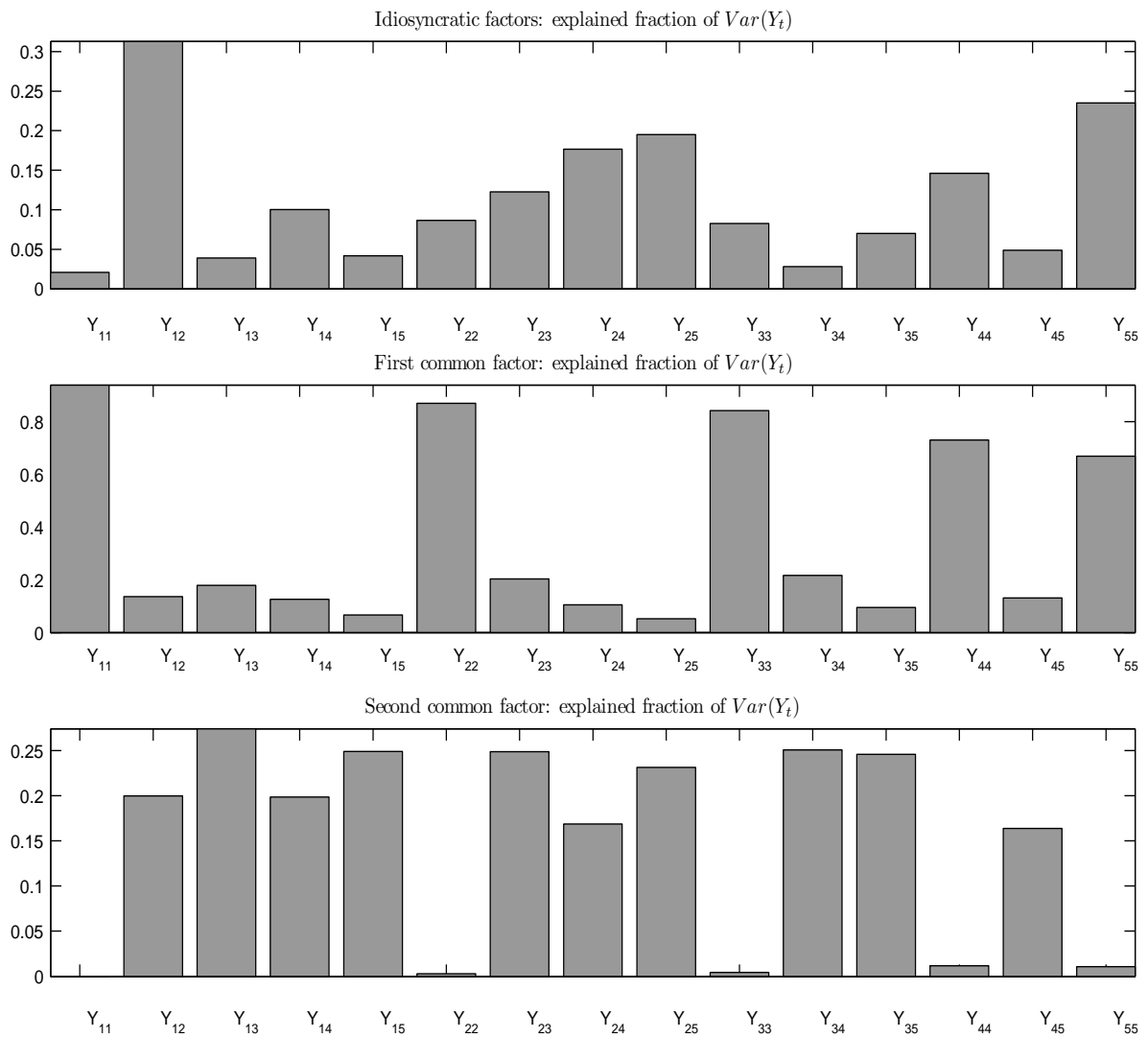


Figure 5.8: Filtered Common Factors: 2-Factor Model, Data Set 1

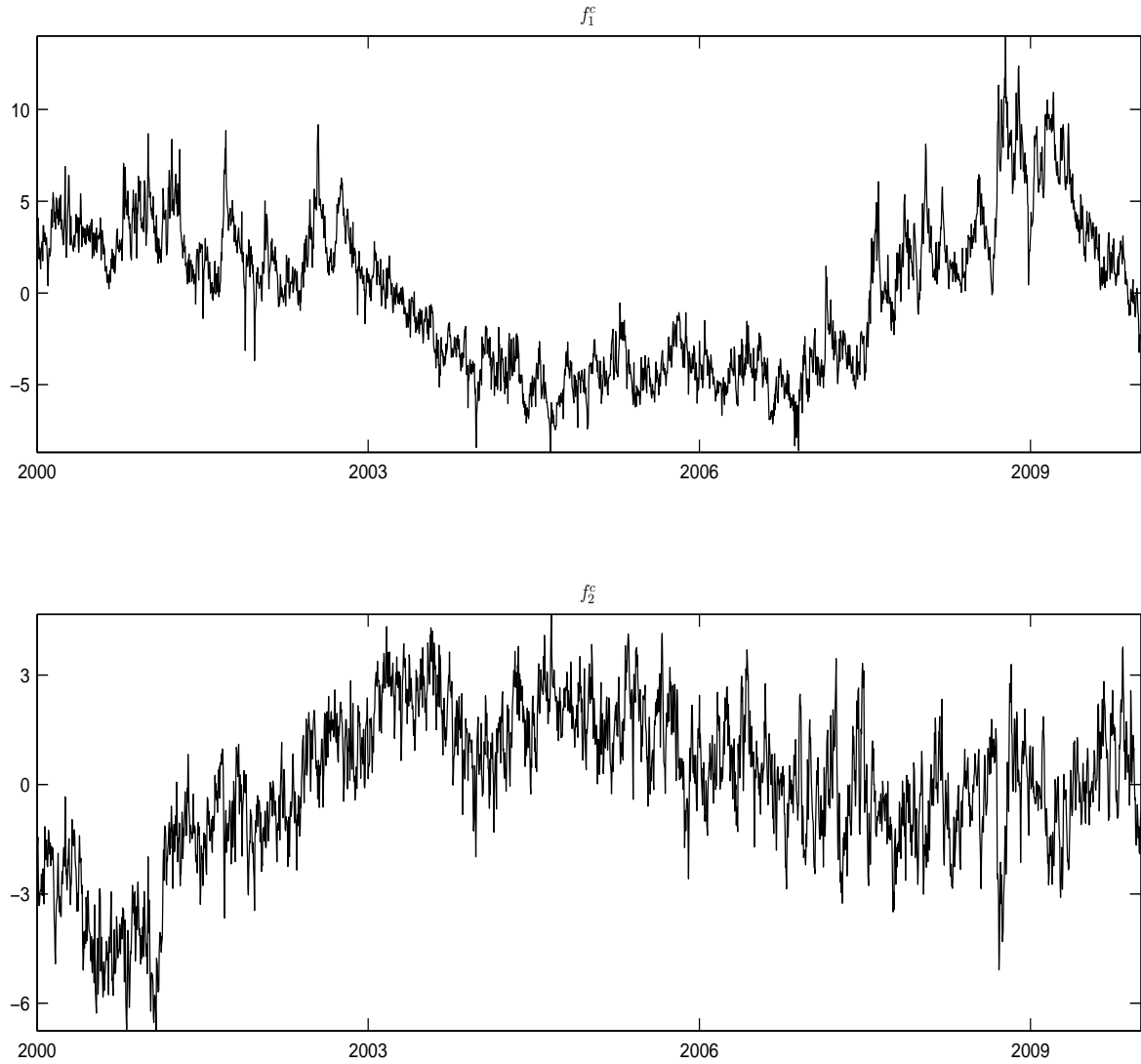


Figure 5.9: Filtered Idiosyncratic Factors: 2-Factor Model, Data Set 1

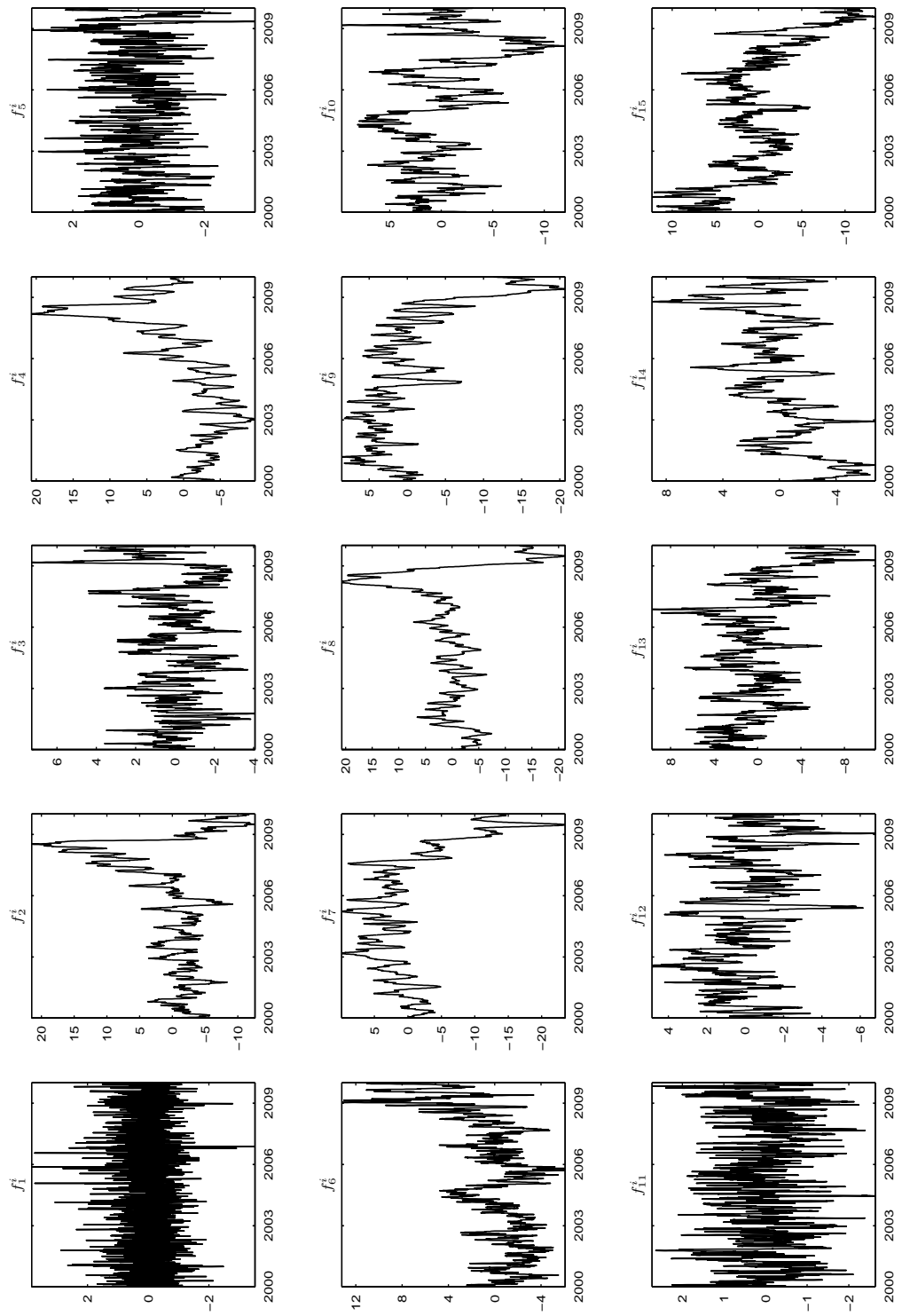
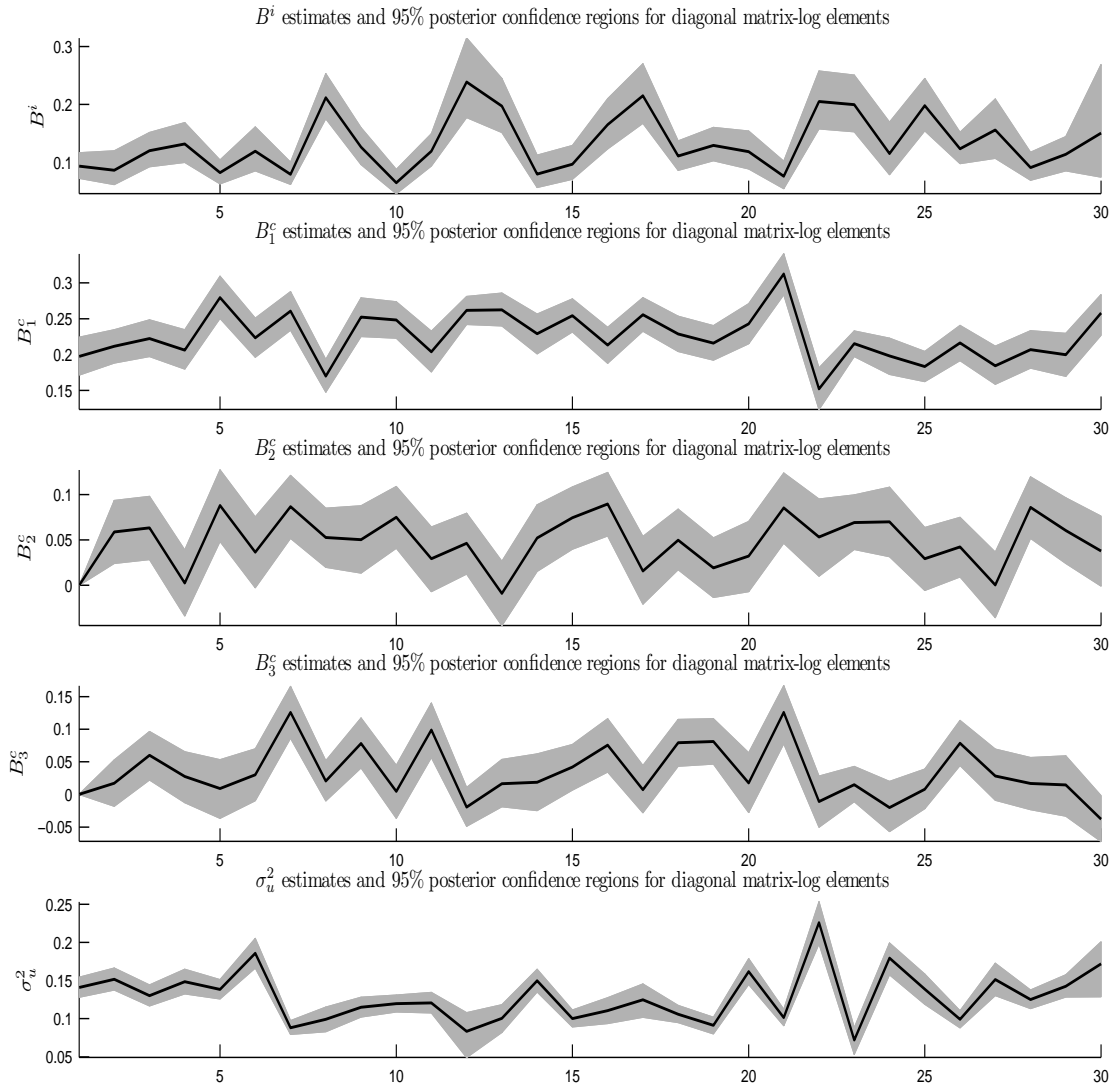
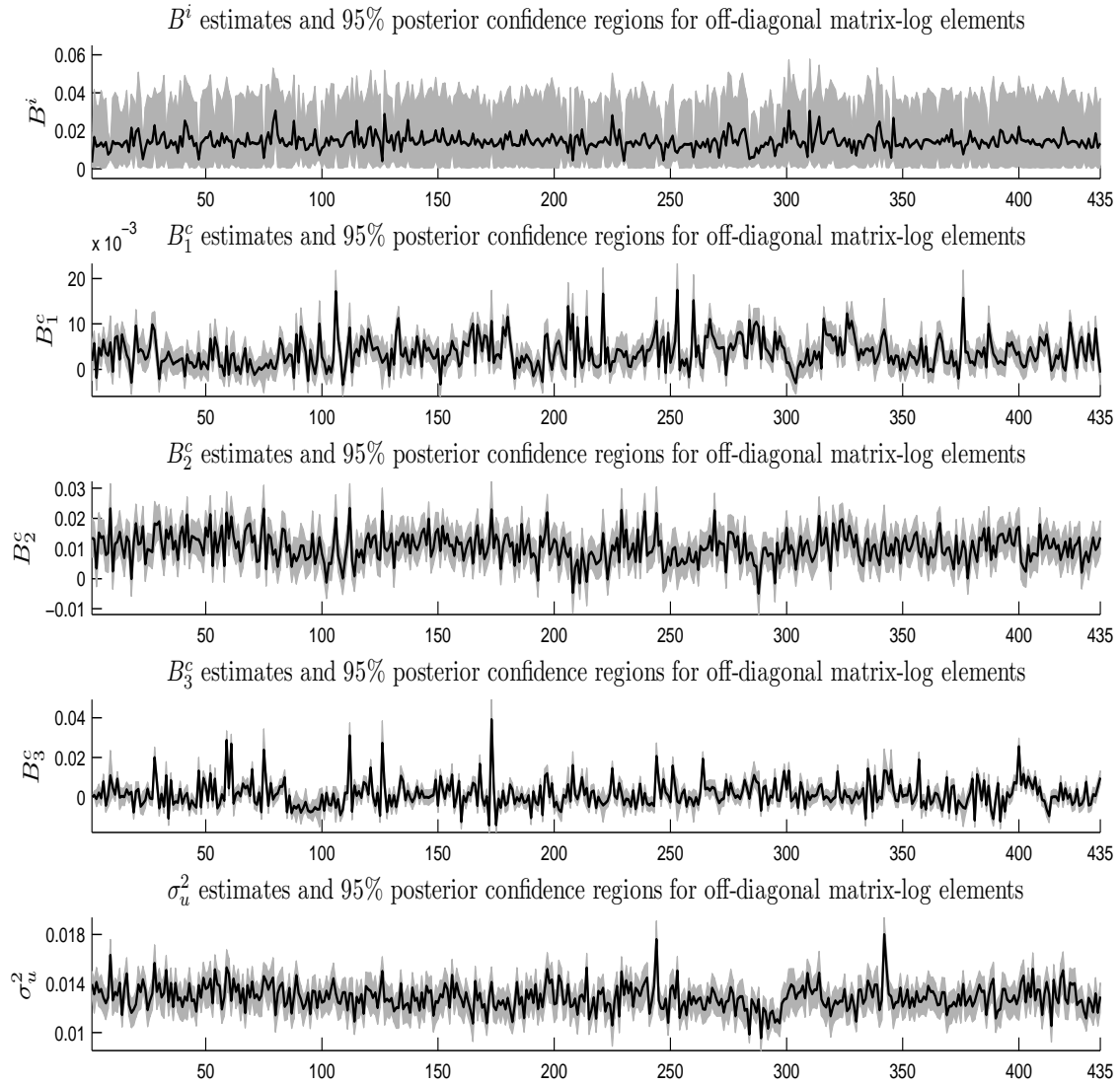


Figure 5.10: Point Estimates and 95% Posterior Confidence Regions for Factor Loadings and Measurement Error Variances of Diagonal Matrix-Log Elements: 3-Factor Model, Data Set 2



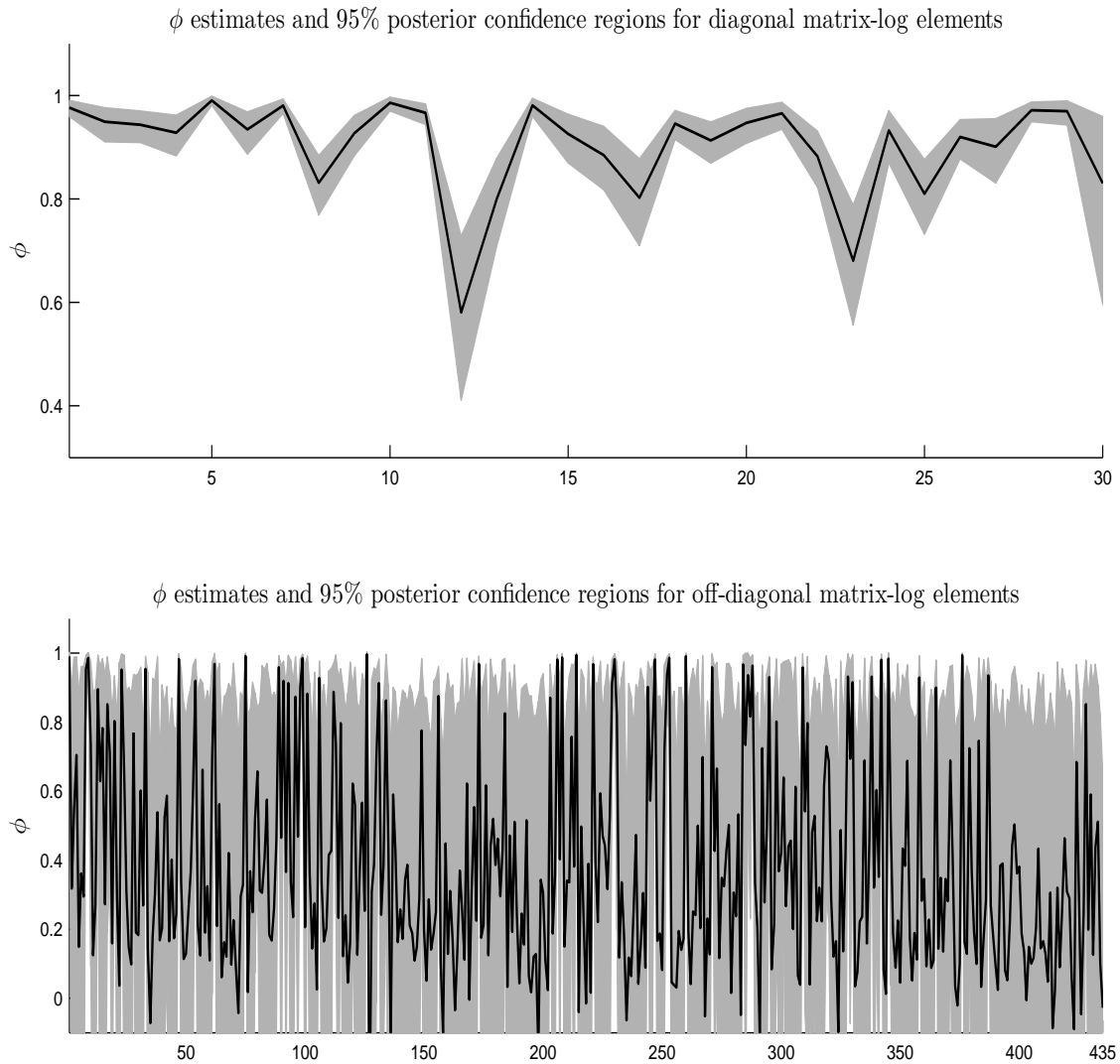
Each plot depicts the 30 point estimates and corresponding 95% posterior confidence regions (gray shaded) for the factor loadings or measurement error variances, respectively, for the $k = 30$ diagonal elements of Y_t .

Figure 5.11: Point Estimates and 95% Posterior Confidence Regions for Factor Loadings and Measurement Error Variances of Off-Diagonal Matrix-Log Elements: 3-Factor Model, Data Set 2



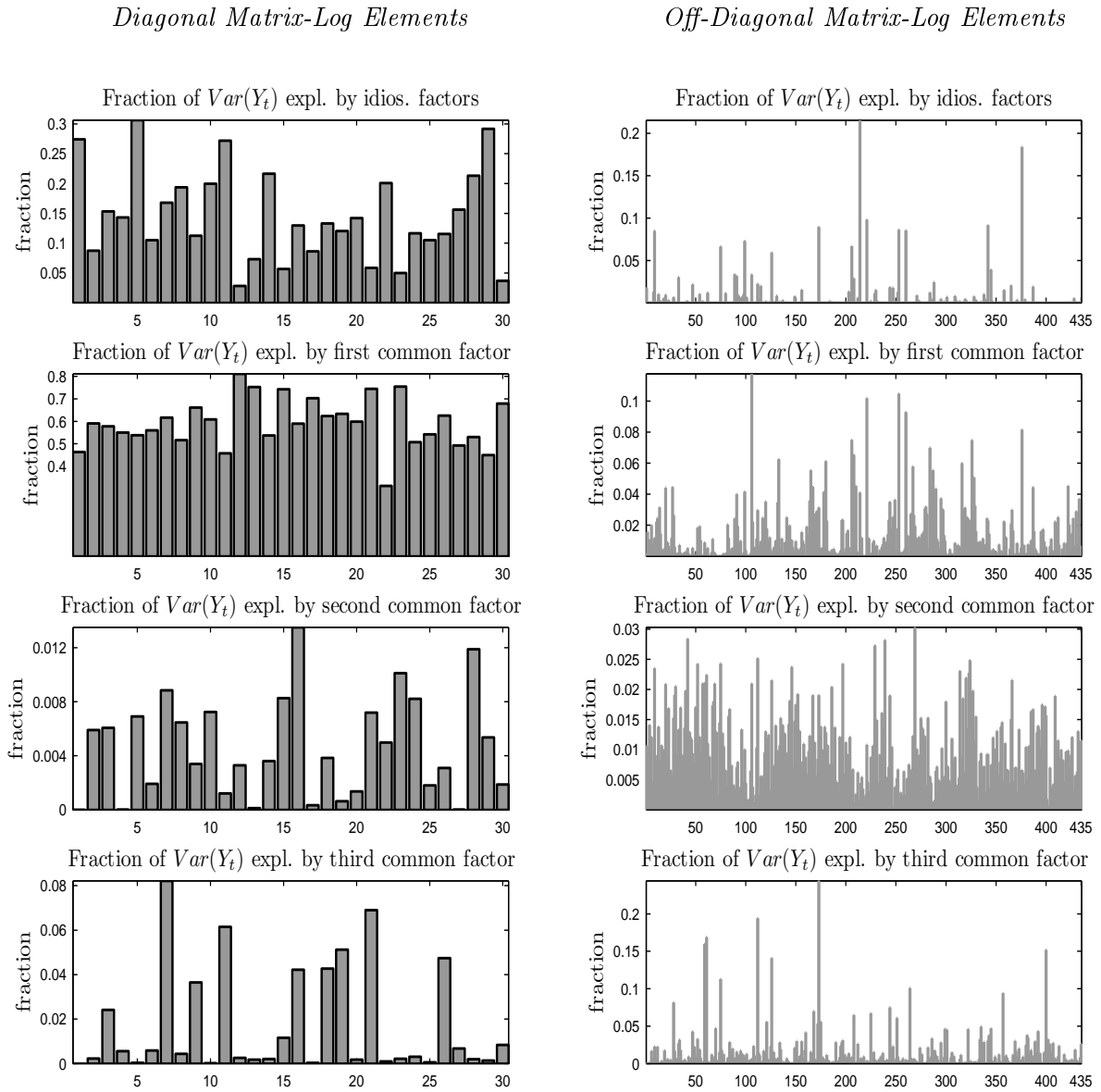
Each plot depicts the 435 point estimates and corresponding 95% posterior confidence regions (gray shaded) for the factor loadings or measurement error variances, respectively, for the $k(k-1)/2 = 435$ off-diagonal elements of Y_t .

Figure 5.12: Point Estimates and 95% Posterior Confidence Regions for Persistency of Idiosyncratic Factors: 3-Factor Model, Data Set 2



The first plot depicts the 30 point estimates and corresponding 95% posterior confidence regions (gray shaded) for the idiosyncratic persistence parameters $\{\phi_j^i\}$ corresponding to the $k = 30$ diagonal elements of Y_t . The second plot depicts the $k(k - 1)/2 = 435$ point estimates and corresponding 95% posterior confidence regions (gray shaded) for the idiosyncratic persistence parameters $\{\phi_j^i\}$ corresponding to the 435 off-diagonal elements of Y_t .

Figure 5.13: Fraction of Total Variance Explained by Factors: 3-Factor Model, Data Set 2



Left panel: fraction of total variance explained by factors for the $k = 30$ diagonal matrix-log elements.
 Right panel: fraction of total variance explained by factors for the $k(k - 1)/2 = 435$ off-diagonal matrix-log elements.

Figure 5.14: Filtered Common Factors: 3-Factor Model, Data Set 2

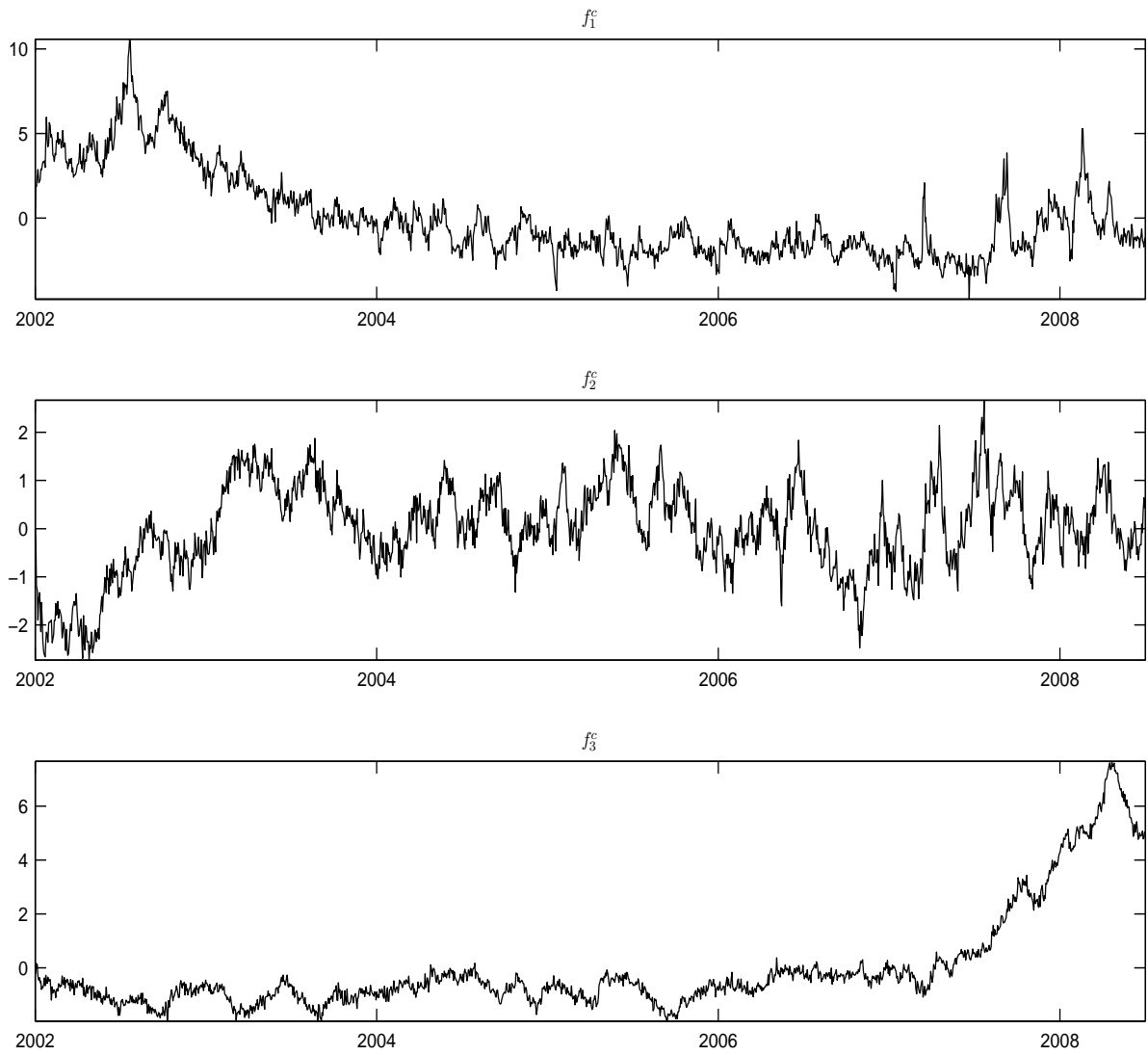
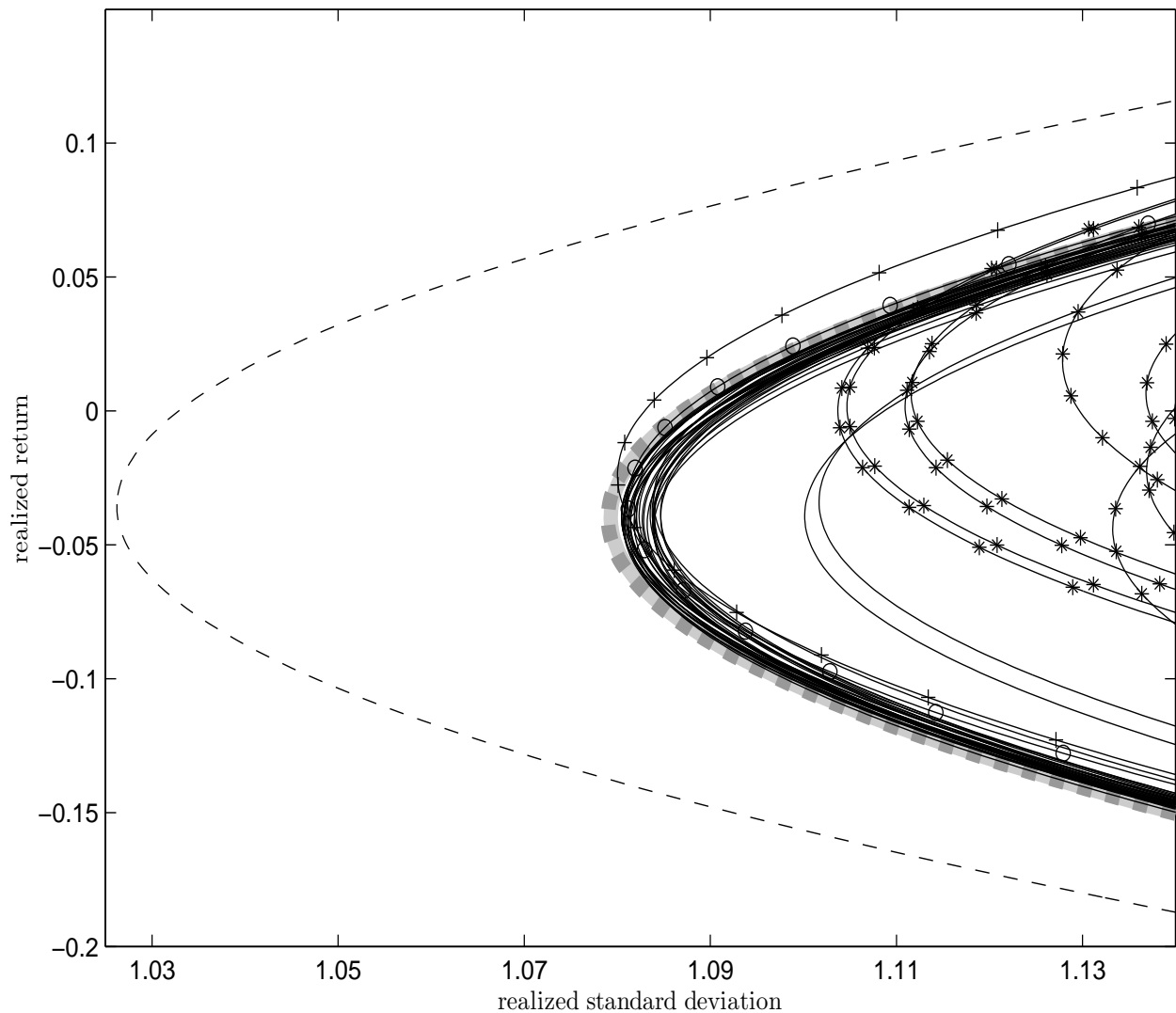
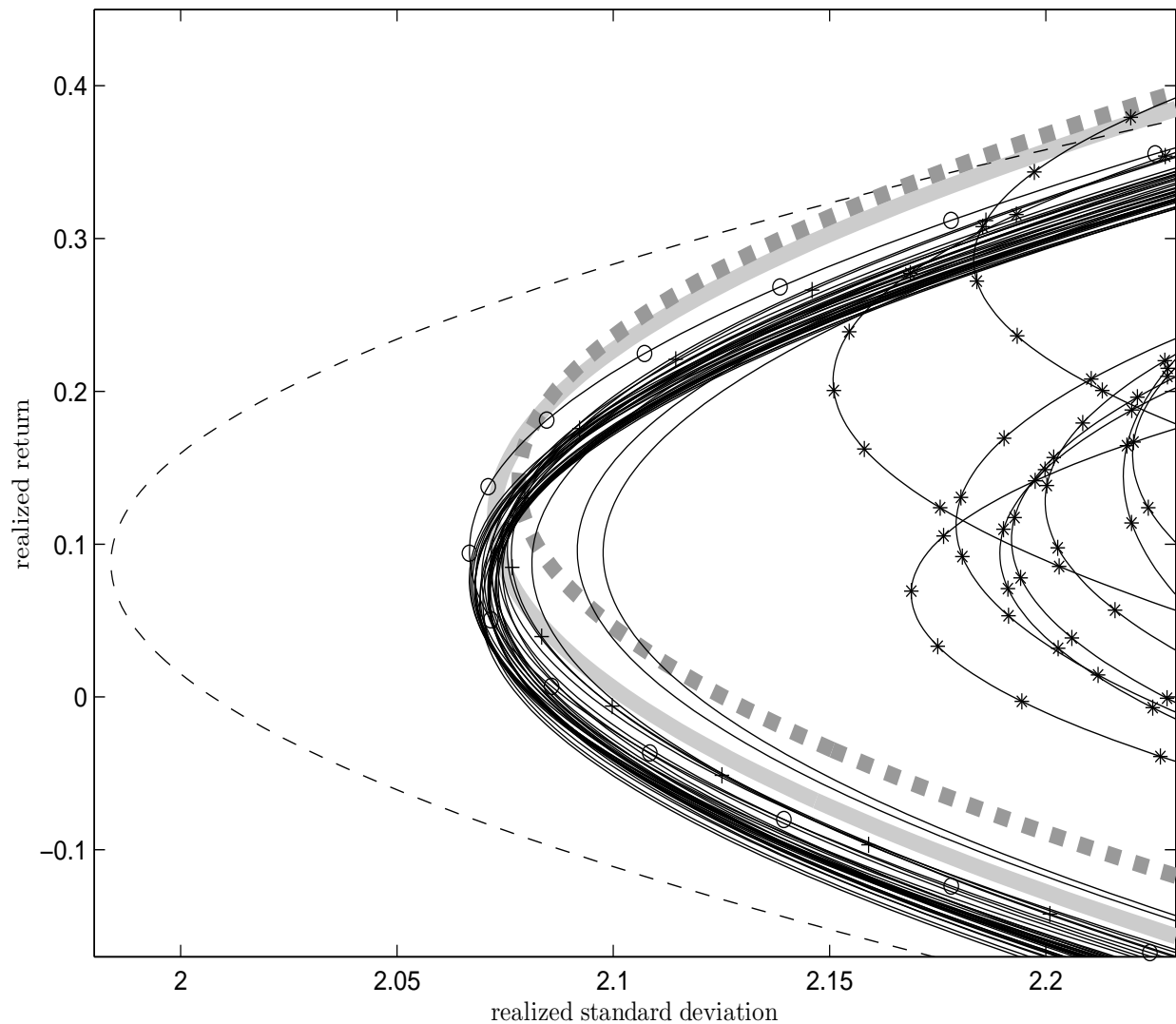


Figure 5.15: Mean-Variance Plots: Data Set 1, Phase I



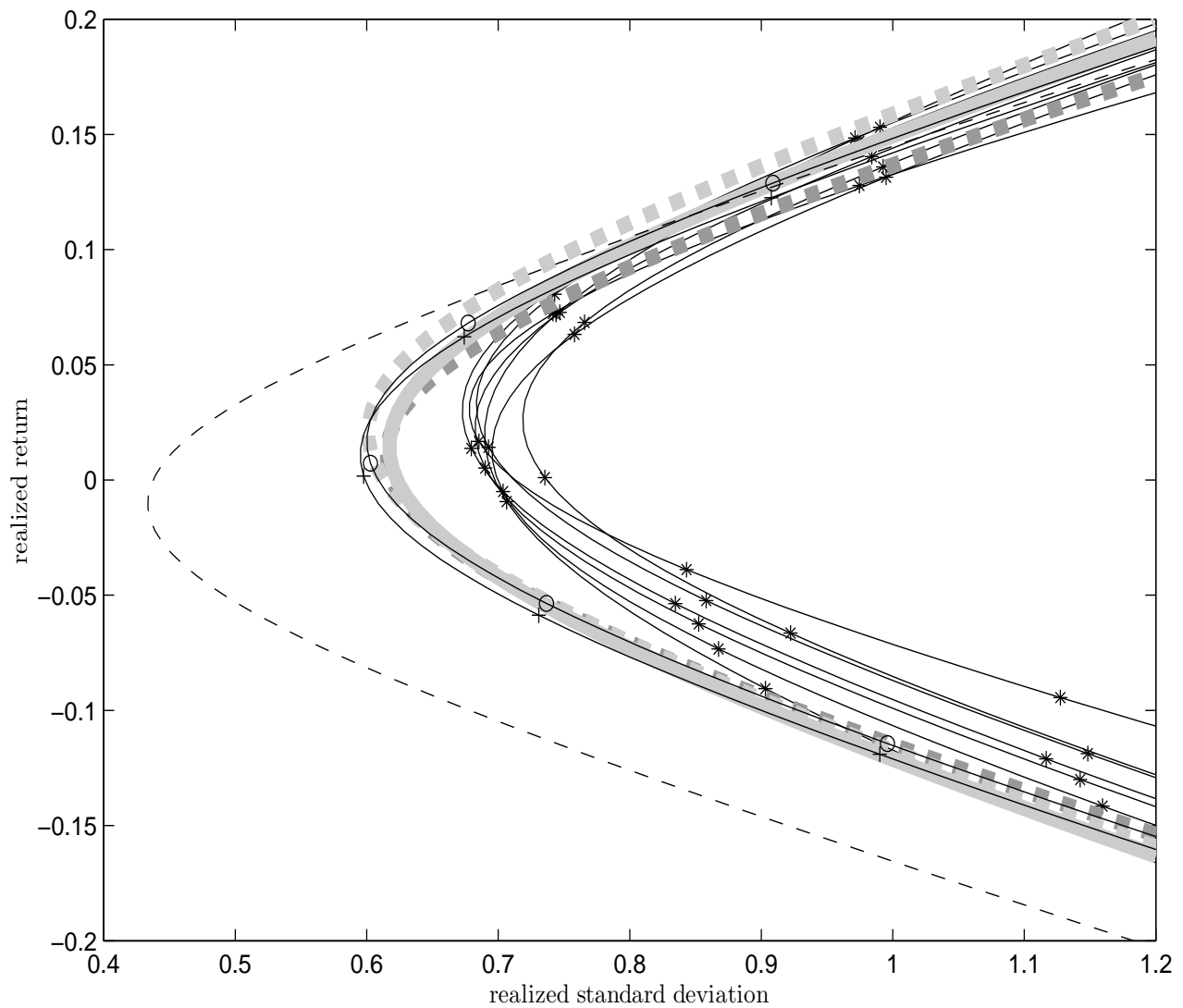
Mean-variance plots of the ex-post realized portfolio return (y-axis in %) against realized standard deviation (on the x-axis in %) for data set 1, out-of-sample phase I. Dashed solid line: ideal forecast based on observed covariance matrix; dashed bold dark-gray line: 1-factor model; bold light-gray line: 2-factor model; *: multivariate GARCH models based on daily returns; +: EWMA; o: DCC-CAW; solid: remaining CAW models. All plots are averages across the 252 out-of-sample periods in phase I.

Figure 5.16: Mean-Variance Plots: Data Set 1, Phase II



Mean-variance plots of the ex-post realized portfolio return (y-axis in %) against realized standard deviation (on the x-axis in %) for data set 1, out-of-sample phase II. Dashed solid line: ideal forecast based on observed covariance matrix; dashed bold dark-gray line: 1-factor model; bold light-gray line: 2-factor model; *: multivariate GARCH models based on daily returns; +: EWMA; o: DCC-CAW; solid: remaining CAW models. All plots are averages across the 251 out-of-sample periods in phase II.

Figure 5.17: Mean-Variance Plots: Data Set 2, Phase I



Mean-variance plots of the ex-post realized portfolio return (y-axis in %) against realized standard deviation (on the x-axis in %) for data set 2, out-of-sample phase I. Dashed solid line: ideal forecast based on observed covariance matrix; dashed bold dark-gray line: 1-factor model; bold light-gray line: 2-factor model; dashed bold light-gray line: 3-factor model; *: multivariate GARCH models based on daily returns; +: EWMA; o: DCC-CAW; solid: remaining CAW models. All plots are averages across the 228 out-of-sample periods in phase I.

Chapter 6

Conclusion

This thesis contributes to recent developments in multivariate volatility modeling. The analysis focuses on stochastic volatility models and the direct modeling of realized (co)variances as precise measures of latent variances and covariances. Several novel time-series models are proposed and analyzed in order to capture the complex serial and cross-sectional dynamics of daily and intra-daily asset return (co)variances and investigate the short-term information transmission on international financial markets as reflected by variance and covariance interdependencies. The proposed volatility models address both low-dimensional as well as high-dimensional volatility modeling. The models' in-sample properties are thoroughly analyzed using model diagnostic tests while the out-of-sample forecasting performance is evaluated using comprehensive out-of-sample experiments including a range of prominent forecasting models from the relevant literature.

Chapter 2 analyzes the theoretical and empirical properties of the Wishart multivariate Stochastic Volatility (WMSV) approach of Philipov and Glickman (2006) and Asai and McAleer (2009) and proposes a new flexible model extension featuring Markov switching (MS) volatility regimes. The proposed MS WMSV model allows for state-dependent volatility/correlation levels and volatility transmission effects across assets. An empirical application to five European stock index return series shows that the proposed regime-switching specification substantially improves the model fit. The MS WMSV model is found to capture sudden changes in the volatility level related to particular events like the 2005 terrorist attacks in London or the Lehman Brothers bust in September 2008, as well as lasting structural changes due the recent subprime crisis. The estimation results indicate the presence of a high-volatility and a low-volatility regime where states of high market volatility correspond to increasing market correlations. This indicates the presence of contagion effects in asset returns as well as vanishing diversification benefits in periods of turmoil. The high-volatility states are accompanied by increasing volatility transmission across assets, which indicates volatility contagion, i.e. crisis-related increases in inter-asset volatility dependencies. The enhanced in-sample properties of the MS WMSV approach are accompanied by according out-of-sample results: A Value-at-Risk (VaR) forecasting experiment shows that the MS WMSV model outperforms a range of competing volatility models from the literature with

respect to unconditional coverage of the 5% VaR level.

Chapter 3 addresses an alternative approach of covariance estimation and modeling which attracted substantial interest in recent years: the direct modeling of consistent estimates of variances and covariances of daily asset returns, so-called realized (co)variances. The realized volatility approach facilitates a more precise measurement and forecasting of daily asset return volatilities compared to MGARCH or MSV models. Chapter 3 proposes a conditional autoregressive Wishart (CAW) model for the analysis of realized covariance matrices of asset returns. The model is designed to represent complex temporal interdependencies across variances and covariances and is based upon an autoregressive moving average structure for the scale matrix of the central Wishart distribution. The model accounts for symmetry and positive definiteness of the predicted covariance matrices without imposing parametric restrictions and can easily be estimated by Maximum Likelihood. In order to explicitly account for long-memory type dependence patterns of realized (co)variances the CAW specification is combined with the mixed data sampling (MIDAS) approach of Ghysels et al. (2005, 2006) and, alternatively, with a heterogeneous autoregressive (HAR) component as used by Corsi (2009) and Bonato et al. (2009). The resulting CAW specifications stay in the spirit of component GARCH models with short- and long-run components and offer an attractive way of capturing complex dependence structures in asset return volatilities. An empirical application to daily realized covariance matrices for the returns of five stocks shows that the CAW model outperforms GARCH-type volatility models in 1-period ahead volatility forecasting, which reflects the gains of the direct modeling of realized (co)variances as opposed to daily return data based GARCH models. In terms of accounting for the observed dynamic behavior of the realized covariances as well as in terms of out-of-sample covariance predictions the MIDAS-CAW specification shows overall better results compared to the baseline CAW and the HAR-CAW alternatives. Furthermore, the MIDAS-CAW model is found to remove most, though not all, of the observed serial dependence in the variances and covariances.

Chapter 4 extends the CAW model in order to investigate the short-term interdependence of the realized variances and covariance of the US Dow Jones and the German stock index DAX. Compared to conditional volatilities obtained from MGARCH or MSV models the use of high-frequency data is expected to result in improved inference on volatility transmission across markets. In order to analyze intra-day volatility transmission patterns a novel sequential phase model is proposed, which accounts for the three distinct geographical intra-day trading periods of the US and German stock market. The model contributes to the literature by explicitly accounting for the chronological ordering of overlapping and non-overlapping trading periods while embodying the realized covariance as an additional indirect transmission channel for volatility shocks and accounting for the contemporaneous interdependence between the (co)variances during partially overlapping trading times. The empirical results show that both own market heat-wave and cross market meteor-shower effects are present across all intra-day trading periods. In general, the impact of heat-wave effects is larger

than of meteor-shower effects. Furthermore, the importance of those effects is found to depend critically on the information currency, that means more recent news appear to be more important than older news. Additionally, the results indicate considerable changes in the short-run volatility transmission mechanism during the recent subprime crisis period with a substantially stronger persistence of volatility shocks. The subprime crisis period shows up with much more pronounced meteor-shower effects of volatility shocks from the U.S. market, where the crisis originates, to the German market volatility.

In order to mitigate the curse of dimensionality in multivariate volatility modeling Chapter 5 illustrates a new flexible latent dynamic factor model for realized covariance matrices. The model is based on the matrix logarithm function which enables the modeling of log-(co)variances in Euclidean space, preserving positive definiteness and symmetry of covariance matrix forecasts without having to impose restrictions on the parameter space. By combining latent heterogeneous autoregressive processes (HAR, see Corsi, 2009) for the common factor structure with idiosyncratic AR(1) factors for series-specific dynamics the model effectively reduces the curse of dimensionality while allowing for rich (co)variance dynamics including a long-memory type of persistence. The simulated Bayesian estimation approach using basic Markov Chain Monte Carlo (MCMC) techniques enables straightforward estimation of the model parameters. An empirical application to realized (co)variances of up to 30 NYSE stocks shows that the factor model successfully accounts for the observed dynamic behavior of up to 465 (co)variance series, where 2 to 3 common factors appear overall sufficient in driving the cross-sectional dynamics. This dimension reduction enables practical applications like the forecasting of optimal portfolio weight vectors in mean/variance portfolio optimization, which typically requires precise (co)variance forecasts in high dimensions in order to effectively exploit diversification benefits. A comprehensive out-of-sample forecasting experiment based on the statistical root mean squared error criterium as well as an economic application to the forecasting of optimal portfolio weight vectors in mean/variance portfolio optimization shows that the factor model outperforms a range of prominent forecasting models from the relevant literature.

Summarizing the findings, the analysis provided in this thesis addresses several important aspects of multivariate volatility modeling and proposes novel approaches in order to assess and predict (co)variance dynamics. A particular focus is devoted to the modeling of strong persistence in multivariate daily return volatilities: Applications of Markov switching regimes and long term volatility components show promising results, in-sample as well as out-of sample. While the approaches are examined separately so far, future research may be devoted to combinations of long- and short-term components where the latter may be driven by Markov switching volatility levels accounting for structural changes linked to periods of turmoil and outstanding market events. The main line of analysis is devoted to the direct modeling of realized covariance matrices. The proposed conditional autoregressive Wishart (CAW) approach represents a particularly flexible, easy to extent and straight forward to

estimate modeling tool, which allows for a precise assessment of volatility transmission effects on international financial markets. An analysis of intra-day volatility transmission effects between the US and the German stock market shows several interesting results: the importance of accounting for the covariance transmission channel and contemporaneous dependence among the variances in order to enable identification of direct causal effects of news, as well as the strong dependence of spillover effects on the currency of the respective information content. A particular finding throughout the thesis is increasing return correlation and intensifying volatility transmission linked to crisis periods e.g. given by the subprime crisis and distinguished events like the terrorists attacks in London, 2005, which caused strong reactions in stock market volatilities. Since this strengthening of inter-market dependencies reinforces global financial crises, it is of crucial importance to provide models assessing these contagion effects in order to understand the impact of financial distress around the globe. A significant part of the present thesis is devoted to the evaluation of the forecasting performance of the proposed volatility models. The analysis addresses the classical mean-squared forecasting error criterion as well as the performance of mean-variance optimal portfolio and Value-at-Risk forecasts. Compared to the commonly applied MGARCH approach the direct modeling of realized (co)variances is found to result in improved 1-period ahead forecasting performance. Since practical applications of multivariate volatility models typically require forecasts of high-dimensional covariance matrices, the thesis investigates the presence of low-dimensional factor structures in (co)variance dynamics. Two to three common factors are found to be overall sufficient in order to drive the cross-sectional dynamics of volatilities and correlations of up to 30 assets, in-sample as well as out-of-sample.

Bibliography

- [1] Andersen, T.G., Bollerslev, T., Diebold, F.X., and P. Labys. 2003. Modeling and forecasting realized volatility. *Econometrica* 71: 579–625.
- [2] Andreou, E., and E. Ghysels. 2002. Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* 17: 579–600.
- [3] Asai, M., and M. McAleer. 2009. The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics* 150: 182–192.
- [4] Asai, M., McAleer, M., and J. Yu. 2006. Multivariate stochastic volatility: a review. *Econometric Reviews* 25: 145–175.
- [5] Audrino, F., and F. Corsi. 2010. Modeling tick-by-tick realized correlations. *Computational Statistics and Data Analysis* 54: 2372–2382.
- [6] Barndorff-Nielsen, O.E., and N. Shephard. 2004. Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica* 72: 885–925.
- [7] Barndorff-Nielsen, O.E., and N. Shephard. 2004b. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2: 1–37.
- [8] Bauer, G.H., and K. Vorkink. 2011. Forecasting multivariate realized stock market volatility. *Journal of Econometrics* 160: 93–101.
- [9] Bauwens, L., Laurent, S., and J.V.K. Rombouts. 2006. Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21: 79–109.
- [10] Bauwens, L., Lubrano, M., and J.F. Richard. 1999. *Bayesian Inference in Dynamic Econometric Models*. Oxford University press: New York.
- [11] Bauwens, L., and G. Storti. 2011. Modelling vast dimensional realized covariance matrices. Working paper.
- [12] Bekaert, G., Ehrmann, M., Fratzscher, M., and A. Mehl. 2011. Global crisis and equity market contagion. Working paper.
- [13] Black, F. 1976. Studies of stock market volatility changes. *Proceedings of the American Statistical Association. Business and Economic Statistics Section*: 177–181.
- [14] Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–327.

- [15] Bollerslev, T. 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics* 72: 498–505.
- [16] Bollerslev, T., and R.F. Engle. 1993. Common persistence in conditional variances. *Econometrica* 61: 167–186.
- [17] Bollerslev, T., Engle, R.F., and J. Wooldridge. 1988. A capital asset pricing model with time varying covariances. *Journal of Political Economy* 95: 116–131.
- [18] Bonato, M., Caporin, M., and A. Ranaldo. 2009. Forecasting realized (co)variances with a block structure Wishart autoregressive model. Working paper.
- [19] Brown, S.J. 1990. Estimating volatility. In S. Figlewski, W. Silber, and M. Subrahmanyam (Eds.), *Financial Options: From Theory to Practice*: 516–537. Irwin Professional Publishing.
- [20] Bubák, V., Kočenda, E., and F. Žikeš. 2011. Volatility transmission in emerging European foreign exchange markets. *Journal of Banking and Finance* 35: 2829–2841.
- [21] Cappiello, L., Engle, R.F., and K. Sheppard. 2006. Asymmetric dynamics in the correlation of global equity and bond returns. *Journal of Financial Econometrics* 4: 537–572.
- [22] Carvalho, C.M., and H.F. Lopes. 2007. Simulation-based sequential analysis of Markov switching stochastic volatility models. *Computational Statistics and Data Analysis* 51: 4526–4542.
- [23] Chiang, M.-H., and L.-M. Wang. 2011. Volatility contagion: a range-based volatility approach. *Journal of Econometrics* 165: 175–189.
- [24] Chib, S. 2001. Markov Chain Monte Carlo methods: computation and inference. In J.J. Heckman, and E.E. Leamer (Eds.), *Handbook of Econometrics* 5: 3569–3648. North Holland.
- [25] Chib, S., Nardari, F., and N. Shephard. 2006. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics* 134: 341–371.
- [26] Chiriac, R., and V. Voev. 2011. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics* 26: 922–947.
- [27] Chiu, T.Y.M., Leonard, T., and K.W. Tsui. 1996. The matrix-logarithmic covariance model. *Journal of the American Statistical Association* 91: 198–210.
- [28] Christie, A.A. 1982. The stochastic behavior of common stock variances: value, leverage and interest rate effects. *Journal of Financial Economics* 10: 407–432.
- [29] Christodoulakis, G.A., and S.E. Satchell. 2002. Correlated ARCH: modelling the time-varying correlation between financial asset returns. *European Journal of Operations Research* 139: 351–370.
- [30] Christoffersen, P. 1998. Evaluating interval forecasts. *International Economic Review* 39: 841–862.

- [31] Colacito, R., Engle, R.F., and E. Ghysels. 2011. A component model for dynamic correlations. *Journal of Econometrics* 164: 45–59.
- [32] Corsi, F. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7: 174–196.
- [33] Cubadda, G., Hecq, A., and F.C. Palm. 2009. Studying co-movements in large multivariate data prior to multivariate modelling. *Journal of Econometrics* 148: 25–35.
- [34] Danielsson, J. 1998. Multivariate stochastic volatility models: estimation and a comparison with VGARCH models. *Journal of Empirical Finance* 5: 155–173.
- [35] Diebold, F.X. 1986. Modeling the persistence of conditional variances: a comment. *Econometric Reviews* 5: 51–56.
- [36] Diebold, F.X., and K. Yilmaz. 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. *Economic Journal* 119: 158–171.
- [37] Dimpfl, T., and R.C. Jung. 2012. Financial market spillovers around the globe. *Applied Financial Economics* 22: 45–57.
- [38] Doz, C., and F. Renault. 2006. Factor stochastic volatility in mean models: a GMM approach. *Econometric Reviews* 25: 275–309.
- [39] Driessen, J., Maenhout, P.J., and G. Vilkov. 2009. The price of correlation risk: evidence from equity options. *The Journal of Finance* 64: 1377–1406.
- [40] Engle, R.F. 2002. Dynamic conditional correlation: a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20: 339–350.
- [41] Engle, R.F., Gallo, G.M., and M. Velucchi. 2012. Volatility spillover in East Asian financial markets: a MEM-based approach. *Review of Economics and Statistics* 94: 222–233.
- [42] Engle, R.F., Ghysels, E., and B. Sohn. 2009. Stock market volatility and macroeconomic fundamentals. Working paper.
- [43] Engle, R.F., Ito, T., and W.-L. Lin. 1990. Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. *Econometrica* 58: 525–542.
- [44] Engle, R.F., and K.F. Kroner. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* 11: 122–150.
- [45] Engle, R.F., and G.J. Lee. 1999. A permanent and transitory component model of stock return volatility. In R.F. Engle and H. White (Eds.), *Cointegration, Causality, and Forecasting, A Festschrift in Honor of Clive W.J. Granger*: 475–497. Oxford University Press.
- [46] Engle, R.F., Ng, V.K., and M. Rothschild. 1990b. Asset pricing with a factor-ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics* 45: 213–238.

- [47] Engle, R.F., and J.G. Rangel. 2008. The spline-GARCH model for low frequency volatility and its global macroeconomic causes. *Review of Financial Studies* 21: 1187–1222.
- [48] Fisher, R.A. 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10: 507–521.
- [49] Forbes, K.J., and R. Rigobon. 2002. No contagion, only interdependence: measuring stock market comovements. *The Journal of Finance* 57: 2223–2261.
- [50] Gallant, A.R., Rossi, P.E., and G. Tauchen. 1993. Nonlinear dynamic structures. *Econometrica* 61: 871–907.
- [51] Gander, M.P.S., and D.A. Stephens. 2007. Stochastic volatility modeling with general marginal distributions: inference, prediction and model selection. *Journal of Statistical Planning and Inference* 137: 3068–3081.
- [52] Gelman, A., Carlin, J.B., Stern, H.S., and D.B. Rubin. 2003. *Bayesian Data Analysis*, 2nd edition. Chapman and Hall/CRC Press: London.
- [53] Geweke, J., and R. Meese. 1981. Estimating regression models of finite but unknown order. *International Economic Review* 22: 55–70.
- [54] Geweke, J., and Zhou, G. 1996. Measuring the price of the arbitrage pricing theory. *Review of Financial Studies* 9: 557–587.
- [55] Ghysels, E., Santa-Clara, P., and R. Valkanov. 2005. There is a risk-return trade-off after all. *Journal of Financial Economics* 76: 509–548.
- [56] Ghysels, E., Santa-Clara, P., and R. Valkanov. 2006. Predicting volatility, how to get most out of returns data sampled at different frequencies. *Journal of Econometrics* 131: 59–95.
- [57] Ghysels, E., Sinko, A., and R. Valkanov. 2007. MIDAS regression, further results and new directions. *Econometric Reviews* 26: 53–90.
- [58] Golosnoy, V., Gribisch, B., and R. Liesenfeld. 2012. The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics* 167: 211–223.
- [59] Gouriéroux, C., Jasiak, J., and R. Sufana. 2009. The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics* 150: 167–181.
- [60] Gray, S.F. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Econometrics* 42: 27–62.
- [61] Haas, M., Mittnik, S., and M.S. Paoletta. 2004. A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics* 2: 493–530.
- [62] Hafner, C.M. 2003. Fourth moment structure of multivariate GARCH models. *Journal of Financial Econometrics* 1: 26–54.
- [63] Hafner, C.M., and O. Linton. 2010. Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics* 159: 55–73.

- [64] Hamao, Y., Masulis, R.W., and V. Ng. 1990. Correlations in price changes and volatility across international stock markets. *Review of Financial Studies* 3: 281–307.
- [65] Hamilton, J.D. 1994. *Time Series Analysis*, Princeton University Press: Princeton.
- [66] Hamilton, J.D., and R. Susmel. 1994. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64: 307–333.
- [67] Harvey, A., Ruiz, E., and N. Shephard. 1994. Multivariate stochastic variance models. *Review of Economic Studies* 61: 247–264.
- [68] Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57: 97–109.
- [69] Huang, X. and G. Tauchen. 2005. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3: 456–499.
- [70] Hull, J., and A. White. 1987. The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42: 281–300.
- [71] Jacquier, E., Polson, N.G., and P.E. Rossi. 1999. Stochastic volatility: univariate and multivariate extensions. Working paper.
- [72] Jin, X., and J.M. Maheu. 2011. Modeling realized covariances and returns. Working paper.
- [73] J.P. Morgan. 1996. RiskMetrics. Technical Document (fourth ed.): New York.
- [74] Kim, C.J., and C.R. Nelson. 1999. *State-Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*. The MIT Press.
- [75] Kim, S., Shephard, N., and S. Chib. 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65: 361–393.
- [76] Krishnan, C.N.V., Petkova, R., and P. Ritchken. 2009. Correlation risk. *Journal of Empirical Finance* 16: 353–367.
- [77] Kyle, A.S. 1985. Continuous auctions and insider trading. *Econometrica* 53: 1315–1336.
- [78] Lamoureux, C.G., and W.D. Lastrapes. 1990. Persistence in variance, structural change and the GARCH model. *Journal of Business and Economic Statistics* 8: 225–234.
- [79] Laurent, S., Rombouts, J.V.K., and F. Violante. 2009. On loss functions and ranking forecasting performances of multivariate volatility models. CIRANO discussion paper.
- [80] Ledoit, O., Santa-Clara, P., and M. Wolf. 2003. Flexible multivariate GARCH modeling with an application to international stock markets. *Review of Economics and Statistics* 85: 735–747.
- [81] Leeb, H., and B.M. Pötscher. 2009. Model selection. In T.G Andersen, R.A. Davis, J.-P. Kreiß, and T. Mikosch (Eds.), *Handbook of Financial Time Series*. Springer: New York.
- [82] Li, Q., and J.S. Racine. 2007. *Nonparametric Econometrics*. Princeton University Press: New Jersey.

- [83] Liesenfeld, R., and J.-F. Richard. 2003. Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance* 10: 505–531.
- [84] Liesenfeld, R., and J.-F. Richard. 2008. Improving MCMC, using efficient importance sampling. *Computational Statistics and Data Analysis* 53: 272–288.
- [85] Lin, W.L. 1992. Alternative estimators for factor GARCH models - a Monte Carlo comparison. *Journal of Applied Econometrics* 7: 259–279.
- [86] Lopes, H.F., and C.M. Carvalho. 2007. Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference* 137: 3082–3091.
- [87] Lopez, J.A., and C.A. Walter. 2001. Evaluating covariance matrix forecasts in a value-at-risk framework. *The Journal of Risk* 3: 69–97.
- [88] Lütkepohl, H. 1996. *Handbook of Matrices*. Wiley: Chichester.
- [89] Lütkepohl, H. 2005. *New Introduction to Multiple Time Series Analysis*. Springer: Berlin.
- [90] Markowitz, H. 1952. Portfolio selection. *The Journal of Finance* 7: 77–91.
- [91] Melvin, M., and B.P. Melvin. 2003. The global transmission of volatility in the foreign exchange market. *Review of Economics and Statistics* 85: 670–679.
- [92] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and E. Teller. 1953. Equation of state calculations by fast computing machines. *Journal of Chemical Physics* 21: 1087–1092.
- [93] Muirhead, R.J. 1982. *Aspects of Multivariate Statistical Theory*. Wiley: New Jersey.
- [94] Noureldin, D., Shephard, N., and K. Sheppard. 2011. Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 27: 907–933.
- [95] Onatski, A. 2010. Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics* 92: 1004–1016.
- [96] Philipov, A., and M.E. Glickman. 2006. Multivariate stochastic volatility via Wishart processes. *Journal of Business and Economic Statistics* 24: 313–328.
- [97] Pitt, M.K., and N. Shephard. 1999. Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association* 94: 590–599.
- [98] Pitt, M.K., and N. Shephard. 1999b. Time varying covariances: a factor stochastic volatility approach. In J.M. Bernardo, J.O. Berger, A.P. David, and A.F.M. Smith (Eds.), *Bayesian Statistics* 6: 547–570. Oxford University Press.
- [99] Protter, P. 2004. *Stochastic Integration and Differential Equations*. Springer: New York.
- [100] Ross, S.A. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13: 341–360.
- [101] Ross, S.M. 2002. *Simulation*. Academic Press.

- [102] Rubin, D.B. 1987. Comment on: The calculation of posterior distributions by data augmentation, by M.A. Tanner and W.H. Wong. *Journal of the American Statistical Association* 82: 543–546.
- [103] Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6: 461–464.
- [104] Sharpe, F. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *The Journal of Finance* 19: 425–442.
- [105] Shephard, N. 1996. Statistical aspects of ARCH and stochastic volatility. In D.R. Cox, D.V. Hinkley, and O.F. Barndorff-Nielsen (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*: 1–67. Chapman & Hall: London.
- [106] Sims, C.A. 1980. Macroeconomics and reality. *Econometrica* 48: 1–48.
- [107] Smith, M., and A. Pitts. 2006. Foreign exchange intervention by the Bank of Japan: Bayesian analysis using a bivariate stochastic volatility model. *Econometric Reviews* 25: 425–451.
- [108] So, M.K.P., Lam, K., and W.K. Li. 1998. A stochastic volatility model with Markov switching. *Journal of Business and Economic Statistics* 16: 244–253.
- [109] Solnik, B., Bourcelle, C., and Y. Le Fur. 1996. International market correlation and volatility. *Financial Analyst Journal* 52: 17–34.
- [110] Tanner, M., and W. Wong. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 82: 528–540.
- [111] Taylor, S.J. 1982. Financial returns modelled by the product of two stochastic processes - a study of daily sugar prices. In O.D. Anderson (Ed.), *Time Series Analysis: Theory and Practice* 1: 203–226. North-Holland, Amsterdam.
- [112] Taylor, S.J. 1986. *Modelling Financial Time Series*. Wiley: Chichester.
- [113] Tsay, R.S. 2005. *Analysis of Financial Time Series*. Wiley: Hoboken, New Jersey.
- [114] Tse, Y.K., and A.K.C. Tsui. 2002. A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* 20: 351–362.
- [115] Wongswan, J. 2006. Transmission of information across international equity markets. *Review of Financial Studies* 19: 1157–1189.
- [116] Yu, J., and R. Meyer. 2006. Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews* 25: 361–384.
- [117] Zhang, L., Mykland, P.A., and Y. Aït-Sahalia. 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 472: 1394–1411.

K O O P E R A T I O N E N

Abgesehen von meiner Person haben Roman Liesenfeld und Vasyl Golosnoy im Rahmen gemeinsamer Forschungsprojekte an den Kapiteln 3 und 4 der vorliegenden Arbeit mitgewirkt. Mein Anteil ist hierbei wie folgt gegeben:

Kapitel 3: The Conditional Autoregressive Wishart Model (Publiziert im “Journal of Econometrics” 167, 2012, S. 211-223)

Mein Beitrag zu dieser Arbeit besteht in

- der Mit-Entwicklung der behandelten Fragestellungen und ihrer inferentiell-statistischen Umsetzung
- der Aufarbeitung des verwendeten Datenmaterials und Entwicklung der verwendeten Computerprogramme in den Programmiersprachen GAUSS und MATLAB
- der Überarbeitung des ersten Manuskripts gemeinsam mit den Mitverfassern.

Kapitel 4: Intra-Daily Volatility Spillovers between the US and German Stock Markets (Eingereicht bei der Zeitschrift “Journal of Financial Econometrics”)

Mein Beitrag zu dieser Arbeit besteht in

- der Mit-Entwicklung der behandelten Fragestellungen und ihrer inferentiell-statistischen Umsetzung
- der Aufarbeitung des verwendeten Datenmaterials und Entwicklung der verwendeten Computerprogramme in der Programmiersprache MATLAB
- der Herleitung der Bedingung für die Existenz des stationären Erwartungswertes für das 3-Phasen Modell
- der Mitarbeit bei der Gliederung und Formulierung der ersten Fassung des Manuskripts, welches darauf gemeinsam mit den Mitverfassern überarbeitet wurde.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich, abgesehen von den auf der vorangegangenen Seite präzisierten Kooperationen mit Roman Liesenfeld und Vasyl Golosnoy, meine Doktorarbeit "Modeling and Forecasting of Multivariate Stock Market Volatility" selbständig und ohne fremde Hilfe angefertigt habe und dass ich alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehnenden Ausführungen meiner Arbeit, besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert habe.

Kiel, den 30.11.2012

Bastian Gribisch

CURRICULUM VITAE

PERSÖNLICHE ANGABEN

Geburtsdatum: 14.08.1981

Staatsangehörigkeit: Deutsch

Geburtsort: Bad Oldesloe

Familienstand: ledig

AUSBILDUNG

Oktober 2008 - heute	Promotionsstudium an der Christian-Albrechts-Universität zu Kiel, Thema der Dissertation: "Modeling and Forecasting of Multivariate Stock Market Volatility"
September 2008	Diplom Volkswirt (Note 1,0), Thema der Diplomarbeit: "Bayesian Analysis of Bivariate Stochastic Volatility Models"
Oktober 2004 - September 2008	Studium der Volkswirtschaftslehre an der Christian-Albrechts-Universität zu Kiel
August 2002 - Juni 2004	Ausbildung zum Bankkaufmann bei der Hamburger Sparkasse AG
August 2001 - Juli 2002	Zivildienst bei der Johanniter Unfallhilfe in Bad Oldesloe
Juni 2001	Abitur (Note 2,0) an der Theodor-Mommsen-Schule in Bad Oldesloe

PRAKTIKA/BERUFSERFAHRUNG

Februar 2007 - März 2007	Praktikum bei der Deutschen Bundesbank, Zentralbereich “Banken und Finanzaufsicht”, Abteilung “Mikro- und Makroprudenzielle Analyse des Bankensektors”: Mitarbeit an den Projekten “Ansteckungseffekte im Interbankenmarkt” und “Zinserträge und Fristentransformation”
September 2008 - heute	Wissenschaftlicher Mitarbeiter am Lehrstuhl für Statistik und empirische Wirtschaftsforschung an der Christian-Albrechts-Universität zu Kiel
Oktober 2006 - Juli 2008	Tutor am Institut für Statistik und Ökonometrie der Christian-Albrechts-Universität zu Kiel

FORSCHUNGSTHEMEN

- Multivariate stochastische Volatilitätsmodelle
- Modellierung des Zeitreihenverhaltens realisierter Volatilitätsmaße
- Volatilitätsprognosen
- Transmission von Aktienmarktvolatilitäten auf internationalen Finanzmärkten

LEHRVERANSTALTUNGEN

- Übung zur Methodenlehre der Statistik I (SS 2011)
- Übung zur Methodenlehre der Statistik II (WS 2008/2009 + WS 2010/2011)
- Übung zur Statistik für Fortgeschrittene I (SS 2009)
- Übung zur Statistik für Fortgeschrittene II (WS 2009/2010)
- Statistik für Fortgeschrittene: Klausurvorbereitungskurs (SS 2010)
- Übung zur Advanced Statistics I (WS 2011/2012 + WS 2012/2013)
- Übung zur Advanced Statistics II (SS 2012)
- Übung zur Mikroökonomie (WS 2008/2009 + SS 2010)
- Übung zu Statistics for Financial Markets
(WS 2009/2010 + WS 2010/2011 + WS 2011/2012 + WS 2012/2013)
- Übung zu Simulation Based Inference (SS 2009 + SS 2011 + SS 2012)

PRÄSENTATIONEN

- Juni 2012: “A latent dynamic factor approach to forecasting multivariate stock market volatility”, Forschungsseminar des Lehrstuhls für Statistik, Wirtschaftswissenschaftliche Fakultät der Universität Augsburg, Augsburg
- Dezember 2011: “Intra-Daily Volatility Spillovers between the U.S. and German Stock Markets”, 5th CSDA International Conference on Computational and Financial Econometrics (CFE'11), University of London, UK, London
- September 2011: “Intra-Daily Volatility Spillovers between the U.S. and German Stock Markets”, Jahrestagung des Vereins für Socialpolitik 2011, Frankfurt am Main
- August 2011: “Intra-Daily Volatility Spillovers between the U.S. and German Stock Markets”, EEA-ESEM 2011, Oslo
- Mai 2011: “The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility”, 2nd Humboldt - Copenhagen Conference on Financial Econometrics, Copenhagen
- Dezember 2010: “The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility”, 3rd International Conference of the ERCIM Working Group on Computing & Statistics (ERCIM'10), University of London, UK, London
- September 2010: “Multivariate Wishart Stochastic Volatility and Changes in Regime”, Statistische Woche, Nürnberg
- September 2010: “The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility”, Jahrestagung des Vereins für Socialpolitik, Kiel
- März 2010: “Multivariate Wishart Stochastic Volatility and Changes in Regime”, DAG-Stat 2010, Dortmund

MITGLIEDSCHAFTEN

- Econometric Society
- Deutsche Statistische Gesellschaft
- Verein für Socialpolitik

GUTACHTER TÄTIGKEIT

- Computational Statistics & Data Analysis

AUSZEICHNUNGEN

- Erich Schneider Preis für das beste Abschlussexamen im SS 2008
- Erich Schneider Preis für die beste Diplomarbeit im SS 2008
- Vortrags-Stipendium der Schweizerischen Nationalbank 2011

EDV-KENNTNISSE

- L^AT_EX, Microsoft Office
- MatLab, GAUSS, EViews, STATA, MAPLE, C++

SPRACHEN

- Deutsch
- Englisch, verhandlungssicher
- Spanisch, Grundkenntnisse

PUBLIKATIONSLISTE

VERÖFFENTLICHUNGEN IN REFERIERTEN FACHZEITSCHRIFTEN

- “The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility”, mit V. Golosnoy und R. Liesenfeld, 2012, Journal of Econometrics 167, 211-223

ARBEITSPAPIERE

- “Intra-Daily Volatility Spillovers between the U.S. and German Stock Markets”, mit V. Golosnoy und R. Liesenfeld
- “A Latent Dynamic Factor Approach to Forecasting Multivariate Stock Market Volatility”
- “Multivariate Wishart Stochastic Volatility and Changes in Regime”