Gribisch, Bastian

Working Paper

Multivariate wishart stochastic volatility and changes in regime

Economics Working Paper, No. 2012-14

Provided in Cooperation with:
Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Gribisch, Bastian (2012) : Multivariate wishart stochastic volatility and changes in regime, Economics Working Paper, No. 2012-14, Kiel University, Department of Economics, Kiel

This Version is available at:
https://hdl.handle.net/10419/66205

Christian-Albrechts-Universität zu Kiel

Department of Economics

Economics Working Paper
No 2012-14

multivariate wishart stochastic volatility and changes in regime

by Bastian Gribisch

issn 2193-2476

Multivariate Wishart Stochastic Volatility and Changes in Regime

Bastian Gribisch *

Institute of Statistics and Econometrics, Christian-Albrechts-Universität Kiel, Germany

(November 6, 2012)

Abstract

This paper generalizes the basic Wishart multivariate stochastic volatility model of Philipov and Glickman (2006) and Asai and McAleer (2009) to encompass regime switching behavior. The latent state variable is driven by a first-order Markov process. The model allows for state-dependent (co)variance and correlation levels and state-dependent volatility spillover effects. Parameter estimates are obtained using Bayesian Markov Chain Monte Carlo procedures and filtered estimates of the latent variances and covariances are generated by particle filter techniques. The model is applied to five European stock index return series. The results show that the proposed regime-switching specification substantially improves the in-sample fit and the VaR forecasting performance relative to the basic model.

JEL classification: C32, C58, G17

Keywords: Multivariate stochastic volatility, Dynamic correlations, Wishart distribution, Markov switching, Markov chain Monte Carlo;

*Corresponding address: Institut für Statistik und Ökonometrie, Christian-Albrechts-Universität zu Kiel, Ohlshausenstraße 40-60, D-24118 Kiel, Germany. Tel.: +49(0)4318802229; fax: +49(0)4318807605. *E-mail address:* bastian.gribisch@stat-econ.uni-kiel.de (B. Gribisch)

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 (see e.g. Hull and White, 1987). However, developing flexible multivariate SV (MSV) specifications proved to be complicated.[1] Proposed MSV models, e.g. employed by Daníelsson (1998), Harvey et al. (1994) 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 that models that allow for time-varying correlations offer a better fit to the data. Factor SV models e.g. applied by Chib et al. (2006), Doz and Renault (2006) and Pitt and Shephard (1999b) 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. Asai and McAleer (2009) and Philipov and Glickman (2006) 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 (see e.g. Bodnar and Okhrin, 2008, and Muirhead, 1982). 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 and correlation dynamics since the scale matrix of the Wishart distribution is modeled conditional on the history of the complete covariance matrix. The desirable properties of the Wishart distribution also contribute to its increasing popularity in the literature on direct modeling of realized (co)variance measures (see e.g. Jin and Maheu, 2012, and Noureldin et al., 2011).

[1] See the excellent overview on multivariate SV models of Asai et al., 2006.

The present paper 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 correlations and state-dependent volatility transmission across assets, so-called volatility spillover effects (see e.g. Gallo and Otranto, 2008). 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. Fast tracking of structural changes in the (co)variance structure helps to avoid this bias. Gray (1996), Haas et al. (2004) and Hamilton and Susmel (1994) 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 distinct 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. Billio and Carporin, 2010, Chiang and Wang, 2011, and Forbes and Rigobon, 2002), 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 volatility 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 rest of the paper is organized as follows. Section 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 3 presents estimation- and model diagnostic results and the VaR forecasting application. Section 4 concludes.

2. Model Specification, Bayesian Inference and Model Diagnostics

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 $(t = 1, \ldots, T)$. The basic WMSV model is given by

$$\xi_t | \Sigma_t \sim N(0, \Sigma_t), \tag{1}$$

$$\Sigma_t^{-1} | \Sigma_{t-1}^{-1} \sim \mathcal{W}_k(\nu, S_t/\nu), \tag{2}$$

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)

$$E[\Sigma_t^{-1}|\Sigma_{t-1}^{-1}] = S_t, \tag{3}$$

$$E[\Sigma_t|\Sigma_{t-1}] = \frac{1}{\nu - k - 1} S_{t-1}^{-1}. \tag{4}$$

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:

$$S_t = \Sigma_{t-1}^{-d/2} A \Sigma_{t-1}^{-d/2}, \tag{5}$$

where A is a positive definite $k \times k$ parameter matrix and d is a scalar persistence parameter.[2] Based on the spectral decomposition $\Sigma_t^{-1} = V_t \Lambda_t V_t'$ we obtain

$$\Sigma_t^{-d/2} = V_t \Lambda_t^{\frac{d}{2}} V_t', \tag{6}$$

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 expression in Eq. (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 WMSV model using the properties of the Wishart distribution: Denoting the $k \times k$ identity matrix by I_k and the lower

[2] The 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: $S_t = A^{1/2}\Sigma_{t-1}^{-d}A^{1/2'}$.

triangular Cholesky factor of A by L, i.e. $A = LL'$, we obtain

$$\Sigma_t^{-1} = \frac{1}{\nu}\Sigma_{t-1}^{-d/2} L \; \mathcal{W}_k(\nu, I_k) \; L' \Sigma_{t-1}^{-d/2}, \tag{7}$$

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

$$\ln|\Sigma_t^{-1}| = -k\ln(\nu) + \ln|A| + d\ln|\Sigma_{t-1}^{-1}| + \ln|\mathcal{W}_k(\nu, I_k)|. \tag{8}$$

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. (4) and (5) show that the interpretation of inter-temporal (co)variance transmission is actually based on 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 $\{d_1 = 0.2,\ d_2 = 0.4,\ d_3 = 0.6,\ d_4 = 0.8,\ d_5 = 0.9\}$ and $\{\nu_1 = 20,\ \nu_2 = 40,\ \nu_3 = 60,\ \nu_4 = 80,\ \nu_5 = 90\}$. The matrices A_i^{-1}, $i = 1, \ldots, 5$, are characterized by overall increasing matrix entries in i on each single position in A_i^{-1}. Let 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 vech(A_1^{-1}) is set to its point estimate 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 Section 3 below). For $i = 2, \ldots, 5$ we obtain

$\text{vech}(A_i^{-1}) = (1.2,\ 2,\ 1.2)' \odot \text{vech}(A_{i-1}^{-1})$, where $\odot$ denotes element-wise multiplication[3]. Figure 1 shows 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 (left panel) 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 (right panel) 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):

$$\text{Cov}(\sigma_{ij,t}^{-1}, \sigma_{lm,t}^{-1} | \Sigma_{t-1}^{-1}) = \frac{1}{\nu}(s_{il,t} \cdot s_{jm,t} + s_{im,t} \cdot s_{jl,t}), \tag{9}$$

for $i, j, l, m = 1, \ldots, k$, 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. 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

[3]The simulation results are found to be robust to variations in the parameter values.

matrix

$$\Pr(s_t|s_{t-1}) = \begin{bmatrix} (1-e_1) & e_1 \\ e_2 & (1-e_2) \end{bmatrix}, \tag{10}$$

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

$$\xi_t|\Sigma_t \sim N\Big(0, \Sigma_t\Big), \tag{11}$$

$$\Sigma_t^{-1}|\Sigma_{t-1}^{-1} \sim \mathcal{W}_k(\nu, S_t/\nu), \qquad S_t = \Sigma_{t-1}^{-d/2} A_{s_t} \Sigma_{t-1}^{-d/2}, \tag{12}$$

together with Eq. (10). According to the simulation results of Section 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 positivity. Note that it is straightforward 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 Carvalho and Lopes (2007) and Lopes and Carvalho (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.

2.3. Estimation and Diagnostics

Following Asai and McAleer (2009) and 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

$$L(\{\xi_t\}_{t=1}^T|\theta^{\text{WMSV}}) = \int_{\Sigma_1,\ldots,\Sigma_T} \ldots \int \prod_t^T P(\xi_t|\Sigma_t) \times P(\Sigma_t|\Sigma_{t-1}, \theta^{\text{WMSV}})\ d\Sigma_1, \ldots, d\Sigma_T. \tag{13}$$

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 posterior 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 variables[4]. Bayesian point estimates are obtained by averaging the Gibbs draws after convergence of the Markov chain[5]. 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

[4]For details on the Gibbs sampling algorithm and Monte Carlo Markov Chain methods see e.g. Bauwens et al. (1999).

[5]I.e. after a certain number of burn-in iterations of the Gibbs sampler.

$\{s_t\}_{t=1}^T$ 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.

After a model has been fitted to the data, diagnostics are applied in order to check the model's ability to reflect (co)variance dynamics and distributional characteristics of the observed return series. Diagnostic tests on (co)variance dynamics are conducted from the vector of standardized Pearson residuals

$$e_t^* \quad = \quad \text{Var}[\xi_t|\mathcal{F}_{t-1}]^{-\frac{1}{2}}\xi_t = E[\Sigma_t|\mathcal{F}_{t-1}]^{-\frac{1}{2}}\xi_t, \tag{14}$$

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:

$$e_t^* = E[\Sigma_t|\mathcal{F}_{t-1}]^{-\frac{1}{2}}\xi_t \cong \left(\frac{1}{M}\sum_{j=1}^{M}\Sigma_t^{(j)}\right)^{-\frac{1}{2}}\xi_t, \tag{15}$$

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

be approximated by

$$Pr(y_{i,t} \leq y_{i,t}^{o}|\mathcal{F}_{t-1}) \cong u_{i,t}^{M} = \frac{1}{M}\sum_{j=1}^{M} \Pr(y_{i,t} \leq y_{i,t}^{o}|\sigma_{ii,t}^{(j)}), \tag{16}$$

where $\sigma_{ii,t}^{(j)}$ denotes the i'th diagonal element of $\Sigma_t^{(j)}$, drawn from $f(\Sigma_t|\mathcal{F}_{t-1})$, $j = 1, \ldots, M$. 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 $i = 1, \ldots, k$ 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.

3. Empirical Application

3.1 Data

The (MS) WMSV models are applied to daily AR(p) pre-filtered stock index log-returns[6] for France, Germany, Italy, Switzerland and the UK from January 2, 2003 to December 31, 2008, leaving a sample of 1565 observations.[7] The return series are illustrated in Figure 3 and descriptive statistics are given in Table 1. All series feature excess kurtosis, insignificant autocorrelation in levels and significant autocorrelation in squared returns. The reported sample correlations indicate a huge degree of co-movement for all five stock indices.

3.2 Estimation Results

3.2.1 Basic WMSV Model

Table 2 presents the estimation results for the basic WMSV model. The chosen prior distributions are overall uninformative and reported in the table. 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

[6] Datastream DS market indices.

[7] The 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.

Ross (2002). All parameter estimates are significant at the 5% level.

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 4 and 5 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 5 shows strong co-movement and significant dynamics in the correlation series.

Table 3 shows diagnostics for the series of Pearson residual cross-products. The Ljung-Box test at 50 lags indicates significant predictability of 14 out of 15 series. 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 6 (left panel) 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 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). The previous findings are supported by QQ-plots depicted in Figure 7 (left panel). The plots show severe deviations from normality in the tails of the residual distribution.

3.2.2 MS WMSV Model

Table 2 shows estimation results and prior distributions for the two-state Markov switching WMSV model. 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 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 8 depicts smoothed state and volatility estimates for France and Germany obtained under the basic WMSV and the MS WMSV model. The figure illustrates the link of the second volatility state to periods of high market volatility, where MS WMSV implied volatility significantly exceeds basic WMSV implied volatility. 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 midst 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 tracking of 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). 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 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 periods of turmoil 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 mar-

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

ket 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 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 3 shows Ljung-Box diagnostic test results for the series of Pearson residual cross-products. 13 out of 15 series are unpredictable at the 1% level. This finding is supported by sample ACFs of squared residual series depicted in Figure 6 (right panel). Compared to the basic WMSV approach the MS framework offers enhanced flexibility in capturing strong persistence of asset return (co)variances. The MS WMSV model captures long-memory like persistence patterns by combining structural shifts in the mean of the volatility process with volatility persistence in each regime. Table 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. Figure 7 (right panel) depicts QQ-plots which confirm the Jarque-Bera test results. 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.

3.3 Value-at-Risk Forecasting Application

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

$$\text{VaR}_{p,t|t-1}(\alpha) = \sqrt{\hat{\sigma}_{p,t|t-1}}\; F^{-1}(\alpha), \tag{17}$$

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 now explicitly based on VaR estimates and include a penalty for model inaccuracy (see Lopez and Walter, 2001).

According to common practice (see e.g. Chib et al., 2006, and Lopez and Walter, 2001) 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 $\xi_t = H_t^{1/2}\upsilon_t$, where $\upsilon_t \sim \mathcal{N}(0, I_k)$ and $H_t^{1/2}$ is the lower triangular Cholesky factor of the conditional covariance matrix H_t, which is specified as

$$H_t = D_0 D_0' + \sum_{i=1}^{p} D_i H_{t-i} D_i' + \sum_{j=1}^{q} G_j [\xi_{t-j}\xi_{t-j}'] G_j', \tag{18}$$

   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 condi-

tional variances $\{h_{ii,t}\}_{i=1}^{k}$. The modeling of dynamic conditional correlations is based on the decomposition

$$H_t = D_t P_t D_t, \tag{19}$$

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

$$P_t = \big(\text{diag}(Q_t)\big)^{-\frac{1}{2}} Q_t \big(\text{diag}(Q_t)\big)^{-\frac{1}{2}}, \tag{20}$$

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

$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha u_{t-1} u_{t-1}' + \beta Q_{t-1}, \tag{21}$$

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

$$u_{i,t} = \frac{\xi_{i,t}}{\sqrt{h_{ii,t}}}, \quad i = 1, \ldots, k. \tag{22}$$

$\bar{Q}$ 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

$$H_t = (1 - \lambda)\xi_{t-1}\xi_{t-1}' + \lambda H_{t-1}. \tag{23}$$

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. pro-

vided 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 coverage test illustrated by Lopez and Walter (2001) and e.g. applied by Chib et al. (2006) and Storti (2006). This test is explicitly incorporated into the Basel bank capital requirements. Defining an indicator variable

$$I_t = \begin{cases} 1 & \text{if } \xi_{p,t} < \text{VaR}_{p,t|t-1}, \\ 0 & \text{if } \xi_{p,t} \geq \text{VaR}_{p,t|t-1}, \end{cases} \tag{24}$$

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

$$LR_{\text{uc}} = 2\{ \ln[\ \hat{\alpha}^\gamma (1-\hat{\alpha})^{T^\star - \gamma}\ ] - \ln[\ \alpha^\gamma (1-\alpha)^{T^\star - \gamma}\ ]\ \}, \tag{25}$$

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

Table 6 presents the forecasting results. The basic WMSV model shows the overall worst VaR forecasting performance across all considered volatility models - the hit-rate amounts to 15%. The significant overestimation of coverage can be traced back to the model's general problem in capturing the leptokurtic distribution of daily asset return data. This shortcoming clearly effects the VaR measure. Extending the basic WMSV model by Markov switching regimes significantly improves the VaR forecasting results. The respective hit-rate amounts to 8.78%, which is closest to the 5% level across all considered forecasting models - according to the test results we cannot reject the null of correct unconditional coverage at the 1% significance level. The fact that most applied volatility models show significant violations of coverage is explained by the overall high volatility level in 2008

which comes as a result of the subprime crisis.

4. Conclusions

This paper proposes a new Markov switching (MS) extension to the basic Wishart MSV (WMSV) model of Asai and McAleer (2009) and Philipov and Glickman (2006). 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.

Appendix

Full Conditional Distributions: Basic WMSV Model

The basic WMSV model is outlined in Eqs. (1), (2) and (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 $\theta^{\text{aug}} = (\theta', \text{vech}(\Sigma_1)', \ldots, \text{vech}(\Sigma_T)')$, we obtain

$$\begin{aligned} P(\theta^{\text{aug}}|\Xi) \quad \propto \quad & \prod_{t=1}^{T} f(\xi_t|\Sigma_t^{-1}) \times f(\Sigma_t^{-1}|\Sigma_{t-1}^{-1}, \theta) \\ & \times \pi_{A^{-1}}(Q_0, \gamma_0) \times \pi_d(0,1) \times \pi_\nu(\alpha_0, \beta_0). \end{aligned} \tag{26}$$

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

$$\begin{aligned} p(\Sigma_t^{-1}|\theta_-^{\text{aug}}) \quad \propto \quad & f(\xi_t|\Sigma_t^{-1}) \times f(\Sigma_t^{-1}|\Sigma_{t-1}^{-1}) \times f(\Sigma_{t+1}^{-1}|\Sigma_t^{-1}) \\ \propto \quad & |\Sigma_t^{-1}|^{(\nu-k-d\nu)/2} \times \exp\{-0.5 \text{ tr}[(S_t^{-1} + \xi_t\xi_t')\Sigma_t^{-1}]\} \\ & \times \exp\{-0.5 \text{ tr}[S_{t+1}^{-1}\Sigma_{t+1}^{-1}]\} \\ \propto \quad & \mathcal{W}_k^\kappa(\Sigma_t^{-1}|\tilde{\nu}, \tilde{S}_t) \times f(\Sigma_t^{-1}), \end{aligned} \tag{27}$$

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

$$\begin{aligned}
\tilde{\nu} &= \nu(1-d)+1, && (28)\\
\tilde{S}_t &= (S_t^{-1}+\xi_t\xi_t')^{-1}, && (29)\\
f(\Sigma_t^{-1}) &= \exp\{-0.5\ \mathrm{tr}[S_{t+1}^{-1}\Sigma_{t+1}^{-1}]\}, && (30)\\
S_t &= \Sigma_t^{-d/2}A\Sigma_t^{-d/2}. && (31)
\end{aligned}$$

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 $p(\Sigma_t^{-1}|\theta_-^{\mathrm{aug}})$. The proposal density is given by $\mathcal{W}_k(\nu,\tilde{S}_t)$.

Full conditional distribution of A^{-1}:

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

$$\begin{aligned}
p(A^{-1}|\theta_-^{\mathrm{aug}}) &\propto \pi_{A^{-1}}(Q_0,\gamma_0)\ \prod_{t=1}^{T} f(\Sigma_t^{-1}|\Sigma_{t-1}^{-1})\\
&\propto \pi_{A^{-1}}(Q_0,\gamma_0)\ |A^{-1}|^{(T\nu)/2}\\
&\quad\times\exp\Big\{-0.5\mathrm{tr}\Big[\nu\sum_{t=1}^{T}\Sigma_{t-1}^{d/2}\Sigma_t^{-1}\Sigma_{t-1}^{d/2}A^{-1}\Big]\Big\}\\
&\propto \pi_{A^{-1}}(Q_0,\gamma_0)\times\mathcal{W}_k^{\kappa}(A^{-1}|\gamma,U), \qquad (32)
\end{aligned}$$

where $U^{-1}=\nu\sum_{t=1}^{T}\Sigma_{t-1}^{d/2}\Sigma_t^{-1}\Sigma_{t-1}^{d/2}$ and $\gamma=T\nu+k+1$. Hence

$$\begin{aligned}
p(A^{-1}|\theta_-^{\mathrm{aug}}) &\propto \pi_{A^{-1}}(Q_0,\gamma_0)\times\mathcal{W}_k^{\kappa}(A^{-1}|\gamma,U)\\
&\propto |A^{-1}|^{(\gamma_0+\gamma-2k-2)/2}\ \exp\{-0.5\ \mathrm{tr}[(Q_0^{-1}+U^{-1})A^{-1}]\}. \qquad (33)
\end{aligned}$$

Therefore $A^{-1}|\theta_-^{\mathrm{aug}}\sim\mathcal{W}_k(\tilde{\gamma},\tilde{U})$, where $\tilde{U}^{-1}=Q_0^{-1}+U^{-1}$ and $\tilde{\gamma}=\gamma_0+\gamma-k-1$.

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

$$\begin{aligned} p(d|\theta_{-}^{\text{aug}}) &\propto \pi_d(0,1)\prod_{t=1}^{T}|\Sigma_{t-1}^{-1}|^{-d\nu/2} \\ &\quad \exp\Big\{-0.5\ \text{tr}\Big[\Big((1/\nu)\Sigma_{t-1}^{-d/2}A\Sigma_{t-1}^{-d/2}\Big)^{-1}\Sigma_t^{-1}\Big]\Big\} \\ &\propto \exp\Big\{d\psi - 0.5\ \text{tr}\Big[Q(d)A^{-1}\Big]\Big\}, \end{aligned} \tag{34}$$

where $\psi = -\frac{\nu}{2}\sum_{t=1}^{T}\ln(|\Sigma_{t-1}^{-1}|)$ and $Q(d) = \sum_{t=1}^{T}\nu\Sigma_{t-1}^{d/2}\Sigma_t^{-1}\Sigma_{t-1}^{d/2}$.

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

$$\begin{aligned} p(\nu|\theta^{\text{aug}}) &\propto \pi_\nu(\alpha_0,\beta_0)\times\prod_{t=1}^{T}f(\Sigma_t^{-1}|\Sigma_{t-1}^{-1}) \\ &\propto \exp\{(\alpha-1)\ln(\nu-k)-\beta(\nu-k)\} \\ &\quad \times\left(\frac{|\nu A^{-1}|^{\nu/2}}{2^{\nu k/2}\prod_{j=1}^{k}\Gamma\Big((\nu-j+1)/2\Big)}\right)^{T} \\ &\quad \times\prod_{t=1}^{T}|Q_t^{-1}|^{\nu/2}\exp\{-0.5\ \text{tr}\Big[Q^{-1}A^{-1}\Big]\}, \end{aligned} \tag{35}$$

where $Q_t^{-1} = \Sigma_{t-1}^{d/2}\Sigma_t^{-1}\Sigma_{t-1}^{d/2}$ and $Q^{-1} = \nu\sum_{t=1}^{T}\Sigma_{t-1}^{d/2}\Sigma_t^{-1}\Sigma_{t-1}^{d/2}$.

Full Conditional Distributions: Markov Switching MWSV Model

The MS WMSV model is outlined in Eqs. (10), (11) and (12). The joint prior distribution is assumed to factor into the product of marginal prior distributions. Given the state sequence $s = (s_1, s_2, \ldots, s_T)'$, 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 $A_{s_t}\forall t = 1,\ldots,T$ instead of A.

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

Denoting $\underline{\Sigma_t^{-1}} = \{\Sigma_1^{-1},\ldots,\Sigma_t^{-1}\}$ and exploiting the Markov property of s_t, the full conditional

density of the state vector s can be factorized as

$$\begin{aligned} p(s|\theta_{-}^{\text{aug}}) &= P(s|\underline{\Sigma}_T^{-1},\theta) \\ &= P(s_T|\underline{\Sigma}_T^{-1},\theta) \times P(s_{T-1}|s_T,\underline{\Sigma}_T^{-1},\theta) \times \cdots \times P(s_1|s_2,\underline{\Sigma}_T^{-1},\theta) \\ &= P(s_T|\underline{\Sigma}_T^{-1},\theta) \times P(s_{T-1}|s_T,\underline{\Sigma}_{T-1}^{-1},\theta) \times \cdots \times P(s_1|s_2,\underline{\Sigma}_1^{-1},\theta). \end{aligned} \tag{36}$$

The conditional probabilities

$$P(s_t|s_{t+1},\underline{\Sigma}_t^{-1},\theta) = \frac{P(s_{t+1}|s_t) \times P(s_t|\underline{\Sigma}_t^{-1},\theta)}{P(s_{t+1}|\underline{\Sigma}_t^{-1},\theta)} \tag{37}$$

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 $\forall t \in \{1,\ldots,T\}$:

$$I \qquad P(s_t,s_{t-1}|\underline{\Sigma}_{t-1}^{-1},\theta) = P(s_t|s_{t-1}) \times P(s_{t-1}|\underline{\Sigma}_{t-1}^{-1},\theta) \tag{38}$$

$$II \qquad P(s_t|\underline{\Sigma}_{t-1}^{-1},\theta) = \sum_{s_{t-1}} P(s_t,s_{t-1}|\underline{\Sigma}_{t-1}^{-1},\theta) \tag{39}$$

$$III \qquad f(\Sigma_t^{-1},s_t|\underline{\Sigma}_{t-1}^{-1},\theta) = f(\Sigma_t^{-1}|s_t,\Sigma_{t-1}^{-1},\theta) \times P(s_t|\underline{\Sigma}_{t-1}^{-1},\theta) \tag{40}$$

$$IV \qquad f(\Sigma_t^{-1}|\underline{\Sigma}_{t-1}^{-1},\theta) = \sum_{s_t} f(\Sigma_t^{-1},s_t|\underline{\Sigma}_{t-1}^{-1},\theta) \tag{41}$$

$$V \qquad P(s_t|\underline{\Sigma}_t^{-1},\theta) = \frac{f(\Sigma_t^{-1},s_t|\underline{\Sigma}_{t-1}^{-1},\theta)}{f(\Sigma_t^{-1}|\underline{\Sigma}_{t-1}^{-1},\theta)}. \tag{42}$$

The whole state sequence $s = (s_1,s_2,\ldots,s_T)'$ can then be sampled backward recursively based on Eq. (36).

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

$$\begin{aligned} p(e_i|\text{rest}) &\propto \pi_{e_i}(\alpha_{i,0},\beta_{i,0}) \times \prod_{j=1}^{g_i} e_i \prod_{j=1}^{h_i}(1-e_i) \\ &\propto e_i^{\alpha_{i,0}-1}(1-e_i)^{\beta_{i,0}-1} \times e_i^{g_i}(1-e_i)^{h_i}, \end{aligned} \tag{43}$$

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 $\beta_i = \beta_{i,0} + h_i$, $i \in \{1, 2\}$.

References

[1] Andreou, E., Ghysels, E., 2002. Detecting multiple breaks in financial market volatility dynamics. Journal of Applied Econometrics 17, 579–600.

[2] Asai, M., McAleer, M., 2009. The structure of dynamic correlations in multivariate stochastic volatility models. Journal of Econometrics 150, 182–192.

[3] Asai, M., McAleer, M., Yu, J., 2006. Multivariate stochastic volatility: a review. Econometric Reviews 25, 145–175.

[4] Bauwens, L., Lubrano, M., Richard, J.F., 1999. Bayesian Inference in Dynamic Econometric Models. Oxford University press, New York.

[5] Bekaert, G., Ehrmann, M., Fratzscher, M., Mehl, A., 2011. Global crisis and equity market contagion. Working paper.

[6] Billio, M., Carporin, M., 2010. Market linkages, variance spillovers, and correlation stability: Empirical evidence of financial contagion. Computational Statistics and Data Analysis 54, 2443–2458.

[7] Black, F. 1976. Studies of stock market volatility changes. Proceedings of the American Statistical Association, Business and Economic Statistics Section, 177–181.

[8] Bodnar, T., Okhrin, Y., 2008. Properties of the singular, inverse and generalized inverse partitioned Wishart distributions. Journal of Multivariate Analysis 99, 2389–2405.

[9] 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.

[10] Carvalho, C.M., Lopes, H.F., 2007. Simulation-based sequential analysis of Markov switching stochastic volatility models. Computational Statistics and Data Analysis 51, 4526–4542.

[11] Chiang, M.-H., Wang, L.-M., 2011. Volatility contagion: a range-based volatility approach. Journal of Econometrics 165, 175–189.

[12] Chib, S., Nardari, F., Shephard, N., 2006. Analysis of high dimensional multivariate stochastic volatility models. Journal of Econometrics 134, 341–371.

[13] Christie, A.A., 1982. The stochastic behavior of common stock variances: value, leverage and interest rate effects. Journal of Financial Economics 10, 407–432.

[14] Daníelsson, J., 1998. Multivariate stochastic volatility models: estimation and a comparison with VGARCH models. Journal of Empirical Finance 5, 155–173.

[15] Diebold, F.X., 1986. Modeling the persistence of conditional variances: a comment. Econometric Reviews 5, 51–56.

[16] Diebold, F.X., Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. Economic Journal 119, 158–171.

[17] Doz, C., Renault, F., 2006. Factor stochastic volatility in mean models: a GMM approach. Econometric Reviews 25, 275–309.

[18] Engle, R.F., 2002. Dynamic conditional correlation: a simple class of multivariate GARCH models. Journal of Business and Economic Statistics 20, 339–350.

[19] Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150.

[20] Forbes, K.J., Rigobon, R., 2002. No contagion, only interdependence: measuring stock market comovements. The Journal of Finance 57, 2223–2261.

[21] Gallo, G.M., Otranto, E., 2008. Volatility spillovers, interdependence and comovements: a Markov switching approach. Computational Statistics and Data Analysis 52, 3011–3026.

[22] Gander, M.P.S., Stephens, D.A., 2007. Stochastic volatility modeling with general marginal distributions: inference, prediction and model selection. Journal of Statistical Planning and Inference 137, 3068–3081.

[23] Gray, S.F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Econometrics 42, 27–62.

[24] Haas, M., Mittnik, S., Paolella, M.S., 2004. A new approach to Markov-switching GARCH models. Journal of Financial Econometrics 2, 493–530.

[25] Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton.

[26] Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64, 307–333.

[27] Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Review of Economic Studies 61, 247–264.

[28] Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. Journal of Finance 42, 281–300.

[29] Jin, X., and J.M. Maheu, 2012. Modelling realized covariances and returns. Working paper.

[30] J.P. Morgan, 1996. RiskMetrics. Technical Document (fourth ed.), New York.

[31] Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.

[32] Lamoureux, C.G., Lastrapes, W.D., 1990. Persistence in variance, structural change and the GARCH model. Journal of Business and Economic Statistics 8, 225–234.

[33] Liesenfeld, R., Richard, J.-F., 2003. Univariate and multivariate stochastic volatility models: estimation and diagnostics. Journal of Empirical Finance 10, 505–531.

[34] Liesenfeld, R., Richard, J.-F., 2008. Improving MCMC, using efficient importance sampling. Computational Statistics and Data Analysis 53, 272–288.

[35] Lopes, H.F., Carvalho, C.M., 2007. Factor stochastic volatility with time varying loadings and Markov switching regimes. Journal of Statistical Planning and Inference 137, 3082–3091.

[36] Lopez, J.A., Walter, C.A., 2001. Evaluating covariance matrix forecasts in a value-at-risk framework. The Journal of Risk 3, 69–97.

[37] Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New Jersey.

[38] Noureldin, D., Shephard, N., and K. Sheppard, 2011. Multivariate high-frequency-based volatility (HEAVY) models. Journal of Applied Econometrics 27, 907–933.

[39] Philipov, A., Glickman, M.E., 2006. Multivariate stochastic volatility via Wishart processes. Journal of Business and Economic Statistics 24, 313–328.

[40] Pitt, M.K., Shephard, N., 1999. Filtering via simulation: auxiliary particle filters. Journal of the American Statistical Association 94, 590–599.

[41] Pitt, M.K., Shephard, N., 1999b. Time varying covariances: a factor stochastic volatility approach, in: Bernardo, J.M., Berger, J.O., David, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 6. Oxford University Press, pp. 547–570.

[42] Ross, S.M., 2002. Simulation. Academic Press.

[43] Smith, M., Pitts, A., 2006. Foreign exchange intervention by the Bank of Japan: Bayesian analysis using a bivariate stochastic volatility model. Econometric Reviews 25, 425–451.

[44] So, M.K.P., Lam, K., Li, W.K., 1998. A stochastic volatility model with Markov switching. Journal of Business and Economic Statistics 16, 244–253.

[45] Solnik, B., Bourcrelle, C., Le Fur, Y., 1996. International market correlation and volatility. Financial Analyst Journal 52, 17–34.

[46] Storti, G., 2006. Minimum distance estimation of GARCH(1,1) models. Computational Statistics and Data Analysis 51, 1803–1821.

[47] Taylor, S.J., 1982. Financial returns modelled by the product of two stochastic processes - a study of daily sugar prices, in: Anderson, O.D. (Ed.), Time Series Analysis: Theory and Practice 1. North-Holland, Amsterdam, pp. 203–226.

[48] Taylor, S.J., 1986. Modelling Financial Time Series. Wiley, Chichester.

[49] Yu, J., Meyer, R., 2006. Multivariate stochastic volatility models: Bayesian estimation and model comparison. Econometric Reviews 25, 361–384.

Table 1. Descriptive Statistics for the Daily Index Log Returns.

Statistic	France	Germany	Italy	Switzerland	UK
Sample	1.00	0.72	0.91	0.89	0.90
correlation	.	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. Estimation Results.

	Basic WMSV Model																
	A^{-1}															ν	d
	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}		
Estimate	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. $\times 10^2$	0.57	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
$q_{.025}$	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
$q_{.975}$	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																
	A_1^{-1}															ν	d
	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}		
Estimate	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. $\times 10^2$	0.90	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
$q_{.025}$	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
$q_{.975}$	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
	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}		
Estimate	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. $\times 10^2$	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
$q_{.025}$	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
$q_{.975}$	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: $[q_{.025}; q_{.975}]$. 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 3. Model Diagnostic Results: Pearson Residuals.

Ljung-Box test statistics for residual cross-products, 50 lags														
$e_1^* \times e_1^*$	$e_1^* \times e_2^*$	$e_1^* \times e_3^*$	$e_1^* \times e_4^*$	$e_1^* \times e_5^*$	$e_2^* \times e_2^*$	$e_2^* \times e_3^*$	$e_2^* \times e_4^*$	$e_2^* \times e_5^*$	$e_3^* \times e_3^*$	$e_3^* \times e_4^*$	$e_3^* \times e_5^*$	$e_4^* \times e_4^*$	$e_4^* \times e_5^*$	$e_5^* \times e_5^*$
Data														
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*
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 particle filtering is based on 100,000 particles. *: Significant at the 1% level.

Table 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 particle filtering is based on 100,000 particles.
*: Significant at the 1% level.

Table 5. Simulation Results for the Fitted MS WMSV Model.

	France	Germany	Italy	Switzerland	UK	France	Germany	Italy	Switzerland	UK
	$E[\mathbf{\Sigma}_t]$, $\mathbf{A}_1$					$E[\mathbf{\Sigma}_t]$, $\mathbf{A}_2$				
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 6. VaR Forecasting Results.

		5% VaR				
Model	(p,q)	Hit-Rate	LR_{uc}	Model	Hit-Rate	LR_{uc}
DCC-GARCH	$(2,1)$	0.0916	0.0054	EWMA	0.0954	0.0026
CCC-GARCH	$(1,1)$	0.0992	0.0012	WMSV	0.1527	< 0.0001
BEKK-GARCH	$(1,1)$	0.0916	0.0054	MS WMSV	0.0878	0.0108
D-BEKK-GARCH	$(1,1)$	0.1107	< 0.0001			

The table reports hit-rates and p-values for the likelihood ratio test of unconditional 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.

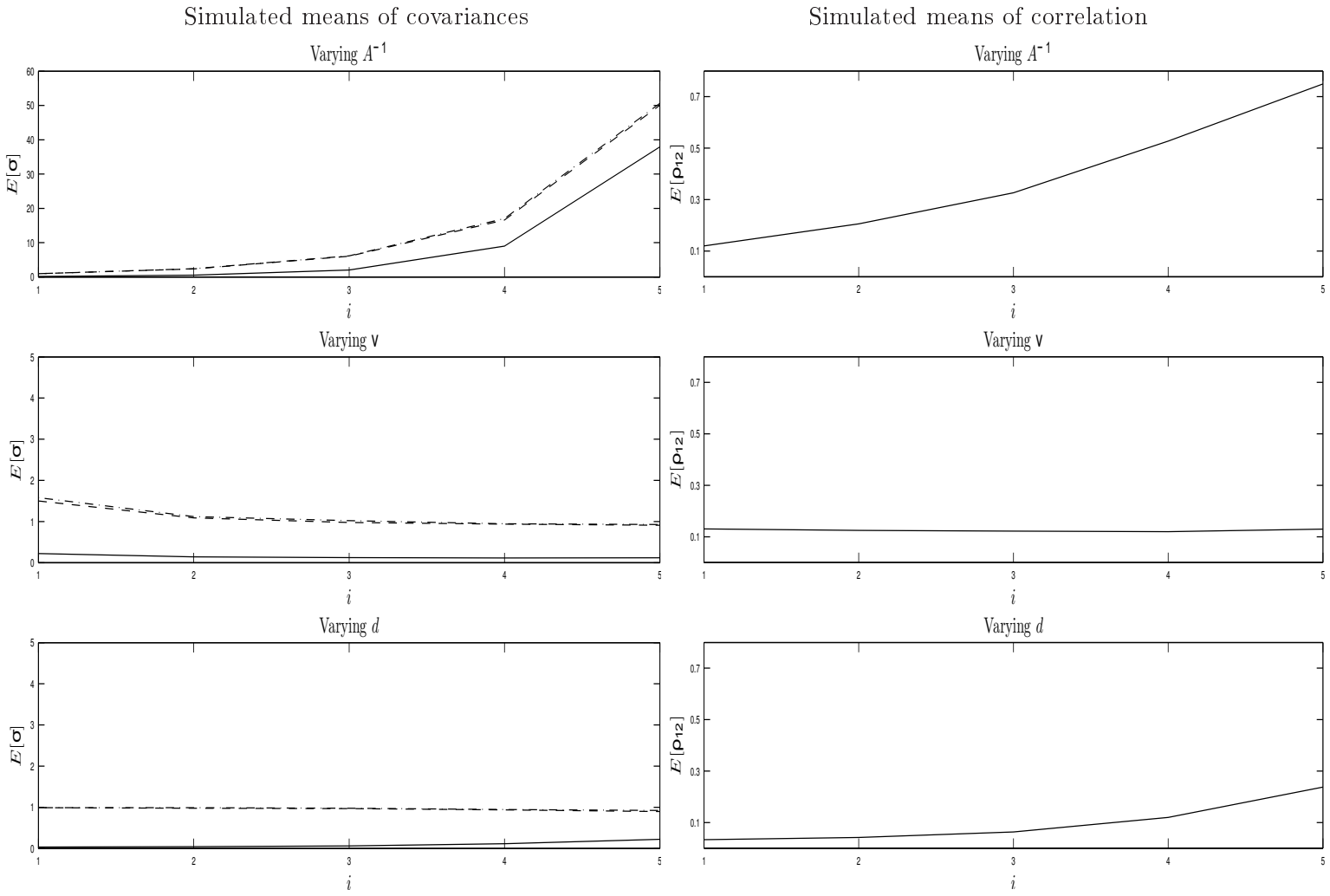

Figure 1. Simulated means of covariances and correlations. Left panel: Simulated means of covariances. Dashed line: σ_{11}*; solid line:* σ_{12}*; dashdotted line:* σ_{22}*. Right panel: simulated means of correlation.* $\rho_{12} = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$*.* i *is the index on the respective parameter sets given in Section 2.1. Simulation sample size:* $T = 20,000$*. All remaining model parameters are kept constant at* $vech(A_1^{-1}) = (0.96,\ 0.02,\ 0.96)'$*,* $\nu_4 = 80$ *and* $d_4 = 0.8$*, respectively.*

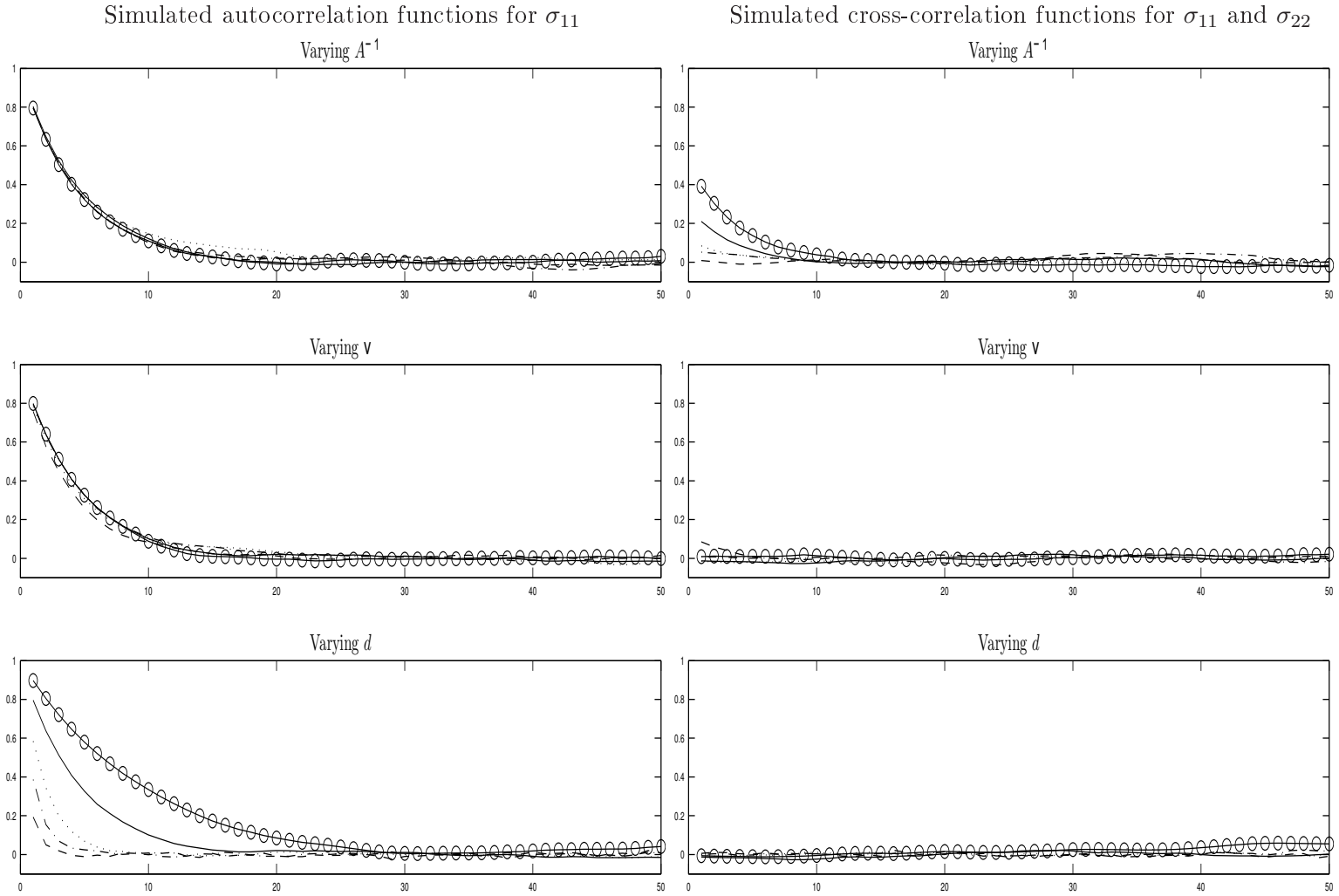

Figure 2. Simulated autocorrelation and cross-correlation functions for volatilities. Left panel: Simulated autocorrelation functions for σ_{11}. Right panel: Simulated cross-correlation functions $Corr[\sigma_{11,t}, \sigma_{22,t-q}]$ for $q = 1, \ldots, 50$. Simulation sample size: $T = 20,000$. Dashed line: $i = 1$; dashdotted line: $i = 2$; dotted line: $i = 3$; solid line: $i = 4$; $\circ$: $i = 5$, where i is the index on the respective parameter sets given in Section 2.1. All remaining model parameters are kept constant at $vech(A_1^{-1}) = (0.96,\ 0.02,\ 0.96)'$, $\nu_4 = 80$ and $d_4 = 0.8$, respectively.

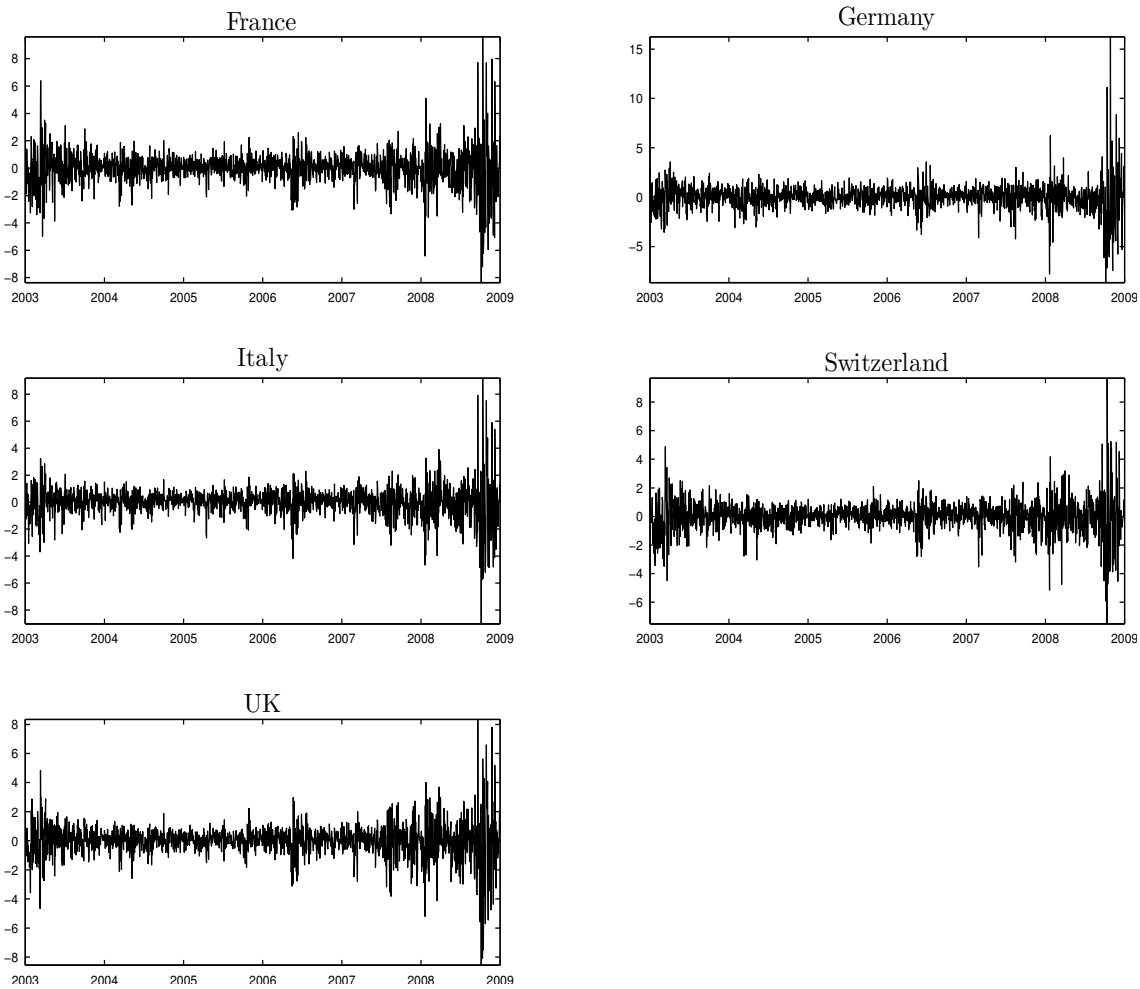

Figure 3. Log-returns of Datastream DS market indices. The number of observations for each series is 1565.

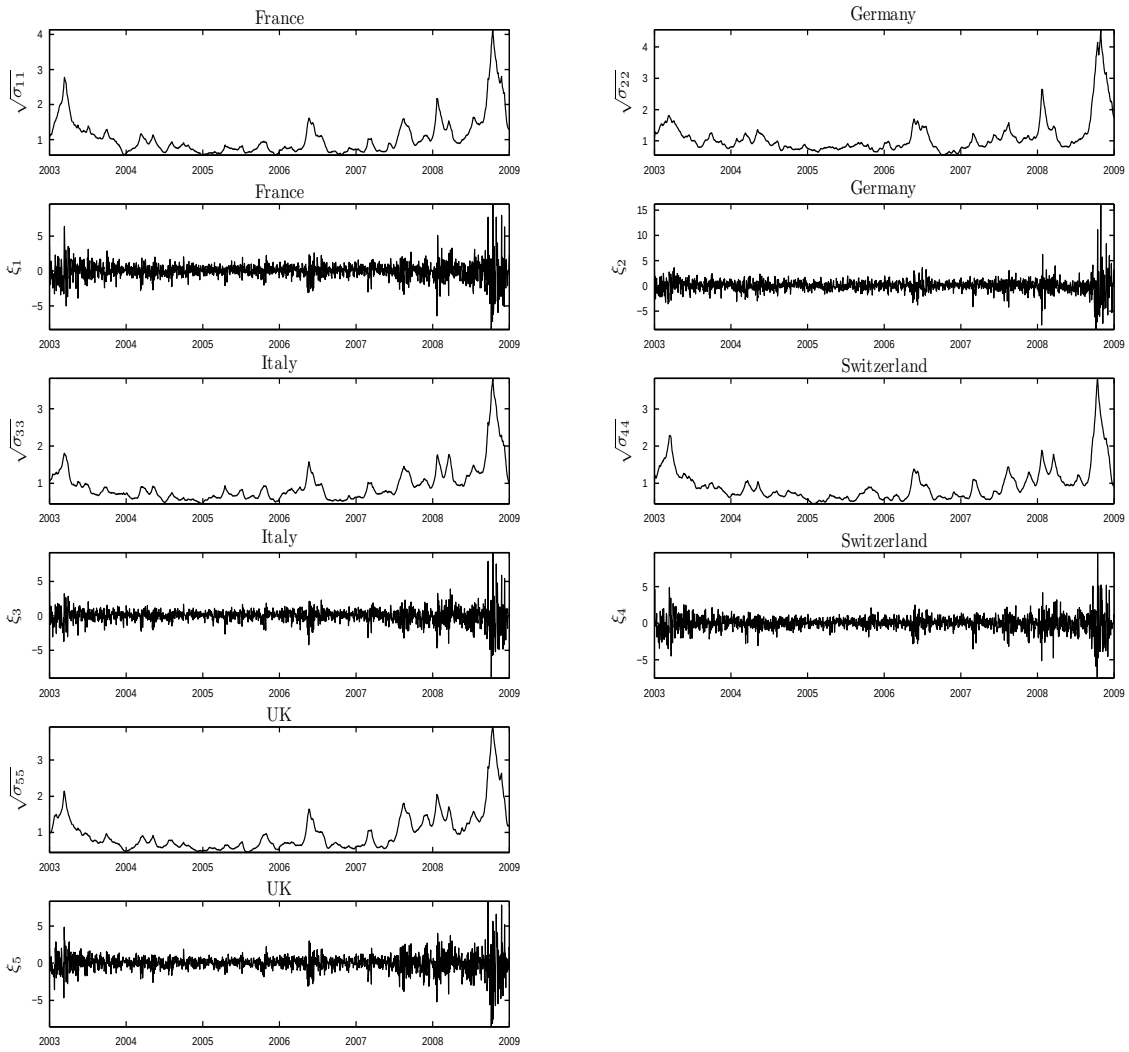

Figure 4. Smoothed volatility estimates and corresponding return series: Basic WMSV model.

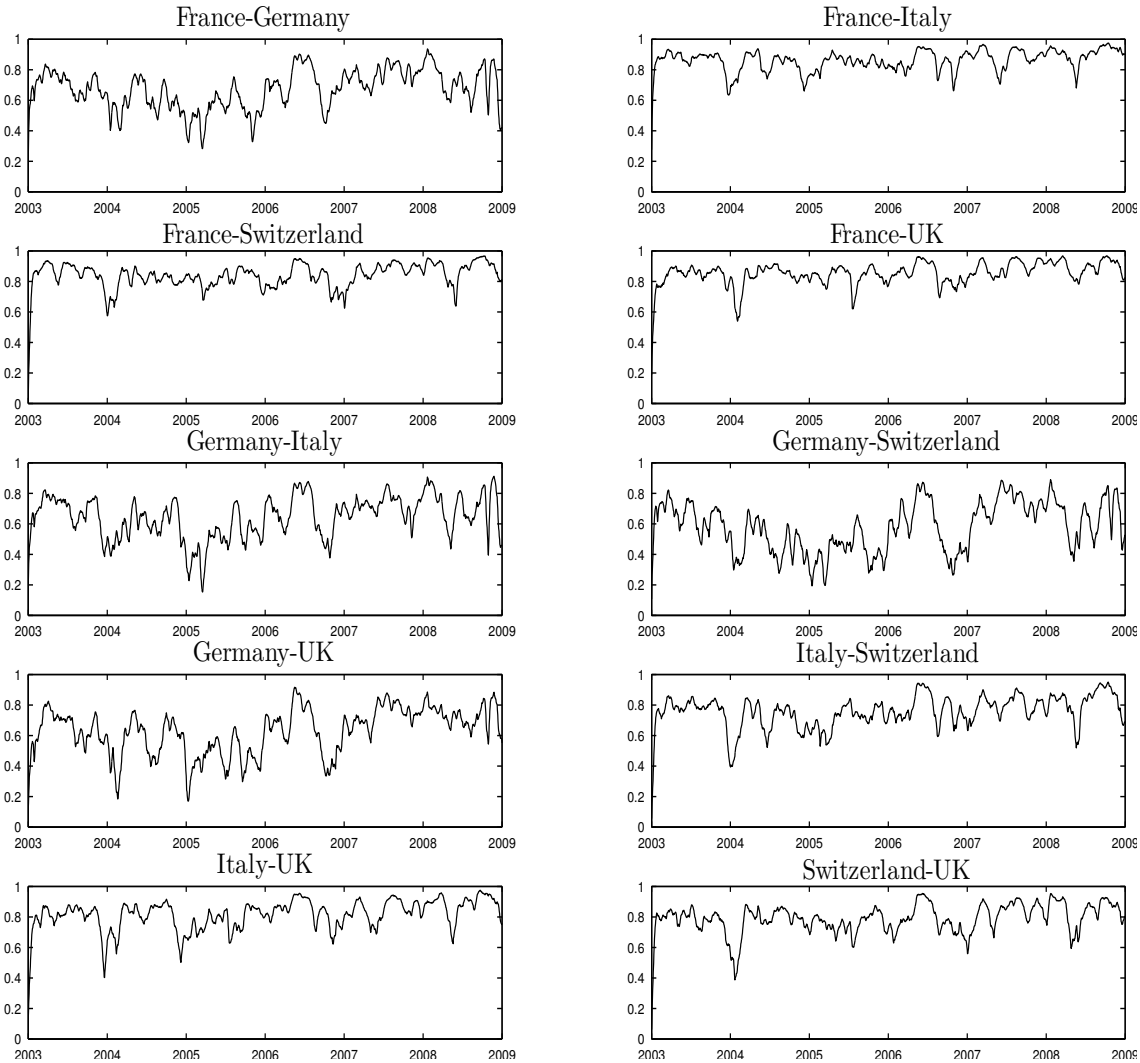

Figure 5. Smoothed correlation estimates: Basic WMSV model.

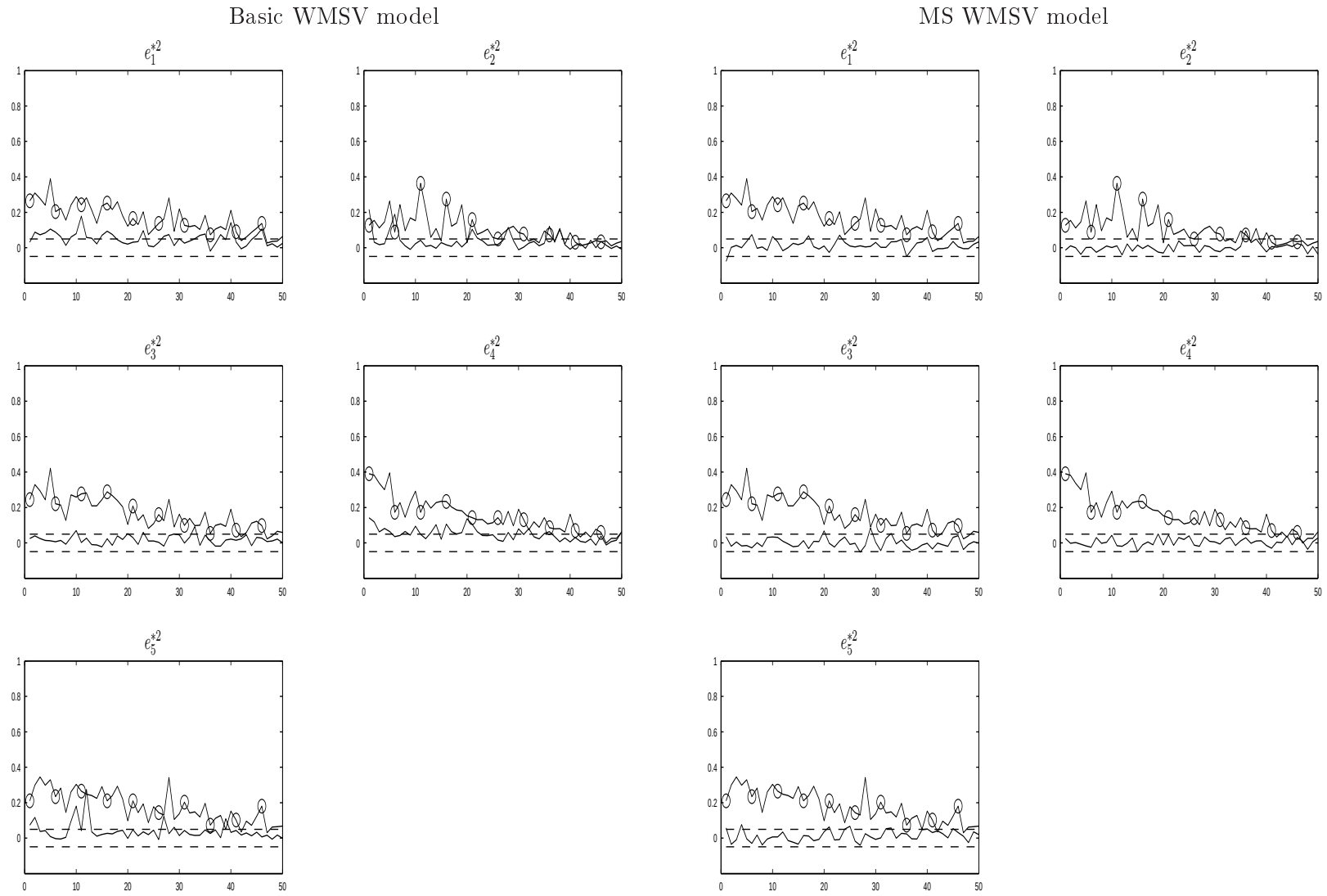

Figure 6. Sample autocorrelation functions of squared residual series. Left panel: Basic WMSV model. Right panel: MS WMSV model. Solid line: WMSV model; ∘: squared return series; dashed line: 95% Bartlett confidence bands for no serial dependence.

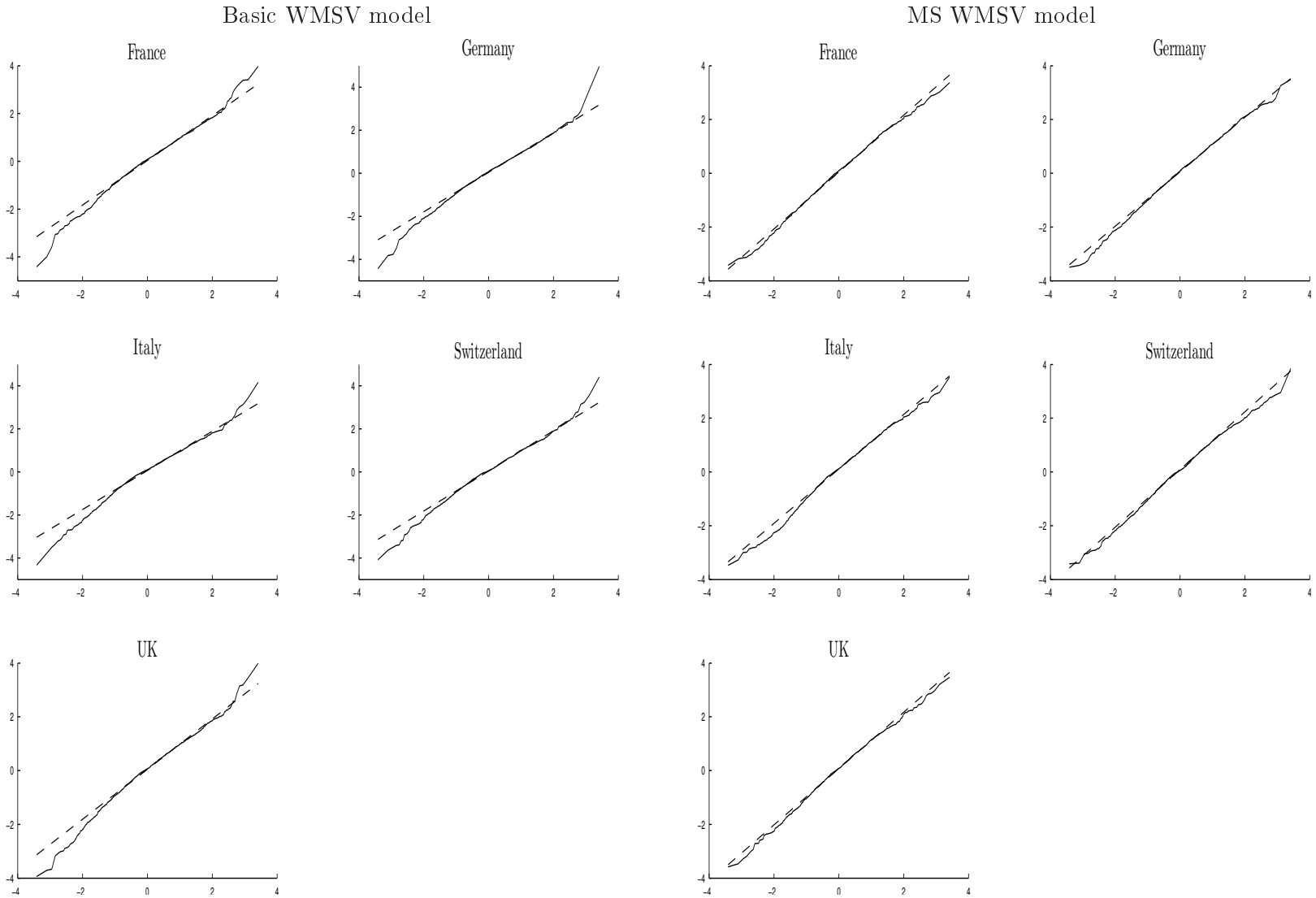

Figure 7. Residual QQ-plots. Left panel: Basic WMSV model. Right panel: MS WMSV model. The qq-plots refer to the residual series e_t on distributional diagnostics.

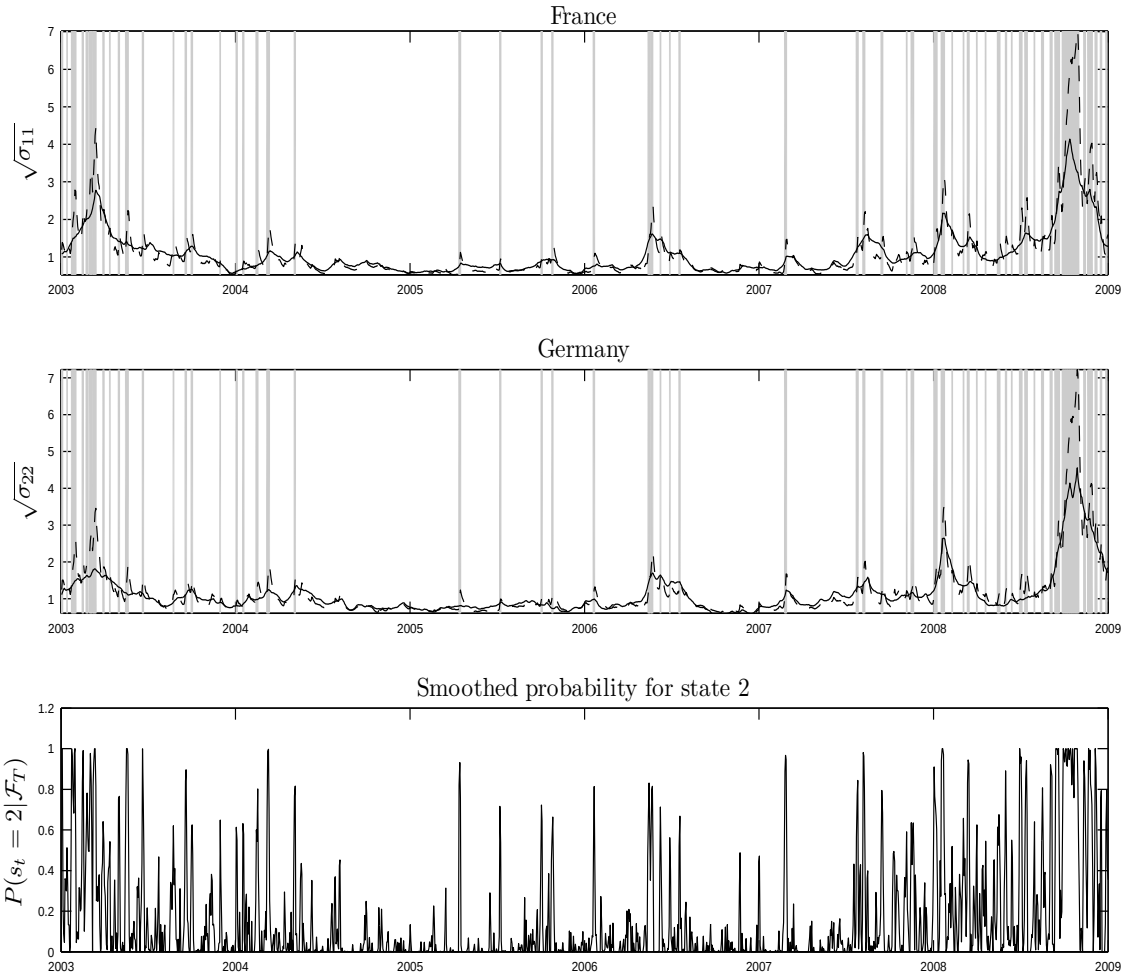

Figure 8. Smoothed volatility and Markov state estimates. Solid line: Basic WMSV model; dashed line: MS WMSV model. The gray shaded areas mark periods where the smoothed state probability exceeds 0.5.