

Gribisch, Bastian; Liesenfeld, Roman

## Conference Paper

# Multivariate Wishart Stochastic Volatility Models

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Advances in Time Series Analysis, No. B6-V2

### Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Gribisch, Bastian; Liesenfeld, Roman (2010) : Multivariate Wishart Stochastic Volatility Models, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Advances in Time Series Analysis, No. B6-V2, Verein für Socialpolitik, Frankfurt a. M.

This Version is available at:

<https://hdl.handle.net/10419/37511>

#### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

#### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Multivariate Wishart Stochastic Volatility Models

Bastian Gribisch\*, Roman Liesenfeld†

February 8, 2010

## Abstract

We generalize the basic Wishart multivariate stochastic volatility model of Philipov & Glickman (2006) to encompass regime switching behavior. The latent state variable is driven by a first-order Markov process. In order to estimate the proposed model we use Bayesian Markov Chain Monte Carlo procedures. For the computation of filtered estimates of the latent variances and covariances we rely upon particle filter techniques. The model is applied to five European stock index returns. Our results show that our proposed regime-switching specification substantially improves the estimates of the conditional covariance matrix and the VaR performance relative to the basic model.

*JEL classification: C11, C15, C32, G15*

*Keywords: Markov Switching, MCMC, Multivariate Stochastic Volatility, Particle Filter, Volatility Spillovers*

## 1 Introduction

Learning about existing inter-market linkages and the extent of integration between different markets is of great importance for economists and investors interested in assessing new investment opportunities, examining the effectiveness of capital markets or analyzing the risk involved in international portfolio diversification. In order to reflect inter-market linkages, contagion and spillover effects on financial markets, the basic univariate frameworks of volatility modeling have to be extended to the multivariate case.

Engle (1982) and Bollerslev (1986) proposed the famous Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) framework to model the conditional variance of asset returns, where the univariate models have quickly been

---

\**Postal address:* Institute of Statistics and Econometrics, Wilhelm-Seelig-Platz 7, 24118 Kiel. Tel.: (+49) 431-880-2229. *E-mail address:* bastian.gribisch@stat-econ.uni-kiel.de.

†*Postal address:* Institute of Statistics and Econometrics, Wilhelm-Seelig-Platz 7, 24118 Kiel. Tel.: (+49) 431-880-2166. *E-mail address:* liesenfeld@stat-econ.uni-kiel.de.

extended to the multivariate case allowing to display time-varying correlations in the context of small asset portfolios and a contagion-based analysis of inter-market spillover effects. The probably most prominent representatives of these models are given by the DCC GARCH model of Engle (2002) and the BEKK GARCH model of Engle & Kroner (1995).

In contrast to the existing ARCH and GARCH models, which represent a purely statistical description of the observed stylized facts of financial asset returns, the SV model dating back to Clark (1973) features a theoretical foundation in the information process driving observed asset prices and allows for greater flexibility in describing the stylized facts of returns and volatilities by introducing a stochastic evolution of the conditional volatility process<sup>1</sup>. Compared to ARCH and GARCH models, SV model specifications therefore appear to constitute a more appropriate framework for discussing effects of international information transmission.

However, generalizing the existing univariate Stochastic Volatility (SV) specifications to a multivariate setting turned out to be rather complicated<sup>2</sup>. Proposed multivariate SV models, e.g. employed by Harvey et al. (1994), Danielsson (1998) and Smith & Pitts (2006), therefore usually feature vectors of log-volatilities interacting through a constant correlation structure. The latter feature, however, is inappropriate in describing effects of time-varying international information transmission which find a representation in time-varying correlations of asset returns. Factor models, employed e.g. by Pitt & Shephard (1999), Liesenfeld & Richard (2003) and Chib et al. (2006), represent an alternative framework capable of accommodating time-varying correlation structures. The return vector is assumed to be driven by a fixed number of latent factors (e.g. common market factors) which in turn feature stochastic volatilities. Hence factor models are able to display effects of international information spillover resulting from single driving factors like e.g. risk factors and have already been successfully employed to analyze contagion effects<sup>3</sup>. The main drawback of the factor model approach, however, is its lack of flexibility. The dynamic evolution of return variances and correlations is actually driven by the same set of underlying latent factors and is therefore not allowed to evolve independently from one another over time.

The apparent problem of modeling dynamic stochastic correlations in a flexible way has been alleviated by Philipov & Glickman (2006), who proposed a Multivariate Stochastic Volatility (MSV) model based on the Wishart distribution for the inverse covariance matrix of a return vector. The approach does not only allow for a direct modeling of time-varying correlations independent from

---

<sup>1</sup>Compare e.g. Danielsson (1998).

<sup>2</sup>Compare Chib et al. (2009) for an overview.

<sup>3</sup>Compare e.g. Lopes & Carvalho (2007).

the time-varying volatilities without assuming an underlying factor structure, but in addition for volatility spillover effects in a flexible MSV framework. However, the ability of the Wishart MSV model in reflecting the dynamic and distributional properties of the return process could potentially be improved.

The present paper analyzes the stochastic properties of the basic Wishart Multivariate Stochastic Volatility (WMSV) model and extends the latter by introducing a Markov Switching (MS) framework, which induces state-dependent volatility spillover and volatility scaling effects. The latter effect is known to compensate the general overestimation of the persistence within the volatility process as e.g. documented by Diebold (1986), Lamoureux & Lastrapes (1990), Hamilton (1994) and Lopes & Carvalho (2007). Both model specifications are applied to five-dimensional return data of European stock indices, where the focus of the analysis is given by international contagion and spillover effects. We furthermore apply model diagnostic tests based on standardized returns, which check the model's ability to accommodate the dynamic correlation structure within the return data and the model's ability to fit the unconditional distribution of the observed return series. Based on a forecasting application we finally test for the unconditional and conditional coverage of the 5% Value-at-Risk (VaR).

The outline of the article is given as follows: Section two illustrates the basic and the MS WMSV model, the Bayesian simulation based estimation scheme, model diagnostics based on standardized returns as well as test procedures for the unconditional and conditional coverage of the 5% VaR based on VaR forecasting. Section three discusses the estimation-, VaR-forecast- and model diagnostic results for both model specifications and section four concludes.

## **2 Model Specification, Bayesian Inference and Model Checking**

The present section details the applied basic Wishart Multivariate Stochastic Volatility (WMSV) model proposed by Philipov & Glickman (2006) and its stochastic and dynamic properties as well as a Markov Switching (MS) WMSV model specification. We additionally illustrate the applied Bayesian simulation based estimation scheme, model diagnostic tests based on standardized returns and model checking via VaR forecasts.

## 2.1 Basic WMSV Model

The Wishart Multivariate Stochastic Volatility (WMSV) model proposed by Philipov & Glickman (2006) describes the stochastic behavior of a  $k$ -dimensional return vector  $\mathbf{Y}_t$  over the time periods  $t = 1, \dots, T$ , where the covariance matrix of  $\mathbf{Y}_t$  conditional on the realization of the past period's covariance matrix follows an inverse Wishart distribution.

The WMSV model for the time evolution of a  $k$ -dimensional return series  $\{\mathbf{Y}_t\}_{t=1}^T$  is given by the stochastic process

$$\mathbf{Y}_t | \Sigma_t \sim N(\mathbf{0}, \Sigma_t), \quad (1)$$

$$\Sigma_t^{-1} | \nu, \mathbf{S}_{t-1} \sim \text{Wish}_k(\nu, \mathbf{S}_{t-1}), \quad (2)$$

where the return vector  $\mathbf{Y}_t$  is assumed to be mean-corrected.  $\Sigma_t$  denotes the covariance matrix of  $\mathbf{Y}_t$  and  $\nu$  ( $\nu > k$ ) and  $\mathbf{S}_t$  (positive definite) are the degrees of freedom (d.o.f.) parameter and the scaling matrix of the Wishart distribution. The mean correction of  $\mathbf{Y}_t$  is e.g. achieved by filtering each return series by autoregressive processes.

To cover the volatility clustering effect and dynamic correlations, the evolution of the Wishart distribution's scaling matrix is modeled time-dependent by introducing a positive definite symmetric parameter matrix  $\mathbf{A}$  and a scalar persistence parameter  $d$ . The resulting parametrization of the scaling matrix  $\mathbf{S}_t$  is then given by

$$\mathbf{S}_t = \frac{1}{\nu} \Sigma_t^{-d/2} \mathbf{A} \Sigma_t^{-d/2}, \quad (3)$$

where the scaling matrix of the Wishart distribution for the inverse covariance matrix  $\Sigma_t^{-1}$  is modeled as a function of the lagged inverse covariance matrix  $\Sigma_{t-1}^{-1}$ .  $\Sigma_t^{-d/2}$  furthermore denotes a specific function of the spectral decomposition of the inverse covariance matrix  $\Sigma_t^{-1}$ . Denoting the matrix of eigenvectors of  $\Sigma_t^{-1}$  by  $\mathbf{V}_t$  and the respective diagonal matrix of eigenvalues  $\mathbf{\Lambda}_t$ , the expression  $\Sigma_t^{-d/2}$  can be decomposed according to

$$\Sigma_t^{-d/2} = \mathbf{V}_t \mathbf{\Lambda}_t^{\frac{d}{2}} \mathbf{V}_t', \quad (4)$$

where the power operator is defined to work element-wise. The quadratic expression in  $\mathbf{S}_t$  secures the positive definiteness needed for the scaling matrix of the Wishart distribution.

The stochastic properties of multivariate SV models in general are hard or even impossible to derive analytically. This also holds for the illustrated Wishart SV model. However, the distribution of  $\Sigma_t^{-1}$  and its inverse conditional on the lagged covariance matrix is Wishart and inverse Wishart,

respectively. The according conditional means are therefore easily obtained:

$$E[\boldsymbol{\Sigma}_t^{-1} | \mathbf{A}, d, \nu, \boldsymbol{\Sigma}_{t-1}] = \nu \mathbf{S}_{t-1} = \boldsymbol{\Sigma}_{t-1}^{-d/2} \mathbf{A} \boldsymbol{\Sigma}_{t-1}^{-d/2} \quad (5)$$

$$E[\boldsymbol{\Sigma}_t | \mathbf{A}, d, \nu, \boldsymbol{\Sigma}_{t-1}] = \frac{1}{\nu - k - 1} \mathbf{S}_{t-1}^{-1} = \frac{\nu}{\nu - k - 1} \boldsymbol{\Sigma}_{t-1}^{d/2} \mathbf{A}^{-1} \boldsymbol{\Sigma}_{t-1}^{d/2}. \quad (6)$$

The model's dynamics are governed by the parameters  $\mathbf{A}$  and  $d$  which play an important role in the model's ability to describe the stylized facts of financial return series. The parameter  $d$  describes the overall persistence of the volatility process and is theoretically bounded between 0 and 1, which can be seen by rewriting the WMSV model by making use of the properties of the Wishart distribution. Denoting the  $k$ -dimensional identity matrix by  $\mathbf{I}_k$  and the Cholesky factor of  $\mathbf{A}$  by  $\mathbf{L}$ , the inverse of the covariance matrix in period  $t$  is given by

$$\boldsymbol{\Sigma}_t^{-1} = \frac{1}{\nu} \boldsymbol{\Sigma}_{t-1}^{-d/2} \mathbf{L} \text{Wish}_k(\nu, \mathbf{I}_k) \mathbf{L}' \boldsymbol{\Sigma}_{t-1}^{-d/2}. \quad (7)$$

Eq. (7) indicates an autoregressive representation for  $\ln |\boldsymbol{\Sigma}_t^{-1}|$  given by

$$\ln |\boldsymbol{\Sigma}_t^{-1}| = -k \ln(\nu) + \ln |\mathbf{A}| + d \ln |\boldsymbol{\Sigma}_{t-1}^{-1}| + \ln |\text{Wish}_k(\nu, \mathbf{I}_k)|. \quad (8)$$

Hence stationarity of the Wishart process is preserved by restricting the parameter  $d$  to the interval between -1 and 1. In practice, however, the parameter  $d$  should additionally be restricted to positivity to rule out stochastic processes for  $\boldsymbol{\Sigma}_t^{-1}$  which alternate between powers of inverses.

Eq. (8) additionally implies a volatility scaling property of the parameter matrix  $\mathbf{A}$  and the d.o.f parameter  $\nu$ , while eq. (6) suggests a MSV model being able to represent volatility spillover effects by linking contemporaneous and past volatilities by means of the parameter matrix  $\mathbf{A}$ . Eq. (6) finally shows that the variance/covariance-related interpretation of inter-temporal effects implied by  $\mathbf{A}$  actually has to be based on the latter's inverse.

While the moments of the covariance matrix  $\boldsymbol{\Sigma}_t$  conditional on  $\boldsymbol{\Sigma}_{t-1}$  and the model's parameter vector  $\boldsymbol{\theta}$  are easily obtained via the properties of the inverse Wishart distribution, these moments are actually of limited interest due to the unobservable conditioning information. Since no closed form analytical expression can be derived, we simulate the corresponding unconditional moments based on a two-dimensional WMSV model and different parameter sets to assess the influence of certain parameter constellations on the unconditional distribution of the returns' covariance matrices.

For each structural model parameter five parameter constellations are considered: In case of the inverse parameter matrix  $\mathbf{A}^{-1}$ , the specifications  $\mathbf{A}_i^{-1}$ ,  $i = 1, \dots, 5$ , are distinguished by overall increasing matrix entries in  $i$  on each single position in  $\mathbf{A}^{-1}$ . In case of  $d$  and  $\nu$  the corresponding

parameter sets are given by  $[d_1 = 0.2, d_2 = 0.4, d_3 = 0.6, d_4 = 0.8, d_5 = 0.9]$  and  $[\nu_1 = 20, \nu_2 = 40, \nu_3 = 60, \nu_4 = 80, \nu_5 = 90]$ .

Figures 1 to 3 illustrate the obtained simulation results with respect to the unconditional mean and standard deviation of  $\Sigma_t$ . The findings are summarized by a volatility- and volatility of volatility scaling property of the inverse parameter matrix  $\mathbf{A}^{-1}$  if each element in  $\mathbf{A}^{-1}$  is increased or decreased, a similar but reversed scaling property of  $\nu$  and finally increasing unconditional means of volatilities and covariances with increasing persistence parameter  $d$ . In case of  $d$ , however, the respective unconditional volatility of the volatility is decreasing with increasing parameter values. All simulation results are *ceteris paribus*.

Simulated unconditional expectations of the implied correlation coefficient are presented in figure 4. Unconditional correlation increases with increasing  $\mathbf{A}^{-1}$  and  $d$ . The effects of varying  $\nu$  on the unconditional mean of correlations, however, appear to be insignificant.

Figures 5 to 7 illustrate the influence of the parameter vector on the model's dynamic properties by presenting simulated unconditional autocorrelation functions of the first return volatility which appear to be solely influenced by varying parameters  $d_i$ . The latter finding confirms the notion of  $d$  as a persistence parameter.

## 2.2 Markov Switching WMSV Model

The Markov Switching (MS) WMSV model specification induces state-dependent volatility scaling and volatility spillover effects by introducing a hidden Markov chain which allows a state-dependent parameter matrix  $\mathbf{A}_{s_t}$  to switch between two realizations,  $\mathbf{A}_1$  and  $\mathbf{A}_2$ , with differing volatility scaling properties. The conditional probabilities of transition between the two states are given by the transition matrix

$$P(s_t | s_{t-1}) = \begin{bmatrix} (1 - e_1) & e_1 \\ e_2 & (1 - e_2) \end{bmatrix}, \quad (9)$$

where  $e_1$  denotes the probability of switching from state 1 to state 2 and  $e_2$  the probability of switching from state 2 to state 1. The resulting MS model is given by

$$\mathbf{Y}_t | \boldsymbol{\Sigma}_t \sim N(0, \boldsymbol{\Sigma}_t) \quad (10)$$

$$\boldsymbol{\Sigma}_t^{-1} | \nu, \mathbf{S}_{t-1} \sim \text{Wish}_k(\nu, \mathbf{S}_{t-1}) \quad (11)$$

$$\mathbf{S}_t = \frac{1}{\nu} \boldsymbol{\Sigma}_t^{-d/2} \mathbf{A}_{s_t} \boldsymbol{\Sigma}_t^{-d/2} \quad (12)$$

$$P(s_t | s_{t-1}) = \begin{bmatrix} (1 - e_1) & e_1 \\ e_2 & (1 - e_2) \end{bmatrix}. \quad (13)$$

Identification is achieved by requiring a positive definite matrix difference  $\mathbf{A}_1 - \mathbf{A}_2$ .

### 2.3 Bayesian Estimation

As proposed by Philipov & Glickman (2006), inference about the WMSV model's parameter vector is based on a Bayesian estimation approach, which becomes especially attractive in case of complex multivariate models including a large number of parameters. High dimensionality of the parameter vector involves potential problems regarding the classical likelihood-based estimation scheme due to the involved numerical maximization of the likelihood function. These complications can be avoided by making use of tractable Bayesian inference techniques.

The goal of Bayesian inference is to obtain knowledge about the joint posterior distribution of the model's parameter vector, whose moments can be used to generate Bayesian point estimates and to assess the according parameter uncertainty. The posterior distribution is proportional to the product of the likelihood function and the joint prior distribution of the parameter vector, where the latter reflects information about the parameter vector prior to the estimation. However, the likelihood function of the WMSV model constitutes a high-dimensional integral given by

$$L(\{\mathbf{Y}_t\}_{t=1}^T | \boldsymbol{\theta}) = \int \dots \int \prod_t^T P(\mathbf{Y}_t | \boldsymbol{\Sigma}_t^{-1}) \times P(\boldsymbol{\Sigma}_t^{-1} | \mathbf{S}_{t-1}, \nu) d\boldsymbol{\Sigma}_1, \dots, d\boldsymbol{\Sigma}_T, \quad (14)$$

where  $\boldsymbol{\theta}$  denotes the model's parameter vector. The integral is not analytically solvable and therefore simulation-based estimation techniques have to be applied.

The Bayesian Monte Carlo Markov Chain (MCMC) simulation based estimation approach became increasingly popular in the last decades and can be readily applied for parameter inference in case of the WMSV model. The MCMC scheme intends to generate draws from the joint posterior distribution of the model's parameter vector via simulating an irreducible and aperiodic Markov



chain. Under some mild regularity conditions, the latter converges to the joint posterior distribution and obtained draws can be used to conduct inference. The construction of the according Markov chain is based on the famous Gibbs sampling algorithm, which entirely consists of iterative drawing from the full conditional distribution of each model parameter, where the parameter vector is augmented by the set of latent variables. Bayesian point estimates are obtained by averaging the Gibbs draws after convergence of the Markov chain<sup>4</sup> whereas the according uncertainty is represented by the Gibbs draws' standard deviation constituting an approximation to the standard deviation of the according model parameter's marginal posterior distribution.

The derivation of the model parameters' full conditional distributions is illustrated in detail 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 drawing issues.

## 2.4 Model Diagnostics

Model diagnostics intending to check the WMSV model's dynamic volatility and correlation properties are based on Pearson residuals which are obtained by estimating the model's parameter vector and pre-multiplying the return vector by the inverse Cholesky factor of the conditional expectation of  $\Sigma_t$  given all return information up to period  $t - 1$ ,

$$\mathbf{e}_t^* = E[\Sigma_t | \underline{\mathbf{Y}}_{t-1}]^{-\frac{1}{2}} \mathbf{Y}_t, \quad (15)$$

where  $E[\Sigma_t | \underline{\mathbf{Y}}_{t-1}]$  is called filtered estimate of  $\Sigma_t$ . Under the null hypothesis of a correctly specified model the obtained Pearson residuals should have zero mean, unit variance and feature no auto-correlation in the first and second order moments<sup>5</sup>, where the latter hypothesis can be tested by applying e.g. the Ljung-Box test for the null-hypothesis of no correlation within the respective time series. A rejection of the null hypothesis, however, implies the model's inability to accommodate the volatility and correlation dynamics within the data and therefore suggests appropriate extensions of the model design.

The filtered covariance estimate  $E[\Sigma_t | \underline{\mathbf{Y}}_{t-1}]$  constitutes a high-dimensional integral, which can be approximated by the sample mean over draws from the respective conditional distribution of  $\Sigma_t$ :

$$\mathbf{e}_t^* = E[\Sigma_t | \underline{\mathbf{Y}}_{t-1}]^{-\frac{1}{2}} \mathbf{Y}_t \cong \left( \frac{1}{M} \sum_{j=1}^M \Sigma_t^j \right)^{-\frac{1}{2}} \mathbf{Y}_t, \quad (16)$$

---

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

<sup>5</sup>Compare Gallant et al. (1991).

where  $\Sigma_t^j$  denotes a draw from  $f(\Sigma_t|\underline{\mathbf{Y}}_{t-1})$ , which is e.g. obtained by applying the standard particle filter algorithm illustrated by Pitt & Shephard (1999).

Besides their dynamic properties, the distribution of the Pearson residuals is unknown. Model diagnostics concerned with distributional assumptions regarding the model structure therefore have to be based on alternative residual data. Kim et al. (1998) propose a specific kind of standardization which under the null hypothesis of a correctly specified model results in independent and identically distributed (iid) standard Normal return residuals. The latter can therefore be applied for checking distributional assumptions. Denoting  $y_{i,t}$  the  $i$ 'th element of the  $k$ -dimensional return vector  $\mathbf{Y}_t$ , the probability  $Pr(\tilde{y}_{i,t+1} \leq y_{i,t+1}|\underline{\mathbf{Y}}_t)$  can be approximated by

$$Pr(\tilde{y}_{i,t+1} \leq y_{i,t+1}|\underline{\mathbf{Y}}_t) \cong u_{i,t+1}^M = \frac{1}{M} \sum_{j=1}^M Pr(\tilde{y}_{i,t+1} \leq y_{i,t+1}|\sigma_{i,t+1}^{2,j}), \quad (17)$$

where  $\sigma_{i,t+1}^{2,j}$  denotes the  $i$ 'th diagonal element of  $\Sigma_{t+1}^j$ , drawn from  $f(\Sigma_{t+1}^{-1}|\underline{\mathbf{Y}}_t)$ ,  $j = 1, \dots, M$ , and a variable with a tilde denotes a random variable. Under the null hypothesis of a correctly specified model the  $\{u_{i,t}^M\}_{t=1}^T$  sequence is iid uniform distributed for all  $i = 1, \dots, k$  and can be mapped into the standard Normal distribution via the inverse of the respective cumulative distribution function (cdf):  $e_{i,t}^M = F_N^{-1}(u_{i,t}^M)$ . These residuals can thereupon be applied for checking distributional assumptions with respect to the model under consideration e.g. by applying the Kolmogorov-Smirnov- or Jarque-Bera test of the null hypothesis of standard Normal distributed residuals.

## 2.5 Value-at-Risk

An alternative approach for comparing alternative WMSV model specifications with respect to their ability of accommodating dynamic correlation and volatility structures as well as distributional properties of the return data is based on the unconditional and conditional coverage of VaR measures of pre-defined portfolios, and has e.g. been applied by Chib et al. (2006). Given a  $k$ -dimensional vector of portfolio weights  $w$  the level  $\alpha$  VaR of a portfolio return  $y_{pt}$  at time  $t$  is computed according to

$$\text{VaR}_{p,t}(\alpha) = \sqrt{\hat{\sigma}_{p,t}} F^{-1}(\alpha), \quad (18)$$

where  $F^{-1}(\alpha)$  denotes the  $\alpha$ -percentile of the cumulative on-step-ahead distribution assumed for portfolio returns and  $\hat{\sigma}_{p,t}$  denotes the model-based portfolio variance forecast calculated using return information up to period  $t - 1$ .

The accuracy of obtained VaR estimates can be checked based on unconditional and conditional coverage tests as illustrated by Lopez & Walter (2001). Defining an indicator variable

$$I_t = \begin{cases} 1 & \text{if } y_{pt} < \text{VaR}_{p,t}, \\ 0 & \text{if } y_{pt} \geq \text{VaR}_{p,t}, \end{cases} \quad (19)$$

and denoting the number of out-of-sample observations by  $T$  and the ‘‘hit-rate’’ by  $\hat{\alpha}$  the hypothesis  $\hat{\alpha} = (1/T) \sum_{t=1}^T I_t = \alpha$  can be tested using the statistic

$$LR_{uc} = 2 \{ \ln [ \hat{\alpha}^\gamma (1 - \hat{\alpha})^{T-\gamma} ] - \log [ \alpha^\gamma (1 - \alpha)^{T-\gamma} ] \}, \quad (20)$$

which is asymptotically  $\chi^2(1)$  distributed with  $\gamma = \sum_{t=1}^T I_t$ . However, the obtained portfolio returns are heteroscedastic. To accommodate this stylized fact 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  $j$  after having been in state  $i$  in the previous period,  $\pi_{01} = T_{01}/(T_{00} + T_{01})$  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

$$LR_{ind} = 2(\ln L_A - \ln L_0), \quad (21)$$

which is also asymptotically  $\chi^2(1)$  distributed. To jointly test the two hypotheses and to test for correct conditional coverage we apply the test statistic

$$L_{cc} = LR_{uc} + LR_{ind}, \quad (22)$$

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

### 3 Empirical Results

The current section presents the data-set and the obtained estimation results for the basic WMSV model and the MS WMSV model specification. The respective model fit is analyzed by the conditional and unconditional coverage of 5% VaR forecasts for the year 2008 as well as model diagnostic tests based on the stochastic properties of standardized returns under the null hypothesis of correct model specification.

The analysis is based on daily AR(p) pre-filtered stock index log-returns<sup>6</sup> for France, Germany, Italy, Switzerland and the UK starting in January 2003 and ending in December 2008 (1565 observations). The return series are illustrated in figure 8 and descriptive statistics are given in table 1, where the Box-Pierce test statistic with 10 lags included is adjusted for serial correlation in the second moment (the so-called volatility clustering effect). All return series feature the stylized facts of financial markets given by high leptokurtosis, insignificant autocorrelation in returns and significant autocorrelation in quadratic returns. The reported sample correlations are quite high and indicate a huge degree of co-movement for all five stock indices.

### 3.1 Basic WMSV Model

Table 2 presents the estimation results for the basic WMSV model. The chosen prior distributions are overall relatively uninformative and reported in the table. All estimation results are based on 50,000 iterations of the Gibbs sampling algorithm and a burn-in phase of 15,000 iterations. Reported point estimates are obtained by averaging the Gibbs output after convergence of the generated Markov chain. The posterior standard deviation and the 95% posterior high density region are estimated by the standard deviation of the respective Gibbs sequence and the according 2.5% and 97.5% quantiles. The MC standard errors intend to give an impression of the numerical efficiency of the simulation based MCMC scheme and are calculated using a correlation consistent Parzen window based spectral estimator for the variance of a sample mean<sup>7</sup> as given by the obtained Bayesian point estimates.

Figure 9 depicts the obtained Gibbs sequences which display fast convergence to the stationary distribution. Histograms of the former are presented in figure 10 and give an impression of the model parameters' marginal posterior distributions.

The reported 95% posterior high density regions imply parameter estimates significantly different from zero with low MC standard errors. The estimated persistence parameter  $d = 0.9467$  indicates a pronounced volatility clustering effect and huge dependencies within the dynamic correlation process. Estimated off-diagonal elements of the sensitivity matrix  $\mathbf{A}^{-1}$  are significantly different from zero. The estimation results therefore suggest significant volatility spillover effects between the five European countries considered. Figure 11 illustrates these effects by simulation-based estimates of cross-country lagged unconditional volatility correlations where France appears to be the driving force of international transmission of uncertainty, whereas German volatility shows the lowest

---

<sup>6</sup>Datastream DS market indices.

<sup>7</sup>Compare Kim et al. (1998).

unconditional correlation with the remaining countries' volatilities.

Smoothed estimates of the dynamic standard deviations and correlations for the whole set of considered log returns are presented in figure 12 and 13 and obtained by averaging the Gibbs output of the covariance and correlation matrices after leaving the burn-in period. The results imply a huge degree of co-movement, i.e. dependence, in volatility, a pronounced pattern of volatility clustering and highly accentuated peaks in volatility at the beginning of 2003, the middle of 2006 and finally a large volatility cluster slowly building up from the middle of 2007. The latter effect is clearly due to the current financial crisis originating from the US subprime market from where uncertainty, reflected by volatility, spread out around the world through various channels of information transmission. The average smoothed cross-correlation estimates involving Germany are significantly lower compared to the remaining cross-correlation estimates and show a pattern of significantly higher cross-country correlations in periods of high market volatility compared to calm periods. According to Forbes & Rigobon (2001), this pattern indicates contemporaneous contagion effects in returns.

Table 5 presents model diagnostic results for both model specifications based on Pearson residuals obtained under the particle filtering scheme. For the basic WMSV framework the presented Ljung-Box test statistics at 50 lags for autocorrelation in the cross-products of the standardized returns are highly significant at every common significance level for almost all cases. However, by comparing the obtained test statistics for the log return series before and after the standardization, a strong reduction in the autocorrelation is observed. The model is able to accommodate the dynamic volatility and correlation pattern, but the quality of the respective fit could clearly be improved.

Table 6 presents diagnostic results based on standardized returns according to Kim et al. (1998). The Kolmogorov-Smirnov and Jarque-Bera test results under the null hypothesis of Normality of the obtained residuals are overall highly significant. The apparent deviations from the Normal distribution are clearly seen by comparing the reported kurtosis statistics with their expected realization under a Normal distribution. However, compared to the kurtosis statistics and test results obtained under the original return data, the significant improvements towards Gaussianity are obvious. Nevertheless the obtained diagnostics imply considerable problems of the model in accommodating the fat tails of the five-dimensional return series.

The model diagnostic results are additionally reflected by obtained 5% VaR-forecast results for an equally weighted European portfolio, presented in table 7. The forecast for the first period of 2008 has been generated by estimating the according WMSV model based on data up to the end of 2007 and forecasting the 5% VaR of the subsequent period's portfolio return. All remaining

forecasts are obtained by moving the whole estimation window by one period and again forecasting the subsequent period's VaR. The forecasting procedure itself is based on the Gibbs scheme as e.g. illustrated by Chib et al. (2006). The obtained results based on the ex post observed hit-rates and according likelihood ratio tests overall indicate significant deviations from the 5% VaR. However it has to be noticed that the forecasting period includes the most volatile time span within the data set. A hit rate significantly different from the 5% VaR therefore does not seem astonishing at all.

### 3.2 Markov Switching WMSV Model

Table 3 presents the obtained estimation results and the according prior distributions for the Markov Switching (MS) WMSV model. The respective Gibbs sequences presented in figures 14 and 15 show convergence after a burn-in phase of 20,000 iterations. Histograms of the Gibbs output are additionally given in figures 16 and 17.

Except for one element, all entries of the  $\mathbf{A}_2^{-1}$  matrix are significantly greater than the corresponding elements of the  $\mathbf{A}_1^{-1}$  matrix. According to the simulation results of section 2.1, the latter finding suggests a more pronounced volatility scaling effect in the second state, which coincides with periods of exceedingly high volatility as illustrated by figure 18. The induced scaling effect prevents a potential underestimation of return volatilities in uncertain periods.

The persistence parameter  $d$  is estimated to be significantly lower compared to the basic WMSV model. This effect is due to the volatility scaling by the MS framework, which alleviates the general overestimation of the persistence parameter.

The significantly higher Bayesian point estimates in case of  $\mathbf{A}_2^{-1}$  linked with the state of high volatility furthermore suggest the existence of more accentuated volatility spillover effects during periods of turmoil. The effects of both parameter matrices on volatility scaling and volatility spillover effects are illustrated by simulation results presented in table 4. The unconditional mean of  $\boldsymbol{\Sigma}_t$  and the implied correlation matrix (not presented in the table) are significantly higher in the second state, and the unconditional correlation between each single asset's volatility at time  $t$  and the remaining assets' volatilities at time  $t - 1$  significantly increases from state one to state two. According cross-correlation functions at different lags are similar to those obtained for the basic WMSV model and therefore not presented here. The obtained estimation results imply enhanced volatility scaling and volatility spillover effects in periods of financial turmoil as indicated by high market volatility, where France again appears to be the driving force of cross-country volatility transmission.

Figure 19 presents smoothed estimates for the time-varying return correlation between France

and Germany. The apparent correlation scaling does not seem to be directly related to the obtained smoothed state estimates. Comparing the average smoothed correlations in each of the two states, however, results in a higher value for the 'volatile' state two. This result is in line with the simulation results of section 2.1 where the unconditional mean of correlations is scaled by overall increasing elements within the parameter matrix  $\mathbf{A}^{-1}$ . This finding is confirmed by the remaining cross-correlation estimates not presented here. During periods of high market volatility the latter again indicate contagion effects in cross-country return correlations driven by Germany.

Pearson residual based model diagnostic results for the MS model specification are presented in table 5. Conducted Ljung-Box tests for the series of cross-products of standardized returns are insignificant at the 1% level for almost all considered cross-products of standardized returns. The MS framework therefore offers a considerable improvement of the WMSV model's dynamic properties which can be traced back to the avoided overestimation of the persistence parameter  $d$ .

With respect to distributional model assumptions, table 6 additionally indicates superior diagnostic results for the Markov Switching WMSV specification. Besides a slight overestimation of the residuals' means and standard deviations the kurtosis results are significantly improved compared to the basic WMSV specification. This effect is due to the accentuated volatility peaks within the MS framework and is additionally represented (although overall insignificant) by the obtained Kolmogorov-Smirnov and Jarque-Bera test statistics.

The obtained 5% VaR-forecast results presented in table 7 furthermore underline the necessity of extending the basic WMSV model by the proposed Markov switching volatility scaling regimes. The obtained hit rate of 9.89% is significantly lower compared to the hit rate of 15.27% obtained under the basic WMSV model. However, all test results of conditional and unconditional coverage again imply a hit rate significantly different from the 5% level. This result is again due to the excessively high volatility in 2008 which is a result of the current financial crisis.

## 4 Conclusion

The present paper proposed and analyzed a Markov Switching (MS) extension to the basic Wishart Multivariate Stochastic Volatility (WMSV) model of Philipov & Glickman (2006). The MS framework is shown to allow for state-dependent volatility scaling and volatility spillover effects. The additional scaling effects avoid a potential underestimation of return volatilities in uncertain periods and alleviate the general overestimation of the persistence within the volatility process.

An empirical application to daily European stock index returns indicates significant spillover and contagion effects in cross-country return correlations and volatilities related to periods of excessive market volatility. Under the MS WMSV model, conducted model diagnostic tests reveal an enhanced quality of the overall model fit with respect to dynamic and distributional properties of the data-generating process. This result is strongly confirmed by obtained coverage test results based on a 5% VaR forecasting application.



## References

- Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroscedasticity', *Journal of Econometrics* pp. 307–326.
- Chib, S., Nardari, F. & Shephard, N. (2006), 'Analysis of high dimensional multivariate stochastic volatility models', *Journal of Econometrics* **134**, 341–371.
- Chib, S., Omori, Y. & Asai, M. (2009), 'Multivariate stochastic volatility', *Handbook of financial time series* pp. 365–400.
- Christoffersen, P. (1998), 'Evaluating interval forecasts', *International Economic Review* **39**, 841–862.
- Clark, P. K. (1973), 'A subordinate stochastic process model with finite variance for speculative prices', *Econometrica, Oxford et al.* pp. 135–155.
- Danielsson, J. (1998), 'Multivariate stochastic volatility models: Estimation and a comparison with VGARCH models', *Journal of Empirical Finance* **5**, 155–173.
- Diebold, F. X. (1986), 'Modeling the persistence of conditional variances: A comment', *Econometric Reviews* (5), 51–56.
- Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation', *Econometrica, Oxford et al.* pp. 987–1007.
- Engle, R. F. (2002), 'Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroscedasticity models', *Journal of business and economic statistics* **20**(3), 339–351.
- Engle, R. F. & Kroner, K. F. (1995), 'Multivariate simultaneous generalized ARCH', *Econometric Theory* (11), 122–150.
- Forbes, K. & Rigobon, R. (2001), 'Measuring contagion: conceptual and empirical issues', *S. Claessens and K. Forbes, International financial contagion, Kluwer academic, New York* pp. 43–66.
- Gallant, A. R., Hsieh, D. A. & Tauchen, G. E. (1991), 'On fitting a recalcitrant series: the pound/dollar exchange rate, 1974-1983', in *Barnett, W. A., Powell, J., Tauchen, G. E. (Eds.)*,

*Nonparametric and Semiparametric Methods in Economics and Statistics. Cambridge University Press, Cambridge* pp. 199–240.

Hamilton, J. D. (1994), *Time series analysis*, Princeton university press, Princeton.

Harvey, A. C., Ruiz, E. & Shephard, N. (1994), ‘Multivariate stochastic variance models’, *Review of Economic Studies* (61), 247–264.

Kim, S., Shephard, N. & Chib, S. (1998), ‘Stochastic volatility: Likelihood inference and comparison with ARCH models’, *Review of Economic Studies, Oxford* pp. 361–393.

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.

Liesenfeld, R. & Richard, J.-F. (2003), ‘Univariate and multivariate stochastic volatility models: Estimation and diagnostics’, *Journal of Empirical Finance* **10**, 505–531.

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.

Lopez, J. A. & Walter, C. A. (2001), ‘Evaluating covariance matrix forecasts in a value-at-risk framework’, *The Journal of Risk* **3**, 69–97.

Philipov, A. & Glickman, M. E. (2006), ‘Multivariate stochastic volatility via wishart processes’, *Journal of business and economic statistics* **24**(3), 313–328.

Pitt, M. K. & Shephard, N. (1999), ‘Filtering via simulation: auxiliary particle filters’, *Journal of the American Statistical Association* **94**(446), 590–599.

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.

## 5 Tables

Table 1: Summary Statistics of 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
Minimum	-8.35	-8.64	-9.01	-7.50	-8.54
Maximum	9.60	16.24	9.19	9.68	8.34
$BP_r^a(10)$	4.98	1.34	2.91	7.91	1.56
p-value	(0.89)	(0.99)	(0.98)	(0.63)	(0.99)
$BP_{r,2}(30)$	2293.70	1137.94	2409.56	2341.55	2583.31
p-value	(0)	(0)	(0)	(0)	(0)

Table 2: Estimation Results: Basic WMSV Model

$\mathbf{A}^{-1}$				
<b>0.8775</b> (7.4644e-04)	<b>0.0241</b> (4.3517e-04)	<b>0.0393</b> (1.6301e-04)	<b>0.0381</b> (6.0392e-04)	<b>0.0363</b> (3.7437e-04)
[0.0057]	[0.0032]	[0.0034]	[0.0037]	[0.0035]
[0.0432; 391.9209]	[0.0179; 391.9209]	[0.0326; 0.0459]	[0.0311; 0.0459]	[0.0295; 0.0432]
.	<b>0.9209</b> (5.3122e-04)	<b>0.0200</b> (2.6193e-04)	<b>0.0141</b> (4.9501e-04)	<b>0.0183</b> (1.8959e-04)
.	[0.0047]	[0.0032]	[0.0031]	[0.0031]
.	[0.9119; 0.9301]	[0.0138; 0.0263]	[0.0081; 0.0204]	[0.0122; 0.0244]
.	.	<b>0.8759</b> (8.7461e-04)	<b>0.0211</b> (4.7998e-04)	<b>0.0311</b> (3.0469e-04)
.	.	[0.0057]	[0.0031]	[0.0033]
.	.	[0.8643; 0.8868]	[0.0150; 0.0272]	[0.0245; 0.0375]
.	.	.	<b>0.8839</b> (8.9553e-04)	<b>0.0278</b> (4.3256e-04)
.	.	.	[0.0054]	[0.0033]
.	.	.	[0.8729; 0.8943]	[0.0215; 0.0343]
.	.	.	.	<b>0.8775</b> (7.8650e-04)
.	.	.	.	[0.0056]
.	.	.	.	[0.8664; 0.8884]
$\nu$	<b>69.0470</b> (0.0102)	$d$	<b>0.9467</b> (5.0902e-04)	
	[0.4809]		[0.0027]	
	[67.8817; 69.9111]		[0.9411; 0.9518]	

Burn-in: 15,000; Gibbs sequences: 50,000; bandwidth for spectral estimator: 35,000; Gamma prior for  $\nu$ :  $\mu = 70$ ,  $\sigma = 10$ ; Wishart prior for  $\mathbf{A}^{-1}$ :  $Q_0 = I_5$ ,  $\gamma_0 = 6$ . In parentheses: MC standard deviation. In square brackets: posterior standard deviation; 95% a posteriori high density region.

Table 3: Estimation Results: MS WMSV Model

		$\mathbf{A}_1^{-1}, \mathbf{A}_2^{-1}$			
$\mathbf{A}_1^{-1}$	<b>0.7511</b> (0.0019)	<b>0.0487</b> (7.9698e-04)	<b>0.0703</b> (9.8845e-04)	<b>0.0702</b> (7.8869e-04)	<b>0.0664</b> (0.0010)
	[0.0090]	[0.0046]	[0.0051]	[0.0044]	[0.0046]
	[0.7351; 0.7691]	[0.0399; 0.0577]	[0.0607; 0.0806]	[0.0616; 0.0788]	[0.0576; 0.0756]
$\mathbf{A}_2^{-1}$	<b>1.2050</b> (0.0038)	<b>0.1071</b> (0.0044)	<b>0.2002</b> (0.0012)	<b>0.1892</b> (0.0080)	<b>0.1858</b> (0.0062)
	[0.0303]	[0.0246]	[0.0201]	[0.0272]	[0.0320]
	[1.1459; 1.2641]	[0.0701; 0.1713]	[0.1639; 0.2425]	[0.1390; 0.2439]	[0.1308; 0.2503]
$\mathbf{A}_1^{-1}$	.	<b>0.8388</b> (0.0026)	<b>0.0374</b> (5.3089e-04)	<b>0.0252</b> (2.8740e-04)	<b>0.0380</b> (0.0014)
	.	[0.0092]	[0.0043]	[0.0043]	[0.0050]
	.	[0.8224; 0.8583]	[0.0292; 0.0459]	[0.0171; 0.0341]	[0.0284; 0.0475]
$\mathbf{A}_2^{-1}$	.	<b>1.3029</b> (0.0068)	<b>0.1063</b> (0.0020)	<b>0.1091</b> (0.0050)	<b>0.0973</b> (0.0048)
	.	[0.0631]	[0.0248]	[0.0257]	[0.0277]
	.	[1.2257; 1.3673]	[0.0586; 0.1561]	[0.0623; 0.1586]	[0.0529; 0.1645]
$\mathbf{A}_1^{-1}$	.	.	<b>0.7474</b> (0.0030)	<b>0.0440</b> (7.7266e-04)	<b>0.0539</b> (0.0015)
	.	.	[0.0110]	[0.0046]	[0.0060]
	.	.	[0.7263; 0.7694]	[0.0352; 0.0531]	[0.0419; 0.0642]
$\mathbf{A}_2^{-1}$	.	.	<b>1.1678</b> (0.0038)	<b>0.0976</b> (0.0072)	<b>0.1879</b> (0.0029)
	.	.	[0.0277]	[0.0263]	[0.0207]
	.	.	[1.1100; 1.2203]	[0.0420; 0.1465]	[0.1466; 0.2288]
$\mathbf{A}_1^{-1}$	.	.	.	<b>0.7598</b> (0.0016)	<b>0.0506</b> (0.0011)
	.	.	.	[0.0096]	[0.0049]
	.	.	.	[0.7418; 0.7794]	[0.0378; 0.0587]
$\mathbf{A}_2^{-1}$	.	.	.	<b>1.2385</b> (0.0129)	<b>0.1362</b> (0.0108)
	.	.	.	[0.0374]	[0.0365]
	.	.	.	[1.1610; 1.3061]	[0.0654; 0.1993]
$\mathbf{A}_1^{-1}$	.	.	.	.	<b>0.7462</b> (0.0018)
	.	.	.	.	[0.0094]
	.	.	.	.	[0.7286; 0.7642]
$\mathbf{A}_2^{-1}$	.	.	.	.	<b>1.2679</b> (0.0037)
	.	.	.	.	[0.0303]
	.	.	.	.	[1.2110; 1.3286]
		$\nu$	$d$	$e_1$	$e_2$
		<b>80.1245</b> (0.0171)	<b>0.8859</b> (0.0017)	<b>0.0836</b> (5.1945e-04)	<b>0.4028</b> (3.4751e-04)
		[0.4941]	[0.0061]	[0.0070]	[0.0098]
		[79.1592; 81.0936]	[0.8750; 0.8972]	[0.0703; 0.0978]	[0.3837; 0.4219]

Burn-in: 20,000; Gibbs sequences: 50,000; bandwidth: 30,000; Gamma prior for  $\nu$ :  $\mu = 80$ ,  $\sigma = 0.5$ ; Wishart prior for  $\mathbf{A}_1^{-1}$  and  $\mathbf{A}_2^{-1}$ :  $Q_0 = I_5$ ,  $\gamma_0 = 6$ . Beta prior for  $e_1$ :  $\mu = 0.09$ ,  $\sigma = 0.01$ . Beta prior for  $e_2$ :  $\mu = 0.4$ ,  $\sigma = 0.01$ . In parentheses: MC standard deviation. In square brackets: posterior standard deviation; 95% a posteriori high density region.

Table 4: Simulation Results, MS WMSV Model: Unconditional Mean and Unconditional (Lagged) Correlation of Return Volatilities

	<i>France</i>	<i>Germany</i>	<i>Italy</i>	<i>Switzerland</i>	<i>UK</i>
$E[\boldsymbol{\Sigma}_t], \mathbf{A}_1$					
<i>France</i>	0.3851	0.2889	0.2844	0.2820	0.2845
<i>Germany</i>	0.2889	0.5101	0.2364	0.2143	0.24146
<i>Italy</i>	0.2844	0.2364	0.3057	0.2258	0.2395
<i>Switzerland</i>	0.2820	0.2143	0.2258	0.3163	0.2353
<i>UK</i>	0.2845	0.2414	0.2395	0.2353	0.3093
$E[\boldsymbol{\Sigma}_t], \mathbf{A}_2$					
<i>France</i>	84.6219	63.0729	71.5678	71.6707	81.9901
<i>Germany</i>	63.0729	58.3977	54.3402	54.8291	61.7737
<i>Italy</i>	71.5678	54.3402	63.6885	60.1072	71.3041
<i>Switzerland</i>	71.6707	54.8291	60.1072	65.6259	69.7737
<i>UK</i>	81.9901	61.7737	71.3041	69.7737	85.2218
$\text{Corr}[\sigma_{jj,t}^2, \sigma_{ii,t-1}^2], \mathbf{A}_1$					
<i>France</i>	0.8884	0.3358	0.5269	0.5090	0.4984
<i>Germany</i>	0.3366	0.8878	0.2821	0.2272	0.2689
<i>Italy</i>	0.5325	0.2845	0.8891	0.4107	0.4702
<i>Switzerland</i>	0.5069	0.2218	0.4091	0.8877	0.4261
<i>UK</i>	0.5006	0.2748	0.4740	0.4314	0.8884
$\text{Corr}[\sigma_{jj,t}^2, \sigma_{ii,t-1}^2], \mathbf{A}_2$					
<i>France</i>	0.8892	0.6569	0.7979	0.7778	0.7762
<i>Germany</i>	0.6569	0.8897	0.6474	0.6202	0.6291
<i>Italy</i>	0.7972	0.6467	0.8915	0.7015	0.7822
<i>Switzerland</i>	0.7747	0.6168	0.7000	0.8903	0.7065
<i>UK</i>	0.7736	0.6284	0.7808	0.7062	0.8922

Simulation sample size:  $T = 20,000$ .  $(i, j)$ -entries refer to  $\text{Corr}[\sigma_{jj,t}^2, \sigma_{ii,t-1}^2]$ , where  $i$  is the line-index and  $j$  is the column-index.  $\nu$  and  $d$  are fixed to the obtained point estimates under the MS framework.

Table 5: Model Diagnostics Based on Pearson Residuals

Country	<i>France</i>	<i>Germany</i>	<i>Italy</i>	<i>Switzerland</i>	<i>UK</i>
<i>Ljung-Box (50) test statistic</i>					
<i>France</i>	75.1851	37.9128	56.8430	55.5294	88.4370
	315.8503	86.3336	84.8437	137.3714	213.1473
	{2653.2}	{1398.4}	{2594.8}	{2571.1}	{2858.8}
<i>Germany</i>	.	44.2132	64.6645	57.9974	48.4857
		262.7033	149.5164	146.5622	203.5875
		{1220.4}	{1229.2}	{1353.1}	{1804.8}
<i>Italy</i>	.	.	61.5855	70.5324	63.4851
			71.3172	140.1393	76.5845
			{2721.3}	{2543.0}	{2903.6}
<i>Switzerland</i>	.	.	.	41.0514	59.9368
				281.4120	115.7524
				{2555.7}	{3031.2}
<i>UK</i>	.	.	.	.	82.5616
					308.1207
					{3016.1}

Ljung-Box test results for cross-products of standardized returns. First number: MS WMSV model; second number: basic WMSV model. In curly brackets: realization of test statistic based on original return series. The 1% critical value of the  $\chi^2_{50}$  distributed test statistic is given by 76.15. Number of particles: 100,000.

Table 6: Diagnostics:  $e$  (Kim et al. (1998)): MS WMSV Model

	<i>France</i>	<i>Germany</i>	<i>Italy</i>	<i>Switzerland</i>	<i>UK</i>
<i>Mean</i>	0.0462	0.0441	0.0668	0.0457	0.0462
	0.0173	0.0177	0.0304	0.0214	0.0189
<i>Std</i>	1.0664	1.0492	1.0680	1.0690	1.0642
	0.9947	0.9900	0.9945	0.9960	0.9936
<i>Kurtosis</i>	2.9483	3.0495	3.0182	2.9214	2.9520
	3.7193	3.9425	3.8131	3.7339	3.6244
	{12.4874}	{24.6755}	{14.4763}	{11.8407}	{12.8685}
<i>KS-Test</i>	0.0446**	0.0417**	0.0658**	0.0484**	0.0469**
	0.0455**	0.0420**	0.0726**	0.0534**	0.0516**
	{0.0559**}	{0.0459**}	{0.0960**}	{0.0789**}	{0.0780**}
<i>JB-Test</i>	8.1067*	6.2948*	22.2527**	9.8345**	12.1775**
	53.8931**	72.1440**	96.1993**	54.4613**	54.3058**
	{5870**}	{30807**}	{8602**}	{5099**}	{6408**}

First number: MS WMSV model; second number: basic WMSV model. In curly brackets: original return series.

Number of particles: 100,000. \*: significant at 5% significance level; \*\*: significant at 1% significance level.

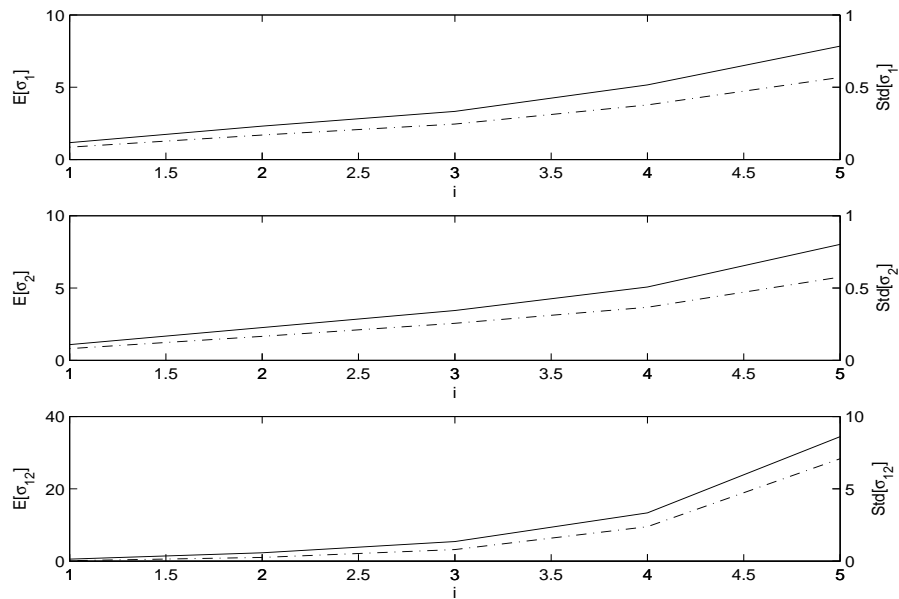
Table 7: 5% VaR Forecasts (2008)

	Hit Rate	$LR_{uc}$	$p_{uc}$	$LR_{cc}$	$p_{cc}$
WMSV	0.1527	38.5194	5.4212e-010	39.6292	2.4810e-009
MS WMSV	0.0989	10.4196	0.0012	11.8902	0.0026



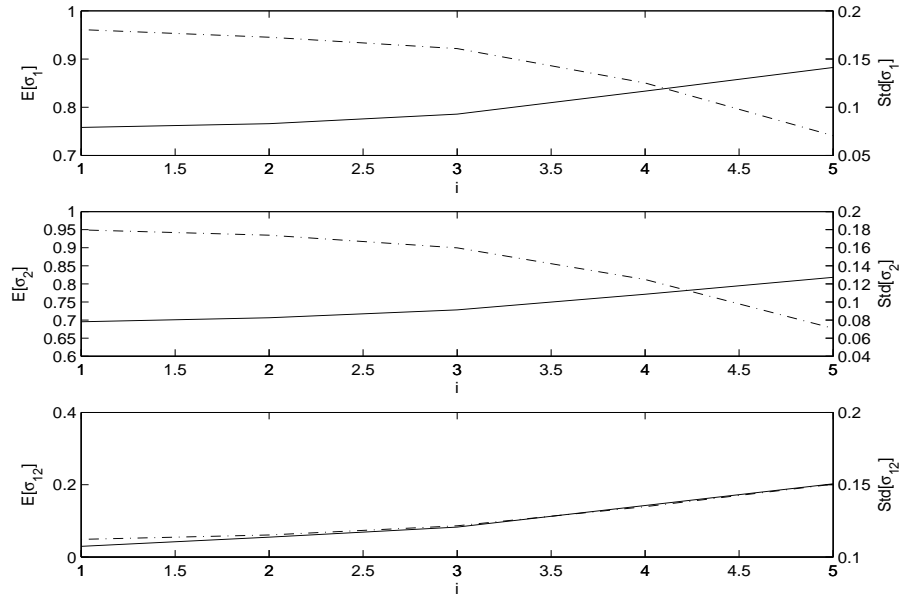
## 6 Figures

Figure 1: Simulated Unconditional Means and Standard Deviations of  $\Sigma_t$  for Varying  $\mathbf{A}^{-1}$  Matrices



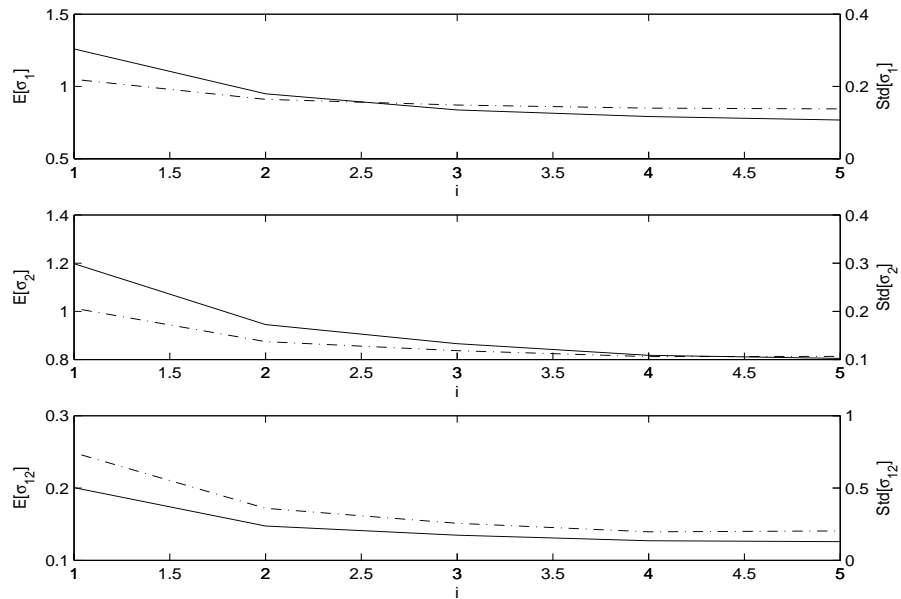
$i$  is the index on the set of considered parameter matrices. Simulation sample size:  $T = 20,000$ . Solid line:  $E[\cdot]$ ; dashed line:  $std[\cdot]$ . All remaining model parameters are kept constant.

Figure 2: Simulated Unconditional Means and Standard Deviations of  $\Sigma_t$  for Varying  $d$



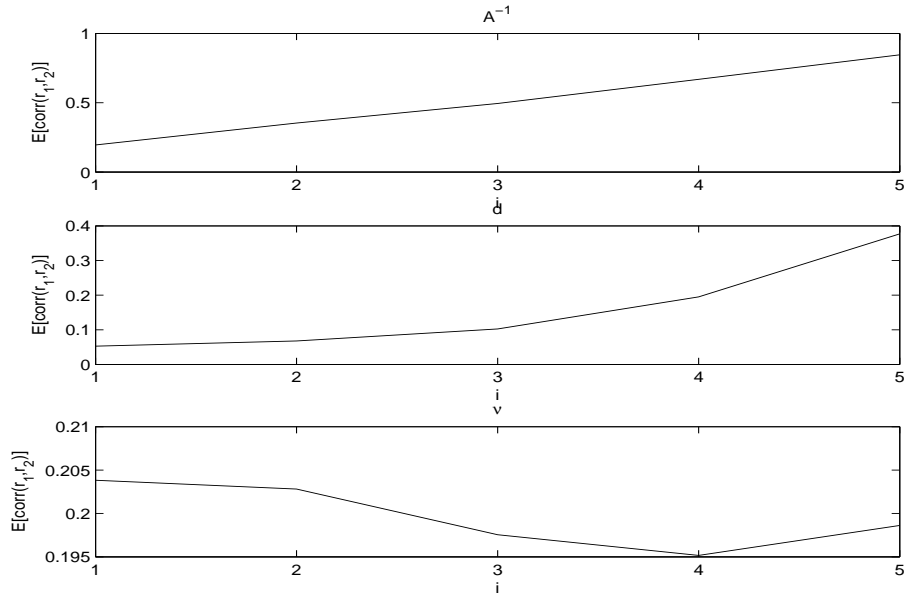
$i$  is the index on the set of considered parameter values. Simulation sample size:  $T = 20,000$ . Solid line:  $E[\cdot]$ ; dashed line:  $\text{std}[\cdot]$ . All remaining model parameters are kept constant.

Figure 3: Simulated Unconditional Means and Standard Deviations of  $\Sigma_t$  for Varying  $\nu$



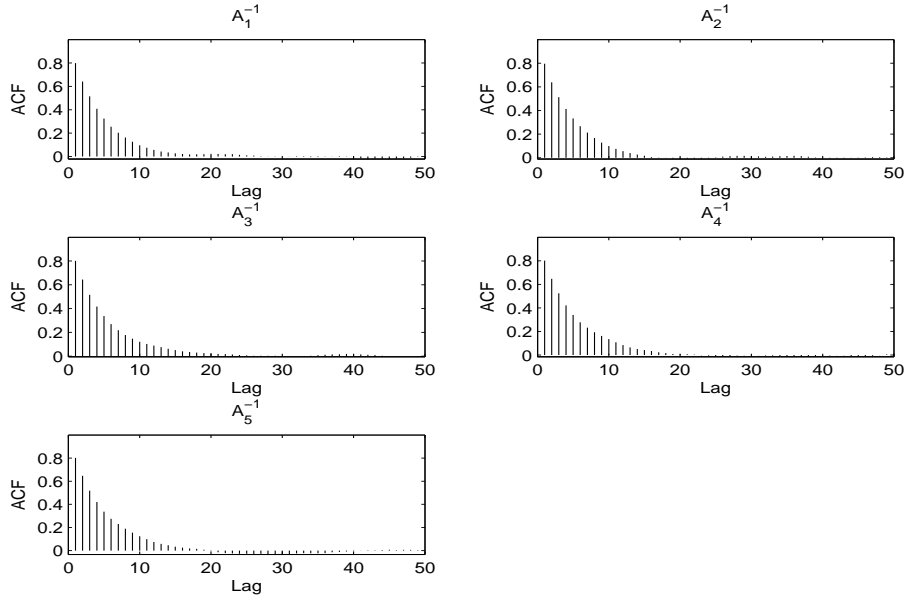
$i$  is the index on the set of considered parameter values. Simulation sample size:  $T = 20,000$ . Solid line:  $E[\cdot]$ ; dashed line:  $\text{std}[\cdot]$ . All remaining model parameters are kept constant.

Figure 4: Simulated Unconditional Means of Return Correlations for Varying Model Parameters



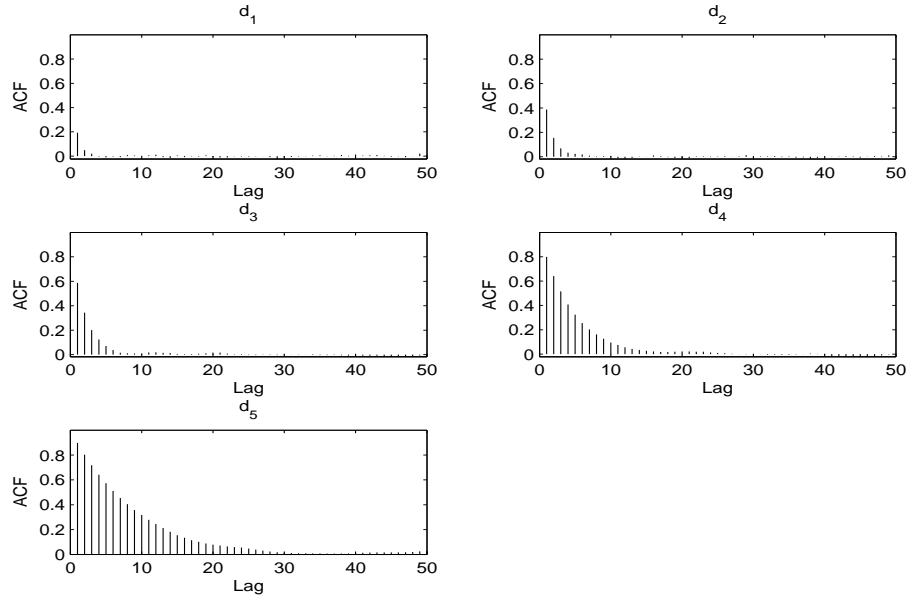
$i$  is the index on the set of considered parameter values. Simulation sample size:  $T = 20,000$ . All remaining model parameters are kept constant.

Figure 5: Simulated Unconditional Autocorrelation Functions of the Variance  $\sigma_1^2$  for Varying  $\mathbf{A}^{-1}$



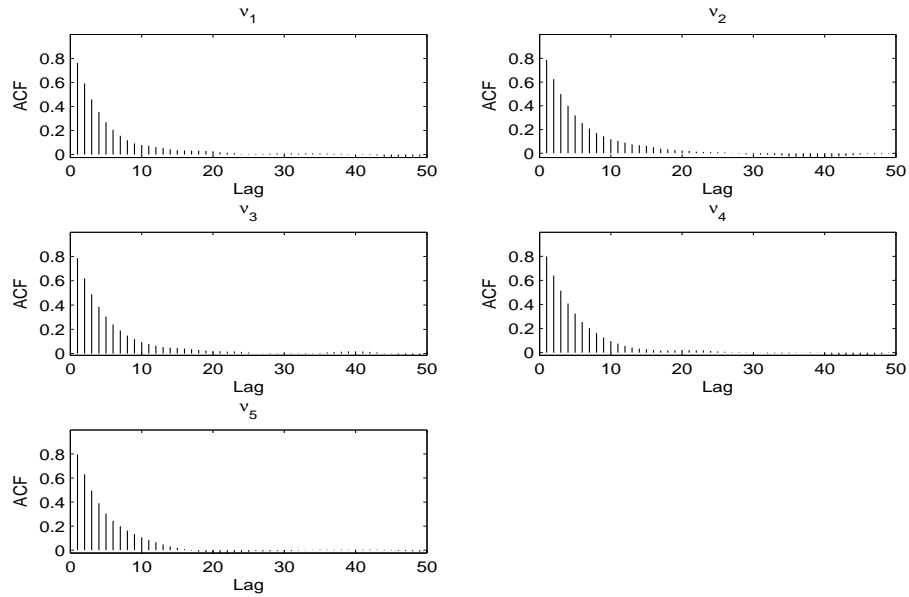
Simulation sample size:  $T = 20,000$ . All remaining model parameters are kept constant.

Figure 6: Simulated Unconditional Autocorrelation Functions of the Variance  $\sigma_1^2$  for Varying  $d$



Simulation sample size:  $T = 20,000$ . All remaining model parameters are kept constant.

Figure 7: Simulated Unconditional Autocorrelation Functions of the Variance  $\sigma_1^2$  for Varying  $\nu$



Simulation sample size:  $T = 20,000$ . All remaining model parameters are kept constant.

Figure 8: Return Series

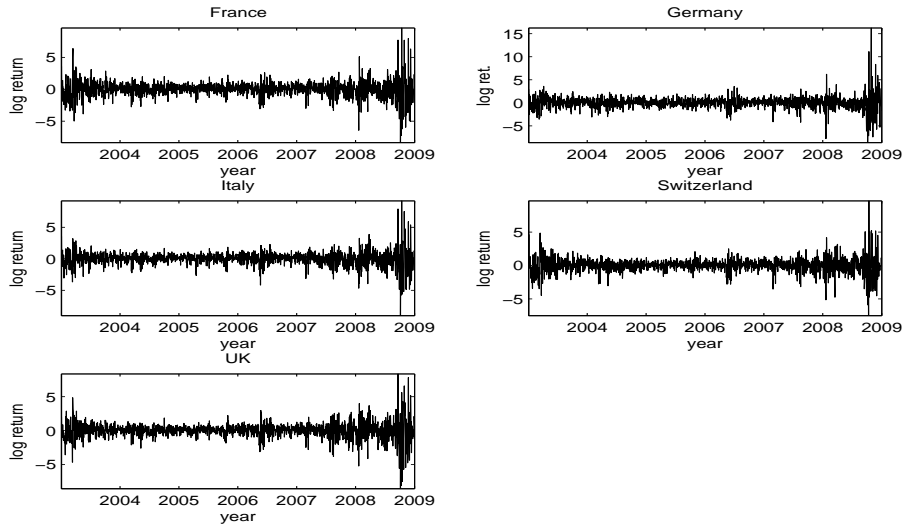


Figure 9: Gibbs Sequences: Basic WMSV Model

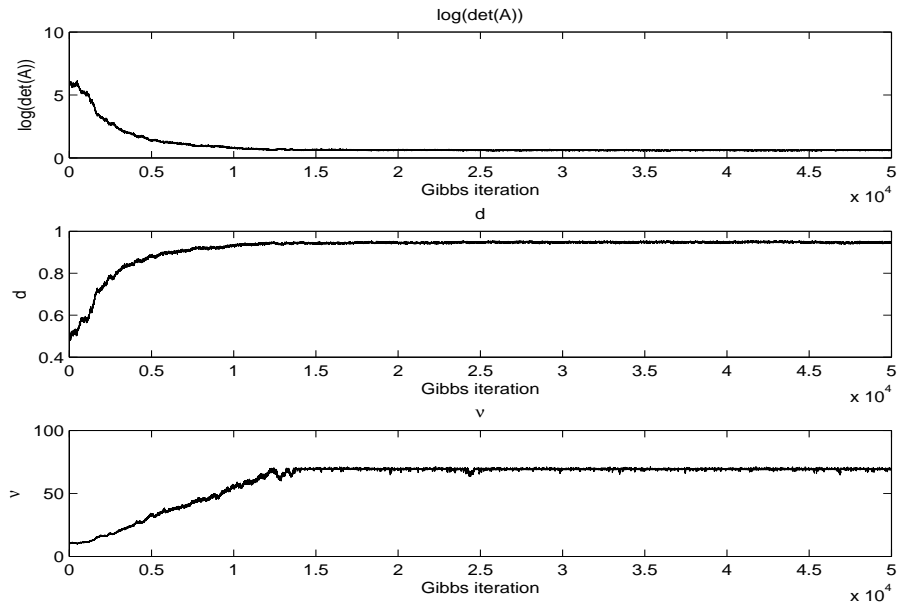


Figure 10: Histograms of Gibbs Sequences after Burn-in: Basic WMSV Model

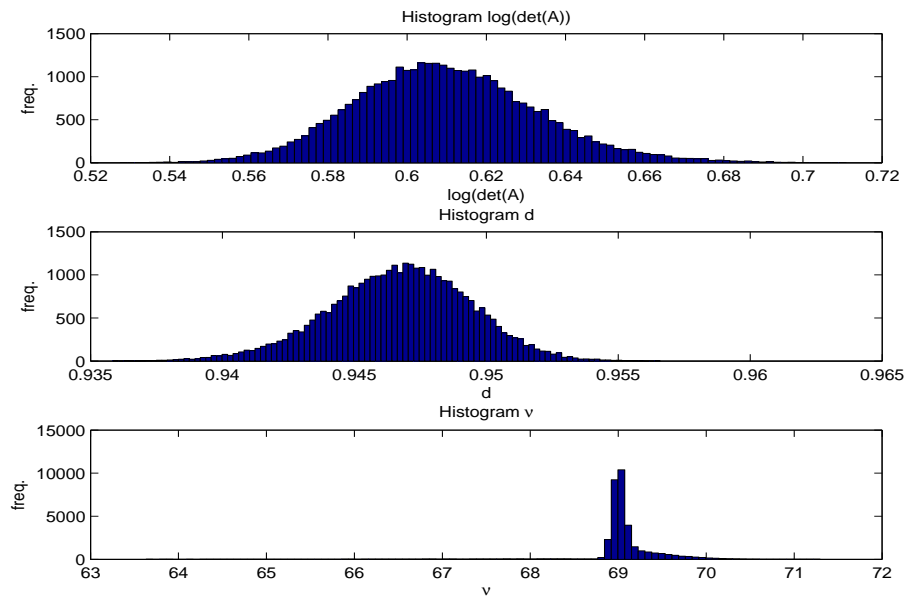
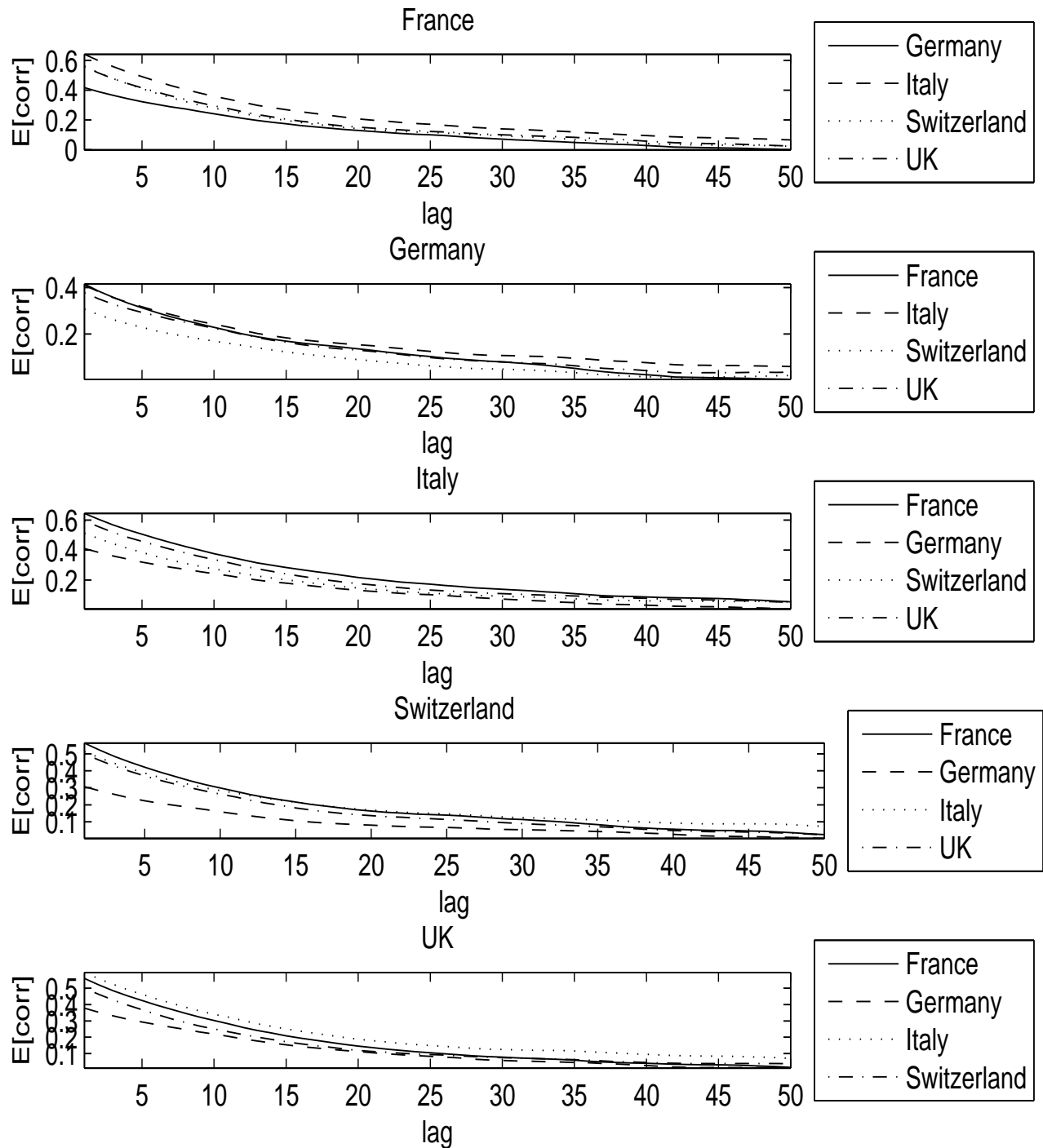


Figure 11: Simulation Based Estimates of Unconditional (Lagged) Volatility Cross-Correlations: Basic WMSV Model



Simulation sample size:  $T = 20,000$ . Each graph shows the cross-correlation of the titled country's volatility with the remaining countries' volatilities at varying lags.

Figure 12: Smoothed Volatility Estimates and Corresponding Return Series: Basic WMSV Model

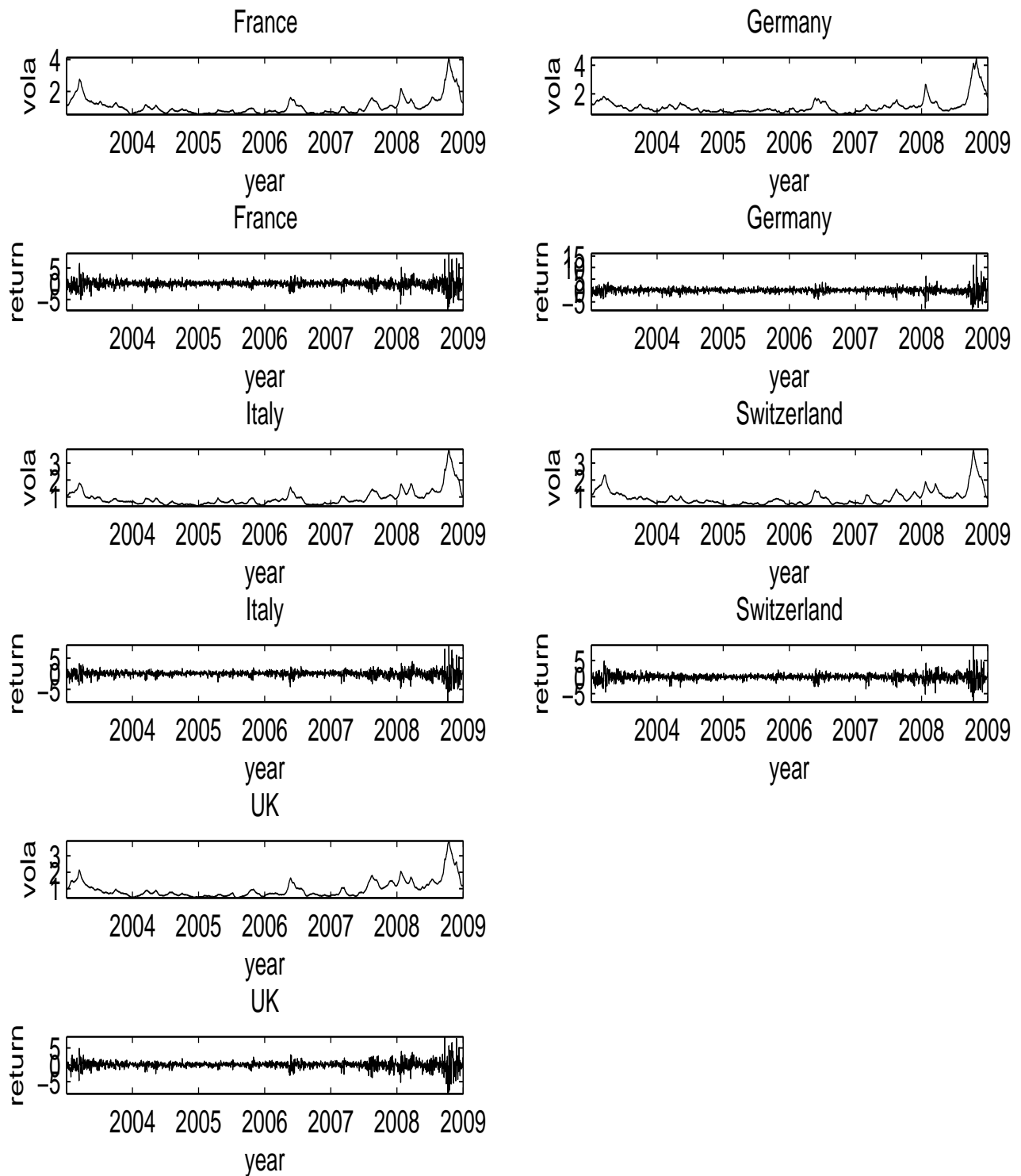




Figure 13: Smoothed Correlation Estimates: Basic WMSV Model

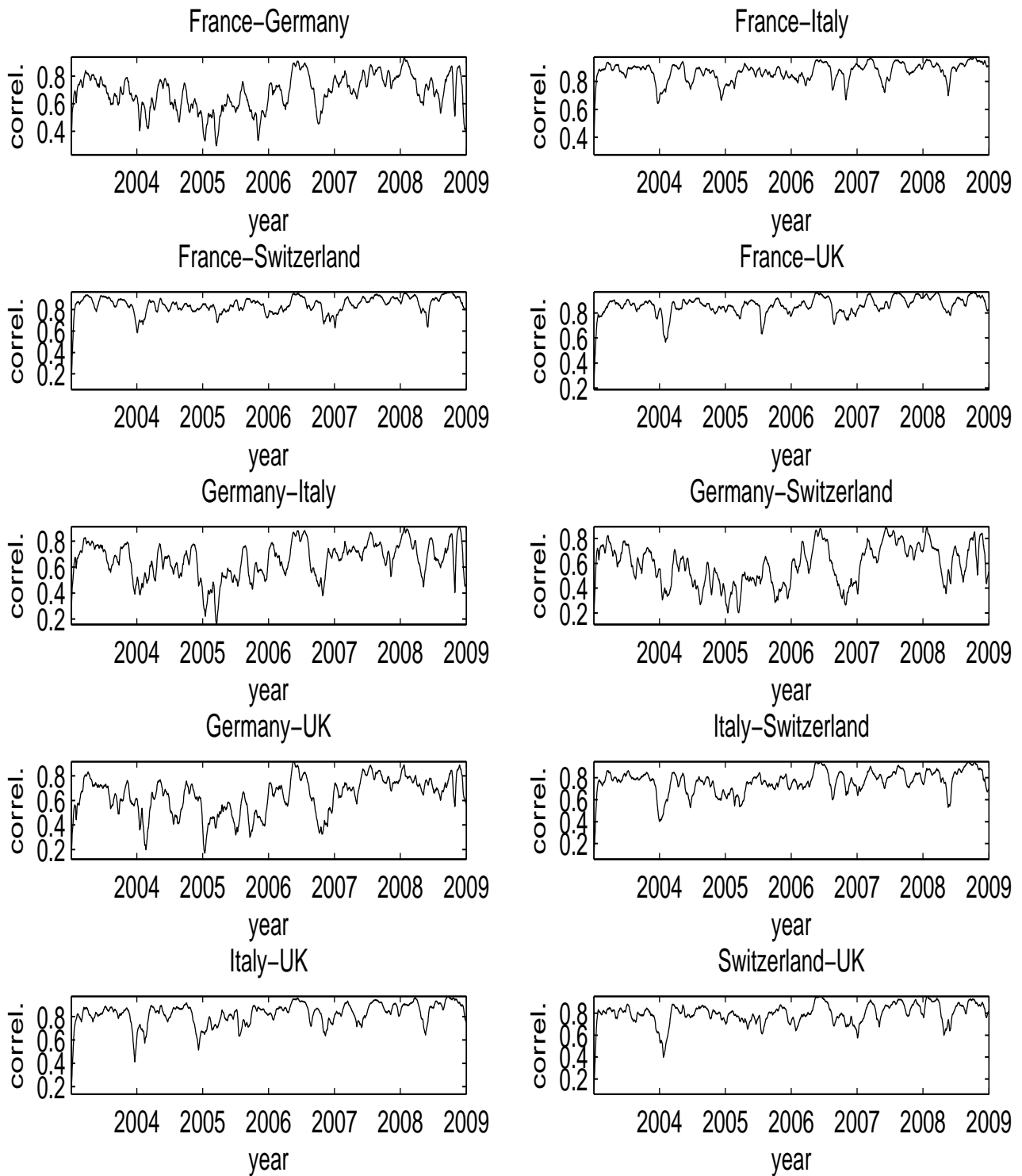


Figure 14: Gibbs Sequences:  $\ln |\mathbf{A}_1|$ ,  $\ln |\mathbf{A}_2|$ ,  $d$ , MS WMSV Model

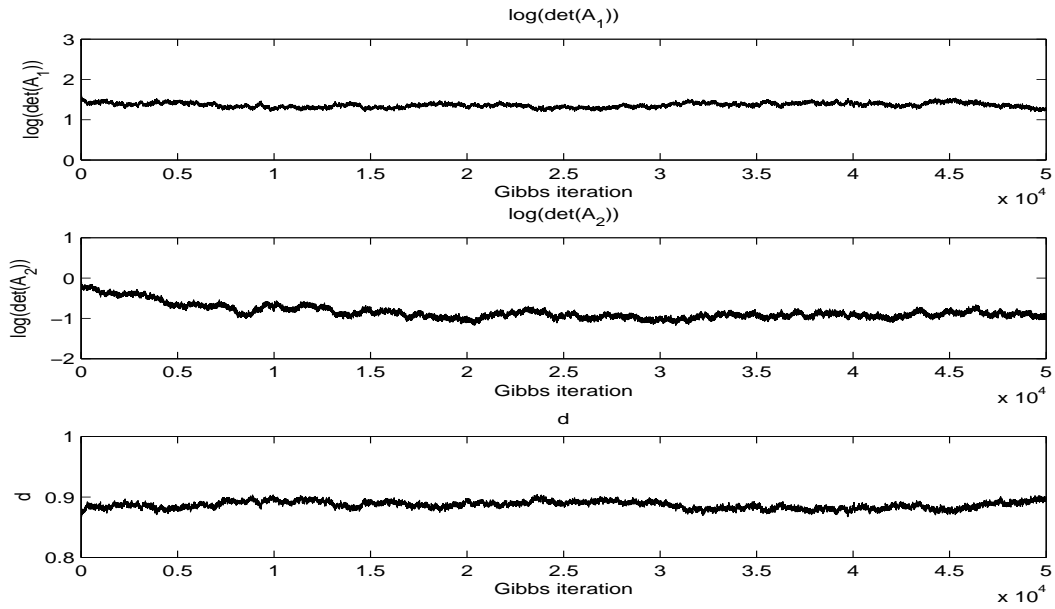


Figure 15: Gibbs Sequences:  $\nu$ ,  $e_1$ ,  $e_2$ , MS WMSV Model

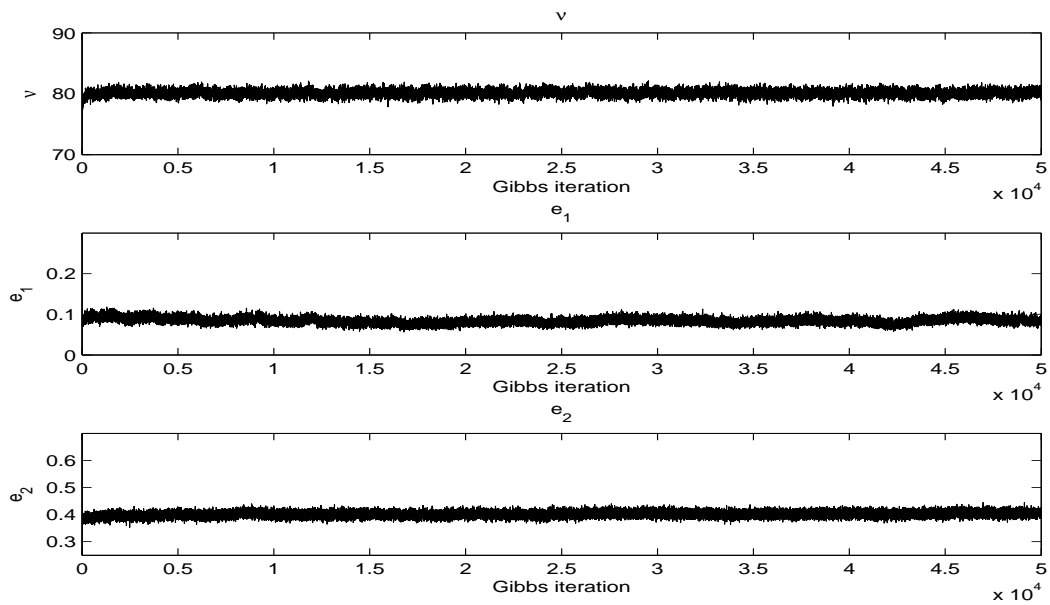


Figure 16: Histograms of Gibbs Sequences after Burn-in:  $\ln |\mathbf{A}_1|$ ,  $\ln |\mathbf{A}_2|$ ,  $d$ , MS WMSV Model

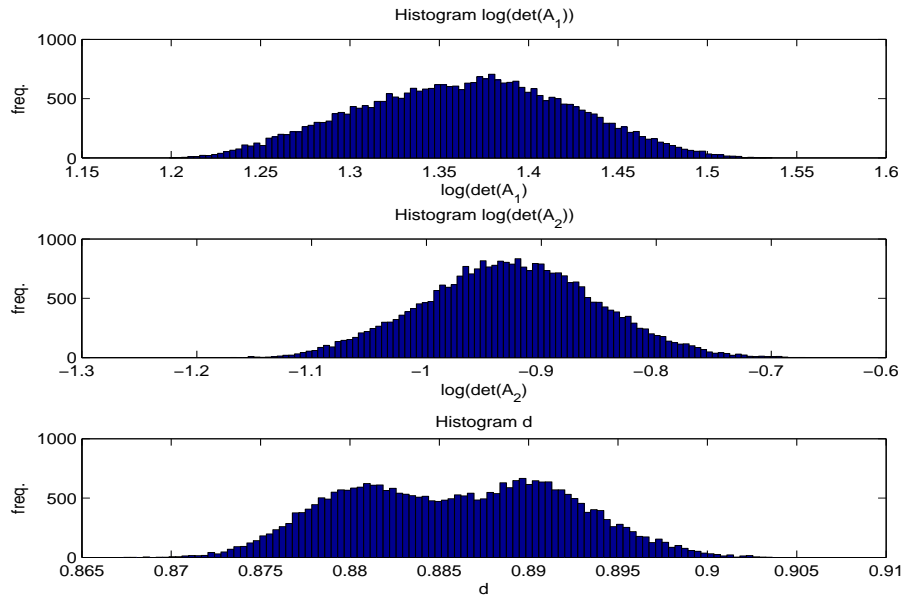


Figure 17: Histograms of Gibbs Sequences after Burn-in:  $\nu$ ,  $e_1$ ,  $e_2$ , MS WMSV Model

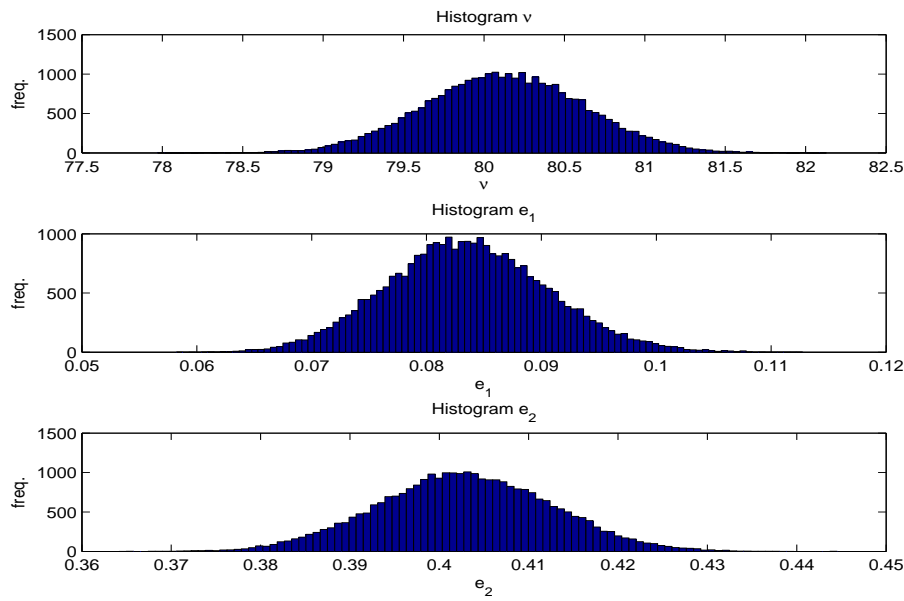
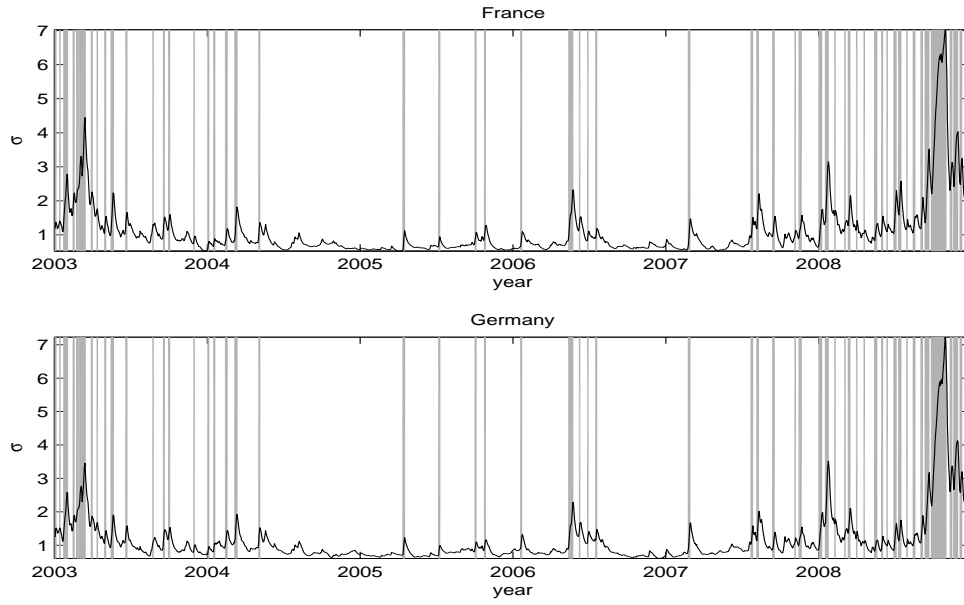
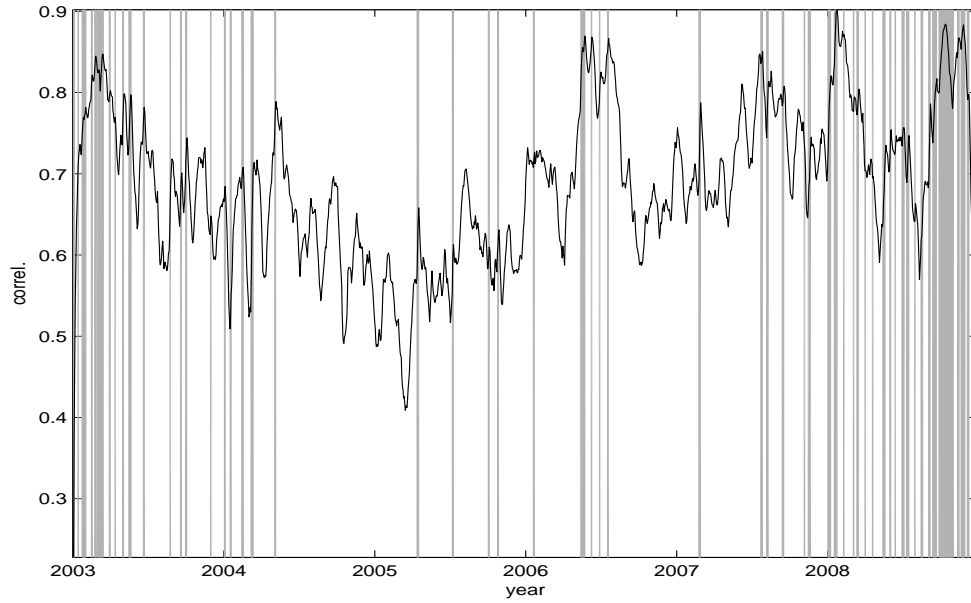


Figure 18: Smoothed Volatility Estimates: MS WMSV Model



In grey: state 2.

Figure 19: Smoothed Correlation Estimates: MS WMSV Model, France and Germany



In grey: state 2.

## 7 Mathematical Appendix

### 7.A Full Conditional Distributions: Basic WMSV Model

According to Bayes' theorem, the full conditional distribution of each element of the model's augmented parameter vector  $\boldsymbol{\theta}^{\text{aug}} = (\mathbf{A}, \nu, d, \{\boldsymbol{\Sigma}_t\}_{t=1}^T)'$  is proportional to the product of the likelihood function and the parameters' joint prior distribution, where the latter (by assumption) factors in the product of the marginal prior distributions:

$$P(\boldsymbol{\theta}^{\text{aug}}|\underline{\mathbf{Y}}) \propto \prod_{t=1}^T P(\mathbf{Y}_t|\boldsymbol{\Sigma}_t, \boldsymbol{\theta}) \times P(\boldsymbol{\Sigma}_t^{-1}|\mathbf{S}_{t-1}, \nu) \\ \times \text{Wish}_k(\mathbf{A}^{-1}; \mathbf{Q}_0, \gamma_0) \times p(d; 0, 1) \times \text{Gamma}(\nu - k; \alpha_0, \beta_0). \quad (23)$$

The full conditional distribution of  $\boldsymbol{\Sigma}_t^{-1}$  is therefore derived by

$$P(\boldsymbol{\Sigma}_t^{-1}|\text{rest}) \propto \text{Wish}(\boldsymbol{\Sigma}_t^{-1}|\nu, \mathbf{S}_{t-1}) \times N(0, \boldsymbol{\Sigma}_t) \times \text{Wish}(\boldsymbol{\Sigma}_{t+1}^{-1}|\nu, \mathbf{S}_t) \quad (24)$$

$$\propto |\mathbf{S}_{t-1}|^{-\nu/2} |\boldsymbol{\Sigma}_t^{-1}|^{(\nu-k-1)/2} \times \exp\{-0.5 \text{tr}[\mathbf{S}_{t-1}^{-1} \boldsymbol{\Sigma}_t^{-1}]\} \\ \times |\boldsymbol{\Sigma}_t|^{-0.5} \exp\{-0.5 \mathbf{Y}_t' \boldsymbol{\Sigma}_t^{-1} \mathbf{Y}_t\} \times |\mathbf{S}_t|^{-\nu/2} \times \exp\{-0.5 \text{tr}[\mathbf{S}_t^{-1} \boldsymbol{\Sigma}_{t+1}^{-1}]\} \quad (25)$$

$$\propto |\boldsymbol{\Sigma}_t^{-1}|^{(\nu-k-1)/2} |\boldsymbol{\Sigma}_t^{-1}|^{0.5} |\boldsymbol{\Sigma}_t^{-1}|^{-(d\nu)/2} \\ \times \exp\{-0.5 [\underbrace{\mathbf{Y}_t' \boldsymbol{\Sigma}_t^{-1} \mathbf{Y}_t}_{=\text{tr}[\mathbf{Y}_t \mathbf{Y}_t' \boldsymbol{\Sigma}_t^{-1}]} + \text{tr}[\mathbf{S}_{t-1}^{-1} \boldsymbol{\Sigma}_t^{-1}]]\} \times \exp\{-0.5 \text{tr}[\mathbf{S}_t^{-1} \boldsymbol{\Sigma}_{t+1}^{-1}]\} \quad (26)$$

$$= |\boldsymbol{\Sigma}_t^{-1}|^{(\nu-k-d\nu)/2} \times \exp\{-0.5 \text{tr}[(\mathbf{S}_{t-1}^{-1} + \mathbf{Y}_t \mathbf{Y}_t') \boldsymbol{\Sigma}_t^{-1}]\} \quad (27)$$

$$\times \exp\{-0.5 \text{tr}[\mathbf{S}_t^{-1} \boldsymbol{\Sigma}_{t+1}^{-1}]\} \quad (28)$$

$$= \text{Wish}(\boldsymbol{\Sigma}_t^{-1}|\tilde{\nu}, \tilde{\mathbf{S}}_{t-1}) \times f(\boldsymbol{\Sigma}_t^{-1}), \quad (29)$$

where

$$\tilde{\nu} = \nu(1-d) + 1; \quad \tilde{\mathbf{S}}_{t-1} = (\mathbf{S}_{t-1}^{-1} + \mathbf{Y}_t \mathbf{Y}_t')^{-1}, \quad (30)$$

$$f(\boldsymbol{\Sigma}_t^{-1}) = \exp\{-0.5 \text{tr}[\mathbf{S}_t^{-1} \boldsymbol{\Sigma}_{t+1}^{-1}]\}, \quad (31)$$

$$\mathbf{S}_t = \frac{1}{\nu} \boldsymbol{\Sigma}_t^{-d/2} \mathbf{A} \boldsymbol{\Sigma}_t^{-d/2}, \quad (32)$$

and the notion 'rest' denotes short-hand notation for the remaining elements of the model's augmented parameter vector  $\boldsymbol{\theta}^{\text{aug}}$ .

Hence the full conditional distribution of  $\boldsymbol{\Sigma}_t^{-1}$  is only known up to an integrating constant. Samples from  $P(\boldsymbol{\Sigma}_t^{-1}|\text{rest})$  can therefore be obtained via the Metropolis-Hastings algorithm with

acceptance probability  $\alpha$  given by

$$\alpha = \frac{P(\boldsymbol{\theta}^*)}{Q(\boldsymbol{\theta}^*)} \times \frac{Q(\boldsymbol{\theta}^{[m-1]})}{P(\boldsymbol{\theta}^{[m-1]})} = \frac{|(\boldsymbol{\Sigma}_t^*)^{-1}|^{(1-\nu d)/2} f(\boldsymbol{\Sigma}_t^{-1*})}{|(\boldsymbol{\Sigma}_t^{[m-1]})^{-1}|^{(1-\nu d)/2} f(\boldsymbol{\Sigma}_t^{-1,[m-1]})}, \quad (33)$$

where

Q: Proposal density given by  $\text{Wish}_k(\boldsymbol{\Sigma}_t^{-1}|\nu, \tilde{\mathbf{S}}_{t-1})$ ,

P: Target density  $P(\boldsymbol{\Sigma}_t^{-1}|\text{rest})$ .

The full conditional distribution of  $\mathbf{A}^{-1}$  is given by a Wishart distribution and can therefore be sampled directly within the Gibbs scheme:

$$P(\mathbf{A}^{-1}|\text{rest}) \propto \text{Wish}_k(\gamma_0, \mathbf{Q}_0) \prod_{t=1}^T P(\boldsymbol{\Sigma}_t^{-1}|\mathbf{S}_{t-1}, \nu) \quad (34)$$

$$\begin{aligned} &\propto \text{Wish}_k(\gamma_0, \mathbf{Q}_0) \prod_{t=1}^T \left| (1/\nu)(\boldsymbol{\Sigma}_{t-1}^{-d/2}) \mathbf{A}(\boldsymbol{\Sigma}_{t-1}^{-d/2}) \right|^{-\nu/2} |\boldsymbol{\Sigma}_t^{-1}|^{(\nu-k-1)/2} \\ &\quad \times \exp \left\{ -0.5 \text{tr} \left[ \left( (1/\nu)(\boldsymbol{\Sigma}_{t-1}^{-d/2}) \mathbf{A}(\boldsymbol{\Sigma}_{t-1}^{-d/2}) \right)^{-1} \boldsymbol{\Sigma}_t^{-1} \right] \right\} \end{aligned} \quad (35)$$

$$\propto \text{Wish}_k(\gamma_0, \mathbf{Q}_0) |\mathbf{A}^{-1}|^{(T\nu)/2} \times \exp \left\{ -0.5 \text{tr} \left[ \nu \sum_{t=1}^T \boldsymbol{\Sigma}_{t-1}^{d/2} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\Sigma}_{t-1}^{d/2} \mathbf{A}^{-1} \right] \right\} \quad (36)$$

$$\propto \text{Wish}_k(\gamma_0, \mathbf{Q}_0) \times \text{Wish}_k(\gamma, \mathbf{Q}), \quad (37)$$

where

$$\mathbf{Q}^{-1} = \nu \sum_{t=1}^T \boldsymbol{\Sigma}_{t-1}^{d/2} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\Sigma}_{t-1}^{d/2}, \quad (38)$$

$$\gamma = T\nu + k + 1, \quad (39)$$

and hence

$$P(\mathbf{A}^{-1}|\text{rest}) \propto \text{Wish}_k(\gamma_0, \mathbf{Q}_0) \times \text{Wish}_k(\gamma, \mathbf{Q}) \quad (40)$$

$$\begin{aligned} &\propto |\mathbf{Q}_0|^{-\gamma_0/2} |\mathbf{A}^{-1}|^{(\gamma_0-k-1)/2} \exp\{-0.5 \text{tr}[\mathbf{Q}_0^{-1} \mathbf{A}^{-1}]\} \\ &\quad |\mathbf{Q}|^{-\gamma/2} |\mathbf{A}^{-1}|^{(\gamma-k-1)/2} \exp\{-0.5 \text{tr}[\mathbf{Q}^{-1} \mathbf{A}^{-1}]\} \end{aligned} \quad (41)$$

$$\propto |\mathbf{A}^{-1}|^{(\gamma_0+\gamma-2k-2)/2} \exp\{-0.5 \text{tr}[(\mathbf{Q}_0^{-1} + \mathbf{Q}^{-1}) \mathbf{A}^{-1}]\} \quad (42)$$

$$\propto \text{Wish}_k(\mathbf{A}^{-1}|\tilde{\mathbf{Q}}, \tilde{\gamma}), \quad (43)$$

where

$$\tilde{\mathbf{Q}}^{-1} = \mathbf{Q}_0^{-1} + \mathbf{Q}^{-1}, \quad (44)$$

$$\tilde{\gamma} = \gamma_0 + \gamma - k - 1. \quad (45)$$

The full conditional distributions of the parameters  $\nu$  and  $d$  are not obtained in closed form. Hence again the Metropolis-Hastings sampling scheme is applied by making use of a truncated normal proposal density. The mean and standard deviation of the latter are given by the optimum and the corresponding Hessian obtained after numerically optimizing the posterior distribution's kernel over the respective parameter space.

$$P(d|\text{rest}) \propto p(d) \prod_{t=1}^T |\Sigma_{t-1}^{-1}|^{-d\nu/2} \exp \left\{ -0.5 \operatorname{tr} \left[ \left( (1/\nu) \Sigma_{t-1}^{-d/2} \mathbf{A} \Sigma_{t-1}^{-d/2} \right)^{-1} \Sigma_t^{-1} \right] \right\} \quad (46)$$

$$\propto \exp \left\{ d\psi - 0.5 \operatorname{tr} \left[ \mathbf{Q}(d) \mathbf{A}^{-1} \right] \right\}, \quad (47)$$

where

$$\psi = -\frac{\nu}{2} \sum_{t=1}^T \ln(|\Sigma_{t-1}^{-1}|), \quad (48)$$

$$\mathbf{Q}(d) = \sum_{t=1}^T \nu \Sigma_{t-1}^{d/2} \Sigma_t^{-1} \Sigma_{t-1}^{d/2}, \quad (49)$$

and

$$p(\nu|\text{rest}) \propto \text{Gamma}(\nu - k|\alpha, \beta) \times \prod_{t=1}^T P(\boldsymbol{\Sigma}_t^{-1}|\mathbf{S}_{t-1}, \nu) \quad (50)$$

$$\begin{aligned} &\propto \exp\{(\alpha - 1) \ln(\nu - k) - \beta(\nu - k)\} \\ &\times \prod_{t=1}^T \left( \frac{|(1/\nu)\boldsymbol{\Sigma}_{t-1}^{-d/2} \mathbf{A} \boldsymbol{\Sigma}_{t-1}^{-d/2}|^{-\nu/2} |\boldsymbol{\Sigma}_t^{-1}|^{(\nu-k-1)/2}}{2^{\nu k/2} \prod_{j=1}^k \Gamma((\nu - j + 1)/2)} \right) \\ &\times \exp\left\{-0.5 \text{tr}\left[\left((1/\nu)\boldsymbol{\Sigma}_{t-1}^{-d/2} \mathbf{A} \boldsymbol{\Sigma}_{t-1}^{-d/2}\right)^{-1} \boldsymbol{\Sigma}_t^{-1}\right]\right\} \end{aligned} \quad (51)$$

$$\begin{aligned} &\propto \exp\{(\alpha - 1) \ln(\nu - k) - \beta(\nu - k)\} \\ &\times \prod_{t=1}^T \left( \frac{|\nu \mathbf{A}^{-1}|^{\nu/2} |(\boldsymbol{\Sigma}_{t-1}^{-1})^d|^{-\nu/2} |\boldsymbol{\Sigma}_t^{-1}|^{(\nu-k-1)/2}}{2^{\nu k/2} \prod_{j=1}^k \Gamma((\nu - j + 1)/2)} \right) \\ &\times \exp\left\{-0.5 \text{tr}\left[\left((1/\nu)\boldsymbol{\Sigma}_{t-1}^{-d/2} \mathbf{A} \boldsymbol{\Sigma}_{t-1}^{-d/2}\right)^{-1} \boldsymbol{\Sigma}_t^{-1}\right]\right\} \end{aligned} \quad (52)$$

$$\begin{aligned} &\propto \exp\{(\alpha - 1) \ln(\nu - k) - \beta(\nu - k)\} \\ &\times \prod_{t=1}^T \left( \frac{|\nu \mathbf{A}^{-1}|^{\nu/2} |(\boldsymbol{\Sigma}_{t-1}^{-1})^d|^{-\nu/2} |\boldsymbol{\Sigma}_t^{-1}|^{\nu/2}}{2^{\nu k/2} \prod_{j=1}^k \Gamma((\nu - j + 1)/2)} \right) \\ &\times \exp\{-0.5 \text{tr}[\mathbf{Q}^{-1} \mathbf{A}^{-1}]\} \end{aligned} \quad (53)$$

$$\begin{aligned} &\propto \exp\{(\alpha - 1) \ln(\nu - k) - \beta(\nu - k)\} \\ &\times \left( \frac{|\nu \mathbf{A}^{-1}|^{\nu/2}}{2^{\nu k/2} \prod_{j=1}^k \Gamma((\nu - j + 1)/2)} \right)^T \\ &\times \prod_{t=1}^T |\mathbf{Q}_t^{-1}|^{\nu/2} \exp\{-0.5 \text{tr}[\mathbf{Q}^{-1} \mathbf{A}^{-1}]\}, \end{aligned} \quad (54)$$

where

$$\mathbf{Q}_t^{-1} = \boldsymbol{\Sigma}_{t-1}^{d/2} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\Sigma}_{t-1}^{d/2}, \quad (55)$$

$$\mathbf{Q}^{-1} = \nu \sum_{t=1}^T \boldsymbol{\Sigma}_{t-1}^{d/2} \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\Sigma}_{t-1}^{d/2}. \quad (56)$$



## 7.B Full Conditional Distributions: Markov Switching MWSV Model

The joint sampling of the whole state sequence  $\{s_t\}_{t=1}^T$  becomes possible by rewriting the full conditional density of the state vector  $\mathbf{s}$  according to

$$P(\mathbf{s}|\text{rest}) = P(\mathbf{s}|\underline{\Sigma}_T^{-1}, \boldsymbol{\theta}) \quad (57)$$

$$= P(s_T|\underline{\Sigma}_T^{-1}, \boldsymbol{\theta}) \times P(s_{T-1}|s_T, \underline{\Sigma}_T^{-1}, \boldsymbol{\theta}) \times \cdots \times P(s_1|s_2, \underline{\Sigma}_T^{-1}, \boldsymbol{\theta}) \quad (58)$$

$$= P(s_T|\underline{\Sigma}_T^{-1}, \boldsymbol{\theta}) \times P(s_{T-1}|s_T, \underline{\Sigma}_{T-1}^{-1}, \boldsymbol{\theta}) \times \cdots \times P(s_1|s_2, \underline{\Sigma}_1^{-1}, \boldsymbol{\theta}), \quad (59)$$

where eq. (59) follows from the Markov property of  $s_t$ . The according conditional probabilities given by

$$P(s_t|s_{t+1}, \underline{\Sigma}_t^{-1}, \boldsymbol{\theta}) = \frac{P(s_{t+1}|s_t) \times P(s_t|\underline{\Sigma}_t^{-1}, \boldsymbol{\theta})}{P(s_{t+1}|\underline{\Sigma}_t^{-1}, \boldsymbol{\theta})} \quad (60)$$

are obtained via the Hamilton filter which - given a starting value for  $P(s_{t-1}|\underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta})$  (e.g. the according stationary probabilities) - proceeds recursively in five steps:

$$I \quad P(s_t, s_{t-1}|\underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta}) = P(s_t|s_{t-1}) \times P(s_{t-1}|\underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta}) \quad (61)$$

$$II \quad P(s_t|\underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta}) = \sum_{s_{t-1}} P(s_t, s_{t-1}|\underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta}) \quad (62)$$

$$III \quad f(\underline{\Sigma}_t^{-1}, s_t|\underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta}) = f(\underline{\Sigma}_t^{-1}|s_t, \underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta}) \times P(s_t|\underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta}) \quad (63)$$

$$IV \quad f(\underline{\Sigma}_t^{-1}|\underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta}) = \sum_{s_t} f(\underline{\Sigma}_t^{-1}, s_t|\underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta}) \quad (64)$$

$$V \quad P(s_t|\underline{\Sigma}_t^{-1}, \boldsymbol{\theta}) = \frac{f(\underline{\Sigma}_t^{-1}, s_t|\underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta})}{f(\underline{\Sigma}_t^{-1}|\underline{\Sigma}_{t-1}^{-1}, \boldsymbol{\theta})}. \quad (65)$$

By making use of conjugate Beta prior distributions  $\{P(e_i; \alpha_i, \beta_i)\}_{i=1}^2$ , the full conditional distributions of the Markov probabilities  $e_i$ ,  $i \in \{1, 2\}$ , are given by

$$P(e_i|\text{rest}) \propto P(e_i; \alpha_i, \beta_i) \times \prod_{j=1}^{g_i} e_i \prod_{j=1}^{h_i} (1 - e_i) \quad (66)$$

$$\propto e_i^{\alpha_i-1} (1 - e_i)^{\beta_i-1} \times e_i^{g_i} (1 - e_i)^{h_i}, \quad (67)$$

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 differ from  $i$  in the subsequent period. The full conditional distribution of  $e_i$  is therefore given by a Beta distribution with hyper-parameters  $\alpha_{i,0} = \alpha_i + g_i$  and  $\beta_{i,0} = \beta_i + h_i$ .

All remaining model parameters are sampled in the same way as presented in the previous appendix while taking the state-dependent parameter matrices  $A_i$  into account.