

Gribisch, Bastian

**Conference Paper**

## A latent dynamic factor approach to forecasting multivariate stock market volatility

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Volatility, No. F01-V2

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Gribisch, Bastian (2013) : A latent dynamic factor approach to forecasting multivariate stock market volatility, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Volatility, No. F01-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/79823>

**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.*

# A latent dynamic factor approach to forecasting multivariate stock market volatility

Bastian Gribisch \*

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

(February 28, 2013)

## Abstract

This paper proposes a latent dynamic factor model for low- as well as high-dimensional realized covariance matrices of stock returns. The approach is based on the matrix logarithm and allows for flexible dynamic dependence patterns by combining common latent factors driven by HAR dynamics and idiosyncratic AR(1) factors. The model accounts for symmetry and positive definiteness of covariance matrices without imposing parametric restrictions. Simulated Bayesian parameter estimates as well as positive definite (co)variance forecasts are obtained using Markov Chain Monte Carlo (MCMC) methods. An empirical application to 5-dimensional and 30-dimensional realized covariance matrices of daily New York Stock Exchange (NYSE) stock returns shows that the model outperforms other approaches of the extant literature both in-sample and out-of-sample.

*JEL classification:* C32, C58, G17

*Keywords:* Latent factor models, Covariance matrix, Matrix logarithm, Realized volatility

---

\*Tel.: +49-431-8802229; fax: +49-431-8807605. *E-mail address:* bastian.gribisch@stat-econ.uni-kiel.de  
(B. Gribisch)

# 1. Introduction

Many financial applications like portfolio management and asset pricing require precise forecasts of the latent covariance matrix of asset returns. For this reason various specifications of multivariate GARCH (MGARCH, see e.g. Bauwens et al., 2006) and stochastic volatility (MSV, see e.g. Asai et al., 2006) models have been proposed to model the (co)variance dynamics. While MGARCH models treat volatility as measurable given past observations, the MSV approach treats the volatility as an inherently latent state. An alternative approach is based on high-frequency return data which enables the computation of a consistent nonparametric estimate of the covariance matrix of low-frequency returns, the so-called realized covariance matrix (see Andersen et al., 2003, and Barndorff-Nielsen and Shephard, 2004). The dynamics of the realized (co)variance series can then be modeled in a direct way promising substantial gains in forecasting the state of the latent volatility process.

Although there exists a vast literature on multivariate GARCH models and a variety of multivariate SV approaches the literature concerned with the direct modeling of realized covariance matrices is rather limited. Besides the requirement of positivity for covariance matrix forecasts, the curse of dimensionality remains the main challenge: Models designed to capture the complex serial and cross-sectional (co)variance dynamics tend to be highly parameterized, since the dimension of the object of interest is proportional to the square of the number of assets. This renders inference on the volatility dynamics complicated even for moderately sized portfolios. Yet empirical applications, e.g. the forecasting of optimal portfolio weight vectors in mean/variance portfolio optimization, typically require forecasts of high-dimensional covariance matrices. Chiriac and Voev (2011) use a fractionally integrated VARMA process to model the elements of the Cholesky factor of the realized covariance matrix. The model suffers from the curse of dimensionality and the authors restrict their empirical application to six assets. Several models proposed for realized covariance matrices are based on the conditional Wishart distribution (see Gouriéroux et al., 2009, Jin and Maheu, 2011, Noureldin et al., 2011, and Golosnoy et al., 2012). Empirical applications of the models are overall limited to portfolios of up to ten assets. Bauer and Vorkink (2011) propose a factor model for the distinct elements of the matrix logarithm of the covariance matrix. The factors are driven by lagged volatilities, lagged returns and other forecasting variables. Although the model could in general be applied to

the forecasting of high-dimensional covariance matrices, the authors restrict their application to the 5-dimensional case and focus on discussing the predictive power of the various forecasting variables. Since the matrix-log is a nonlinear function bias-correction methods have to be applied for the forecasting of volatilities. The non-linear nature of the model furthermore complicates the analysis of the impact of the forecasting variables on volatility. Recently Bauwens and Storti (2011) proposed a conditional autoregressive Wishart (CAW) model featuring DCC dynamics, which allows for the forecasting of high-dimensional covariance matrices. Their empirical application comprises realized (co)variances for 50 NYSE stocks. Although the CAW-DCC model tackles the curse of dimensionality, this achievement does not come without a cost: the model imposes heavy restrictions on the correlation dynamics.

In the present paper I propose a novel flexible latent dynamic factor model for realized covariance matrices. The factor specification is motivated by persistent common dynamics of realized (co)variance series. The model is based on the matrix logarithm function which enables the modeling of log-(co)variances in Euclidean space, similar to the approach of Bauer and Vorkink (2011), preserving positive definiteness and symmetry of covariance matrix forecasts without having to impose restrictions on the parameter space. By modeling the dynamics of the common factors as heterogeneous autoregressive processes (HAR, see Corsi, 2009) and assuming AR(1) processes for the idiosyncratic dynamics the model mitigates (though not eliminates) the curse of dimensionality while allowing for rich (co)variance dynamics and can be readily applied to the forecasting of high-dimensional covariance matrices (say,  $\leq 30$  assets). In contrast to the observation driven approach of Bauer and Vorkink (2011) the latent factor model offers enhanced flexibility by allowing for idiosyncratic (co)variance dynamics and modeling the latent covariance matrix via a parameter-driven state-space approach in the spirit of SV models for asset returns. The simulated Bayesian estimation approach using Markov Chain Monte Carlo (MCMC) techniques enables straightforward estimation of the model parameters and forecasting of covariance matrices without having to rely on bias correction methods as in Bauer and Vorkink (2011). Since the elements of the matrix logarithm of a covariance matrix can be interpreted as approximations to logarithmic variances and correlations the factor model allows to investigate the presence of joint risk-factors related to market risk and

diversification risk and offers a direct link to the recent asset pricing literature (see e.g. Krishnan et al., 2009, and Driessen et al., 2009). In order to assess the model’s forecasting performance I conduct a comprehensive out-of-sample experiment including a range of prominent forecasting models from the relevant literature. Besides a statistical evaluation based on the mean squared error criterion I also address the practitioners point of view by investigating the performance of mean-variance optimal portfolios selected using the various forecasting models. An application to two data sets of 5 and 30-dimensional covariance matrices of NYSE traded stocks shows that the model outperforms proposed volatility models of the extant literature for low as well as high-dimensional covariance matrices both in-sample and out-of-sample.

The rest of the paper is organized as follows. Section 2 reviews the matrix logarithm and introduces the factor model, its estimation and model diagnostic tests. The empirical application to NYSE data is presented in Section 3. Section 4 concludes. Details on parameter estimation are provided in the Appendix.

## 2. The Volatility Model

### 2.1 The Matrix Logarithm

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

$$R_t = \text{expm}(Y_t) = \sum_{q=0}^{\infty} \frac{Y_t^q}{q!}, \quad (1)$$

where  $Y_t^0$  is the identity matrix and  $Y_t^q$  denotes standard matrix multiplication of  $Y_t$   $q$  times. From the spectral decomposition  $R_t = L_t D_t L_t'$  we directly obtain  $Y_t = \text{logm}(R_t) = L_t \ln(D_t) L_t'$ , where  $\ln(D_t)$  denotes a diagonal matrix of log-eigenvalues and  $L_t$  the corresponding matrix of eigenvectors.

---

<sup>1</sup>The realized covariance matrix  $R_t$  is computed by adding up the outer products of high-frequency (e.g. 5-minute) log-return vectors within a given day  $t$  (for details see Section 3.1 below).

Taking the matrix logarithm of a real, positive definite matrix  $R_t$  results in a real, symmetric matrix  $Y_t$  and applying the matrix exponential function to a real symmetric matrix results in a real symmetric positive definite matrix (see Chiu et al., 1996, Lemma 1).

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

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

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

$$= \sum_{q=0}^{\infty} \frac{1}{q!} y_{ii,t}^q + \nu_{ii,t} \quad (3)$$

$$= \exp(y_{ii,t}) + \nu_{ii,t}, \quad (4)$$

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

$$r_{ij,t} = \rho_{ij,t} \sqrt{r_{ii,t} r_{jj,t}} \cong \rho_{ij,t} \exp\left(\frac{1}{2}(y_{ii,t} + y_{jj,t})\right). \quad (5)$$

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

$$r_{ij,t} \cong \rho_{ij,t} \left(1 + \frac{1}{2}(y_{ii,t} + y_{jj,t})\right) \quad \forall i \neq j. \quad (6)$$

Truncating the power series expansion of Eq. (1) at the second order, we obtain

$$\begin{aligned} r_{ij,t} &\cong y_{ij,t} + \frac{1}{2} \sum_{z=1}^k y_{iz,t} y_{zj,t} \\ &\cong y_{ij,t} + \frac{1}{2} y_{ij,t} (y_{ii,t} + y_{jj,t}) = y_{ij,t} \left(1 + \frac{1}{2}(y_{ii,t} + y_{jj,t})\right) \quad \forall i \neq j, \end{aligned} \quad (7)$$

where the second equation follows from the first by setting  $y_{iz,t} y_{zj,t} = 0$  for  $z \notin \{i, j\}$ . Eqs. (6) and (7) imply  $y_{ij,t} \cong \rho_{ij,t}$ . The quality of this approximation depends on the quality of the log-variance approximation by the diagonal elements of the matrix logarithm and the overall variance level (see the 1'st order TSE around  $\frac{1}{2}(y_{ii,t} + y_{jj,t}) = 0$  leading to Eq. 6). Section 3.1 below analyzes the quality of the approximation for a time-series of 5-dimensional covariance matrices. The results indicate that the elements of the logarithmic covariance matrices capture the dynamics of correlations and log-variances to a great extent.

## 2.2 The Dynamic Factor Model

The  $k \times k$  realized covariance matrix  $R_t$  consistently estimates the latent integrated covariance matrix  $\Sigma_t$  of the  $k$ -dimensional period- $t$  log-return vector  $\xi_t$ .<sup>2</sup> The literature on realized volatility modeling concludes that realized variances and covariances as well as their logarithmic counterparts feature a common long-memory type of dependence pattern (see e.g. Bauer and Vorkink, 2011, Chiriac and Voev, 2011, and Figures 1 and 3). Motivated by persistent common dynamics in logarithmic realized variances and covariances, I assume a persistent latent common factor structure for the  $n$  distinct elements of the matrix logarithm  $x_t = \text{vech}(\log(\Sigma_t))$ . Given the series of vectorized logarithmic

---

<sup>2</sup>See e.g. Andersen et al., 2003, Barndorff-Nielsen and Shephard, 2004, and the very general assumptions on the log-price process therein.

realized covariance matrices  $\{y_t\}_{t=1}^T$  the resulting state-space model reads

$$x_t = a + B^c f_t^c + w_t \quad (8)$$

$$y_t = x_t + u_t, \quad u_t \stackrel{iid}{\sim} N(0, \Sigma_u), \quad (9)$$

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

$$B^c = \begin{pmatrix} b_{1,1}^c & b_{1,2}^c & \dots & b_{1,p}^c \\ b_{2,1}^c & b_{2,2}^c & \dots & b_{2,p}^c \\ \vdots & & & \vdots \\ b_{n,1}^c & b_{n,2}^c & \dots & b_{n,p}^c \end{pmatrix}, \quad (10)$$

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

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

$$f_{t,j}^c = \alpha_j^c + \phi_{j,1}^c f_{t-1,j}^c + \phi_{j,2}^c \sum_{i=1}^5 f_{t-i,j}^c + \phi_{j,3}^c \sum_{i=1}^{10} f_{t-i,j}^c + \phi_{j,4}^c \sum_{i=1}^{20} f_{t-i,j}^c + \eta_{t,j}^c, \quad (11)$$

where  $\eta_{t,j}^c \sim N(0, \sigma_{c,j}^2)$  and  $j = 1, \dots, p$ . The HAR model results in a restricted AR(20) representation



for the common factor dynamics. Corsi (2009) and Audrino and Corsi (2010) find that HAR processes offer enhanced in-sample fit and out-of-sample forecasting performance in modeling log-volatilities and correlations, which are effectively approximated by the matrix logarithm. Due to the aggregation of flexible HAR dynamics for the  $p$  common factors, the factor structure is expected to accommodate a large variety of dependence patterns. Residual persistence is therefore expected to be short-lived and series-specific dynamics in  $w_t$  are assumed to be driven by AR(1) processes:

$$f_{t,j}^i = \alpha_j^i + \phi_j^i f_{t-1,j}^i + \eta_{t,j}^i, \quad \eta_{t,j}^i \sim N(0, \sigma_{i,j}^2), \quad (12)$$

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

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

An important part of factor analysis is devoted to the interpretation of the common factors. From an asset pricing perspective we expect systematic variance dynamics reflected by the volatility of the latent market portfolio appearing e.g. in the CAPM asset pricing model (see Sharpe, 1964). A respective variance factor indicates un-diversifiable market risk. Krishnan et al. (2009) and Driessen et al. (2009) analyze the pricing of market-wide time-varying diversification benefits: So-called ‘‘correlation risk’’ is captured by a market-wide correlation factor. Investors would pay a premium for assets that perform well in states of high asset correlation, since increasing correlations imply lower diversification benefits and typically increasing market volatility. Driessen et al. (2009) assume a market-wide correlation factor and observe a significant pricing of correlation risk, which furthermore removes well-known biases in option pricing models. Krishnan et al. (2009) mention the

importance of controlling for the market variance and asset-specific volatility when estimating correlation risk: If asset returns follow a one-factor model, the model-implied correlations are increasing in the asset betas and market variance and decreasing in idiosyncratic asset volatility, everything else held equal. The proposed factor model allows for investigating the presence of both systematic correlation and market risk.

### 2.3 Model Estimation and Diagnostics

Since the proposed factor model belongs to the class of linear Gaussian state space models, Maximum Likelihood estimation and Bayesian inference using Monte Carlo Markov Chain (MCMC) methods are straightforward to implement. In contrast to the ML approach, the Bayesian estimation scheme offers the advantage of avoiding high-dimensional numerical optimization of the log-likelihood function. In addition, Bayesian estimation easily accommodates nonlinear forecasting of covariance matrices within the MCMC sampling scheme. Standard Kalman filter based Maximum Likelihood estimation, in contrast, implies biased volatility forecasts due to the nonlinear matrix exponential function. I therefore apply Bayesian estimation with conjugate prior distributions for all model parameters. Forward Filtering Backward Sampling (FFBS) serves for joint full conditional sampling of the latent factors (see Kim and Nelson, 1999). Details on the (overall uninformative) prior distributions, the implementation of the Gibbs sampling algorithm and the forecasting of covariance matrices are provided in the Appendix.

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

$$e_t = \text{Var}[y_t|\mathcal{F}_{t-1}]^{-1/2} ( y_t - E[y_t|\mathcal{F}_{t-1}] ), \quad (13)$$

where  $\mathcal{F}_{t-1}$  is the information set including lagged observations up to period  $t-1$  and  $\text{Var}[y_t|\mathcal{F}_{t-1}]^{-1/2}$  denotes the inverse Cholesky factor of  $\text{Var}[y_t|\mathcal{F}_{t-1}]$ . For a correctly specified model the standardized residuals  $e_{ij,t}$  in  $e_t$  are serially and cross-sectionally uncorrelated. The modified Portmanteau test

statistic represents a standard tool for detecting significant serial and cross-correlation in the residual series of multivariate econometric models. The modified Portmanteau statistic at  $l$  lags is

$$\bar{Q}_l = T^2 \sum_{i=1}^l (T-i)^{-1} \text{tr}(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}), \quad (14)$$

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

### 3. Empirical Application

#### 3.1 Data

The proposed dynamic factor model is applied to a 5-dimensional and a 30-dimensional data set of daily realized covariance matrices of equity market returns. The underlying stocks are traded at the New York Stock Exchange (NYSE) and listed in Table 1. The daily realized covariance matrix is computed as  $R_t = \sum_{j=1}^M \xi_{t,j} \xi_{t,j}'$ , where  $\xi_{t,j}$  is the vector of returns for the  $k = 5$  or  $k = 30$  stocks computed for the  $j$ th 5-minute interval of trading day  $t$  between 9:30 a.m. and 4:00 p.m. The ordering of the assets in the vector  $\xi_{t,j}$  corresponds to the ordering of the assets in Table 1. Following Chiriac and Voev (2011) the realized (co)variance measure is further refined by averaging over 30 subsampling subgrids per day in order to exploit the data richness more efficiently and to cope with market microstructure noise. The sample period of the first data set (“Data Set 1”) starts at January 1, 2000, and ends on December 31, 2009, covering 2514 trading days including the sub-prime crisis. The data has already been studied by Golosnoy et al. (2012) and represents an updated version of the data set evaluated by Chiriac and Voev (2011). The application of the proposed factor model to 5-dimensional covariance matrices allows for an in-sample and out-of sample comparison to various

---

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

volatility models from the relevant literature, where applications to higher dimensions, say larger than 10, are practically impossible. The second data set (“Data Set 2”) extends the first by 25 additional stocks randomly selected from the S&P 100. Since intra-day price data for the additional stocks is not freely available for the whole sample period of data set 1, the second data set is restricted to the period from February 2, 2002, to May 30, 2008, covering 1564 trading days.

Figure 1 shows time series plots of the realized variances and covariances of the first data set. It reveals strong persistence and a common U-shaped pattern in the variance and covariance series. During the early 2000s, in the aftermath of the dot-com bubble, and during the recent sub-prime crisis starting in 2008 the level of the variances and covariances is significantly higher than in the middle part of the sample. Descriptive statistics are provided in Table 2. The empirical distribution of the variances and the covariances is highly skewed to the right and highly leptokurtic. The respective autocorrelation functions (ACFs) plotted in Figure 2 die out at a very slow rate indicating very strong serial correlation. Figures 3 and 4 illustrate the corresponding matrix-logarithmic time series and autocorrelation functions. Unsurprisingly, while the original series feature huge isolated volatility peaks, the logarithm greatly reduces the scale of these events letting the series appear much more homogeneous. The sample ACFs show persistent serial correlation, which is particularly pronounced for the diagonal elements of the matrix logarithm. Furthermore, the series feature distinct dynamic patterns for the diagonal and off-diagonal matrix-log elements. As discussed in Section 2.1 the diagonal elements can be interpreted as approximations to logarithmic variances while the off-diagonal elements can be interpreted as approximations to correlations. Figure 5 illustrates the quality of this approximation by comparing log-variance and correlation series with the matrix logarithm. The matrix-logarithmic series capture the dynamics of the original log-variance and correlation series to a great extent. The second data set extends the first and comprises 465 distinct series, where the according statistical properties are similar to the 15 time series of the first data set.

## 3.2 Estimation Results

### 3.2.1 Data Set 1: 5-Dimensional Covariance Matrix

Applying the Onatski criterion to data set 1 results in a maximum number of  $p = 2$  common dynamic factors. I therefore estimate the factor model from Equations (8) to (12) including one and two common HAR factors. The MCMC scheme is based on 40,000 Gibbs iterations and a burn-in of 2,000 iterations. In order to assess the numerical accuracy of the estimates, MC standard errors are calculated using a correlation consistent Parzen window based spectral estimator for the variance of the sample mean (see Kim et al., 1998). The ratio of MC standard error to posterior standard deviation addresses the proportion of variation in the estimates due to simulation relative to the variation induced by the data. All numerical standard errors are within the range of 0.006-7% of the posterior standard deviations, indicating an acceptable balance of numerical and statistical uncertainty (see e.g. Kim et al., 1998). The chosen prior distributions are overall uninformative and given in the Appendix. The parameter estimates indicate weak stationarity of the data generating processes for all considered models.

I now turn to the in-sample analysis. In order to enable a comparison to competing models from the relevant literature, I also estimate the MIDAS-CAW(3,3) model analyzed by Golosnoy et al. (2012)<sup>4</sup> and the DCC-CAW model of Bauwens and Storti (2011). The models are illustrated in Section 3.3.1 below.

Table 3 shows Portmanteau diagnostic test results for the estimated Pearson residual series. Note that the diagnostics for the factor models are based on residual series obtained for logarithmic (co)variance data, while the test results for the CAW models are based on residual series obtained for the original (co)variance data. Hence the results cannot be compared directly, but indicate how the models fit the dynamics of the respective original or logarithmic time series. Due to the rich serial and cross-sectional dynamics of the  $n = 15$  (co)variance series the diagnostic tests indicate significant residual predictability for all considered models. Yet a comparison to the obtained test statistics for the raw data shows that all models successfully account for a major portion of the highly persistent

---

<sup>4</sup>In the empirical analysis provided by Golosnoy et al. (2012) the MIDAS-CAW(3,3) model was found to offer the overall best in-sample fit within the range of considered CAW specifications.

(co)variance dynamics. The diagnostics for the 2-factor model indicate significant improvements compared to both competing models. This finding is particularly remarkable in relation to the flexible MIDAS-CAW approach. The dynamics implied by the MIDAS-CAW(3,3) model are driven by 168 (often insignificant) parameters as opposed to 82 parameters for the 2-factor model<sup>5</sup>. The previous findings are confirmed by additional F-test results for residual predictability presented in Table 4. Considering the 1% significance level and 50 lags of daily (logarithmic) realized (co)variances, the two-factor model successfully accounts for the predictability of twelve out of 15 logarithmic (co)variance series as opposed to five (co)variance series for the MIDAS-CAW and four (co)variance series for the DCC-CAW model. Figure 6 shows sample autocorrelation functions of the 15 residual series obtained for the fitted 2-factor model. The plots indicate that the model dramatically reduces the serial correlation in the raw data (compare to Figure 4). Summarizing the model diagnostic results, a parsimoniously parameterized factor model with two common factors and idiosyncratic dynamics offers a good fit to the complex serial and cross-sectional dynamics of the underlying logarithmic (co)variance data. Compared to the MIDAS-CAW approach the main source of parsimony is the reduction of cross-sectional dependence to loadings on a few persistent common factors.

Motivated by the model diagnostic results the subsequent analysis focuses on the 2-factor specification. Table 5 shows the parameter estimates. All estimated loadings are significantly different from zero. The estimates of the HAR parameters imply significance of the first, second and fourth HAR component. The estimated characteristic roots of the HAR-implied AR(20) processes are given by .9942 and .9893, respectively, and imply weak stationarity though strong persistence. The estimated AR(1) coefficients of the idiosyncratic factors reveal strong series specific serial correlation. Figure 7 shows bar plots of the fraction of total variance explained by the factors. Besides indicating the importance of idiosyncratic dynamics the figure shows that the first common factor is mainly associated with the diagonal elements of  $Y_t$ , while the second common factor appears to be almost exclusively driving off-diagonal dynamics. Referring to the properties of the matrix logarithm discussed in Section 2.1 this finding allows for the interpretation of common factors as market risk and correlation risk factor, in line with the asset pricing literature (see Section 2.2). Filtered estimates

---

<sup>5</sup>The Schwarz-preferred MIDAS-CAW(2,2) model still comprises 118 parameters and offers slightly worse model diagnostic results compared to the MIDAS-CAW(3,3) model.

are depicted in Figure 8 and confirm this interpretation. Comparing the second plot of Figure 8 with the time series in Figure 3 shows that the dynamics of the correlation factor mainly capture the persistently high correlation level in 2003 to 2007 when market volatility was comparably low. The second bar-plot in Figure 7 shows that the market risk factor has explanatory power for the correlation series, which are approximated by the off-diagonal elements in  $Y_t$ . This result corresponds to the common finding that high market volatility tends to be accompanied by strong correlation (see e.g. Solnik et al., 1996). Figure 9 depicts filtered estimates of the idiosyncratic factors. The time-series plots show that series-specific dynamics are mainly caused by the recent sub-prime crisis resulting in strong distinct reactions of a few logarithmic (co)variance series in 2007 to 2009 letting the dynamics of the corresponding idiosyncratic factors appear non-stationary. This finding indicates potential crisis related breaks in the idiosyncratic volatility and correlation structure which would motivate further research e.g. addressing Markov Switching regimes in volatility/correlation levels (see e.g. Lopes and Carvalho, 2007), which are beyond the scope of this paper.

### 3.2.2 Data Set 2: 30-Dimensional Covariance Matrix

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

Table 3 shows model diagnostic results. Since (MIDAS-)CAW models are generally not tractable for more than ten assets, the diagnostics are limited to the factor models and the DCC-CAW approach. The DCC-CAW results indicate significant residual predictability at any conventional significance level. The factor model residuals, in contrast, pass the test of the Null of no serial and cross-correlation at the 1% significance level for 75 as well as 100 lags. The large cross-sectional dimension precludes further testing for predictability of single residual series. Comparing the Portmanteau diagnostic test results for the factor model residuals obtained for data set 1 to the respective

test results obtained for data set 2 reveals a better model fit to the observed logarithmic (co)variance dynamics in case of data set 2. Since the second data set does not cover the full extent of the sub-prime crisis this finding indicates a general problem of fitting the complex (co)variance dynamics in this particular period.

Motivated by the model diagnostic results the subsequent analysis focuses on the 3-factor model. Table 6 shows estimates of the HAR parameters, which imply significance of the first, second and fourth HAR component. The characteristic roots are given by .9906, .9723 and .9981 indicating weak stationarity and high persistence of the joint (co)variance dynamics. Figure 10 depicts the fraction of total variance of the  $\{y_t\}_{t=1}^T$  series explained by the factors. The first row of plots in Figure 10 illustrates the importance of idiosyncratic variation for the series of (approximate) logarithmic variances with e.g. 30% explained variation for the American Express volatility and 26% explained variation for the volatility of the Dell stock. Significant explanatory power is also found for idiosyncratic factors of particular off-diagonal matrix-log elements (up to 22% explained variation), where the respective series mostly refer to approximate correlations involving Citigroup, JP Morgan Chase & Co., Intel, Dell and Microsoft, which have all been particularly affected by the sub-prime crisis inducing partly idiosyncratic volatility and correlation dynamics.

The particularly high fraction of variation of the diagonal matrix-log elements explained by the first common factor (30-80%, see Figure 10, second row of plots) motivates the interpretation as market risk factor. As already observed for data set 1, the market risk factor also captures significant variation of particular off-diagonal matrix-log series (up to 12%). The explanatory power of the second common factor is overall limited (up to 3% of total variation), but it appears to be mainly attributed to the off-diagonal elements in  $Y_t$  (compare the two plots in the third row of Figure 10). The factor is therefore interpreted as correlation risk factor. Figure 11 shows filtered factor estimates which confirm the previous factor interpretations (compare to Figure 8). In addition, the figure sheds light on the role of the third common factor, which appears to cover joint dynamics specifically linked to the sub-prime crisis inducing pronounced volatility and correlation peaks. Particularly high fractions of total variation explained by the third common factor are found for approximate log-variances of Citigroup (8%), JP Morgan Chase & Co. (7%), Intel (5%), and Microsoft (5%) and



corresponding off-diagonal matrix-log elements (5-25% explained variation). As mentioned above, these stocks have been particularly affected by the sub-prime crisis.

### 3.3 Forecasting Results

I now compare the 1-day ahead forecasting performance of the dynamic factor model with alternative forecasting models from the relevant literature. Forecasts are denoted by  $\hat{R}_{t+1} = E(R_{t+1}|\mathcal{F}_t)$ . In addition to a statistical evaluation of the models' forecasting capabilities based on a root mean squared error (RMSE) criterion I follow Chiriac and Voev (2011) in also addressing potential economic benefits associated with accurate volatility forecasts. This is accomplished via evaluating the performance of portfolio optimization strategies based on volatility forecasts. For the first data set I follow the lines of Golosnoy et al. (2012), and select two out-of-sample windows: The first window is selected to be prior to the recent sub-prime crisis and covers the period from July 2, 2007 through June 30, 2008, with relatively low volatility (see the dark-gray shaded areas in Figures 1 and 3). The second window starts at July 1, 2008 and ends June 30, 2009 (see the light-gray shaded areas in Figures 1 and 3). The window covers the sub-prime crisis featuring a very high volatility level. The second data set ends at May 30, 2008, and allows for a slightly truncated version of the first forecasting window. All models are re-estimated daily and new forecasts are generated based on the updated parameter estimates. The set of competing models is given by the CAW, MIDAS-CAW, HAR-CAW, DCC-CAW, BEKK-GARCH, DCC-GARCH, and the EWMA approach, which are illustrated below.

#### 3.3.1 Competing Models

The baseline CAW( $p, q$ ) model proposed by Golosnoy et al. (2012) assumes a conditional Wishart distribution for the covariance matrix  $R_t$ ,

$$R_t|\mathcal{F}_{t-1} \sim \mathcal{W}_k(\nu, S_t/\nu), \tag{15}$$

where  $\mathcal{W}_k$  denotes the law of a  $k$ -dimensional Wishart distribution and  $\nu > k$  is the scalar degree of freedom.  $S_t/\nu$  is the  $k \times k$  symmetric, positive definite scale matrix, such that the conditional mean

is given by  $E(R_t|\mathcal{F}_{t-1}) = S_t$ , which is assumed to follow a linear recursion of order  $(p,q)$

$$S_t = CC' + \sum_{i=1}^p B_i S_{t-i} B_i' + \sum_{j=1}^q A_j R_{t-j} A_j', \quad (16)$$

where  $C$  is a  $k \times k$  lower-triangular matrix and  $A_j, B_i$  are  $k \times k$  parameter matrices. Golosnoy et al. (2012) additionally propose Mixed Data Sampling (MIDAS)-CAW and HAR-CAW specifications, which are specifically designed to capture the long-run movements of the (co)variances. The MIDAS-CAW model decomposes volatility into a secular component  $M_t$  and a mean-reverting short-run component  $S_t^*$ :

$$E(R_t|\mathcal{F}_{t-1}) = S_t = C_t S_t^* C_t', \quad \text{with} \quad M_t = C_t C_t', \quad (17)$$

where  $C_t$  is the lower-triangular Cholesky factor of the secular component  $M_t$ . The short-run component  $S_t^*$  is then assumed to follow a covariance-stationary CAW( $p, q$ ) process with  $E(S_t^*) = I_k$ . The long-run component  $M_t$  is specified as an extension of the MIDAS polynomial proposed by Engle et al. (2009), which aggregates weighted realized covariance matrices computed over  $L$  lags of  $m$ -period realized (co)variances, where the authors select  $L = 12$  and  $m = 20$ . The HAR-CAW model assumes the following specification for the scale matrix  $S_t$ :

$$S_t = CC' + AR_{t-1}A' + A^{(w)} \bar{R}_{t-1}^{(w)} A^{(w)'} + A^{(bw)} \bar{R}_{t-1}^{(bw)} A^{(bw)'} + A^{(m)} \bar{R}_{t-1}^{(m)} A^{(m)'}, \quad (18)$$

with  $\bar{R}_{t-1}^{(x)}$  denoting the realized covariance computed over a time window  $x = \{w, bw, m\}$ , where  $w$  stands for the weekly (5 days),  $bw$  for the biweekly (10 days), and  $m$  for the monthly (20 days) horizon.  $A$  and  $A^{(x)}$  are  $k \times k$  parameter matrices. Although the autoregressive parameter matrices can be restricted to the diagonal case, the illustrated CAW specifications are hardly applicable to more than ten assets (see Golosnoy et al., 2012). Diagonal CAW models completely exclude cross-sectional (co)variance dynamics.

I furthermore consider two multivariate GARCH approaches for asset returns, which are frequently applied in the literature. The BEKK-GARCH( $p,q$ ) model of Engle and Kroner (1995) for a vector

of  $k$  daily stock returns, denoted by  $\xi_t$ , assumes  $\xi_t = \Sigma_t^{1/2} v_t$ , where  $v_t \sim \mathcal{N}(0, I_k)$  and  $\Sigma_t^{1/2}$  is the Cholesky factor of the conditional covariance matrix  $\Sigma_t$ , which is then specified as

$$\Sigma_t = D_0 D_0' + \sum_{i=1}^p D_i \Sigma_{t-i} D_i' + \sum_{j=1}^q G_j [\xi_{t-j} \xi_{t-j}'] G_j', \quad (19)$$

where  $D_0$  is a lower triangular  $k \times k$  matrix and  $D_i, G_j$  are  $k \times k$  matrices. Similar to the CAW models, empirical applications of the BEKK-GARCH( $p, q$ ) model and respective diagonal specifications are restricted to lower dimensional covariance matrices.

The DCC-GARCH model of Engle (2002) constitutes a multivariate GARCH approach applicable to high-dimensional covariance matrices. The DCC-GARCH( $p, q$ ) model assumes conditional normality for the return vector  $\xi_t$  and GARCH( $p, q$ ) dynamics for the conditional variances  $\{\sigma_{ii,t}^2\}_{i=1}^k$ . The modeling of dynamic conditional correlations is based on the decomposition

$$\Sigma_t = D_t P_t D_t, \quad (20)$$

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

$$P_t = (\text{diag}(Q_t))^{-\frac{1}{2}} Q_t (\text{diag}(Q_t))^{\frac{1}{2}}, \quad (21)$$

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

$$Q_t = (1 - \alpha - \beta) \bar{Q} + \alpha u_{t-1} u_{t-1}' + \beta Q_{t-1}, \quad (22)$$

where  $\alpha$  and  $\beta$  are positive scalar parameters and  $u_t$  is the  $k$ -dimensional vector of standardized residuals with elements

$$u_{i,t} = \frac{\xi_{i,t}}{\sigma_{ii,t}}, \quad i = 1, \dots, k. \quad (23)$$

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

Recently Bauwens and Storti (2011) proposed a conditional autoregressive Wishart (CAW) model

featuring DCC dynamics, which can be readily applied to the forecasting of high-dimensional covariance matrices. The model assumes a conditional Wishart distribution for the realized covariance matrix  $R_t$  and decomposes the scale matrix  $S_t$  as

$$S_t = D_t P_t D_t, \quad (24)$$

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

$$P_t = (1 - \alpha - \beta)\bar{R} + \alpha Z_{t-1} + \beta P_{t-1}, \quad (25)$$

where  $Z_t$  denotes the realized correlation matrix at time  $t$ ,

$$Z_t = (\text{diag}(R_t))^{-\frac{1}{2}} R_t (\text{diag}(R_t))^{-\frac{1}{2}}, \quad (26)$$

and  $\alpha$  and  $\beta$  are positive scalar parameters.  $\bar{R}$  denotes the unconditional correlation matrix which is consistently estimated by the sample mean of realized correlation matrices. Bauwens and Storti (2011) assume independent HAR dynamics for the conditional variances  $\{s_{ii,t}\}_{i=1}^k$ .

The Exponentially Weighted Moving Average (EWMA) approach finally represents a simple forecasting model being applicable even for high-dimensional realized covariance matrices. The EWMA model, which is often used in risk management systems like RiskMetrics (see J.P. Morgan, 1996) to forecast variances and covariances, is given by

$$E(R_t | \mathcal{F}_{t-1}) = (1 - \lambda)R_{t-1} + \lambda E(R_{t-1} | \mathcal{F}_{t-2}), \quad (27)$$

where I set  $\lambda$  to its typical value given by 0.94.

Detailed discussions on obtaining forecasts given the volatility models illustrated above are provided by Golosnoy et al. (2012) and Chiriac and Voev (2011).

### 3.3.2 Statistical Evaluation

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

$$FN = \frac{1}{T^{\text{fore}}} \sum_t \|R_{t+1} - \hat{R}_{t+1}\| = \frac{1}{T^{\text{fore}}} \sum_t \left[ \sum_{i,j} (r_{ij,t+1} - \hat{r}_{ij,t+1})^2 \right]^{1/2}, \quad (28)$$

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

Table 7 shows the forecasting results for data set 1 and Table 8 presents those for data set 2. In forecasting the 5-dimensional covariance matrices of data set 1 the 2-factor model outperforms the competing models in terms of forecasting precision prior to the subprime crisis as well as in the crisis period. For data set 2 the set of competing models is substantially reduced since only the EWMA approach, the DCC-GARCH and the DCC-CAW model can be successfully applied to high-dimensional (co)variance forecasting. The lowest average Frobenius norm is obtained for the DCC-CAW approach closely followed by the 3-factor model. Note that the forecasting results for models based on daily asset return data are overall clearly inferior to the forecasting results obtained by volatility models using realized covariance matrices.

### 3.3.3 Economic Evaluation

In order to assess the economic value of the obtained volatility forecasts I follow Chiriac and Voev (2011) in constructing portfolios which maximize the utility of a risk-averse investor. Here I assume a second degree polynomial utility function and/or a conditional return distribution which is completely characterized by its first two moments (e.g. the normal distribution). The investor's portfolio optimization problem then reduces to the minimization of portfolio volatility via selecting the according asset weights while fixing a given expected return (Markowitz, 1952).

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

problem

$$\hat{\omega}_{t+1} = \arg \min_{\omega_{t+1}} \omega'_{t+1} \hat{R}_{t+1} \omega_{t+1} \quad \text{s.t.} \quad \omega'_{t+1} E(\xi_{t+1}|\mathcal{F}_t) = \mu_p \quad \text{and} \quad \omega'_{t+1} \iota = 1, \quad (29)$$

where  $\omega_{t+1}$  is the  $k \times 1$  vector of portfolio weights chosen at  $t$  and held until  $t+1$ ,  $\iota$  is an  $k \times 1$  vector of ones, and  $\mu_p$  is the target return.  $\xi_{t+1}$  denotes the 1-day ahead asset return vector. I assume serially uncorrelated daily asset returns which is typically met in practice and set  $E(\xi_{t+1}|\mathcal{F}_t) \stackrel{!}{=} \mu$ , where  $\mu$  is approximated by the sample mean of returns. In order to assess the predictive accuracy of the considered models I compare the ex-post realization of the conditional portfolio mean and standard deviation. I therefore solve the minimization problem of Equation (19) resulting in an optimal weight vector  $\hat{\omega}_{t+1}$  for each model and compute  $\xi_{t+1}^p = \hat{\omega}'_{t+1} \xi_{t+1}$  and  $\sigma_{t+1}^p = \sqrt{\hat{\omega}'_{t+1} R_{t+1} \hat{\omega}_{t+1}}$  for  $t = T^*, T^* + 1, T^* + 2, \dots$ , where  $T^*$  denotes the number of in-sample observations. Solving the optimal portfolio problem for various levels of the target portfolio return  $\mu_p$  results in a predicted efficiency frontier, which characterizes the best mean-variance trade-off achievable by using a particular forecasting model. A suitable benchmark scenario is obtained by constructing the efficiency frontier using the ideal forecast  $\hat{R}_{t+1} = R_{t+1}$ .

Figures 12 to 14 show the obtained efficiency frontiers for the two data sets and forecasting windows averaged over the respective forecasting periods. As expected, the results show a by far lower achievable portfolio variance for a given portfolio mean in case of the ideal benchmark forecasts. Among the considered forecasting models the factor approach shows overall remarkably good ex-post mean-variance tradeoffs. The EWMA and the DCC-CAW models are nevertheless strong competitors. Selecting the global minimum variance portfolio as a natural reference point, the factor models stay unmatched in the rather calm forecasting phase I of data set 1 but are marginally outperformed in the turbulent phase II of data set 1 and the calm phase of data set 2. The results show considerable gains by the direct modeling of realized (co)variances opposed to daily return data based GARCH models where forecast-based ex post mean-variance tradeoffs are overall strictly inferior.

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

evidence in favor of the factor model approach illustrated in this chapter.

## 4. Conclusions

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

The empirical application to 5- and 30-dimensional realized covariance matrices of NYSE-traded stocks shows that the factor model successfully accounts for the observed dynamic behavior of up to 465 (co)variance series, where 2 to 3 common factors appear overall sufficient in order to drive the cross-sectional dynamics. This finding implies significant dimension reduction in multivariate volatility modeling without substantially affecting in-sample fit. A comprehensive out-of-sample forecasting experiment based on statistical and economical evaluation criteria finally shows that the factor model specifications outperform a range of prominent forecasting models from the relevant literature.

## Acknowledgements

I thank Giuseppe Storti and Valeri Voev for providing the data used in this paper.



## Appendix: Gibbs Sampling Algorithm

I now illustrate the Gibbs sampling algorithm for obtaining simulated Bayesian point estimates of the parameters of the factor model presented in Section 2.2. According to Bayes' theorem the full conditional distribution of each sub-vector of the model's augmented parameter vector  $\theta^{aug} = (a', \text{vec}(B^c)', \text{diag}(B^i)', \text{diag}(\Sigma_u)', \phi', f')$  is proportional to the product of the likelihood function and the sub-vector's joint prior distribution. Here the vector  $\phi$  embraces the factors' AR coefficients and  $f$  summarizes all  $q = p + n$  factors for all time periods. I assume that the joint prior distribution can be factorized into the product of marginal prior distributions. Given an initialization for the parameter vector  $\theta^{aug}$  the Gibbs sampling algorithm now consists in iterative simulation of parameter values from the full conditional distributions of sub-vectors of  $\theta^{aug}$ . The algorithm thereby generates a Markov chain which converges to the posterior of the model parameters<sup>6</sup>. The Gibbs draws conducted until convergence (so-called burn-in phase) are discarded and only the remaining draws are used for estimation purposes. Posterior mean and Posterior standard deviation are then approximated by the respective sample moments. I now state the applied full conditional distributions.

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

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

$$\mu_i = \Sigma_i \left( \Sigma_{0,i}^{-1} \mu_{0,i} + \widehat{\Sigma}_{i,\text{ols}}^{-1} \widehat{\beta}_{i,\text{ols}} \right), \quad \Sigma_i = \left( \Sigma_{0,i}^{-1} + \widehat{\Sigma}_{i,\text{ols}}^{-1} \right)^{-1}, \quad (30)$$

where  $\widehat{\beta}_{i,\text{ols}} = (X_i'X_i)^{-1}X_i'y_i$  and  $\widehat{\Sigma}_{i,\text{ols}} = \sigma_{u,i}^2(X_i'X_i)^{-1}$  denote the ordinary least squares estimates of  $\beta_i$  and the residual variance. In the empirical application of Section 3.2 I chose the following

---

<sup>6</sup>For details on the Gibbs sampling algorithm and Monte Carlo Markov Chain methods see e.g. Bauwens et al. (1999).

hyper-parameters:  $\mu_{0,i} = 0.2$  and  $\Sigma_{0,i} = 0.04 \times I_{q+1}$ , implying a prior standard-deviation of 0.2.

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

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

$$\gamma_i = \left(\frac{T}{2} + \gamma_{0,i}\right), \quad \delta_i = \left(\frac{u_i' u_i}{2} + \delta_{0,i}^{-1}\right)^{-1}. \quad (31)$$

In the empirical application of Section 3.2 I chose the following hyper-parameters:  $\gamma_{0,i} = 2.04$  and  $\delta_{0,i} = 4.81$ , implying a prior mean of 0.2 and a prior standard-deviation of 1.

**Full conditional sampling of  $\phi$ :**

Given the factors  $f$  and assuming conjugate Gaussian priors, the factor persistencies are sampled analogously to  $\beta_i$ . In the empirical application of Section 3.2 I chose the following hyper-parameters:  $\mu_{0,i} = 0.5 \times \iota_{\tilde{q}}$  and  $\Sigma_{0,i} = I_{\tilde{q}}$ , where  $\tilde{q} = 1$  for idiosyncratic factors and  $\tilde{q} = 4$  for common HAR factors;  $i \in \{1, \dots, q\}$ .  $\iota_{\tilde{q}}$  denotes a  $\tilde{q}$ -dimensional column vector of ones.

**Full conditional distribution of the factors  $f$ :**

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

$$y_t = a + Zs_t + u_t, \quad u_t \stackrel{iid}{\sim} N(0, \Sigma_u) \quad (32)$$

$$s_t = Hs_{t-1} + R\eta_t, \quad \eta_t \stackrel{iid}{\sim} N(0, I_q), \quad (33)$$

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

period  $t$ . Hence  $m = n + 20p$ . Accordingly

$$Z = \begin{pmatrix} B, 0_{n \times 19p} \end{pmatrix}, \quad H = \begin{pmatrix} \text{diag}(\phi_1^i, \dots, \phi_n^i) & 0 & 0 & 0 & \dots & 0 & 0 \\ & \Phi_1 & \Phi_2 & \Phi_3 & \dots & \Phi_{19} & \Phi_{20} \\ 0_{20p \times n} & I_p & 0 & 0 & \dots & 0 & 0 \\ & 0 & I_p & 0 & \dots & 0 & 0 \\ & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ & 0 & 0 & 0 & \dots & I_p & 0 \end{pmatrix}, \quad (34)$$

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

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

$$\begin{aligned} P(\underline{s}_T | \underline{y}_T) &= P(s_T | \underline{y}_T) \times P(s_{T-1} | s_T, \underline{y}_T) \times P(s_{T-2} | s_{T-1}, s_T, \underline{y}_T) \\ &\quad \times P(s_{T-3} | s_{T-2}, s_{T-1}, s_T, \underline{y}_T) \times \dots \times P(s_1 | s_2, s_3, \dots, s_T, \underline{y}_T) \\ &= P(s_T | \underline{y}_T) \times P(s_{T-1} | s_T, \underline{y}_{T-1}) \times P(s_{T-2} | s_{T-1}, \underline{y}_{T-2}) \\ &\quad \times P(s_{T-3} | s_{T-2}, \underline{y}_{T-3}) \times \dots \times P(s_1 | s_2, y_1) \\ &= P(s_T | \underline{y}_T) \times \prod_{t=1}^{T-1} P(s_t | s_{t+1}, \underline{y}_t), \end{aligned} \quad (35)$$

where I omit dependence on the parameter vector for the sake of readability. The second step of the derivations in Eq. (25) is due to the state-space framework of Eqs (22) and (23) implying a Markov property for  $s_t$  and no additional information beyond  $s_{t+1}$  and  $\underline{y}_t$  relevant for predicting  $s_t$ . The decomposition of the joint full conditional density of  $\underline{s}_T$  in Eq. (25) implies that the whole  $\underline{s}_T$  sequence can be drawn jointly via recursive sampling from  $P(s_t | s_{t+1}, \underline{y}_t)$ . Note that the state-space model includes state-equations, which are identities. This results in a conditional variance of  $s_t$  given  $s_{t-1}$  being not positive definite. Therefore only the first  $q$  elements of  $s_t$  denoted by  $s_t^*$  can be conditioning factors in the full conditional sampling of  $\underline{s}_T$  and hence  $P(\underline{s}_T | \underline{y}_T) = P(s_T | \underline{y}_T) \times \prod_{t=1}^{T-1} P(s_t | s_{t+1}^*, \underline{y}_t)$

(see Kim and Nelson, 1999, p. 194 ff.). Due to the model's linear Gaussian nature we directly obtain a conditional normal distribution for  $s_t$  given  $s_{t+1}^*$  and  $\underline{y}_t$  with mean and variance given by

$$s_{t|t, s_{t+1}^*} = s_{t|t} + P_{t|t} H^{*'} (H^* P_{t|t} H^{*'} + I_q)^{-1} (s_{t+1}^* - H^* s_{t|t}) \quad (36)$$

$$P_{t|t, s_{t+1}^*} = P_{t|t} - P_{t|t} H^{*'} (H^* P_{t|t} H^{*'} + I_q)^{-1} H^* P_{t|t}, \quad (37)$$

where  $H^*$  comprises the first  $q$  rows of  $H$  and I refer to standard Kalman filtering notation in denoting  $s_{t|t} = E[s_t | \underline{y}_t]$  and  $P_{t|t} = \text{Var}[s_t | \underline{y}_t]$ . Both  $s_{t|t}$  and  $P_{t|t}$  are readily available from the Kalman filter algorithm  $\forall t$ . The prior derivations imply that the whole  $\underline{s}_T$  sequence can be drawn jointly full conditional via recursive sampling from conditional normal distributions with moments given in Eqs. (26) and (27).

### Forecasting:

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

## References

- [1] Andersen, T.G., Bollerslev, T., Diebold, F.X., and P. Labys. 2003. Modeling and forecasting realised volatility. *Econometrica* 71: 579-625.
- [2] Asai, M., McAleer, M., and J. Yu. 2006. Multivariate stochastic volatility: A review. *Econometric Reviews* 25: 145-175.
- [3] Audrino, F., and F. Corsi. 2010. Modeling tick-by-tick realized correlations. *Computational Statistics and Data Analysis* 54: 2372-2382.
- [4] Barndorff-Nielsen, O.E., and N. Shephard. 2004. Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica* 72: 885-925.
- [5] Bauer, G.H., and K. Vorkink. 2011. Forecasting multivariate realized stock market volatility. *Journal of Econometrics* 160: 93-101.
- [6] Bauwens, L., Laurent, S., and J.V.K. Rombouts. 2006. Multivariate GARCH models: A survey. *Journal of Applied Econometrics* 21: 79-109.
- [7] Bauwens, L., Lubrano, M., and J.F. Richard. 1999. *Bayesian inference in dynamic econometric models*, Oxford University press: New York.
- [8] Bauwens, L., and G. Storti. 2011. Modelling vast dimensional realized covariance matrices. Working paper.
- [9] Chib, S., Nardari, F., and N. Shephard. 2006. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics* 134: 341-371.
- [10] Lopes, H.F., and C.M. Carvalho. 2007. Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference* 137: 3082-3091.
- [11] Chiriac, R., and V. Voev. 2011. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics* 26: 922-947.
- [12] Chiu, T.Y.M., Leonard, T., and K.W. Tsui. 1996. The matrix-logarithmic covariance model. *Journal of the American Statistical Association* 91: 198-210.
- [13] Corsi, F. 2009. A simple approximative long-memory model of realized volatility. *Journal of Financial Econometrics* 7: 174-196.
- [14] Driessen, J., Maenhout, P.J., and G. Vilkov. 2009. The price of correlation risk: evidence from equity options. *The Journal of Finance* 64: 1377-1406.
- [15] Engle, R.F. 2002. Dynamic conditional correlation: A simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20: 339-350.
- [16] Engle, R.F., Ghysels, E., and B. Sohn. 2009. Stock market volatility and macroeconomic fundamentals. Working paper.

- [17] Engle, R.F., and K.F. Kroner. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* 11: 122-150.
- [18] Geweke, J., and Zhou, G. 1996. Measuring the price of the arbitrage pricing theory. *Review of Financial Studies* 9: 557-587.
- [19] Golosnoy, V., Gribisch, B., and R. Liesenfeld. 2012. The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics* 167: 211-223.
- [20] Gouriéroux, C., Jasiak, J., and R. Sufana. 2009. The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics* 150: 167-181.
- [21] J.P. Morgan. 1996. RiskMetrics, Technical Document (fourth ed.): New York.
- [22] Jin, X., and J.M. Maheu. 2011. Modeling realized covariances and returns. Working paper.
- [23] Kim, C.J., and C.R. Nelson. 1999. State-space models with regime switching: classical and gibbs sampling approaches with applications. The MIT Press.
- [24] Kim, S., Shephard, N., and S. Chib. 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65: 361-393.
- [25] Krishnan, C.N.V., Petkova, R., and P. Ritchken. 2009. Correlation risk. *Journal of Empirical Finance* 16: 353-367.
- [26] Ledoit, O., Santa-Clara, P., and M. Wolf. 2003. Flexible multivariate GARCH modeling with an application to international stock markets. *Review of Economics and Statistics* 85: 735-747.
- [27] Lütkepohl, H. 2005. *New Introduction to Multiple Time Series Analysis*, Springer: Berlin.
- [28] Markowitz, H. 1952. Portfolio selection. *The Journal of Finance* 7: 77-91.
- [29] Noureldin, D., Shephard, N., and K. Sheppard. 2011. Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, in press.
- [30] Onatski, A. 2010. Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics* 92: 1004-1016.
- [31] Sharpe, F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19: 425-442.
- [32] Solnik, B., Bourcelle, C., and Y. Le Fur. 1996. International market correlation and volatility. *Financial Analyst Journal* 52: 17-34.

Table 1. NYSE Traded Stocks of Data Set 1 and Data Set 2.

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

\*: Stock belongs to data set 1.

Table 2. Descriptive statistics.

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

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



Table 3. Portmanteau Diagnostic Test Results.

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

\* indicates significance at the 1% level.

Table 4. *P-values of Residual F-tests.*

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

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

Table 5. Estimation Results: 2 Factor Model, Data Set 1.

Series	$y_{11}$	$y_{21}$	$y_{31}$	$y_{41}$	$y_{51}$	$y_{22}$	$y_{32}$	$y_{42}$	$y_{52}$	$y_{33}$	$y_{43}$	$y_{53}$	$y_{44}$	$y_{54}$	$y_{55}$
$B^i$	.2414	.0186	.0183	.0081	.0271	.1233	.0097	.0090	.0108	.0831	.0285	.0196	.1150	.0116	.1084
$B_1^c$	.2909	.0183	.0165	.0137	.0087	.3053	.0190	.0124	.0085	.2700	.0174	.0111	.2038	.0130	.2126
$B_2^c$	0	.0438	.0404	.0340	.0334	.0338	.0415	.0311	.0352	.0385	.0370	.0354	.0510	.0286	.0533
$\sigma_u^2$	.0605	.0142	.0128	.0142	.0122	.0733	.0125	.0133	.0118	.1015	.0116	.0127	.1058	.0139	.0959
$f^i$	$\phi_1^i$	$\phi_2^i$	$\phi_3^i$	$\phi_4^i$	$\phi_5^i$	$\phi_6^i$	$\phi_7^i$	$\phi_8^i$	$\phi_9^i$	$\phi_{10}^i$	$\phi_{11}^i$	$\phi_{12}^i$	$\phi_{13}^i$	$\phi_{14}^i$	$\phi_{15}^i$
$f_1^c$	.4399	.9879	.9062	.9876	.8009	.9585	.9888	.9916	.9885	.9740	.7628	.9218	.9576	.9496	.9793
	$\phi_{1,1}^c$	$\phi_{1,2}^c$	$\phi_{1,3}^c$	$\phi_{1,4}^c$	$\psi_1$										
	.6677	.0410	.0045*	.0034*	.9942										
$f_2^c$	$\phi_{2,1}^c$	$\phi_{2,2}^c$	$\phi_{2,3}^c$	$\phi_{2,4}^c$	$\psi_2$										
	.6151	.0250	.0006*	.0108	.9893										

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

Table 6. Estimation Results: 3 Factor Model, Data Set 2.

$f_1^c$	$\phi_{1,1}^c$	$\phi_{1,2}^c$	$\phi_{1,3}^c$	$\phi_{1,4}^c$	$\psi_1$	$f_2^c$	$\phi_{2,1}^c$	$\phi_{2,2}^c$	$\phi_{2,3}^c$	$\phi_{2,4}^c$	$\psi_2$
	.6152	.0433	.0080*	.0033*	.9906		.3043	.0575	.0189*	.0057*	.9723
$f_3^c$	$\phi_{3,1}^c$	$\phi_{3,2}^c$	$\phi_{3,3}^c$	$\phi_{3,4}^c$	$\psi_3$						
	.1967	.0228*	.0292*	.0195	.9981						

\*: 95% posterior confidence region includes the null. Number of Gibbs sequences: 40,000; Burn-in: 10,000.  $\psi_i$ : characteristic root of the  $i$ 'th common factor's restricted AR(20) process (HAR).

Table 7. Statistical Evaluation of Forecasting Accuracy: Data Set 1.

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

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

*Table 8. Statistical Evaluation of Forecasting Accuracy: Data Set 2, Phase 1.*

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

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

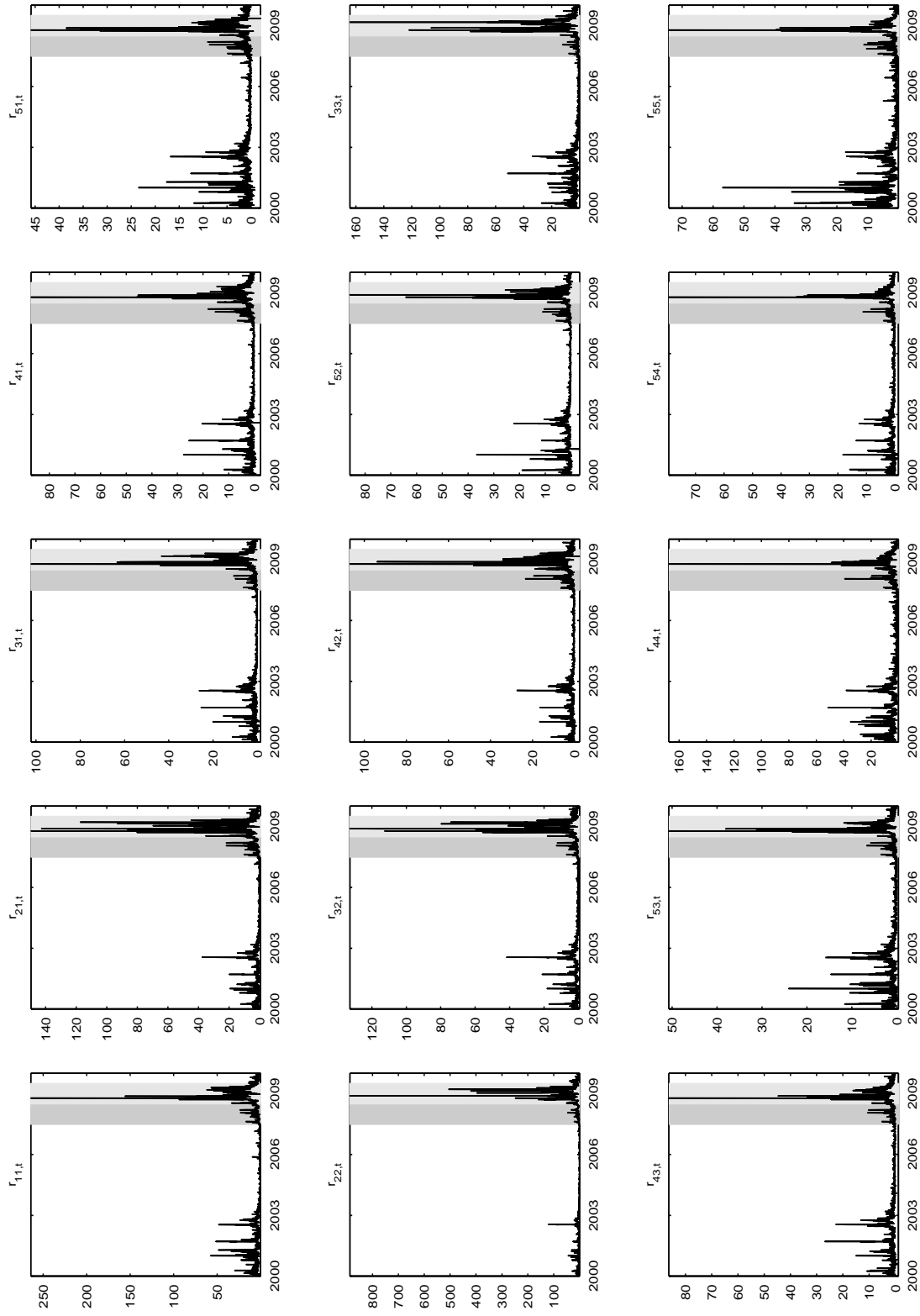


Figure 1. Time series of daily realized variances and covariances  $r_{ij,t}$  for *AXP* ( $i = 1$ ), *C* ( $i = 2$ ), *GE* ( $i = 3$ ), *HD* ( $i = 4$ ), and *IBM* ( $i = 5$ ) stock, data set 1; the gray shaded areas mark the two out-of-sample windows used in the forecast experiment.

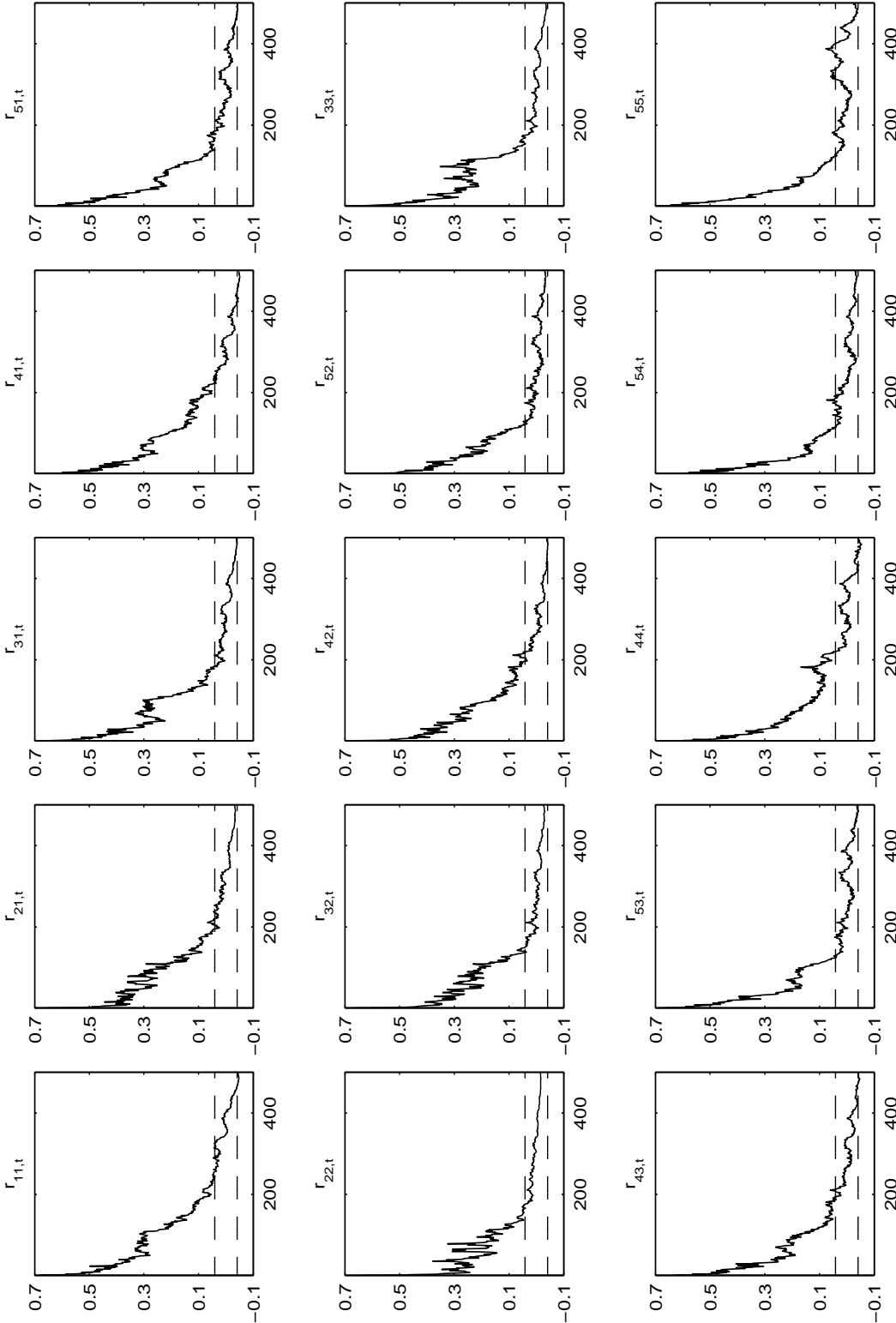


Figure 2. Sample autocorrelation function of daily realized variances and covariances  $r_{ij,t}$  for AXP ( $i = 1$ ), C ( $i = 2$ ), GE ( $i = 3$ ), HD ( $i = 4$ ), and IBM ( $i = 5$ ) stock, data set 1; the dashed lines indicate the 95% Bartlett confidence bands for no serial dependence.



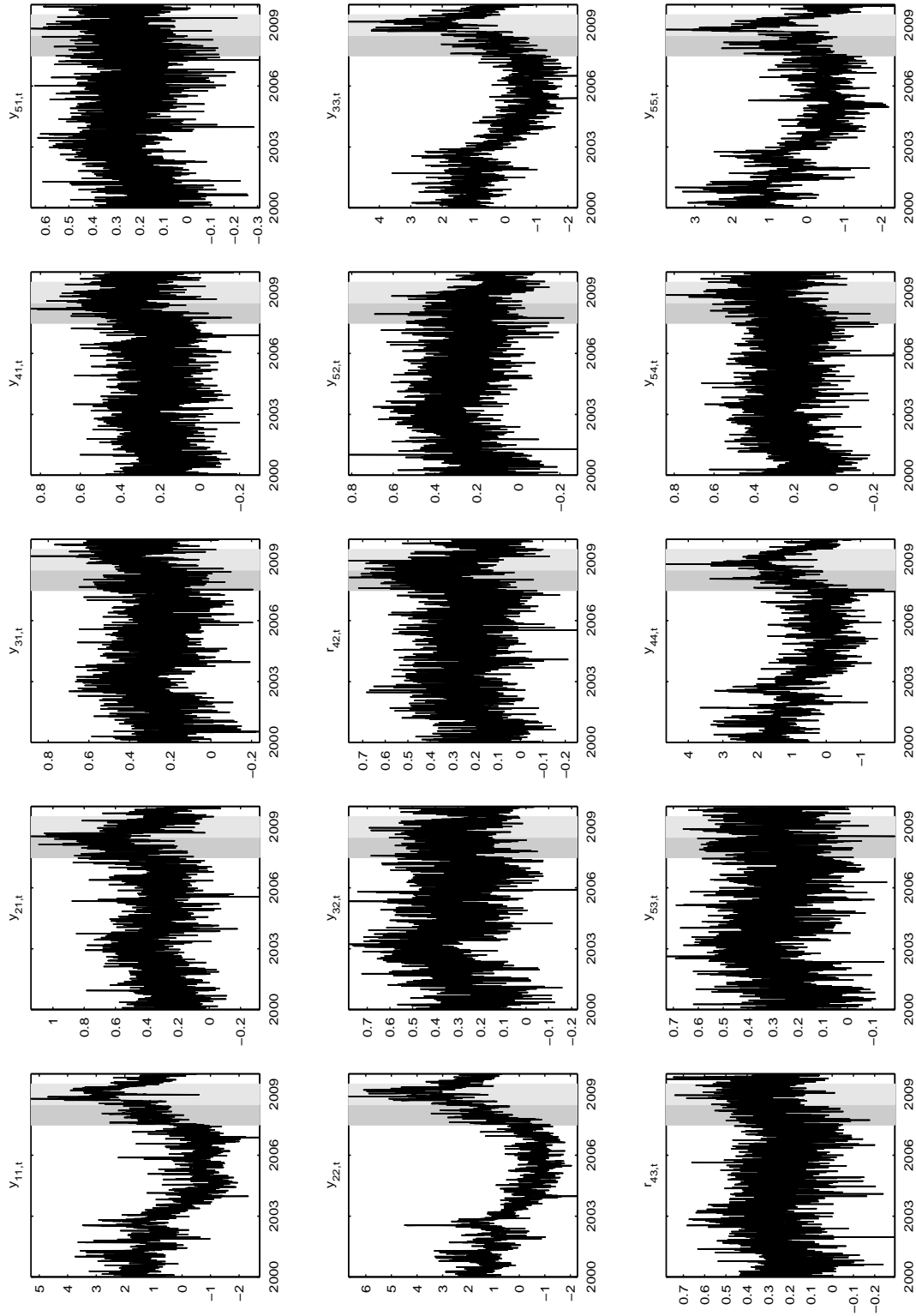


Figure 3. Time series of the matrix logarithm  $y_{i,j,t}$  of daily realized variances and covariances  $r_{i,j,t}$  for AXP ( $i = 1$ ), C ( $i = 2$ ), GE ( $i = 3$ ), HD ( $i = 4$ ), and IBM ( $i = 5$ ) stock, data set 1; the gray shaded areas mark the two out-of-sample windows used in the forecast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

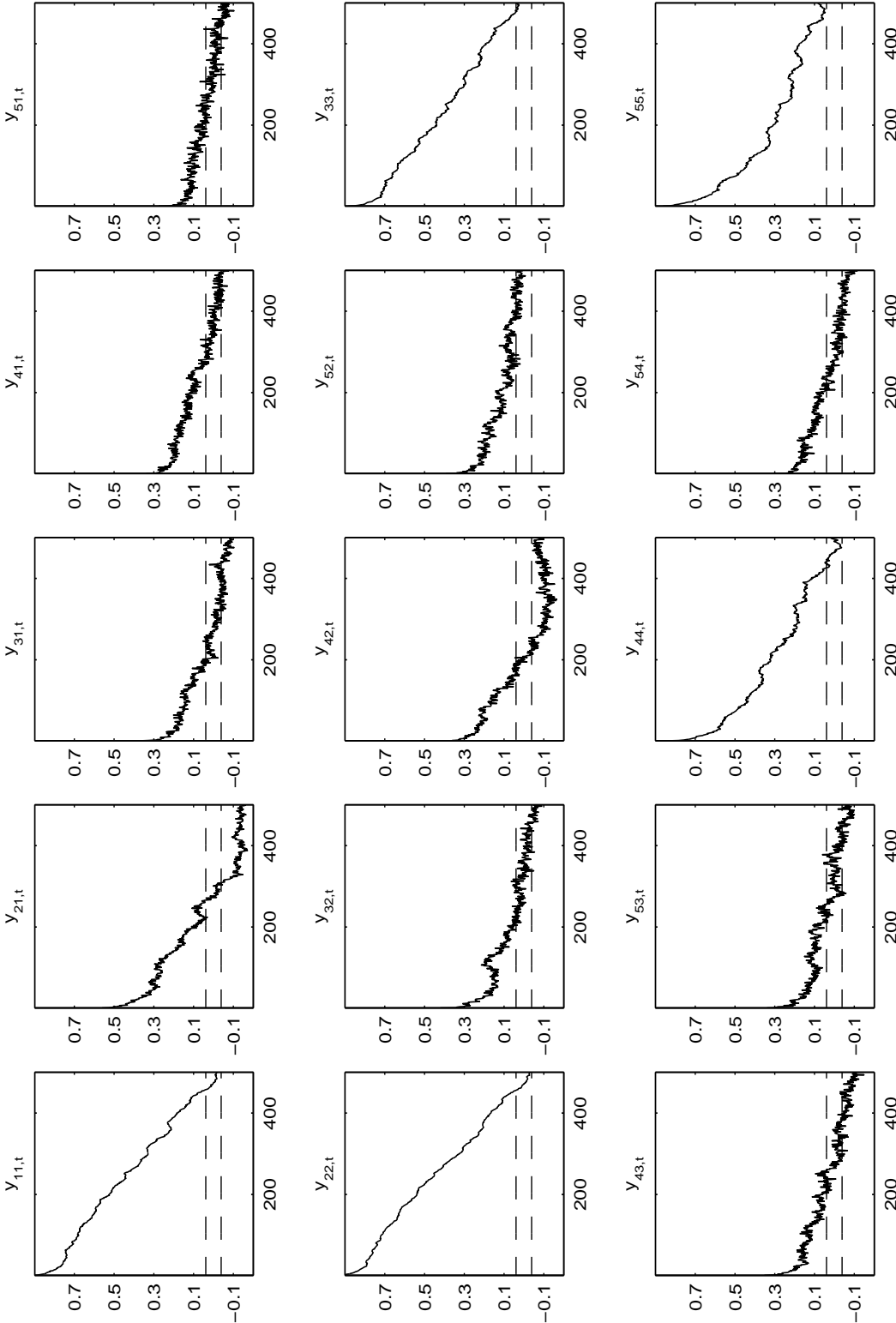


Figure 4. Sample autocorrelation function of the matrix logarithm  $y_{ij,t}$  of daily realized variances and covariances  $r_{ij,t}$  for AXP ( $i=1$ ), C ( $i=2$ ), GE ( $i=3$ ), HD ( $i=4$ ), and IBM ( $i=5$ ) stock, data set 1; the dashed lines indicate the 95% Bartlett confidence bands for no serial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

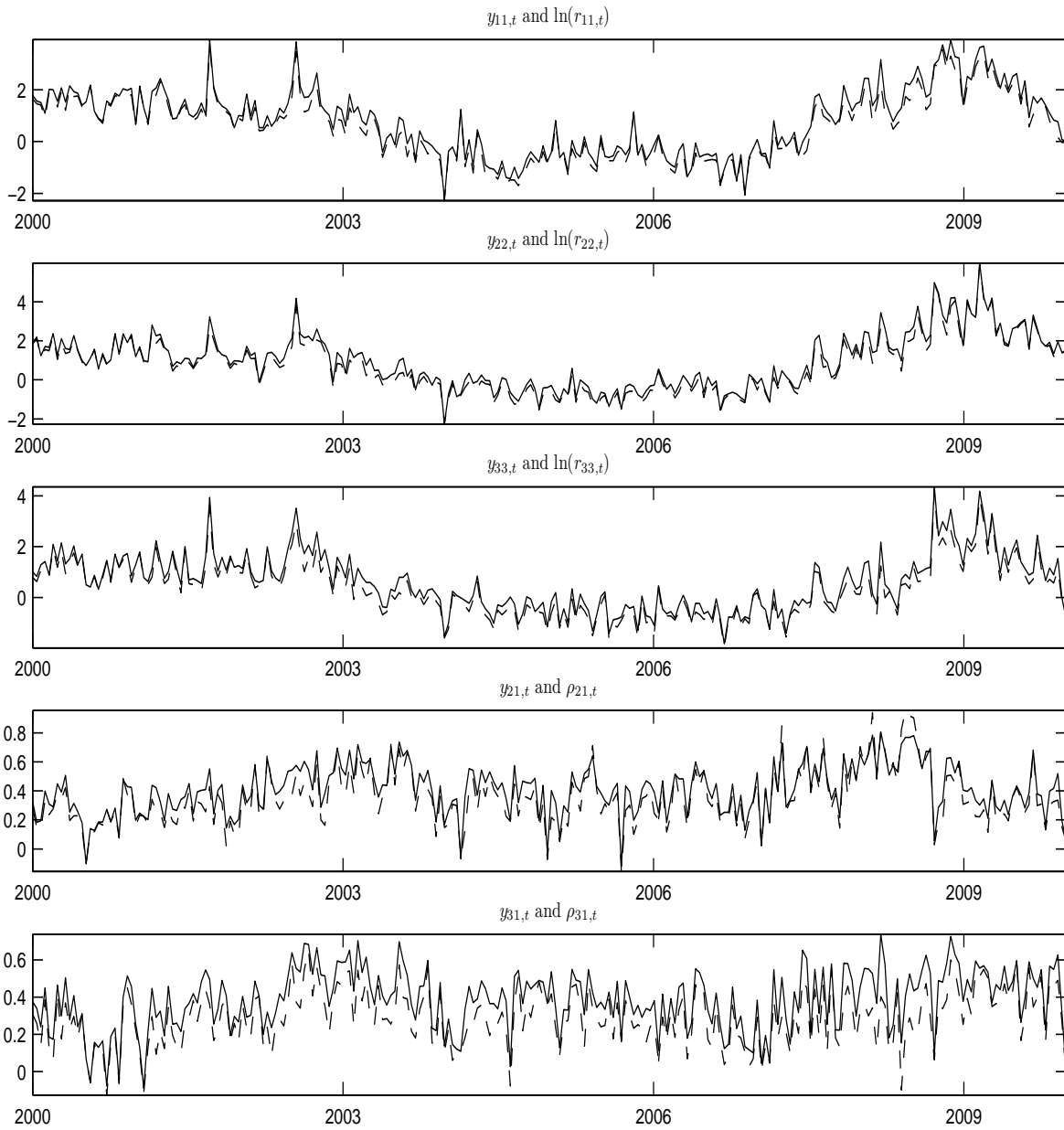


Figure 5. Logarithmic variances and correlations and corresponding matrix-logarithmic approximations for AXP ( $i = 1$ ), C ( $i = 2$ ) and GE ( $i = 3$ ), data set 1. Dashed lines: matrix-log. For reasons of legibility every 100'th observation is plotted.

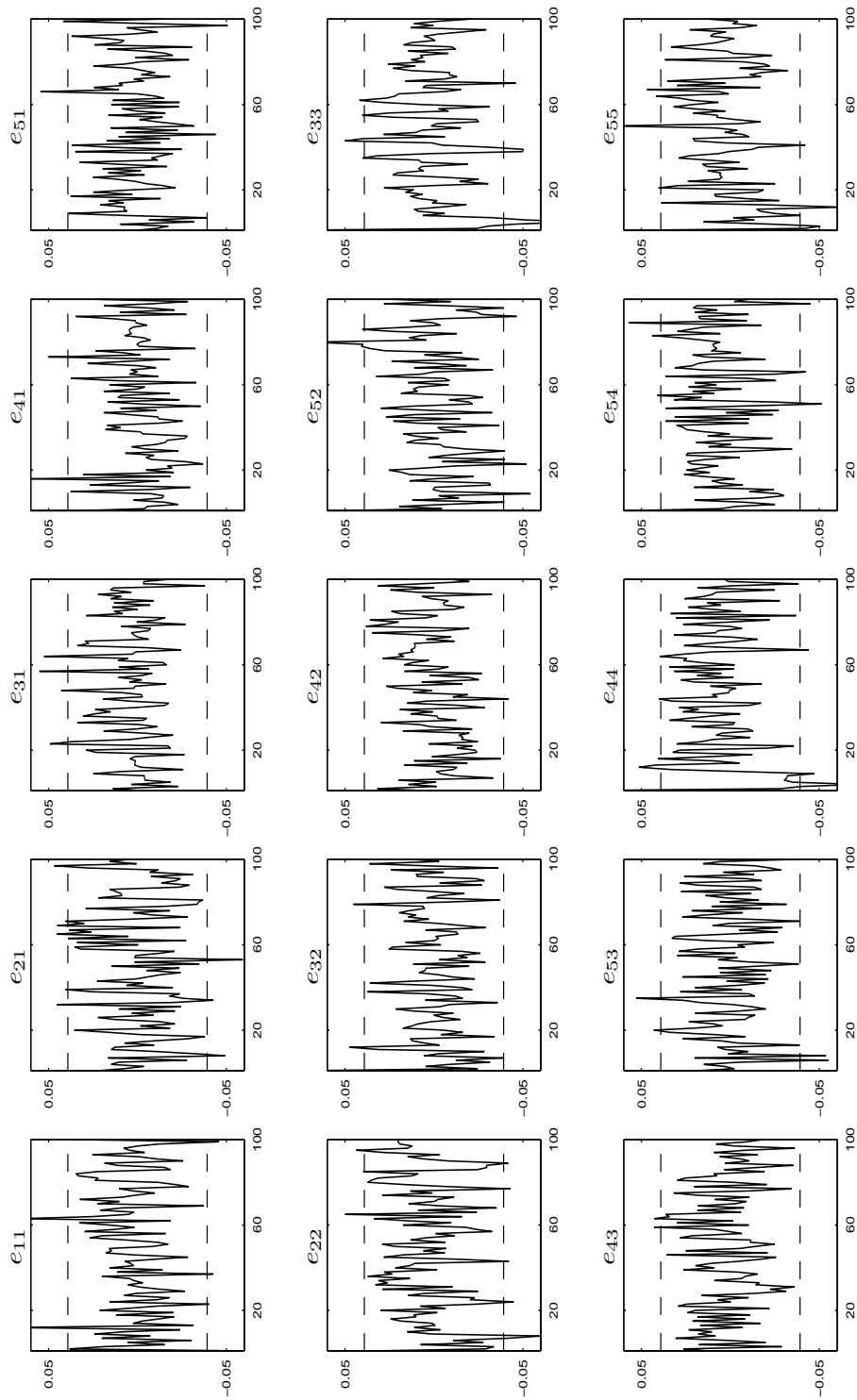


Figure 6. Sample autocorrelation function of the standardized residuals  $e_{ij,t}$  from the model with 2 common factors; data set 1. The dashed lines indicate 95% Bartlett confidence bands for no serial dependence.

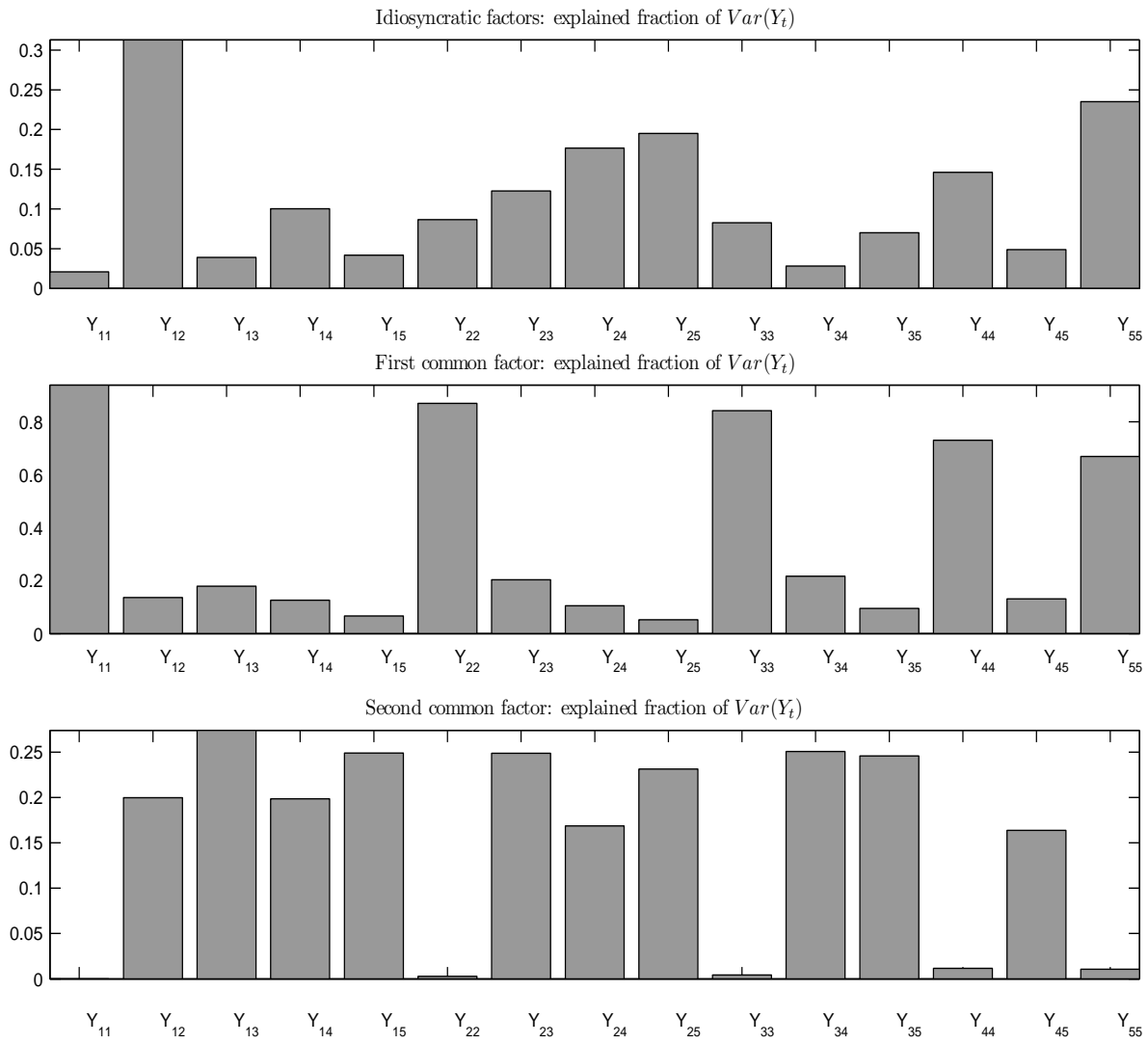


Figure 7. Fraction of total variance explained by the factors. Factor model including 2 common and 15 idiosyncratic factors; data set 1.

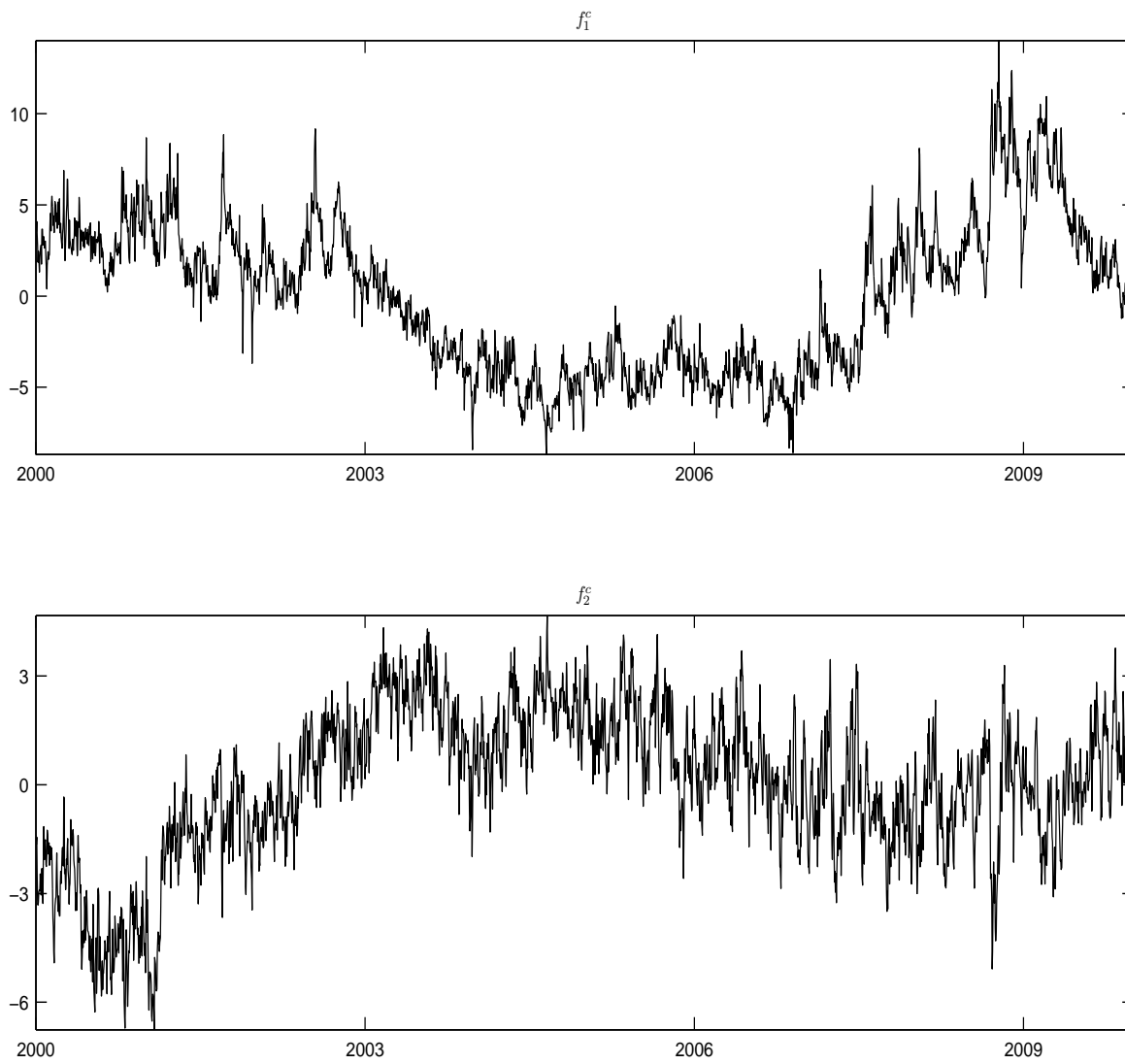


Figure 8. Filtered common factors. Factor model including 2 common and 15 idiosyncratic factors; data set 1.

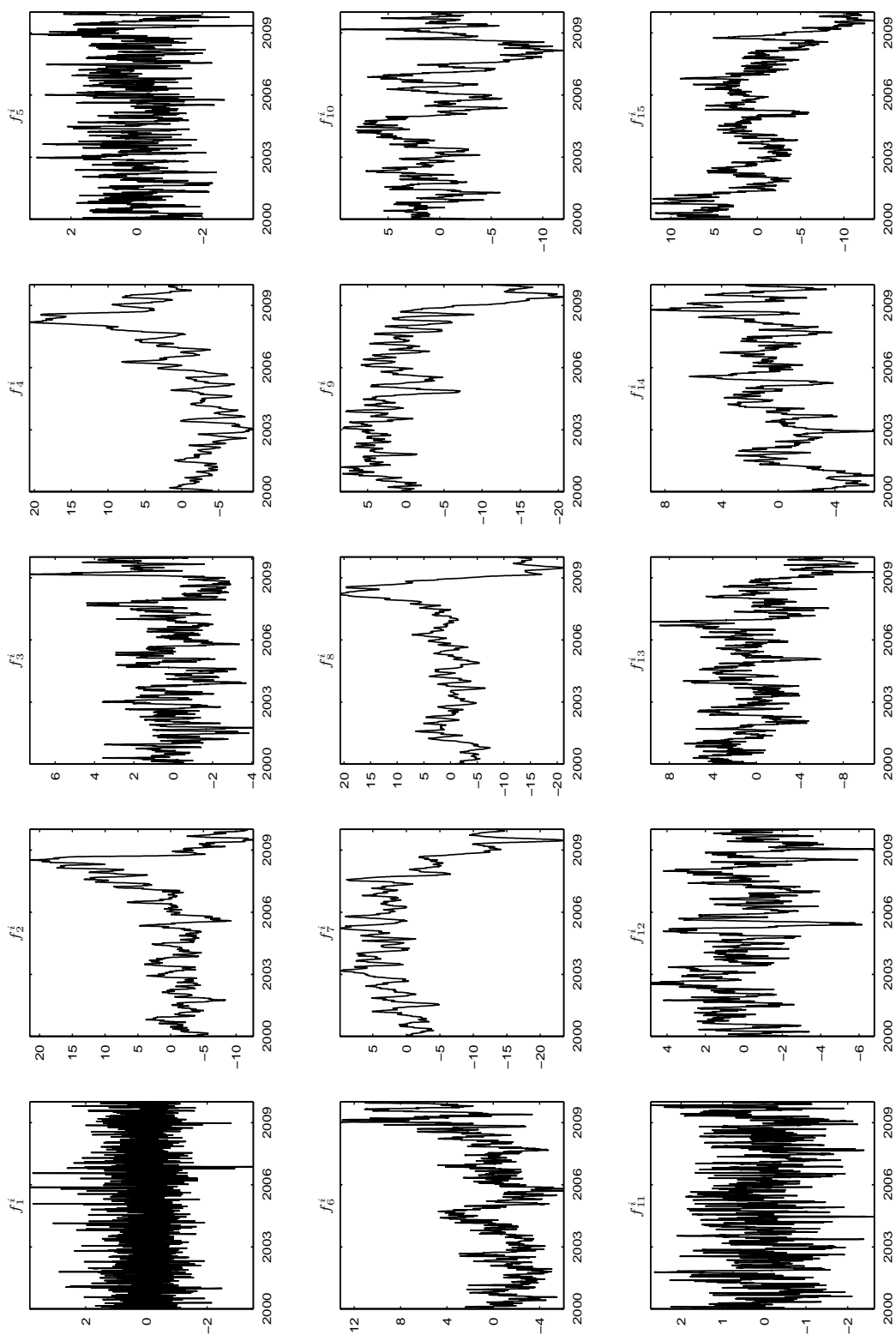


Figure 9. Filtered idiosyncratic factors. Factor model including 2 common and 15 idiosyncratic factors; data set 1.

*Diagonal Matrix-Log Elements*

*Off-Diagonal Matrix-Log Elements*

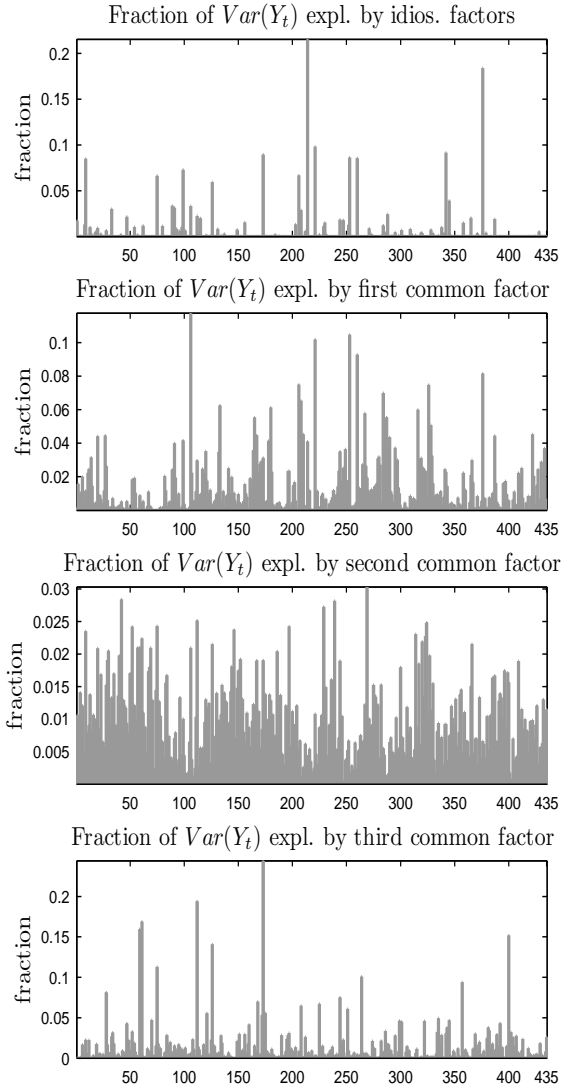
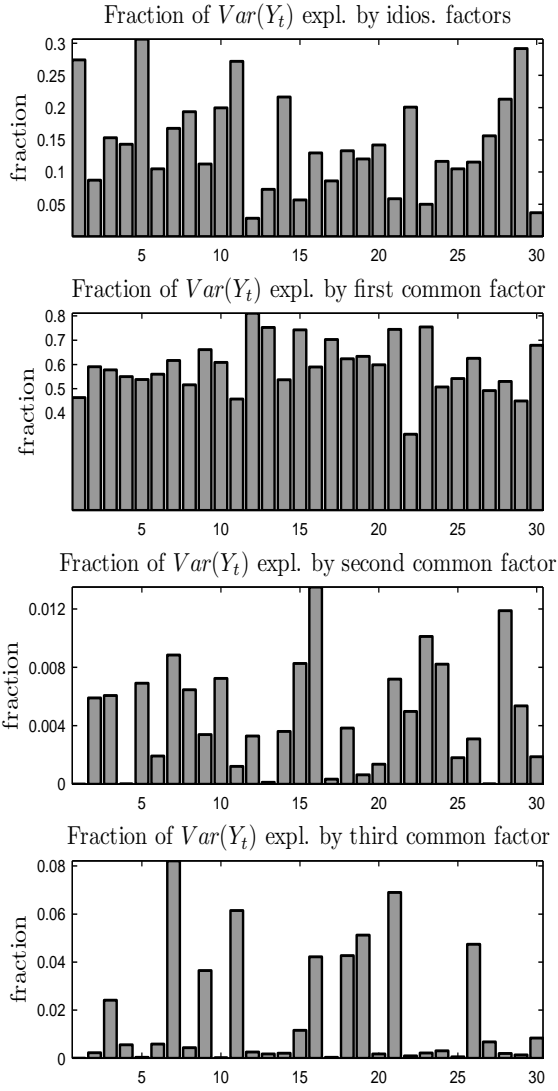


Figure 10. Fraction of total variance explained by the factors. Factor model including 3 common and 465 idiosyncratic factors; data set 2. Left panel: fraction of total variance explained by factors for the  $k = 30$  diagonal matrix-log elements. Right panel: fraction of total variance explained by factors for the  $k(k - 1)/2 = 435$  off-diagonal matrix-log elements.



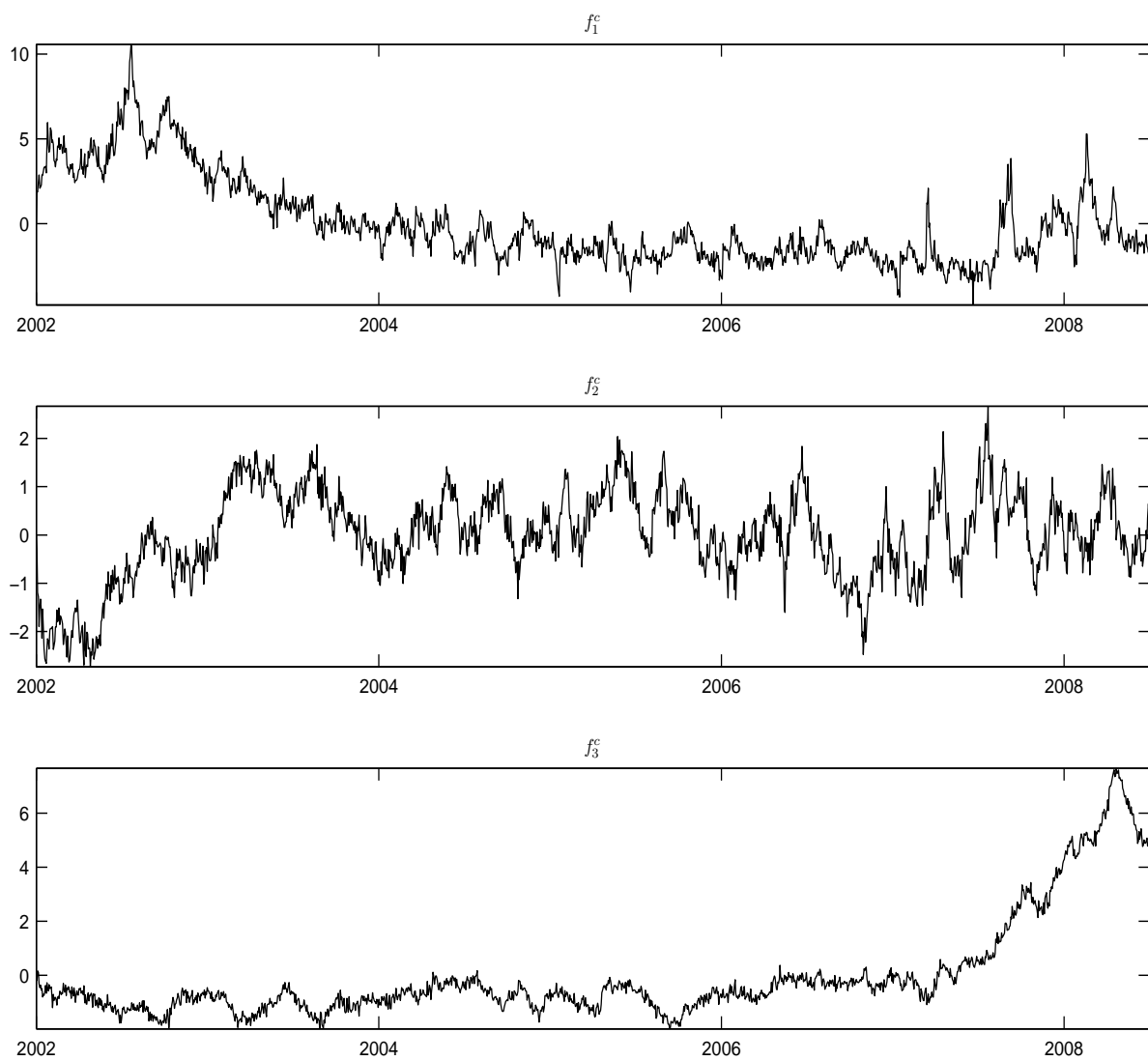


Figure 11. Filtered common factors. Factor model including 3 common and 465 idiosyncratic factors; data set 2.

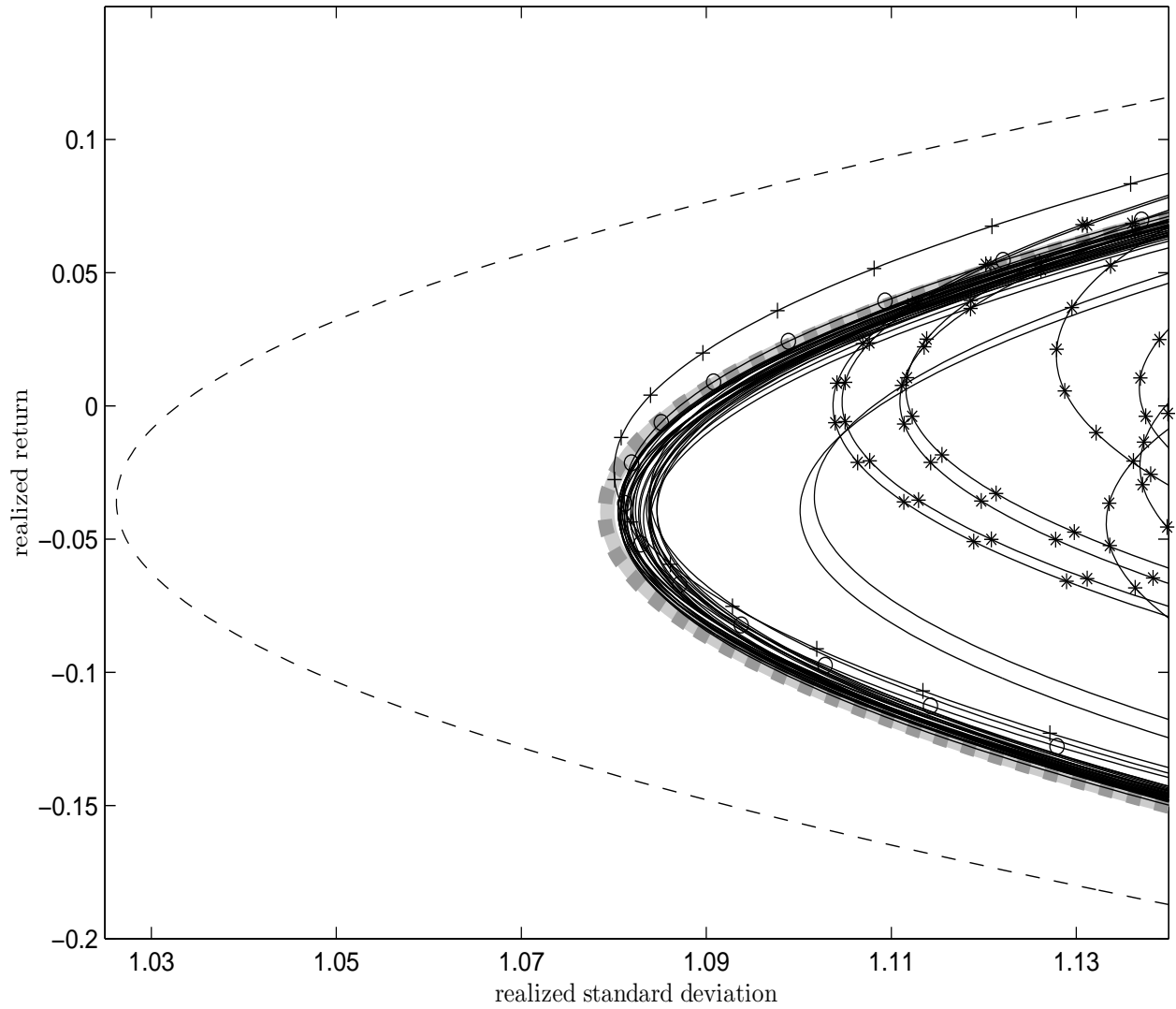


Figure 12. Mean-variance plots of the ex-post realized portfolio return (y-axis in %) against realized standard deviation (on the x-axis in %) for data set 1, out-of-sample phase I. Dashed solid line: ideal forecast based on observed covariance matrix; dashed bold dark-gray line: 1-factor model; bold light-gray line: 2-factor model; \*: multivariate GARCH models based on daily returns; +: EWMA; o: DCC-CAW; solid: remaining CAW models. All plots are averages across the 252 out-of-sample periods in phase I.

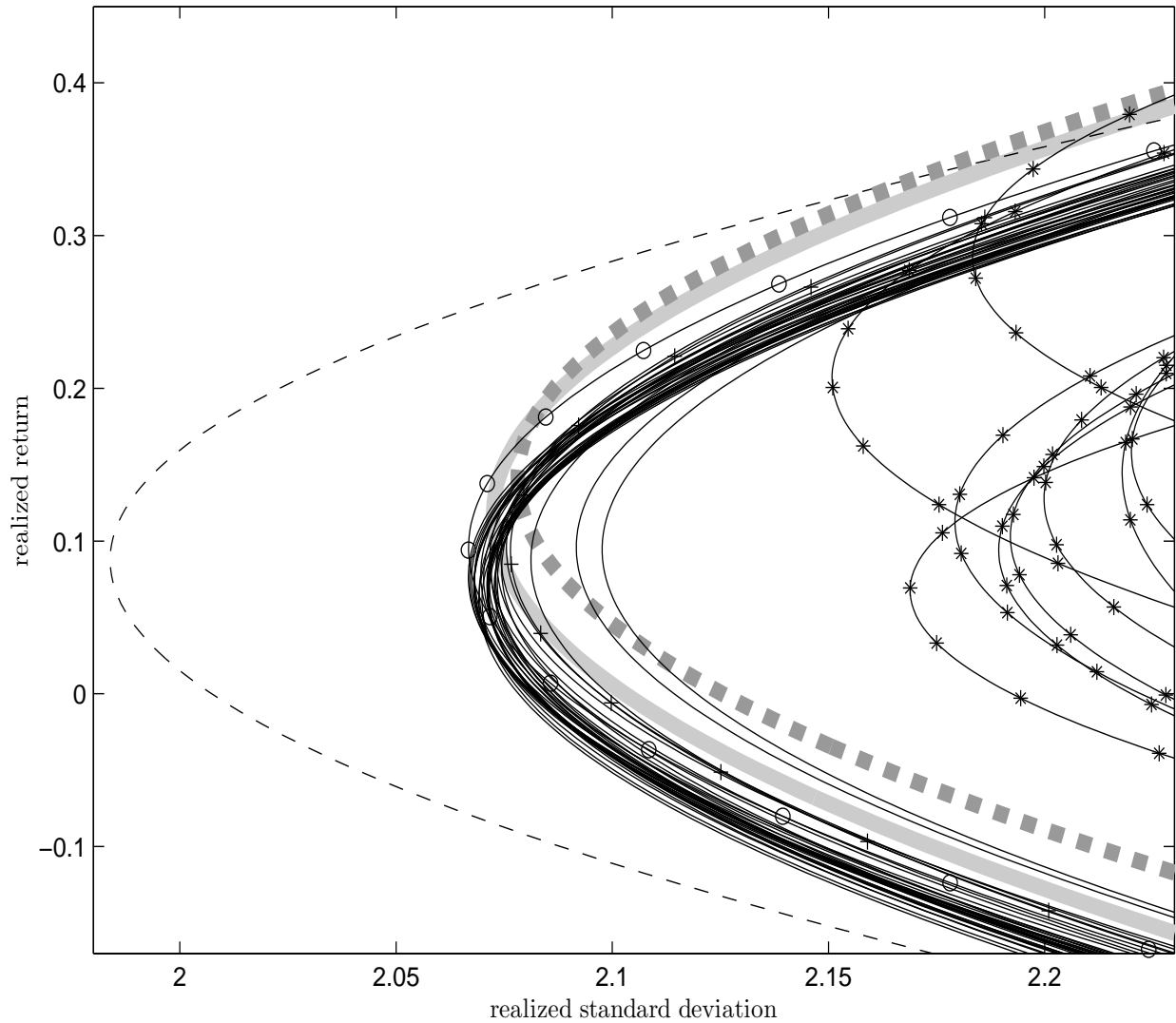


Figure 13. Mean-variance plots of the ex-post realized portfolio return (y-axis in %) against realized standard deviation (on the x-axis in %) for data set 1, out-of-sample phase II. Dashed solid line: ideal forecast based on observed covariance matrix; dashed bold dark-gray line: 1-factor model; bold light-gray line: 2-factor model; \*: multivariate GARCH models based on daily returns; +: EWMA; o: DCC-CAW; solid: remaining CAW models. All plots are averages across the 251 out-of-sample periods in phase II.

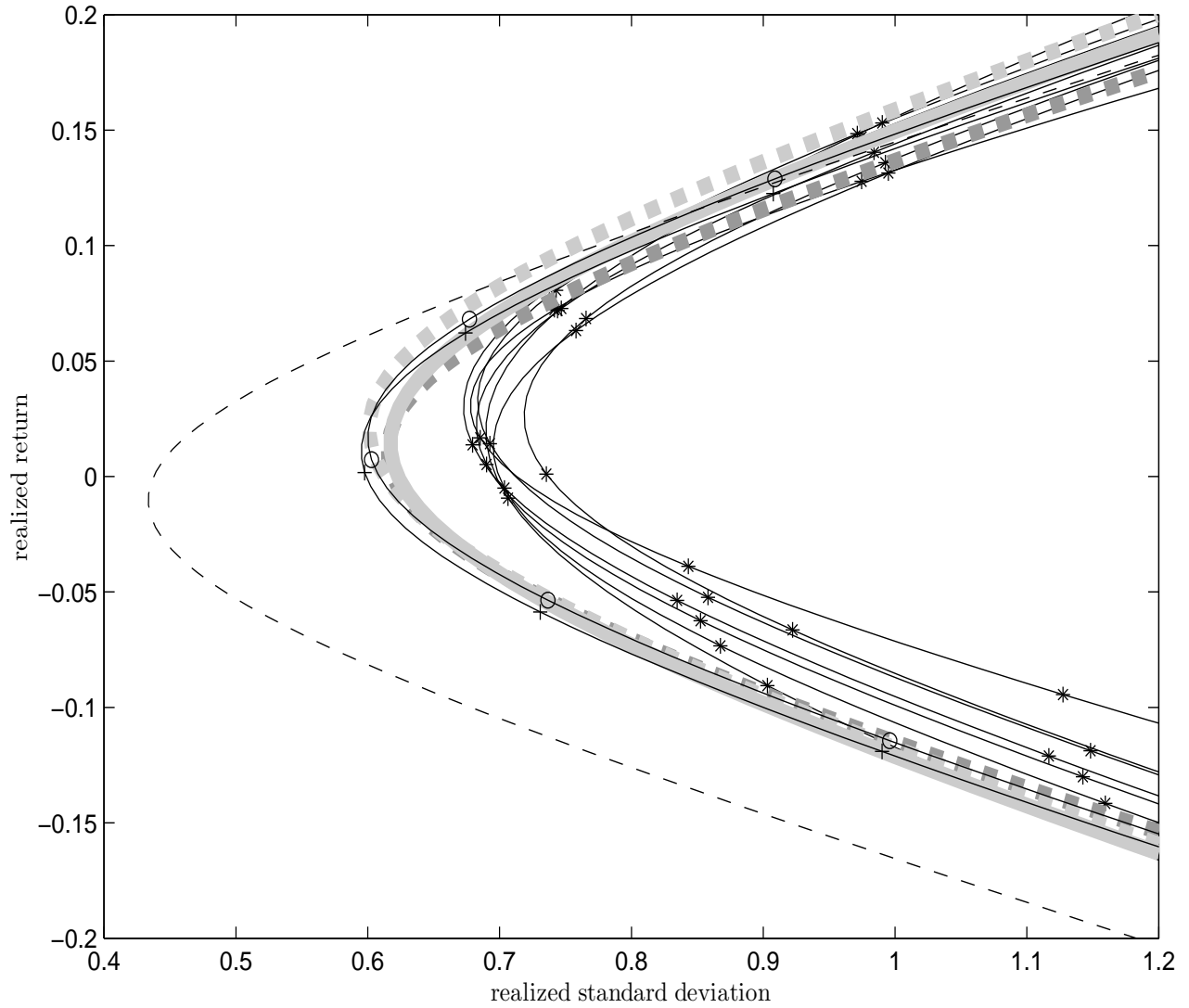


Figure 14. Mean-variance plots of the ex-post realized portfolio return (y-axis in %) against realized standard deviation (on the x-axis in %) for data set 2, out-of-sample phase I. Dashed solid line: ideal forecast based on observed covariance matrix; dashed bold dark-gray line: 1-factor model; bold light-gray line: 2-factor model; dashed bold light-gray line: 3-factor model; \*: multivariate GARCH models based on daily returns; +: EWMA; o: DCC-CAW; solid: remaining CAW models. All plots are averages across the 228 out-of-sample periods in phase I.