University of Toronto
Department of Economics

Working Paper 382

Modelling Realized Covariances

By Xin Jin and John M Maheu

November 10, 2009

Modelling Realized Covariances*

Xin Jin[†] John M. Maheu[‡]

October 2009

Abstract

This paper proposes a new dynamic model of realized covariance (RCOV) matrices based on recent work in time-varying Wishart distributions. The specifications can be linked to returns for a joint multivariate model of returns and covariance dynamics that is both easy to estimate and forecast. Realized covariance matrices are constructed for 5 stocks using high-frequency intraday prices based on positive semi-definite realized kernel estimates. We extend the model to capture the strong persistence properties in RCOV. Out-of-sample performance based on statistical and economic metrics show the importance of this. We discuss which features of the model are necessary to provide improvements over a traditional multivariate GARCH model that only uses daily returns.

key words: eigenvalues, dynamic conditional correlation, predictive likelihoods, MCMC.
JEL: C11, C32, C53, G17

*We are grateful for many helpful comments from Christian Gourieroux, Tom McCurdy, Cathy Ning and seminar participants at Ryerson University. Maheu thanks the Social Sciences and Humanities Research Council of Canada for financial support.

[†]Department of Economics, University of Toronto, reynold.jin@utoronto.ca
[‡]Department of Economics, University of Toronto and RCEA, jmaheu@chass.utoronto.ca

1 Introduction

This paper proposes a new dynamic model of realized covariance (RCOV) matrices based on recent work in time-varying Wishart distributions.[1] The specifications can be linked to returns for a joint multivariate model of returns and covariance dynamics that is both easy to estimate and forecast. The models are compared to a dynamic conditional correlation GARCH model that uses only daily returns. Capturing the persistence in RCOV matrices is critical to improved performance. The quality of multiperiod density forecasts and as well as multiperiod minimum variance portfolios are discussed.

Multivariate volatility modeling is a key input into portfolio optimization, risk measurement and management. There has arose a voluminous literature on how to approach this problem. The two popular approaches based on return data are multivariate GARCH (MGARCH) and multivariate stochastic volatility (MSV). Bauwens et al. (2006) provide a recent survey of MGARCH modeling while Asai et al. (2006) review the MSV literature. Despite the important advances in this literature there remain significant challenges. In practise the covariance of returns is unknown and is either projected onto past data in the case of MGARCH or is assumed to be latent in the case of MSV. For MSV sophisticated simulation methods must be used to deal with the unobserved nature of the conditional covariances. However, if an accurate measure of the covariance matrix could be obtained many of these difficulties could be avoided.

Recently, a new paradigm has emerged in which the latent covariance of returns is replaced by an accurate estimate based upon intraperiod return data. The estimator is non-parametric in the sense that we can obtain an accurate measure of daily ex post covariation without knowing the underlying data generating process. Realized covariance (RCOV) matrices open the door to standard time series analysis. See Andersen et al. (2003), Barndorff-Nielsen and Shephard (2004b) and Bandi and Russell (2005b) for the theoretical foundations and Andersen et al. (2009) and McAleer and Medeiros (2008) for surveys of the literature.

The purpose of this paper is to propose a new model for time-varying RCOV matrices. Among the few models in the literature for RCOV matrices is the Wishart autoregressive model of Gourieroux, Jasiak, and Sufana (2009). The process is defined by the Laplace transform and naturally leads to method of moments estimation (see also Chiriac (2006)) while the transition density is a noncentered Wishart. In a different approach Bauer and Vorkink (2007) decompose the RCOV matrix by a log-transformation and then use various time-series approaches to model the elements. Chiriac and Voev (2008) use a 3-step procedure, by first decomposing the RCOV matrices into Cholesky factors and modelling them with a VARFIMA process before transforming them back.

This paper is related to this literature but differs in that our model builds on the MSV Wishart specifications of Asai and McAleer (2009) and Philipov and Glickman (2006). These models specify a standard Wishart transition density for the covariance of returns.[2] We begin with a similar specification that replaces the latent covariance matrix by its RCOV analogue

[1] The Wishart distribution is a generalization of the univariate gamma distribution to nonnegative-definite matrices.

[2] An advantage to working with the Wishart distribution is that the pdf and simulation methods for random draws are readily available, while this is not the case for the noncentral Wishart distribution (Gauthier and Possamai 2009).

estimated following Barndorff-Nielsen et al. (2008) to obtain an *observable* MSV model.[3] Relative to the Wishart MSV models the estimation is considerably simplified.

The empirical analysis of 5 stocks show the strong persistence of the daily time series of RCOV elements. Our basic model does not perform well since it does not capture this feature. We propose an extension of the basic model that introduces components along the lines of Corsi (2009). A component is defined as a sample average of past RCOV matrices based on a particular window of data. Different windows of data give different components. These models provide significant improvements over a multivariate Dynamic Conditional Correlation MGARCH (DCC) model of Engle (2002) that uses only daily returns.

The models are estimated from Bayesian perspective. We show how to estimate the length of data windows that enter into the component version of the model. The second component is associated with 2 weeks of data while the third component is associated with about 3 months of past data. The component models provide a dramatic improvement in capturing the time series autocorrelations of the smallest and largest eigenvalues of the RCOV matrices.

Besides providing new tractable models for multivariate observable SV we also evaluate the models over a term structure of density forecasts of returns and a term structure of global minimum-variance portfolios.[4] This allows for the comparison of models and performance over short to medium investment horizons out-of-sample. Parameter uncertainty is integrated out of these problems.

The gains from using high-frequency data are substantial. The 2 and 3 component models provide the best multiperiod density forecasts as well as the smallest portfolio variance. The best RCOV models provide significant improvements in density forecasts of returns over the DCC model for up to 3 months out-of-sample. The gains from using high frequency data for global minimum variance portfolio selection are important up to 3 weeks out. After this the DCC model provides comparable performance.

In summary, we provide a new approach to modeling multivariate returns that consists of joint models of returns and RCOV matrices. We find that it is critical to include the component extensions to obtain improved performance relative to a DCC model. Compared to MSV models, our specifications provide the additional flexibility of SV at a considerably lower computational cost. This paper is organized as follows. In Section 2, we review the theory and the procedures of constructing the RCOV estimator and the data. In Section 3, our basic model for RCOV is introduced after briefly discussing two representative models of volatility based on daily returns. Section 4 explains the estimation procedure. An extension of the basic RCOV model to allow for components is introduced in Section 5 before reporting the estimation results in Section 6, followed by model comparison. Section 7 concludes.

[3]Estimation of RCOV this way has several benefits including imposing the positive definiteness and accounting for the bias that market microstructure and nonsynchronous trading can have.

[4]Maheu and McCurdy (2009) introduced the term structure of density forecasts for returns using joint models for returns and realized volatility for individual assets. We extend this to include multivariate assets and global minimum variance portfolios.

2 Realized Covariance

2.1 RCOV Construction

Suppose the k-dimensional efficient log-price $Y(t)$, follows a continuous time diffusion process defined as follows:

$$Y(t) = \int_0^t a(u)\mathrm{d}u + \int_0^t \Phi(u)\mathrm{d}W(u), \tag{1}$$

where $a(t)$ is a vector of drift components, $\Phi(t)$ is the instantaneous volatility matrix, and $W(t)$ is a vector of standard independent Brownian motions. The quantity of interest here is $\int_0^\tau \Phi(u)\Phi'(u)du$, known as the integrated covariance of $Y(t)$ over the interval $[0,\tau]$. It is a measure of the ex-post covariation of $Y(t)$. For simplicity, we normalize τ to be 1. Results from stochastic process theory (e.g. Protter (2004)) imply that the integrated covariance of $Y(t)$,

$$\int_0^1 \Phi(u)\Phi'(u)du, \tag{2}$$

is equal to its quadratic variation over the same interval,

$$[Y](1) \equiv \mathrm{plim}_{n\to\infty}\sum_{j=1}^{n}\{Y(t_j)-Y(t_{j-1})\}\{Y(t_j)-Y(t_{j-1})\}' \tag{3}$$

for any sequence of partitions $0 = t_0 < t_1 < \ldots < t_n = 1$ with $\sup_j\{t_{j+1}-t_j\} \to 0$ for $n \to \infty$.

An important motivation for our modeling approach is Theorem 2 from Andersen, Bollerslev, Diebold and Labys (2003). They show that the daily log-return follows,

$$Y(1) - Y(0)|\sigma\{a(v),\Phi(v)\}_{0\le v\le 1} \sim N\left(\int_0^1 a(u)\mathrm{d}u, \int_0^1 \Phi(u)\Phi'(u)du\right),$$

where $\sigma\{a(v),\Phi(v)\}_{0\le v\le 1}$ denotes the sigma-field generated by $\{a(v),\Phi(v)\}_{0\le v\le 1}$. In our empirical work we will assume the drift term is approximately 0 while the integrated covariance can be replaced by an accurate estimate using high-frequency intraday data. We discuss the estimation of this next.

We are interested in obtaining an estimator of the quadratic variation of Y over a day, which is a measure of the ex post daily covariation. This estimator is referred to as realized covariance, or RCOV. We require RCOV to be positive definite. One way of constructing RCOV for a particular day t is

$$\widehat{RCOV}_t = \sum_{i=1}^{n_t} r_{i,t}r'_{i,t},$$

where n_t is the number of intraday log-returns for day t, $r_{i,t}$ is the i^{th} intraday return: $r_{i,t} = Y_{t,i} - Y_{t,i-1}$, $i = 1, 2, \ldots n_t$. In the absence of market microstructure noise, Barndorff-Nielsen and Shephard (2004b) shows that $\widehat{RCOV}_t$ is a consistent estimator of quadratic covariation as $n_t \to \infty$.

In the real world there is microstructure noise that affects the log price process, and intraday prices for different stocks are not observed at the same time, nor the same frequency.

In other words, the price process is not synchronized, introducing another source of bias, known as the Epps effect. In the presence of microstructure noise, Bandi and Russell (2005b) show that the realized covariation estimator given above is not consistent. They propose a method for selecting the optimal sampling frequency as a trade-off between bias and efficiency. Instead of using all the intraday prices available, prices are sampled at a fixed frequency, say every 15 minutes. This method does not come without a cost. A majority of intraday information is discarded and a smaller sample size (the number of intraday prices to construct RCOV) usually means larger variation in the estimator.

We follow the procedure in Barndorff-Nielsen et al. (2008) (BNHLS) to construct RCOV using the high-frequency stock returns. BNHLS propose a multivariate realized kernel to estimate the ex-post covariation of log-prices. They show this new estimator is consistent, guaranteed to be positive semi-definite, and robust to measurement noise of certain types and can also handle non-synchronous trading. To synchronize the data, they use the idea of *refresh time*. A kernel estimation approach is used to minimize the effect of the microstructure noise, and to ensure positive semi-definiteness. They choose the Parzen weight function for the kernel. We review these key ideas.

The econometrician observes the log price process $X = \left(X^{(1)}, X^{(2)}, \ldots, X^{(k)}\right)'$, which is generated by Y, but is contaminated with market microstructure noise. Prices arrive at different times and at different frequencies for different stocks over the unit interval, $t \in [0, 1]$.

Suppose the observation times for the i-th stock are written as $t_1^{(i)}, t_2^{(i)}, \ldots, i = 1, 2, \ldots, k$. Let $N_t^{(i)}$ count the number of distinct data points available for the i-th asset up to time t. The observed history of prices for the day is $X^{(i)}(t_j^{(i)})$, for $j = 1, 2, \ldots, N_1^{(i)}$, i.e, the j-th price update for asset i is $X^{(i)}(t_j^{(i)})$, it arrives at $t_j^{(i)}$. The steps to computing daily RCOV are the following.

1. Synchronizing the data.
   The first key step is to deal with the non-synchronous nature of the data. The idea of refresh time is used here. Define the first refresh time as $\tau_1 = \max\left(t_1^{(1)}, \ldots, t_1^{(k)}\right)$, and then subsequent refresh times as $\tau_{j+1} = \max\left(t^{(1)}_{N^{(1)}_{\tau_j}+1}, \ldots, t^{(k)}_{N^{(k)}_{\tau_j}+1}\right)$. τ_1 is the time it has taken for all the assets to trade, i.e. all their posted prices have been updated at least once. τ_2 is the first time when all the prices are again updated, etc. From now on, we will base our analysis on this new conformed time clock $\{\tau_j\}$, and treat the entire k-dimensional vector of price updates as if it is observed at these refreshed times $\{\tau_j\}$. The number of observations of the synchronized price vector is $n + 1$, which is no larger than the number of observations of the stock with the fewest price updates. Then, the synchronized high frequency return vector is defined as $x_j = X(\tau_j) - X(\tau_{j-1}), j = 1, 2, \ldots, n$, where n is the number of refresh return observations for the day.

2. Compute the positive semi-definite realized kernel. Having synchronized the high frequency vector returns $\{x_j\}, j = 1, 2, \ldots, n$, daily $RCOV_t$ is calculated as,

$$RCOV_t = \sum_{s=-n}^{n} f\left(\frac{s}{S+1}\right)\Gamma_s. \tag{4}$$

The selection of the bandwidth S is discussed in BNHLS while

$$f(x) = \begin{cases} 1-6x^2+6x^3 & 0 \le x \le 1/2 \\ 2(1-x)^3 & 1/2 \le x \le 1 \\ 0 & x > 1. \end{cases}$$

Γ_s is the s-th realized autocovariance:

$$\Gamma_s = \begin{cases} \sum_{j=|s|+1}^{n} x_j x'_{j-s}, & s \ge 0 \\ \\ \sum_{j=|s|+1}^{n} x_{j-s} x'_j, & s < 0. \end{cases}$$

We apply this multivariate realized kernel estimation to our high-frequency data, obtaining a series of daily $RCOV_t$ matrices, which will then be fitted by our proposed Wishart Model. The j-th diagonal element of $RCOV_t$ is called realized volatility[5] and is an ex post measure of the variance for asset j. Realized correlation between asset i and j is $RCOV_{t,ij}/\sqrt{RCOV_{t,ii}RCOV_{t,jj}}$ where $RCOV_{t,ij}$ is the element from the i-th row and j-th column.

2.2 Data

We use high-frequency stock prices for 5 assets, namely Standard and Poor's Depository Receipt (SPY), General Electric Co. (GE), Citigroup Inc.(C), Alcoa Inc. (AA) and Boeing Co. (BA). The sample period runs from 1998/12/04 – 2007/12/31 delivering 2281 days. We reserve the data back to 1998/01/02 (219 observations) as conditioning data for the components models. The data are obtained from the TAQ database. We use transaction prices and closely follow Barndorff-Nielsen et al. (2008) to construct daily RCOV matrices. The data is cleaned as follows. First, trades before 9:30 AM or after 4:00 PM are removed as well as any trades with a zero price. We delete entries with a corrected trade condition, or an abnormal sale condition.[6] Finally, any trade that has a price increase (decrease) of more than 5% followed by a price decrease (increase) of more than 5% is removed. For multiple transactions that have the same time stamp the price is set to the median of the transaction prices. From this cleaned data we proceed to compute the refresh time and the realized kernel discussed in the previous section. The daily return r_t, is the continuously compounded return from the open and close prices and matches RCOV. Table 1 reports the average number of daily transaction for each stock. The average number of transactions based on the refresh time is much lower at 1835. This represents just under 5 transactions per minute. Based on this our sample is quite liquid.

Table 2 shows the sample covariance from daily returns along with the average RCOV. Figure 1 displays daily returns while the corresponding realized volatilities (RV) are in Figure 2.

[5] Also called realized variance in the literature.

[6] Specifically we remove a trade with CORR $\neq 0$, or a trade that has COND letter other than E or F in the TAQ database.

3 Models

One challenge in modelling volatility in a multivariate setting is the positive-definite restriction of the covariance matrices. To ensure positive-definiteness, most models either impose conditions on the parameters (Engle and Kroner (1995)), or reparameterize the volatility matrices by, for example, a Cholesky decomposition (Tsay (2005)), both of which complicate the models in a non-trivial way. First we review two representative multivariate volatility models that use daily returns, the first one is the Dynamic Conditional Correlation MGARCH model introduced by Engle (2002), the second one, which motivates our model to be proposed later, is the Wishart Inverse Covariance Model (WIC) in the MSV family. An advantage of the Wishart model is that it is a distribution over positive-definite matrices.

3.1 Models of daily returns

3.1.1 Dynamic Conditional Correlation Model (DCC)

A typical DCC model is as follows:

$$
\begin{aligned}
r_t|I_{t-1} &\sim N(0, D_t R_t D_t) &(5)\\
D_t &= \mathrm{diag}(\sigma_{i,t}) &(6)\\
\sigma^2_{i,t} &= \omega_i + \kappa_i r^2_{i,t-1} + \lambda_i \sigma^2_{i,t-1}, i = 1, \dots, k &(7)\\
\epsilon_t &= D_t^{-1} r_t &(8)\\
Q_t &= \mathrm{Corr}(r)(1-\alpha-\beta) + \alpha\epsilon_{t-1}\epsilon'_{t-1} + \beta Q_{t-1} &(9)\\
R_t &= \mathrm{diag}(Q_t)^{-1/2} Q_t \mathrm{diag}(Q_t)^{-1/2} &(10)
\end{aligned}
$$

where r_t is a k-dimensional daily return series, $I_{t-1} = \{r_1, ..., r_{t-1}\}$, $\mathrm{Corr}(r)$ is the sample correlation matrix. D_t, R_t, Q_t are all $k \times k$ matrices. The parameters are $\omega_1, \dots, \omega_k$, $\kappa_1, \dots, \kappa_k$, $\lambda_1, \dots, \lambda_k, \alpha, \beta$. In this specification, the unconditional mean of Q_t is just equal to $\mathrm{Corr}(r)$, the sample correlation. This is called correlation targeting and in this way the number of parameters is greatly reduced from $\frac{k^2+5k}{2} + 2$ to $3k + 2$. Equation (7) governs the dynamics of the conditional variances of each individual return by a univariate GARCH process; equation (9) governs the dynamics of the time-varying conditional correlation of the whole return vector. Because $\mathrm{Corr}(r)$ is symmetric positive definite, and $\epsilon_t\epsilon'_t$ is symmetric positive semi-definite, the conditional correlation matrices are guaranteed to be symmetric positive definite.

3.1.2 Wishart Inverse Covariance (WIC) Model

Based on the WIC multivariate stochastic volatility model in Asai and McAleer (2009) consider the following specification:

$$
\begin{aligned}
r_t|\Omega_t &\sim N(0, \Omega_t) &(11)\\
\Omega_t^{-1}|\nu, S_{t-1} &\sim Wishart_k(\nu, S_{t-1}) &(12)\\
S_t &= \frac{1}{\nu}(\Omega_t^{-d/2}) A (\Omega_t^{-d/2}) &(13)
\end{aligned}
$$

where Ω_t is the latent conditional covariance matrix of the return vector at time t.[7] $Wishart_k(\nu, S_{t-1})$ denotes a Wishart distribution of dimension k with time-varying scale matrix S_{t-1} and degree of freedom parameter $\nu \geq k$. The density function has the form

$$Wishart_k(\Omega_t^{-1}|\nu, S_{t-1}) = \frac{|\Omega_t^{-1}|^{\frac{\nu-k-1}{2}}|S_{t-1}^{-1}|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}}\pi^{k(k-1)/4}\prod_{j=1}^{k}\Gamma(\frac{\nu+1-j}{2})}\exp\left(-\frac{1}{2}Tr(\Omega_t^{-1}S_{t-1}^{-1})\right). \tag{14}$$

Philipov and Glickman (2006) proposes a similar model, with their specification using $S_t = \frac{1}{\nu}(A^{1/2})\Omega_t^{-d}(A^{1/2})$.[8]

The parameters are d, ν, and A. d is a scalar parameter, A is a positive definite symmetric parameter matrix. Since the support of a Wishart distribution are symmetric positive definite matrices, given its scale parameter being symmetric positive definite, which is ensured by the quadratic expression of S_t, the covariance matrices are guaranteed to be symmetric positive definite. Asai and McAleer (2009) shows that the condition for stationarity is given by $|d| < 1$.

This model has been estimated using Bayesian methods and daily returns. By far the most challenging and time consuming aspect is the sampling of latent Ω_t^{-1} efficiently.[9] By using the observed RCOV to replace the covariance we completely side-step this difficult issue.

3.2 Basic Model of RCOV

Motivated by Philipov and Glickman (2006) and Asai and McAleer (2009), we propose to model the dynamics of RCOV by a time-varying Wishart distribution, using a similar specification to Asai and McAleer (2009). To be more specific, let $\Sigma_t \equiv RCOV_t$, then the joint model of return vector r_t and its $RCOV_t$ is

$$r_t|\Sigma_t \sim N(0, \Sigma_t) \tag{15}$$

$$\Sigma_t|\nu, S_{t-1} \sim Wishart_k(\nu, S_{t-1}) \tag{16}$$

$$S_t = \frac{1}{\nu}(\Sigma_t^{d/2})A(\Sigma_t^{d/2}) \tag{17}$$

The parameters here, A, a symmetric and positive definite matrix, and two scalars d and ν, all have exactly the same meanings as in the WIC specification above. The law of motion for the RCOV series in the model described by equation (17) will ensure positive definiteness in a very natural way. However, in contrast to Philipov and Glickman (2006) and Asai and

[7] In Asai and McAleer (2009), Ω_t governs the correlation dynamics while the asset variances are modeled by univariate SV models. They estimate the model is two steps and also discuss alternative parameterizations.

[8] If V is a symmetric positive definite matrix then it can be decomposed as $V = EDE^{'}$, where E is an orthogonal matrix and D is a diagonal matrix of eigenvalues. Raising V to a scalar power α is defined as $V^\alpha = ED^\alpha E^{'}$ where D^α has diagonal elements d_i^α. See Greene (2002).

[9] In the MCMC sampler of Asai and McAleer, adjusted to sample covariances, we found that increasing the asset number to 5 rendered the acceptance rate so close to 0 that the model could not be estimated. Philipov and Glickman (2006) document the same problem of low acceptance frequencies for covariances as the number of assets increases.

McAleer (2009) our model governs Σ_t and not the precision Σ_t^{-1}.[10] By the properties of the Wishart distribution, the conditional expectation of Σ_t is:

$$E(\Sigma_t|\Sigma_{t-1}) = \nu S_{t-1} = (\Sigma_{t-1}^{d/2})A(\Sigma_{t-1}^{d/2}). \tag{18}$$

The RCOV inverse then follows the inverse-Wishart distribution with the conditional expectation being:

$$E(\Sigma_t^{-1}|\Sigma_{t-1}) = (\nu - k - 1)^{-1}S_{t-1}^{-1} = \frac{\nu}{\nu - k - 1}(\Sigma_{t-1}^{-d/2})A^{-1}(\Sigma_{t-1}^{-d/2}). \tag{19}$$

The elements of A determine how each element of Σ_t is related to elements of Σ_{t-1}. For example, if $W \sim Wishart_k(\nu, I_k)$ then

$$\Sigma_t = \frac{1}{\nu}(\Sigma_{t-1}^{d/2})A^{1/2}W(A^{1/2})'(\Sigma_{t-1}^{d/2}). \tag{20}$$

The scalar parameter d measures the overall influence of past RCOV on current RCOV. This parameter is closely related to the degree of persistence present in the RCOV series, with d larger the stronger the persistence. A and d together govern the relationship between RCOVs of two adjacent periods. Suppose A is the identity matrix and $d = 1$, then by equation (18), $E(\Sigma_t|\Sigma_{t-1}) = \nu S_{t-1} = \Sigma_{t-1}$, which is a random walk in matrix form. If $d = 0$, then $E(\Sigma_t|\Sigma_{t-1}) = A$, so the RCOV matrix follows an $i.i.d.$ Wishart distribution over time. The degree of freedom parameter ν, determines how close the RCOVs are centered around their conditional mean, with larger ν meaning the random matrices generated by the Wishart distribution are more concentrated around the mean.

Since we know $\{\Sigma_t\}_{t=1}^T$, as RCOV is calculated using intraday data, we do not need equation (15) of return dynamics to estimate the parameters. This is different from the MGARCH and MSV models where the variance is latent and parameters can only be inferred through returns. Equation (15) links the dynamics of returns to that of the RCOV, so the model gives us a law of motion for both the return vector and RCOV at the daily frequency, hence we can simulate the model out-of-sample and perform multiperiod density forecasts of returns.

4 Model Estimation

We apply a standard Bayesian approach to estimate our model, using MCMC methods for posterior simulation. Since the posterior distribution is unknown, the idea behind MCMC simulation is to construct a Markov Chain that has as its limiting distribution the posterior distribution of the parameters of interest. Features of the posterior density can then be estimated consistently based on the samples obtained from posterior simulation. For example, we can estimate the posterior mean of model parameters by the sample average of the MCMC draws. For more details on MCMC methods see Chib (2001). The steps in estimation used here are similar to Asai and McAleer (2009).

To apply Bayesian inference, we need to first assign priors to the parameters. $A^{-1} \sim Wishart_k(\gamma_0, Q_0)$, a Wishart distribution with $k \times k$ symmetric positive definite scale matrix

[10] We found this to lead to significant improvements in model performance.

Q_0 and γ_0 degrees of freedom. We set $Q_0 = I$ and $\gamma_0 = k+1$ to reflect a proper but relatively uniformative prior. d follows the uniform prior $d \sim U(-1, 1)$, and $\nu \sim \exp(\lambda_0)I_{\nu>k}$, an exponential distribution with support truncated to be greater than k. To make the prior flat, λ_0 is set to 100. We assume independence among the prior distributions of our parameters, so the joint prior for A^{-1}, d, ν is just the product of the individual priors.

$$p(A^{-1}, d, \nu) = Wishart_k(A^{-1}|\gamma_0, Q_0) \times p(d) \times \exp(\lambda_0). \tag{21}$$

Given the parameters the data density is the product of the Wishart densities,

$$p(\{\Sigma_t\}_{t=1}^T|A, d, \nu) = \prod_{t=1}^T Wishart_k(\Sigma_t|\nu, S_{t-1}). \tag{22}$$

According to Bayes' rule, the joint posterior distribution of the parameters is,

$$p(A^{-1}, d, \nu|\{\Sigma_t\}_{t=1}^T) \propto Wishart_k(A^{-1}|\gamma_0, Q_0) \times p(d) \times \exp(\lambda_0)I_{\nu>k} \times \prod_{t=1}^T Wishart_k(\Sigma_t|\nu, S_{t-1}). \tag{23}$$

To directly sample all parameters from this complicated distribution is not possible. Instead, we use MCMC methods to iteratively sample from the conditional posterior distribution of each parameter conditional on the other parameters. Samples from this procedure will converge to samples from the the joint posterior distribution. We iterate sampling from the following conditional distributions.

- $A^{-1}|\nu, d, \{\Sigma_t\}_{t=1}^T$
- $d|A, \nu, \{\Sigma_t\}_{t=1}^T$
- $\nu|A, d, \{\Sigma_t\}_{t=1}^T$

Sampling $A^{-1}|\nu, d, \{\Sigma_t\}_{t=1}^T$ is a typical Gibbs sampling step from a Wishart density while the remaining two densities are unknown and require a Metropolis-Hastings step. For these we employ a random walk proposal. See the appendix for full details.

Drawing from these 3 distributions constitutes one sweep of the sampler. After dropping an initial set of draws as burnins we collect M draws to obtain $\{\theta^{(i)}\}_{i=1}^M$, where $\theta = (A^{-1}, d, \nu)$. Then simulation consistent estimates of posterior moments can be obtained as sample averages of the draws. For instance, the posterior mean of θ can be estimated as $M^{-1}\sum_{i=1}^M \theta^{(i)}$.

5 A Component Model of RCOV

Persistence in volatility is one of the most important stylized features of financial assets. As we shall see in the results section, the basic Wishart RCOV model has difficulty capturing the persistence properties of the data, which leads us to explore a component extension of the model. Andersen, Bollerslev and Diebold (2007), Corsi (2009), Maheu and McCurdy (2009) among others use the Heterogeneous AutoRegressive (HAR) model of realize variance in the univariate case in order to capture long-memory like features of volatility parsimoniously.

Motivated by this work, we consider an extension of our basic model to allow for components, which are averages of past Σ_t, to multiplicatively affect the scale parameter matrix, S_t.

The Wishart-RCOV(K) model with $K \geq 1$ components is defined as,

$$
\begin{aligned}
\Sigma_t|\nu, S_{t-1} &\sim Wishart_k(\nu, S_{t-1}) \\
S_t &= \frac{1}{\nu}\left[\prod_{j=K}^{1} \Gamma_{t,\ell_j}^{\frac{d_j}{2}}\right] A \left[\prod_{j=1}^{K} \Gamma_{t,\ell_j}^{\frac{d_j}{2}}\right] && (24) \\
\Gamma_{t,\ell} &= \frac{1}{\ell}\sum_{i=0}^{\ell-1} \Sigma_{t-i} && (25) \\
1 = \ell_1 < &\cdots < \ell_K. && (26)
\end{aligned}
$$

The components enter as a sample average of past Σ_t raised to a different matrix power d_k.[11] The component terms $\Gamma_{t,l}$ allow for more persistence in the location of Σ_t while the different values of d_j allow the effect to be dampened or amplified. In (24) the order of the product operator is important and differs in the two terms. The window width ℓ of each component $\Gamma_{t,l}$ could be preset or be estimated. The case of one component $K = 1, \ell_1 = 1$, is identical to the model discussed in Section 3.2. The remaining parameters A and ν are the same as those discussed before.

The priors on d_j are all $U(-1, 1)$. In the following we confine the analysis to a $K = 1, 2$ and 3 component models. Rather than preset ℓ_2, and ℓ_3, we estimate them. For the Wishart-RCOV(2) model the prior for ℓ_2 is uniform discrete with support $\{2, 3, \ldots, 200\}$, while for Wishart-RCOV(3) both ℓ_2 and ℓ_3 have the same prior with the additional restriction that $\ell_2 < \ell_3$ for identification.

Most of the estimation details are identical to the one component model discussed in Section 4 with the addition of the lag length parameters affecting the components. For example, the 2-component model has the conditional posterior density:

$$
p(\ell_2|A, d_1, d_2, \nu, \{\Sigma_t\}) \;\propto\; p(\ell_2)\exp\left(-\frac{d_2\nu\psi}{2\ell_2} - \frac{1}{2}Tr\left(\nu A^{-1}Q(\ell_2)^{-1}\right)\right) \tag{27}
$$

where

$$
\psi = \sum_{t=1}^{T} \log|\Gamma_{t-1,\ell_2}|, \quad Q(\ell_2)^{-1} = \sum_{t=1}^{T} \Sigma_{t-1}^{-\frac{d_1}{2}}(\Gamma_{t-1,\ell_2}^{-\frac{d_2}{2}})\Sigma_t(\Gamma_{t-1,\ell_2}^{-\frac{d_2}{2}})\Sigma_{t-1}^{-\frac{d_1}{2}}
$$

and $p(\ell_2)$ has the density of a discrete uniform prior over $\{2, 3, \ldots, 200\}$. To sample from the conditional posterior we use a simple random walk proposal. The proposal distribution is a Poisson random variable multiplied by a random variable that takes on values 1 and -1 with equal probability. The density of the proposal is

$$
q(\ell) = \begin{cases} \frac{\lambda^\ell e^{-\lambda}}{2\ell!} & \ell = 1, 2, \cdots \\ e^{-\lambda} & \ell = 0 \\ -\frac{\lambda^\ell e^{-\lambda}}{2\ell!} & \ell = -1, -2, \cdots \end{cases}
$$

[11] We also examined a *geometric* average version using the following specification: $\Gamma_{t,\ell}^d \equiv \Sigma_{t-\ell+1}^{\frac{d}{\ell}}\Sigma_{t-\ell+2}^{\frac{d}{\ell}}\cdots\Sigma_t^{\frac{d}{\ell}}$. So past RCOVs affect current RCOV multiplicatively, which is in accordance with our basic model. We found this geometric average version, while it has similar performance in almost every aspect, is computationally more costly. We will hence focus our results on the sample average version.

In the empirical work $\lambda = 2$. Given the value ℓ_2 in the Markov chain, the new proposal $\ell_2' \sim q(\ell_2)$ is accepted with probability

$$\min\left\{\frac{p(\ell_2'|A, d_1, d_2, \nu, \{\Sigma_t\}_{t=1}^T)}{p(\ell_2|A, d_1, d_2, \nu, \{\Sigma_t\}_{t=1}^T)}, 1\right\}. \tag{28}$$

In a similar way ℓ_3 is sampled for the Wishart-RCOV(3) model.

6 Results

All data is used for estimation of the DCC and Wishart-RCOV(K) models. Parameter estimates for the DCC based on daily returns appear in Table 3 and are typical of results found in the literature. Estimates for the RCOV models are shown in Table 4. Estimates of the elements of A^{-1} are found in Table 5 for Wishart-RCOV(3). In each case, the first 1000 draws are discarded as burnin in posterior simulation and the next 5000 MCMC draws are used for inference. The numerical standard errors are based on the Newey-West estimator for the long run variance with a lag length of 1000. The *inefficiency factors* are the ratio of the long-run variance estimate to the sample variance where the latter assumes an i.i.d. sample. This serves as an indicator of how well the chain mixes. The lower the value is, the closer the sampling is to i.i.d. The 95 percent density intervals are constructed using the 2.5th percentile and 97.5th percentile of the MCMC draws for the corresponding parameters.

The inefficiency measures indicate that the MCMC chain mixes well. Both the d and μ parameter in the Wishart-RCOV(1) model are accurately estimated. For instance, d has a posterior mean of 0.6677 with a 0.95 density interval of $(0.6600, 0.6756)$. There is no evidence from this model that RCOV follows a martingale. Note that the Wishart-RCOV(3) specification nests both the 1 and 2 component versions by setting $d_2 = d_3 = 0$ and $d_3 = 0$, respectively. Table 4 estimates indicate that the Wishart-RCOV(3) model significantly diverges from its nested counterparts with posterior means for d_2 of 0.4502 and d_3 of 0.2651 and both having densities intervals very far from 0.

To evaluate the in-sample fit of the model, we plot in Figure 3 the time-series of the determinant of the RCOV matrices, $\{|\Sigma_t|\}_{t=1}^T$, obtained from the data.[12] Then, for the 1-component model, we plot the determinants of the expected RCOV using the estimated parameters. The in-sample fitted value is $E(\Sigma_t|\Sigma_{t-1}) = \nu S_{t-1}$. The two series in Figures 3 and 4 move together closely, indicating a good fit of the basic model.

To investigate the time-series persistence in RCOV, we plot the sample autocorrelation function (ACF) of both the largest and smallest eigenvalues of the RCOV series observed, and compare it to the posterior mean ACF obtained from the Wishart-RCOV(K) models.[13] The interpretation of the eigenvalues follows Gourieroux, Jasiak, and Sufana (2009). The largest (smallest) eigenvalue is equal to the maximum (minimum) risk from the portfolio with variance $\omega'\Sigma_t\omega$, given standardized portfolio weights $\omega'\omega = 1$. In Figure 5 and 6 both

[12] Muirhead (1982) labels the determinant of the covariance matrix from a multivariate distribution as a *generalized variance* which measures the overall spread of a distribution.

[13] The posterior mean of the ACF of the eigenvalues is obtained as follows. A parameter draw is taken from the posterior density and used to simulate 2281 observations of Σ_t. From this the ACF is computed and saved. This is repeated many times and the average ACF is displayed.

the largest and smallest eigenvalues show strong persistence and are different from zero even 400 lags out. The one component Wishart-RCOV(1) is completely unable to match the data.[14] The $K = 2$, and 3 component models provide significant improvements. The kinks in the ACF function from the simulated models occur at exactly the window width of the components. For instance, the Wishart-RCOV(3) has posterior mean of ℓ_2 equal to 9 and ℓ_3 equal to 64, and the change in the slope of the ACF can be seen at these locations. As we shall see in the remaining results and those in the next section, capturing the persistence in the RCOV matrices is critical to providing better in-sample and out-of-sample predictions.

Table 6 reports the mean square error (MSE) for our RCOV models against the observed RCOV series. For each model, the MSE is calculated in the following way: for each period t, we calculate the expected RCOV according to $E(\Sigma_t|\Sigma_{t-1}) = \nu S_{t-1}$, and subtract it from the observed RCOV, Σ_t. We then square each element of the difference matrix and sum them. As a comparison, the MSE for another 3 models: a naive model that sets the expected RCOV as last period's RCOV (i.e. $E(\Sigma_t|\Sigma_{t-1}) = \Sigma_{t-1}$); a model that sets the predicted RCOV equal to the sample covariance computed from daily returns; and a DCC model are included. The DCC model (Section 3.1.1) is estimated using daily returns to produce an in-sample estimate of the conditional covariance $D_t R_t D_t$. The DCC model provides an improvement over the sample covariance estimate and also beats the naive RCOV model. However, each of the Wishart-RCOV(K) models provide significant improvements relative to any method that uses daily returns. The 3-component has the smallest MSE but this is only marginally smaller that the 2-component version.

To compare how the model tracks correlations Figures 7 and 8 display *realized correlation* computed from the elements of the RCOV matrix along with the fitted estimates from the Wishart-RCOV(3) specification and the DCC model. For the models, the fitted correlation is extracted from $E(\Sigma_t|\Sigma_{t-1})$ for the RCOV model and obtained from $D_t R_t D_t$ for the DCC model. The first figure is the correlation between SPY and GE and the second is the correlation between GE and Citigroup. Both models track the realized correlation closely with the RCOV model displaying a clear advantage. In some episodes the DCC wanders away from the realized correlations. This is not surprising since the DCC can only infer correlations from noisy daily returns.

6.1 Density Forecasts

In this section, we compare our RCOV models to the benchmark DCC model, focusing on their out-of-sample performance. We compare each candidate's forecast ability of future returns through predictive likelihood, which is a popular approach in the literature (Maheu and McCurdy (2009), Amisano and Giacomini (2007), Bao, Lee, and Saltoglu (2007), Weigend and Shi (2000)). From a Bayesian perspective the predictive likelihoods are a key input into model comparison through predictive Bayes factors (Geweke (2005)). Following Maheu and McCurdy (2009) we evaluate a term structure of a model's density forecasts of returns. This is the cumulative predictive likelihood based on out-of-sample data for $h = 1, ..., H$ period ahead density forecasts of returns. Using returns results in a common benchmark for the RCOV model and the DCC model which uses only daily returns in estimation. The

[14]Similar results are obtained for the ACF of individual elements of the RCOV matrices.

Wishart-RCOV(K) models are linked to returns through equation (15).

For a candidate model $\mathcal{A}$, we compute the following cumulative predictive likelihood:

$$\hat{p}_h^{\mathcal{A}} = \sum_{t=T_0-h}^{T-h} \log(p(r_{t+h}|I_t, \mathcal{A})),$$

for $h = 1, 2, \ldots, H$ and $T_0 < T$. For each h, $\hat{p}_h^{\mathcal{A}}$ measures the forecast performance based on the *same common set of returns*: $r_{T_0}, ..., r_T$. Therefore, $\hat{p}_1^{\mathcal{A}}$ is comparable with $\hat{p}_{10}^{\mathcal{A}}$ and allows us to measure the decline in forecast performance as we move from 1 day forecasts to 10 day forecasts using model $\mathcal{A}$. We are also interested in comparing $\hat{p}_h^{\mathcal{A}}$ for a fixed h with another specification, $\hat{p}_h^{\mathcal{B}}$. Better models will have larger cumulative predictive likelihood values.

For the RCOV models $I_t = \{r_1, \Sigma_1, ..., r_t, \Sigma_t\}$ while for the DCC model $I_t = \{r_1, ..., r_t\}$. The predictive likelihood $p(r_{t+h}|I_t, \mathcal{A})$, is the h-period ahead predictive density for model $\mathcal{A}$ evaluated at the realized return r_{t+h},

$$p(r_{t+h}|I_t, \mathcal{A}) = \int p(r_{r+h}|\Omega_{t+h}, \mathcal{A}) p(\Omega_{t+h}|\theta, I_t, \mathcal{A}) p(\theta|I_t, \mathcal{A}) \mathrm{d}\theta \mathrm{d}\Omega_{t+h}. \tag{29}$$

Parameter uncertainty is integrated out of the density. In the DCC model $\Omega_t \equiv D_t R_t D_t$ while for the Wishart-RCOV(K) models $\Omega_t \equiv \Sigma_t$, while θ is the respective parameter vector. The integration is approximated as

$$\int p(r_{t+h}|\Omega_{t+h}, \mathcal{A}) p(\Omega_{t+h}|\theta, I_t, \mathcal{A}) p(\theta|I_t, \mathcal{A}) \mathrm{d}\theta \mathrm{d}\Omega_{t+h} \approx \frac{1}{M} \sum_{i=1}^{M} p(r_{t+h}|\Omega_{t+h}^{(i)}, \mathcal{A}), \tag{30}$$

where $\Omega_{t+h}^{(i)} \sim p(\Omega_{t+h}|\theta^{(i)}, I_t, \mathcal{A})$, and $\theta^{(i)} \sim p(\theta|I_t, \mathcal{A})$. $p(r_{t+h}|\Omega_{t+h}^{(i)}, \mathcal{A})$ is the pdf of a multivariate Normal with mean 0 and covariance $\Omega_{t+h}^{(i)}$ evaluated at r_{t+h}. $\{\theta^{(i)}\}_{i=1}^{M}$ are the MCMC draws from the posterior distribution $p(\theta|I_t, \mathcal{A})$ for the model. Note that for each term $p(r_{t+h}|I_t, \mathcal{A})$ in the out-of-sample period we re-estimate the model to obtain a new set of draws from the posterior to compute (30).

Given a model $\mathcal{A}$ with predictive likelihood $p(r_{T_0}, ..., r_T|\mathcal{A})$, and model $\mathcal{B}$ with predictive likelihood $p(r_{T_0}, ..., r_T|\mathcal{B})$, the Bayes factor in favor of model $\mathcal{A}$ versus model $\mathcal{B}$ is $BF_{\mathcal{AB}} = \frac{p(r_{T_0},...,r_T|\mathcal{A})}{p(r_{T_0},...,r_T|\mathcal{B})}$. Kass and Raftery (1995) suggest interpreting the evidence for $\mathcal{A}$ as: not worth more than a bare mention if $0 \leq BF_{\mathcal{AB}} < 3$; positive if $3 \leq BF_{\mathcal{AB}} < 20$; strong if $20 \leq BF_{\mathcal{AB}} < 150$; and very strong if $BF_{\mathcal{AB}} \geq 150$.

Figure 9 plots $\hat{p}_h$ for all the models against $h = 1, 2, \ldots, H = 60$, giving each model a cumulative log-predictive likelihood term structure. Included are 3 Wishart-RCOV models along with the DCC and a model that sets the conditional covariance to the sample analogue computed from returns. Finally, a perfect foresight model (labelled RCOV in the figure) which assumes the RCOV matrix Σ_{t+h} is known. Of course we cannot use this model in practise, but it provides an upper bound on performance.

The out-of-sample data begins at T_0=2006/03/31 and ends at 2007/12/31 for a total of 441 observations. This is true for each model and each forecast horizon h. With the exception of the perfect foresight model, all specifications have a downward sloping term structure. Intuitively, forecasting further out is more difficult.

The DCC model completely dominates the basic 1-component Wishart-RCOV(1) model. Despite the better measure of latent covariation, the RCOV dynamics in the basic model are too simple and neglect to capture the persistence properties of the eigenvalues discussed above. The 2-component Wishart-RCOV(2) model shows superior forecast ability over the DCC model up to 10 periods ahead, while the 3-component Wishart-RCOV(3) model dominates all other candidates consistently across the entire forecast horizon. Note that according to predictive Bayes factors the improvement of the Wishart-RCOV(3) model over the DCC model is large. For instance, the log-Bayes factors in favor of the Wishart-RCOV(3) specification are: 56.34 ($h = 1$), 20.82 ($h = 30$), and 23.56 ($h = 60$). We conclude that the RCOV models based on high-frequency intraday data can provide substantial improvements over the DCC model estimated from daily returns. For these improvements to be realized it is critical that the persistence properties of RCOV matrices be modeled correctly.

6.2 Economic Evaluation

In this section, we evaluate the out-of-sample performance of the models from a portfolio optimization perspective. We focus on the simple problem of finding the global minimum variance portfolio, so the issue of specifying the expected return is avoided. The h-period ahead global minimum variance portfolio (GMVP) is computed as the solution to

$$\min_{w_{t+h|t}} w'_{t+h|t}\Omega_{t+h|t}w_{t+h|t}$$
$$\text{s.t.} \quad w'_{t+h|t}\iota = 1.$$

$\Omega_{t+h|t}$ is the predictive mean of the covariance matrix at time $t+h$ given time t information for a particular model. Each model is re-estimated at each data point in the out-of-sample period. From the posterior draws the predictive mean of the covariance matrix at time $t+h$ is simulated along the lines of the previous subsection. $w_{t+h|t}$ is the portfolio weight, and ι is a vector with all the elements equal to 1. The optimal portfolio weight is

$$w_{t+h|t} = \frac{\Omega^{-1}_{t+h|t}\iota}{\iota'\Omega^{-1}_{t+h|t}\iota}.$$

It can be shown (Engle and Colacito (2006)) that if the portfolio weights, w_t, are constructed from the true conditional covariance, then the variance of a portfolio computed using the GMVP from any other model must be larger.

We evaluate model performance starting at 2006/03/31 to 2007/12/31 for a total of 441 observations for $h = 1, ..., H = 60$. The specifications considered are: Wishart-RCOV(K), $K = 1, 2, 3$, the DCC model, a model that always uses fixed equal portfolio weights, a model using fixed portfolio weights based on the sample covariance of daily returns, and a perfect foresight model using observed Σ_{t+h} of each period. The last model serves as a proxy for the ideal case where the portfolio is constructed using the "true" variance. We report the sample variances of the GMVPs across models in Figure 10. As in the density forecast exercise we use a common set of returns to evaluate the performance over different h. As a result, the upwards sloping portfolio variances indicates that time-series information is most useful for short term forecasts. Generally, the Wishart-RCOV(3) model provides the lowest portfolio variances and is closest to the perfect foresight model except for the first 3 weeks when the 2 component model is slightly better. The RCOV component models deliver variance reductions for about 3 weeks out, after which the DCC model has approximately the same variance. The DCC model provides improvements over both the equal weight portfolio and the sample covariance model.

7 Conclusion

This paper proposes to model the dynamics of realized covariance matrices (RCOV) based on recent work in time varying Wishart distributions. Realized covariance matrices are constructed for 5 stock returns using high-frequency intraday prices based on positive semi-definite realized kernel estimation introduced by Bardorff-Nielson et al. (2008). We explore the time series properties of our RCOV models and propose component models to capture persistence. Out-of-sample performance of our models are compared to that of a benchmark DCC model that only uses daily returns. The best RCOV models provide significant improvements over the DCC model in terms of density forecasts of returns for up to 3 months out-of-sample. The gains from using high frequency data for global minimum variance portfolios is important up to 3 weeks out.

8 Appendix

8.1 Wishart-RCOV Estimation

Given the priors listed in Section 4 the conditional posterior distributions are as follows.

$$\begin{aligned}
p(A^{-1}|\nu, d, \{\Sigma_t\}_{t=1}^T) &\propto Wishart_k(A^{-1}|\gamma_0, Q_0) \times \prod_{t=1}^{T} Wishart_k(\Sigma_t|\nu, S_{t-1}) \\
&\propto \frac{|A^{-1}|^{\frac{\gamma_0-k-1}{2}}|Q_0^{-1}|^{\frac{\gamma_0}{2}}}{2^{\frac{\gamma_0 k}{2}}\prod_{j=1}^{k}\Gamma(\frac{\gamma_0+1-j}{2})}\exp\left(-\frac{1}{2}Tr(A^{-1}Q_0^{-1})\right) \\
&\quad \times \prod_{t=1}^{T}\frac{|\Sigma_t|^{\frac{\nu-k-1}{2}}|S_{t-1}^{-1}|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}}\prod_{j=1}^{k}\Gamma(\frac{\nu+1-j}{2})}\exp\left(-\frac{1}{2}Tr(\Sigma_t S_{t-1}^{-1})\right) \\
&\quad \times \exp\left(-\frac{1}{2}Tr(A^{-1}Q_0^{-1}) - \frac{1}{2}\sum_{t=1}^{T}Tr(\Sigma_t\nu\Sigma_{t-1}^{-d/2}A^{-1}\Sigma_{t-1}^{-d/2})\right) \\
&\propto |A^{-1}|^{\frac{T\nu+\gamma_0-k-1}{2}}\exp\left(-\frac{1}{2}Tr(A^{-1}Q_0^{-1}) - \frac{1}{2}\sum_{t=1}^{T}Tr(\nu A^{-1}\Sigma_{t-1}^{-d/2}\Sigma_t\Sigma_{t-1}^{-d/2})\right) \\
&\propto |A^{-1}|^{\frac{T\nu+\gamma_0-k-1}{2}}\exp\left(-\frac{1}{2}Tr[A^{-1}(Q_0^{-1} + \nu\sum_{t=1}^{T}\Sigma_{t-1}^{-d/2}\Sigma_t\Sigma_{t-1}^{-d/2})]\right) \\
&\propto Wishart_k(A^{-1}|\widetilde{\gamma}, \widetilde{Q}) \qquad (31)
\end{aligned}$$

Where $\widetilde{Q}^{-1} = \nu \sum_{t=1}^{T} \Sigma_{t-1}^{-d/2} \Sigma_t \Sigma_{t-1}^{-d/2} + Q_0^{-1}$, $\widetilde{\gamma} = T\nu + \gamma_0$.

For d we have,

$$\begin{aligned} p(d|A,\nu,\{\Sigma_t\}_{t=1}^T) &\propto p(d) \times \prod_{t=1}^{T} Wishart_k(\Sigma_t|\nu, S_{t-1}) \\ &= p(d) \prod_{t=1}^{T} \frac{|\Sigma_t|^{\frac{\nu-k-1}{2}} |S_{t-1}^{-1}|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})} \exp\left(-\frac{1}{2} Tr(\Sigma_t S_{t-1}^{-1})\right) \\ &\propto p(d) \exp\left(-\frac{d\nu\psi}{2} - \frac{1}{2} Tr(\nu A^{-1} Q(d)^{-1})\right), \end{aligned} \tag{32}$$

where $\psi = \sum_{t=1}^{T} \log|\Sigma_{t-1}|$, and $Q(d)^{-1} = \sum_{t=1}^{T} \Sigma_{t-1}^{-d/2} \Sigma_t \Sigma_{t-1}^{-d/2}$. To sample from this density we do the following. If d is the previous value in the chain we propose $d' = d + u$ where $u \sim N(0, \sigma^2)$ and accept d' with probability

$$\min \left\{ \frac{p(d'|A,\nu,\{\Sigma_t\}_{t=1}^T)}{p(d|A,\nu,\{\Sigma_t\}_{t=1}^T)}, 1 \right\}. \tag{33}$$

and otherwise retain d. σ^2 is selected to achieve a rate of acceptance between 0.3-0.5.

Finally, ν has the conditional posterior density

$$\begin{aligned} p(\nu|A,d,\{\Sigma_t\}_{t=1}^T) &\propto p(\nu) \times p(\{\Sigma_t\}_{t=1}^T|A,d,\nu) \\ &= p(\nu) \prod_{t=1}^{T} \frac{|\Sigma_t|^{\frac{\nu-k-1}{2}} |S_{t-1}^{-1}|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})} \exp\left(-\frac{1}{2} Tr(\Sigma_t S_{t-1}^{-1})\right) \\ &= p(\nu) \frac{\prod_{t=1}^{T} |\Sigma_t|^{\frac{\nu}{2}} \times \nu^{\frac{T\nu k}{2}} |A^{-1}|^{\frac{T\nu}{2}} \times \prod_{t=1}^{T} |\Sigma_{t-1}|^{-\frac{d\nu}{2}}}{2^{\frac{T\nu k}{2}} (\prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2}))^T} \exp\left(-\frac{1}{2} Tr(\nu A^{-1} Q^{-1})\right) \\ &\propto \exp\left(-\lambda_0 \nu + \frac{T\nu}{2} \log|A|^{-1} + \frac{T\nu k}{2} \log\frac{\nu}{2} - T \sum_{j=1}^{k} \log\Gamma\left(\frac{\nu+1-j}{2}\right)\right) \\ &\quad \times \exp\left(\frac{\nu}{2} \sum_{t=1}^{T} \log|\Sigma_t| - \frac{d\nu}{2} \sum_{t=1}^{T} \log|\Sigma_{t-1}| - \frac{1}{2} Tr(\nu A^{-1} Q^{-1})\right) \end{aligned} \tag{34}$$

where $Q^{-1} = \sum_{t=1}^{T} \Sigma_{t-1}^{-d/2} \Sigma_t \Sigma_{t-1}^{-d/2}$. This is a nonstandard distribution which we sample using a Metropolis-Hastings step with a random walk proposal analogous to the sampling in the previous step above.

8.2 DCC Estimation

The parameters are $(\omega_1, \ldots, \omega_k, \kappa_1, \ldots, \kappa_k, \lambda_1, \ldots, \lambda_k, \alpha, \beta) = \Theta$. All parameters are assigned an independent Normal prior with mean 0 and variance 100, truncated to the interval $(0, 1)$ with the following restrictions:

$$\omega_i > 0,\ \kappa_i \geq 0,\ \lambda_i \geq 0,\ \kappa_i + \lambda_i < 1,\ i = 1, ..., k,\ \alpha \geq 0,\ \beta \geq 0,\ \alpha + \beta < 1. \tag{35}$$

The joint prior $p(\Theta)$ is just the product of the individual priors. The likelihood function $p(\{r_t\}_{t=1}^T|\Theta)$ is:

$$p(\{r_t\}_{t=1}^T|\Theta) = (2\pi)^{\frac{Tk}{2}} \prod_{t=1}^{T} |D_t R_t D_t|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2} \sum_{t=1}^{T} r_t'(D_t R_t D_t)^{-1} r_t \right) \tag{36}$$

The posterior of the parameters $p(\Theta|\{r_t\}_{t=1}^T)$ is:

$$p(\Theta|\{r_t\}_{t=1}^T) \propto p(\Theta)(2\pi)^{\frac{Tk}{2}} \prod_{t=1}^{T} |D_t R_t D_t|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2} \sum_{t=1}^{T} r_t'(D_t R_t D_t)^{-1} r_t \right) \tag{37}$$

To sample from the joint posterior distribution $p(\Theta|\{r_t\}_{t=1}^T)$, we do the following steps: We first adopt a single move sampler. For each iteration, the chain cycles through the conditional posterior densities of the parameters in a fixed order. For each parameter, a random walk with normal proposal is applied. After dropping an initial set of draws as burnins, we collect M draws and use them to calculate the sample covariance matrix of the joint posterior. Then, a block sampler is used to jointly sample the full posterior. The proposal density is a multivariate normal random walk with the covariance matrix set to the sample covariance, obtained from the draws of the single-move sampler, scaled by a scalar. When the model is re-estimated out-of-sample as a new observation arrives the previous sample covariance is used as the next covariance in the multivariate normal random walk. This results in fast efficient sampling.

Reference

1. Amisano, G., R. Giacomini (2007): “Comparing Density Forecasts via Weighted Likelihood Ratio Tests”, *Journal of Business and Economic Statistics*, 25(2), 177-190.

2. Andersen, T., T. Bollerslev, F. X. Diebold (2007), “Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility”, *Review of Economics and Statistics*, 89, 701-720.

3. Andersen, T., T. Bollerslev, F. X. Diebold (2009), “Parametric and Nonparametric Volatility Measurement”, in *Handbook of Financial Econometrics*, VOL 1, Eds Ait-Sahalia and Hansen, Elsevier.

4. Andersen, T., T. Bollerslev, F. X. Diebold, P. Labys (2003), “Modeling and Forecasting Realized Volatility”, *Econometrica*, 71, 529-626.

5. Asai, M., M. McAleer (2009), “The Structure of Dynamic Correlations in Multivariate Stochastic Volatility Models”, *Journal of Econometrics*, 150, 182-192.

6. Asai, M., M. McAleer, J. Yu (2006), “Multivariate Stochastic Volatility: A Review”, *Econometric Reviews*, 25, 145-175.

7. Bandi, F. M., Russell, J. R. (2005b), “Realized covariation, realized beta and microstructure noise”, Unpublished paper. Graduate School of Business, University of Chicago.

8. Bao, Y., T.-H. Lee, and B. Saltoglu (2007), “Comparing Density Forecast Models”, *Journal of Forecasting*, 26(3), 203-225.

9. Barndorff-Nielsen, O. E., Hansen, P. H., Lunde, A., Shephard, N. (2008), “Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading”, working paper, SSRN: http://ssrn.com/abstract=1154144

10. Barndorff-Nielsen, O. E., Shephard, N. (2004b), “Econometric analysis of realised covariation: High Frequency based covariance, regression and correlation in financial economics”, *Econometrica*, 72, 885-925.

11. Bauer, G. H., Vorkink, K. (2007), “Multivariate Realized Stock Market Volatility”, Working Paper, Bank of Canada.

12. Bauwens, L., S. Laurent, J. Rombouts (2006), “Multivariate GARCH Models: a Survey”, *Journal of Applied Econometrics*, 21, 79-109.

13. Chib S. (2001) “Markov Chain Monte Carlo Methods: Computation and Inference”, in *Handbook of Econometrics*, Heckman and Leamer Eds., Elsevier Science.

14. Chiriac, R., Voev, V. (2008): “Modelling and Forecasting Multivariate Realized Volatility”, SSRN: http://ssrn.com/abstract=1174290

15. Chiriac, R. (2006), "Estimating Realized Volatility Wishart Autoregressive Model", Working Paper, University of Konstanz.

16. Corsi, F. (2009), "A Simple Approximate Long-Memory Model of Realized Volatility",*Journal of Financial Econometrics*, Spring 7: 174 - 196.

17. Engle, R. F., Colacito, R. (2006), "Testing and Valuing Dynamic Correlations for Asset Allocation", *Journal of Business and Economic Statistics*, 24, 239-250.

18. Engle, R. F. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivarate Generalized Autoregressive Conditional Heteroskedasticity Models", *Journal of Business and Economic Statistics*, 20, 339-350.

19. Engle, R. F., Kroner, K. F. (1995), "Multivariate Simultaneous Generalized ARCH", *Econometric Theory*, 11, 122-150.

20. Gauthier, P. and D. Possamai (2009) "Efficient Simulation of the Wishart model", http://ssrn.com/abstract=1474728

21. Geweke J. (2005) *Contemporary Bayesian Econometrics and Statistics*, Wiley.

22. Gourieroux, C., J. Jasiak, R. Sufana (2009), "The Wishart Autoregressive Process of Multivariate Stochastic Volatility", *Journal of Econometrics*, 150, 167-181.

23. Greene, W. H. (2002), *Econometric Analysis* (5th Edition), Prentice Hall.

24. Kass R. E. and A. E. Raftery (1995) "Bayes Factors", *Journal of the American Statistical Association*, 90(430), 773-795.

25. Maheu, J. M., T. H. McCurdy (2009), "Do high-frequency measures of volatility improve forecasts of return distributions?", forthcoming, *Journal of Econometrics.*

26. McAleer M. and M. Medeiros (2008) "Realized Volatility: A Review", *Econometric Reviews*, 27(1-3), pages 10-45.

27. Muirhead, R. J., (1982), *Aspects of Multivariate Statistical Theory*, Wiley, New York.

28. Philipov, A., M. E. Glickman (2006), "Multivariate Stochastic Volatility via Wishart Process", *Journal of Business and Economic Statistics*, 24(3), 313-328.

29. Protter, P. (2004), *Stochastic Integration and Differential Equations*, Springer-Verlag, New York.

30. Tsay, R. S., (2005), *Analysis of Financial Time Series*, Wiley.

31. Weigend, A. S., S. Shi (2000), "Predicting Daily Probability Distributions of S&P500 Returns", *Journal of Forecasting*, 19, 375-392.

Table 1: Average daily number of transactions and average daily refresh time (RT) observations per day

SPY	GE	C	AA	BA	RT
6985	7479	6121	3279	3745	1835

This table reports the average daily number of transactions (after data cleaning) for Standard and Poor's Depository Receipt (SPY), General Electric Co. (GE), Citigroup Inc.(C), Alcoa Inc. (AA) and Boeing Co. (BA). The total number of days is 2281. RT reports the average number of daily observations according to the refresh time.

Table 2: Summary statistics: Daily returns and RCOV

	Sample covariance from daily returns					Average of realized covariances				
	SPY	GE	C	AA	BA	SPY	GE	C	AA	BA
SPY	0.963	1.078	1.172	0.834	0.751	0.907	0.972	1.099	0.822	0.718
GE		2.410	1.500	1.062	0.931		2.327	1.250	0.897	0.796
C			2.826	1.014	0.931			3.176	0.982	0.835
AA				3.900	0.993				3.921	0.734
BA					2.933					2.910

This table reports the sample covariance from daily returns and the sample average of the realized covariances. The data are Standard and Poor's Depository Receipt (SPY), General Electric Co. (GE), Citigroup Inc.(C), Alcoa Inc. (AA) and Boeing Co. (BA). Total observations is 2281.

Table 3: Estimation results for DCC

	SPY	GE	C	AA	BA
ω	0.0044(0.0012)	0.0037(0.0013)	0.0107(0.0029)	0.0189(0.0073)	0.0271(0.0096)
κ	0.0331(0.0047)	0.0273(0.0045)	0.0522(0.0066)	0.0349(0.0049)	0.0512(0.0072)
λ	0.9623(0.0057)	0.9713(0.0047)	0.9460(0.0070)	0.9606(0.0064)	0.9403(0.0099)
α			0.0095(0.0010)		
β			0.9881(0.0015)		

This table reports the posterior mean and the posterior standard deviation in parenthesis for the DCC model of Section 3.1.1.

Table 4: Estimation results

Parameter	Mean	NSE	0.95 DI	Ineff
Wishart-RCOV(1)				
d	0.6677	0.0001	$(0.6600, 0.6756)$	3.1170
ν	11.7707	0.0025	$(11.6259, 11.9313)$	4.7304
Wishart-RCOV(2)				
d_1	0.2791	0.0005	$(0.2618, 0.2914)$	21.0949
d_2	0.6539	0.0004	$(0.6404, 0.6706)$	10.6056
ν	14.4389	0.0034	$(14.2661, 14.6320)$	6.4359
ℓ_2	13.6644	0.0438	$(12.0000, 14.0000)$	18.0278
Wishart-RCOV(3)				
d_1	0.2553	0.0004	$(0.2415, 0.2671)$	17.6101
d_2	0.4502	0.0006	$(0.4303, 0.4715)$	17.0676
d_3	0.2651	0.0006	$(0.2413, 0.2858)$	15.5695
ν	14.6679	0.0032	$(14.4736, 14.8603)$	5.3509
ℓ_2	9.0280	0.0219	$(8.0000, 10.0000)$	13.5203
ℓ_3	64.1822	0.0294	$(63.0000, 67.0000)$	3.0019

This table reports the posterior mean, its numerical standard error (NSE), a 0.95 density interval (DI) and the inefficiency factor for model parameters.

Table 5: Posterior mean and standard deviation of lower triangular elements of A^{-1} for Wishart-RCOV(3)

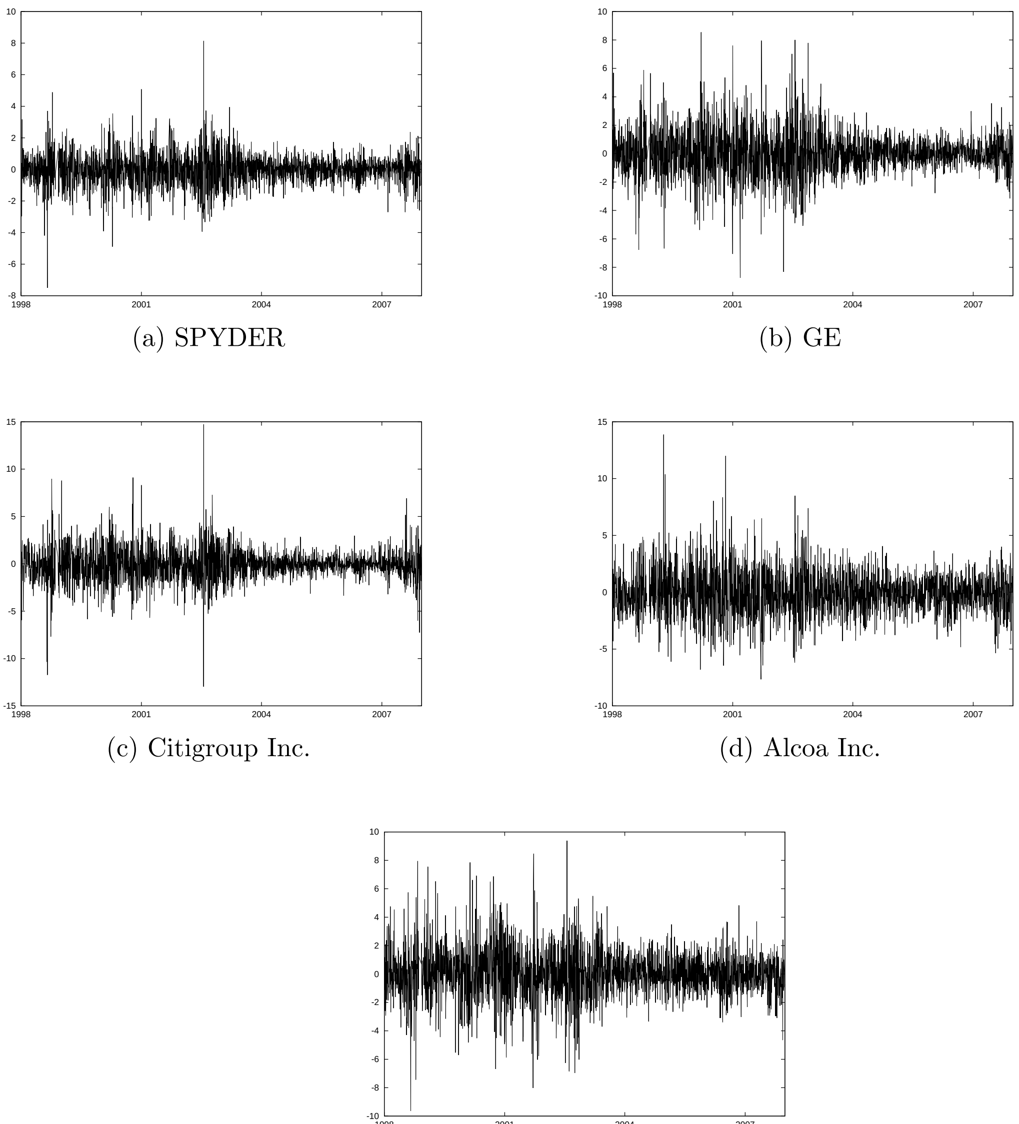
0.9548				
(0.0092)				
−0.0112	0.9256			
(0.0057)	(0.0074)			
−0.0152	−0.0082	0.9136		
(0.0056)	(0.0052)	(0.0074)		
−0.0049	−0.0048	−0.0035	0.8980	
(0.0052)	(0.0050)	(0.0050)	(0.0082)	
−0.0024	−0.0015	−0.0014	−0.0029	0.9026
(0.0051)	(0.0050)	(0.0050)	(0.0050)	(0.0076)

This table reports the posterior mean, and the posterior standard deviation in parentheses for the lower triangle of A^{-1}.

Table 6: Mean square error for different models

3-comp	2-comp	1-comp	$E(\Sigma_t \mid \Sigma_{t-1}) = \Sigma_{t-1}$	DCC	Sample covariance
70.21	70.80	75.12	102.03	85.20	118.52

Given a model's in-sample fitted value $\hat{\Sigma}_t$ and the data Σ_t this table reports $\sum_{t=1}^{T} \|\Sigma_t - \hat{\Sigma}_t\|^2$, where $\|\cdot\|$ denotes the Frobenius matrix norm.



(a) SPYDER

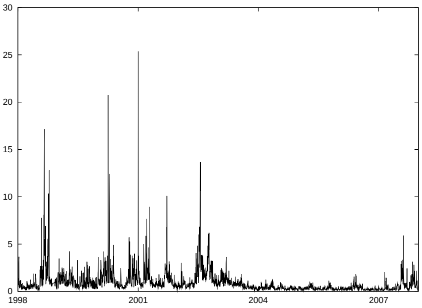
(b) GE

(c) Citigroup Inc.

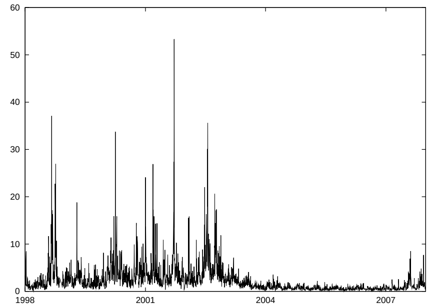
(d) Alcoa Inc.

(e) Boeing Co.

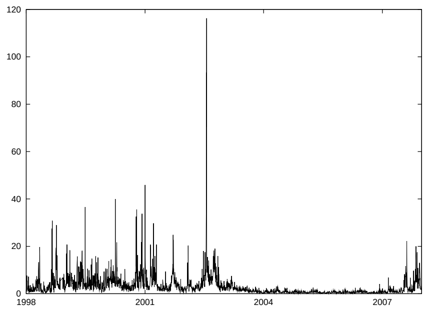
Figure 1: Daily returns



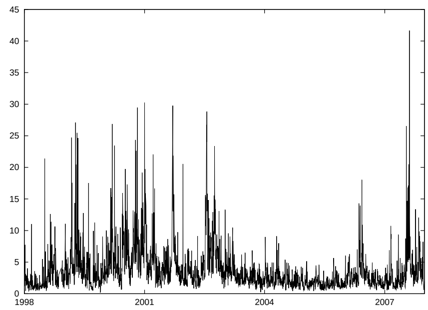
(a) SPYDER



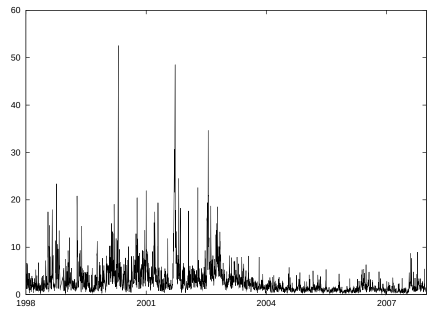
(b) GE



(c) Citigroup Inc.



(d) Alcoa Inc.



(e) Boeing Co.

Figure 2: RV for individual assets

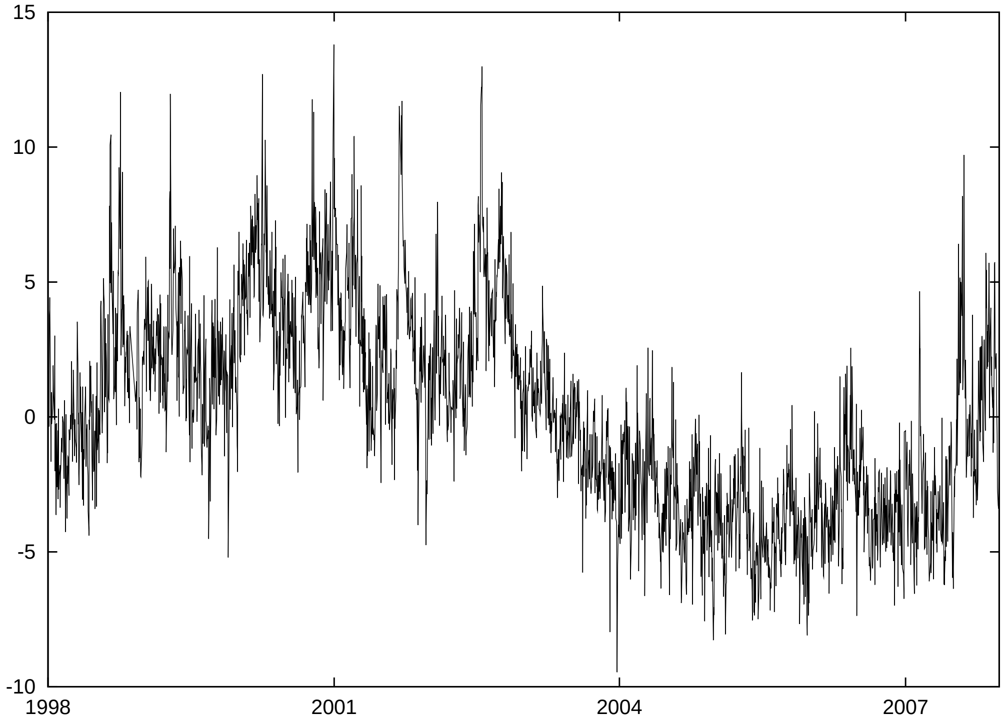


Figure 3: Determinant of RCOV

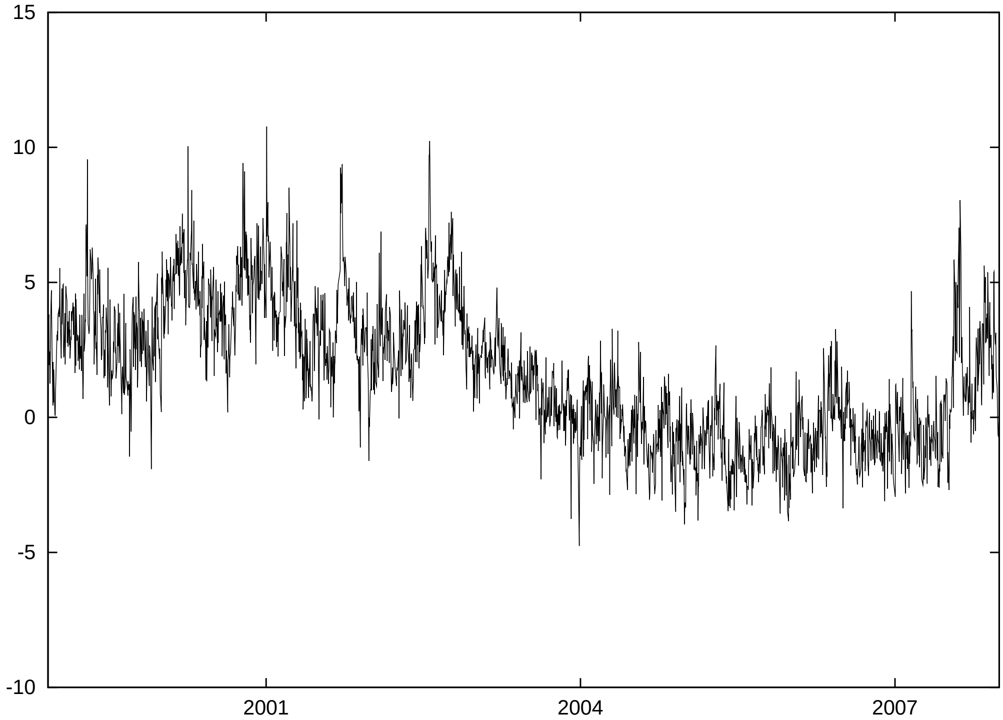


Figure 4: Determinant of “fitted RCOVs”, Wishart-RCOV(1)

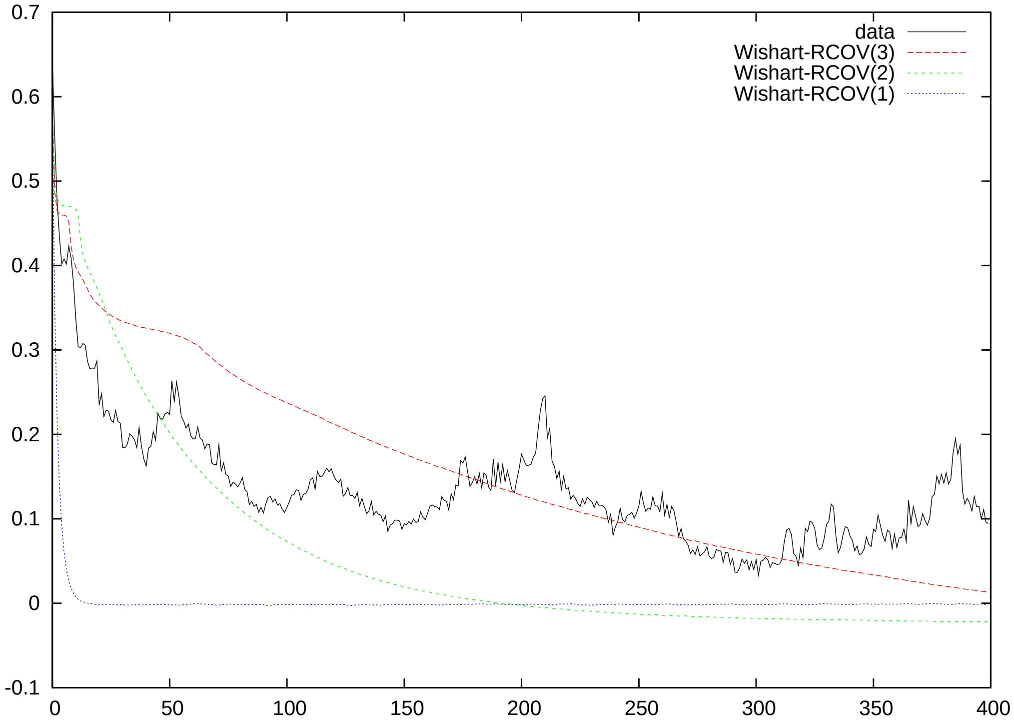


Figure 5: Sample autocorrelation functions of the largest eigenvalues of RCOVs

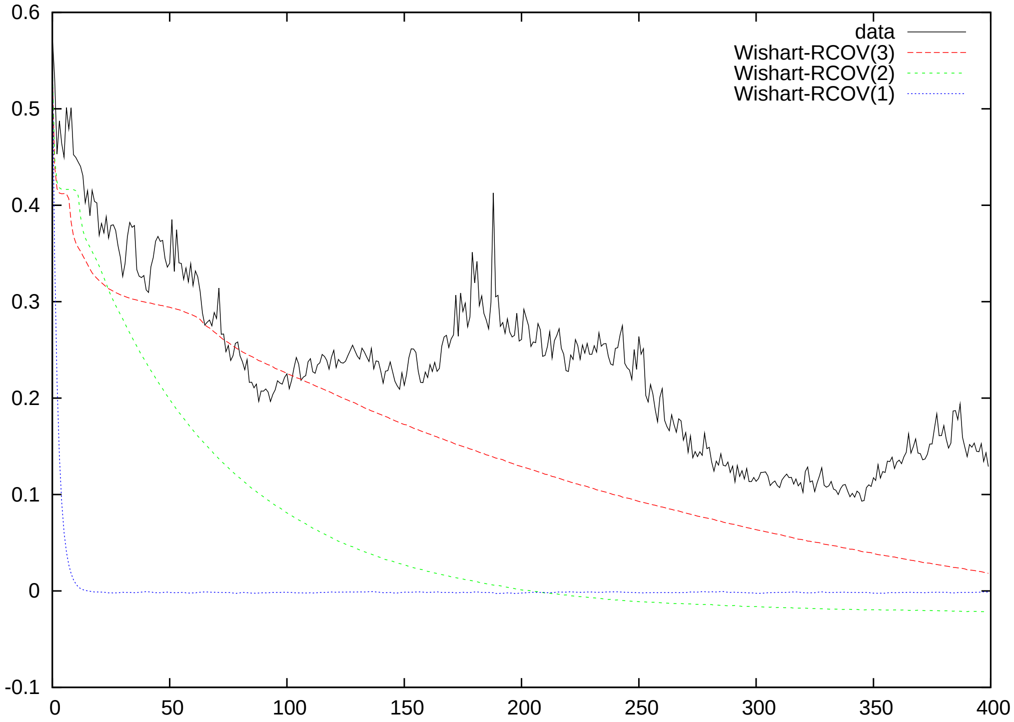


Figure 6: Sample autocorrelation functions of the smallest eigenvalues of RCOVs

Figure 7: Correlation between SPYDER and GE

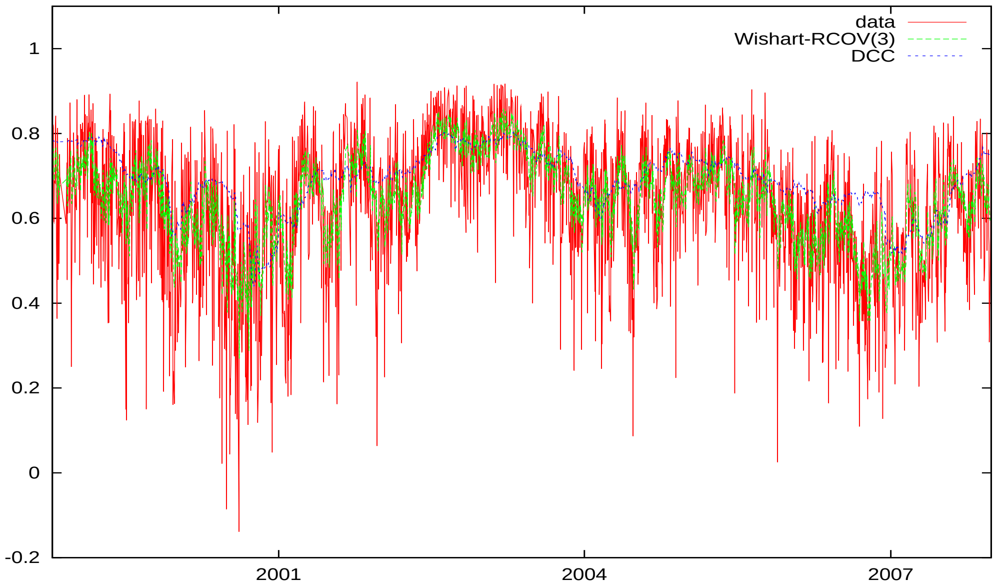
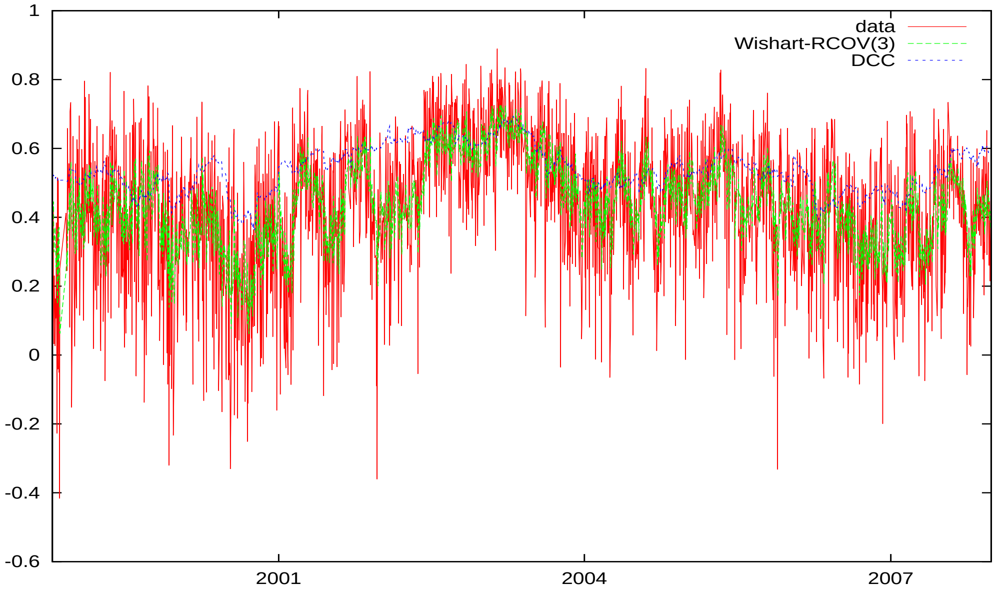


Figure 8: Correlation between GE and Citigroup Inc.



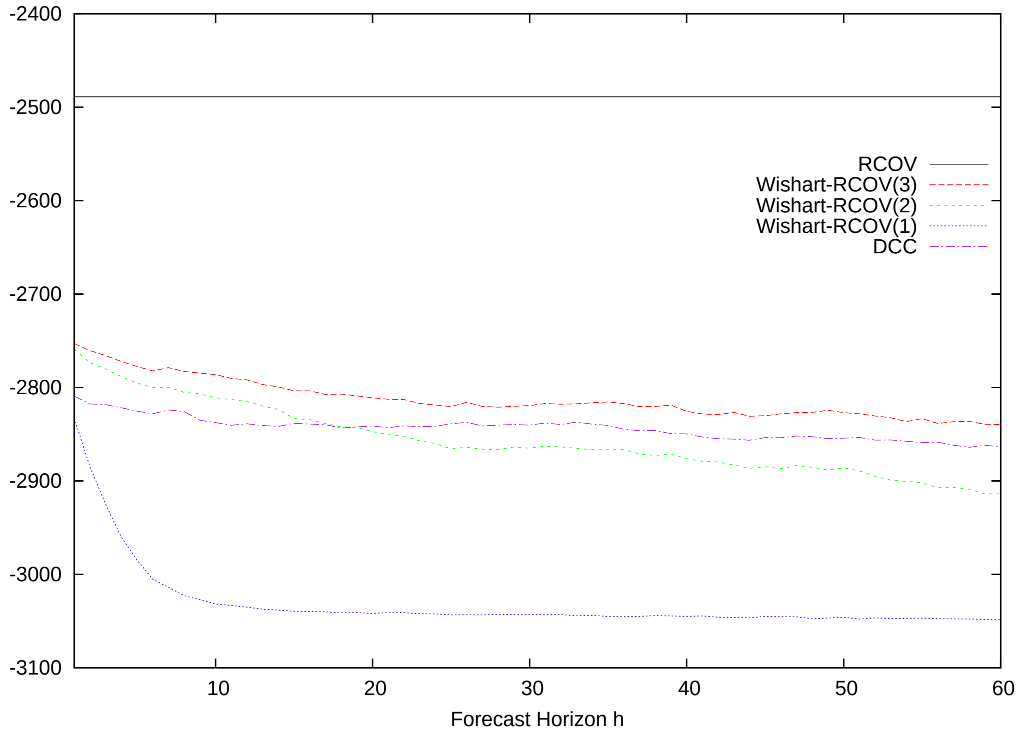


Figure 9: Term structure of cumulative log-predictive likelihood for different models

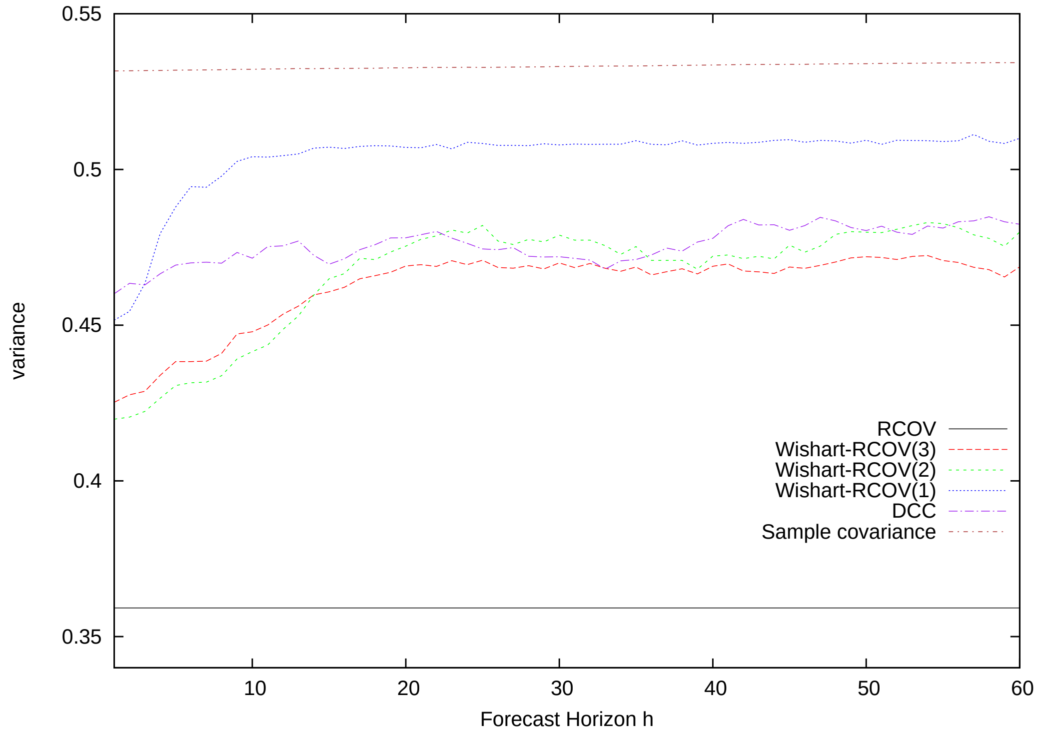


Figure 10: Sample variances of GMVPs across models against forecast horizon