Modelling Realized Covariances and Returns*

Xin Jin† John M. Maheu‡

February 2012

Abstract

This paper proposes new dynamic component models of returns and realized covariance (RCOV) matrices based on time-varying Wishart distributions. Bayesian estimation and model comparison is conducted with a range of multivariate GARCH models and existing RCOV models from the literature. The main method of model comparison consists of a term-structure of density forecasts of returns for multiple forecast horizons. The new joint return-RCOV models provide superior density forecasts for returns from forecast horizons of 1 day to 3 months ahead as well as improved point forecasts for realized covariances. Global minimum variance portfolio selection is improved for forecast horizons up to 3 weeks out.

key words: Wishart distribution, predictive likelihoods, density forecasts, realized covariance targeting, MCMC.
JEL: C11, C32, C53, G17

1 Introduction

This paper proposes new dynamic component models of returns and realized covariance (RCOV) matrices based on time-varying Wishart distributions.[1] Bayesian estimation and model comparison are discussed. While the current literature has focused on the forecasting of realized covariances, this paper demonstrates the benefits to forecasts of the return distribution from the joint modelling of returns and RCOV. We expand and empirically investigate several alternative models that have not been subjected to joint return-RCOV modelling.

*A previous version of this paper was titled "Modelling Realized Covariances". We are grateful for many helpful comments from the Editor Torben Andersen, an Associate Editor and two anonymous referees, Luc Bauwens, Christian Gourieroux, Roxana Halbleib, Ilze Kalnina, Tom McCurdy, Cathy Ning and seminar participants at CREATES, Ryerson University, the Applied Financial Time Series conference HEC Montreal, RCEA Bayesian workshop and CFE'10 London U.K. Maheu thanks the Social Sciences and Humanities Research Council of Canada for financial support.

†Department of Economics, University of Toronto, Canada, reynold.jin@utoronto.ca

‡Department of Economics, University of Toronto, Canada and RCEA Italy, jmaheu@chass.utoronto.ca

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

Multivariate volatility modelling 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 survey of MGARCH modelling 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 nonparametric 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.

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 noncentral Wishart. In a different approach Bauer and Vorkink (2011) decompose the RCOV matrix by a log-transformation and then use various time-series approaches to model the elements. Chiriac and Voev (2010) 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. Noureldin et al. (2011) introduce new multivariate models of volatility based on GARCH type parameterizations using high-frequency data.

MSV Wishart specifications for the covariance of returns are proposed in Asai and McAleer (2009) and Philipov and Glickman (2006a). These models specify a standard Wishart transition density for the inverse covariance matrix of returns.[2] In contrast to modelling the Cholesky factor or log-transformation of RCOV, contemporaneous covariances between elements in the RCOV matrix are straightforward to interpret and model using a Wishart law of motion.

Our approach is also related to the independent work of Golosnoy et al. (2012) and Asai and So (2010) who propose alternative dynamic Wishart models for stock market volatility. These papers focus on RCOV dynamics while our interest is in the joint modelling of returns and realized covariances and density forecasts. In addition, we propose component models in which the lag length of a component is estimated.

The RCOV is estimated using the realized kernels of Barndorff-Nielsen et al. (2008).[3]

[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).

[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.

The empirical analysis of 5 stocks show the strong persistence of the daily time series of RCOV elements. We propose new Wishart specifications with components to capture the persistence properties in realized covariances. 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. Two types of time-varying Wishart models are considered. The first assumes the components affect the scale matrix in an additive fashion while the second has the components enter by a multiplicative term. The additive specification performs the best in our analysis.

The models are estimated from a Bayesian perspective. We show how to estimate the length of data windows that enter into the components of the models. For each of the RCOV models 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 deliver a dramatic improvement in capturing the time series autocorrelations of the eigenvalues of the RCOV matrices.

Besides providing new tractable models for returns and realized covariances we also evaluate the models over a term structure of density forecasts of returns and a term structure of global minimum-variance portfolios.[4] It is important to consider density forecasts of returns since this is the quantity that in principle enters into all financial decisions such as risk measurement and management. In general the covariance of future returns is not a sufficient statistic for the density of returns.[5] Daily returns are common to both the MGARCH and return-RCOV models and provide a common metric to compare models that use high and low-frequency data. In contrast to the value-at-risk measures that focus on the tails of a distribution, the cumulative log-predictive likelihoods measure the accuracy of the whole return distribution. A term structure of forecasts from 1 to 60 days ahead is considered in order to assess model forecast strength at many different horizons.

An important lesson from our work is that the use of realized covariances, which exploit high-frequency intraday data, do not necessarily deliver superior density forecasts of returns. Although our preferred model does, several of the models studied in this paper that use realized covariances do not provide any improvements relative to a MGARCH model with Student-t innovations estimated from daily returns. The functional form of the dynamics of realized covariances is critical to obtaining a better characterization of the distribution of returns. Our results on density forecasts of returns are a new contribution to the literature.

Another contribution of this paper is to extend the RCOV models of Bonato et al. (2009) and Chiriac and Voev (2010) to joint return-RCOV models that we estimate by a full likelihood approach. The models serve as a comparison to the new specifications. Our additive component Wishart model provides superior density forecasts of returns and point forecasts of realized covariances. The improvements are from 1 day ahead forecasts to 3 months ahead. For global minimum variance portfolio selection most of the RCOV models give improvements beyond a MGARCH model for up to 3 weeks ahead.

The joint return-RCOV model based on Bonato et al. (2009) performs poorly relative to

[4]Maheu and McCurdy (2011) 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.

[5]For instance, the predictive density of returns in the models integrate out both parameter uncertainty and uncertainty regarding future RCOV values, making the density highly non-Gaussian.

the other specifications. The model is a HAR type parametrization[6] based on the Wishart autoregressive model of Gourieroux et al. (2009). In these models the source of time variation in the conditional mean of RCOV is the noncentrality matrix. By proposing another non-central Wishart model which has the same form of the condition mean we show that the data strongly favor time variation in the conditional mean coming through the scale matrix and not the noncentrality matrix.

In summary, we provide a new approach to modelling multivariate returns that consists of joint models of returns and RCOV matrices. Our preferred specification is a Gaussian-Wishart model in which the scale matrix in the Wishart density has additive components. This specification allows the RCOV entering the Gaussian distribution for returns to be scaled up or down and thus can accommodate any systematic biases that RCOV may have. We find that it is critical to include and estimate the components to obtain improved performance relative to MGARCH models and other existing models of RCOV. We expand and empirically investigate several alternative models that have not been subjected to joint return-RCOV modelling. Model forecasts uniformly benefit from either covariance or realized covariance targeting. Models are compared based on the quality of multiple horizon density forecasts of returns, point forecasts of RCOV and portfolio selection. Our preferred model delivers improvements in all three areas of comparison.

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, several models for returns and RCOV are introduced including several benchmark multivariate GARCH models of volatility based on daily returns. Section 4 explains the estimation procedure while the computation of density forecasts are found in Section 5. Out-of-sample model comparison results are reported in Section 6, followed by full sample estimates. Section 7 extends the models to allow return-volatility asymmetry and overnight returns. Section 8 concludes. The Appendix contains details on stationarity conditions of our favored model and posterior simulation methods.

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:

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

where $a(t)$ is a vector of drift components, $\Pi(t)$ is the instantaneous volatility matrix, and $W(t)$ is a vector of standard independent Brownian motions.[7] The quantity of interest here is $\int_0^\tau \Pi(u)\Pi'(u)du$, 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

[6]See Corsi (2009) for the heterogeneous autoregressive (HAR) model for realized volatility.

[7]Jumps are not considered in this paper. How to model individual asset jumps and common jumps among several assets is an open question which we leave for future work.

$Y(t)$,

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

is equal to its quadratic variation over the same interval,

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

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 modelling approach is Theorem 2 from Andersen, Bollerslev, Diebold and Labys (2003). They show that the daily log-return follows,

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

where $\sigma\{a(v), \Pi(v)\}_{0\le v\le 1}$ denotes the sigma-field generated by $\{a(v), \Pi(v)\}_{0\le v\le 1}$. This result corresponds to the case with no leverage effect but later on we model an asymmetric effect between lagged returns and RCOV. 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 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, can accommodate endogenous measurement noise 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. 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: refresh time sampling.
   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 of returns $\{x_j\}, j = 1, 2, \ldots, n$, daily $RCOV_t$ is calculated as,

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

where $f()$ is the Parzen kernel and Υ_s is the sample autocovariance of x_j. For full details along with the selection of bandwidth $\mathcal{S}$, see BNHLS.

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.[8] The j-th diagonal element of $RCOV_t$ is called realized variance 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.

One issue with the realized kernel estimator of BNHLS is that the larger the number of assets the less intraday data is available due to the synchronizing scheme of the refresh time sampling. Intuitively, the more stocks involved, the longer it takes for all of them to trade. Furthermore, if the asset mix displays high variation in the frequencies of the intraday price observations across stocks, (e.g., both liquid and illiquid stocks are included), the loss in data would increase further, as the sampling points are determined by the slowest trading asset. Hautsch et al. (2010) introduce one approach to overcome this issues. In our empirical study where dimension is moderate ($k = 5$) and the assets are liquid, the original realized kernel estimator of BNHLS proves to be sufficient.

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.[9] Finally, any trade that has a price increase (decrease) of

[8] Throughout the paper realized covariance (RCOV) is used instead of realized kernel.

[9] 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.

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. Daily returns, conditional on RCOV, are approximately Gaussian. For instance, standardized returns have skewness close to 0 and excess kurtosis between 0 and 1 while Ljung-Box statistics show no evidence of autocorrelation.

3 Models

3.1 New Joint Models of Returns and Realized Covariances

Compared to existing approaches which model factors of RCOV matrices (Cholesky factors, Chiriac and Voev (2010), principle components, Bauer and Vorkink (2011)) an advantage of the Wishart distribution is that it has support over symmetric positive definite matrices and allows for the joint modelling of all elements of a covariance matrix. Conditional moments between realized variances and covariances have closed form expressions.

Motivated by Philipov and Glickman (2006a) and Asai and McAleer (2009), we propose to model the dynamics of RCOV by a time-varying Wishart distribution. This choice is similar to Gourieroux, Jasiak, and Sufana (2009) however they use a noncentral Wishart distribution. We have also explored the inverse Wishart density as another distribution to govern the dynamics of realized covariances but found the Wishart provided superior performance.[10]

Two models are presented in which the scale matrix of the Wishart distribution follows an additive and multiplicative structure. Both models feature components, which is important to providing gains against standard multivariate GARCH models, and accounting for persistence in RCOV elements.

The approach to modelling components is related to Andersen, Bollerslev and Diebold (2007), Corsi (2009), Maheu and McCurdy (2011) among others which uses the Heterogeneous AutoRegressive (HAR) model of realized variance in the univariate case in order to capture long-memory like features of volatility parsimoniously.

[10]Note that the choice of the distribution governing the dynamics of Σ_t is unrelated to the Bayesian conjugate analysis that uses the Wishart as a conjugate prior for Σ_t^{-1} for Gaussian observations.

3.1.1 An Additive Component Wishart Model

Let $\Sigma_t \equiv RCOV_t$, then the Wishart-RCOV-A(K) model with $K \geq 1$ components is defined as,

$$
\begin{aligned}
r_t|\Sigma_t &\sim N(0, \Sigma_t^{1/2}\Lambda(\Sigma_t^{1/2})') && (5)\\
\Sigma_t|\nu, S_{t-1} &\sim Wishart_k(\nu, S_{t-1}) && (6)\\
\nu S_t &= B_0 + \sum_{j=1}^{K} B_j \odot \Gamma_{t,\ell_j} && (7)\\
\Gamma_{t,\ell} &= \frac{1}{\ell}\sum_{i=0}^{\ell-1} \Sigma_{t-i} && (8)\\
B_j &= b_j b_j', \qquad j = 1, \ldots, K && (9)\\
1 = \ell_1 < &\cdots < \ell_K. && (10)
\end{aligned}
$$

$Wishart_k(\nu, S_{t-1})$ denotes a Wishart distribution over positive definite matrices of dimension k with $\nu > k-1$ degrees of freedom and scale matrix S_{t-1}.[11] $\odot$ denotes the Hadamard product of two matrices. Parameters are B_0, ν, $b_1, \ldots, b_K$, $\ell_2, \ldots, \ell_K$. B_0 is a $k \times k$ symmetric positive definite matrix, and b_j's are $k \times 1$ vectors making B_j rank 1. This specification ensures S_t is symmetric positive definite.

Λ is a symmetric positive definite matrix and allows the covariance of returns to deviate from the RCOV measure. Λ is estimated and provides a simple way to scale Σ_t up or down and thus accommodate any systematic biases that RCOV may have. Except for the first component, each $\Gamma_{t,\ell}$ is an average of past Σ_t over ℓ observations. Rather than preset the components to weekly and monthly terms each ℓ is estimated. The components are found to be critical to providing improvements to forecasts.

By the properties of the Wishart distribution, the conditional expectation of Σ_t is:

$$E(\Sigma_t|\Sigma_{1:t-1}) = \nu S_{t-1} = B_0 + \sum_{j=1}^{K} B_j \odot \Gamma_{t-1,\ell_j}, \tag{11}$$

where $\Sigma_{1:t-1} = \{\Sigma_1, \ldots, \Sigma_{t-1}\}$.[12] Conditional moments are straightforward to obtain and interpret. The conditional variance of element (i, j) is

$$\text{Var}(\Sigma_{t,ij}|S_{t-1}, \nu) = \frac{1}{\nu}[\tilde{S}^2_{t-1,ij} + \tilde{S}_{t-1,ii}\tilde{S}_{t-1,jj}] \tag{12}$$

where $\tilde{S}_{t-1,ij}$ is element (i, j) of (11). The conditional variance is increasing in $\tilde{S}_{t-1,ij}$, $\tilde{S}_{t-1,ii}$, and $\tilde{S}_{t-1,jj}$. The conditional covariance between elements has a similar form,

$$\text{Cov}(\Sigma_{t,ij}, \Sigma_{t,km}|S_{t-1}, \nu) = \frac{1}{\nu}[\tilde{S}_{t-1,ij}\tilde{S}_{t-1,km} + \tilde{S}_{t-1,ii}\tilde{S}_{t-1,jj}] \tag{13}$$

[11] For a detailed discussion of the Wishart distribution including various Edgeworth type expansions see Kollo and von Rosen (2005).

[12] The inverse of RCOV follows the inverse-Wishart distribution with the conditional expectation being:

$$E(\Sigma_t^{-1}|\Sigma_{1:t-1}) = (\nu - k - 1)^{-1} S_{t-1}^{-1}.$$

The degree of freedom parameter ν determines how tight the density of Σ_t is centered around its conditional mean, with larger ν meaning the random matrices are more concentrated around νS_{t-1}. Thus, the modelling of the scale matrix and ν are the key factors in affecting the conditional moments of Σ_t. If $B_1 = \ldots, B_K = 0$ then $\Sigma_t \sim i.i.d.Wishart_k(\nu, B_0/\nu)$. On the other hand if $B_0 = 0$, $B_1 = \iota\iota'$ and $K = 1$ we obtain $E(\Sigma_t|\Sigma_{1:t-1}) = \Sigma_{t-1}$.

Each element (i, j) of the scale matrix is a function only of element (i, j) of lagged Σ_t. Many other parametrization could be considered but this specification is reasonably parsimonious and performs well in the empirical work. In related independent work Golosnoy et al. (2012) consider a similar model for RCOV matrices without components but with an autoregressive structure. They provide important results on the unconditional moments and stationarity condition for our time-varying Wishart model. For instance, the unconditional mean of Σ_t is finite if all elements of $\sum_{j=1}^{K} B_j$ are less than 1 in modulus. For the existence of the unconditional second moment (and hence the stationarity condition), see Appendix 9.1.

Instead of estimating B_0, we implement RCOV targeting by setting $B_0 = (\iota\iota' - B_1 - \cdots - B_K) \odot \overline{\Sigma}_t$, where $\overline{\Sigma}_t$ is the sample mean of Σ_t. This ensures that the long-run mean of Σ_t is equal to $\overline{\Sigma}_t$. In estimation we reject any posterior draws in which B_0 is not positive definite.

3.1.2 A Multiplicative Component Wishart Model

Related to the SV model Philipov and Glickman (2006a) we propose a multiplicative Wishart model. The Wishart-RCOV-M(K) model with $K \geq 1$ components is defined as,

$$
\begin{aligned}
r_t|\Sigma_t &\sim N(0, \Sigma_t^{1/2}\Lambda(\Sigma_t^{1/2})') && (14)\\
\Sigma_t|\nu, S_{t-1} &\sim Wishart_k(\nu, S_{t-1}) && (15)\\
S_t &= \frac{1}{\nu}\left[\prod_{j=K}^{1}\Gamma_{t,\ell_j}^{\frac{d_j}{2}}\right] A \left[\prod_{j=1}^{K}\Gamma_{t,\ell_j}^{\frac{d_j}{2}}\right] && (16)\\
\Gamma_{t,\ell} &= \frac{1}{\ell}\sum_{i=0}^{\ell-1}\Sigma_{t-i} && (17)\\
1 = \ell_1 < &\cdots < \ell_K. && (18)
\end{aligned}
$$

A is a positive definite symmetric parameter matrix and d_j is a positive scalar.

The components enter as a sample average of past Σ_t raised to a different matrix power $d_k/2$.[13] The first component is assumed to be a function of only Σ_t, $\ell_1 = 1$. 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 (14) the order of the product operator is important and differs in the two terms.

To discuss some of the features of this model consider the special case with $K = 1$ component, $S_t = \frac{1}{\nu}(\Sigma_t^{d_1/2})A(\Sigma_t^{d_1/2})$. By the properties of the Wishart distribution, the

[13] We also examined a *geometric* average version using the following specification: $\Gamma_{t,\ell}^d \equiv \Sigma_{t-\ell+1}^{\frac{d}{\ell}}\Sigma_{t-\ell+2}^{\frac{d}{\ell}}\cdots\Sigma_t^{\frac{d}{\ell}}$. 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.

conditional expectation of Σ_t is:

$$E(\Sigma_t|\Sigma_{1:t-1}) = \nu S_{t-1} = (\Sigma_{t-1}^{d_1/2})A(\Sigma_{t-1}^{d_1/2}). \quad (19)$$

Additional conditional moments for Σ_t follow the Wishart-RCOV-A(K) discussion above.

The scalar parameter d_1 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 larger d_1 the stronger the persistence. Suppose A is the identity matrix and $d_1 = 1$, then by equation (19), $E(\Sigma_t|\Sigma_{1:t-1}) = \nu S_{t-1} = \Sigma_{t-1}$, which is a random walk in matrix form. If $d_1 = 0$, then $E(\Sigma_t|\Sigma_{1:t-1}) = A$, so the RCOV matrix follows an *i.i.d.* Wishart distribution over time.

By expanding to several components each with a different window lag length ℓ_j and parameter d_j, we obtain a richer model to capture the time series dependencies in realized covariances. Unfortunately, we do not know the unconditional moments for this model with K components, nevertheless, our Bayesian estimation and model comparison approach does not depend on this.

3.2 Benchmark return-RCOV Models

In this section we extend existing specifications for RCOV dynamics by Chiriac and Voev (2010) and Bonato et al. (2009) to joint return-RCOV models to compare with our new models.

3.2.1 Cholesky-VARFIMA$(1, m, 1)$

Apply the Cholesky decomposition to Σ_t such that $\Sigma_t = L_t L_t'$, where L_t is lower triangular. Let $Z_t = \text{vech}(L_t')$ be the $\frac{k(k+1)}{2} \times 1$ vector obtained by stacking the upper triangular elements of L_t'. Chiriac and Voev (2010) propose to model Z_t as a vector autoregressive fractionally integrated moving average (VARFIMA(p, m, q)) model. A restricted VARFIMA(1,m,1) specification is shown to forecast Σ_t well. Extending this model to include returns we have,

$$\begin{aligned} r_t|\Sigma_t &\sim N(0, \Sigma_t^{1/2}\Lambda(\Sigma_t^{1/2})') && (20)\\ (1-\delta L)(1-L)^m I[Z_t - c] &= (1-\psi L)\xi_t, \qquad \xi_t \sim N(0, \Xi). && (21) \end{aligned}$$

There is a common long-memory parameter m to each element of the Cholesky decomposition. The parameters here are δ, m, ψ, c, Ξ. δ, m, ψ are scalars, c is a $\frac{k(k+1)}{2} \times 1$ vector of constants, and Ξ is a $\frac{k(k+1)}{2} \times \frac{k(k+1)}{2}$ symmetric positive definite matrix. Regarding the mean vector c, we follow Chiriac and Voev (2010) to set it at the sample mean of Z_t in estimation, which leaves the number of parameters to be estimated equal to $3 + \frac{k(k+1)}{2}(\frac{k(k+1)}{2} + 1)/2$.[14] Chiriac and Voev (2010) apply the conditional maximum likelihood method developed in Beran (1995). With the same spirit in our Bayesian setting, we follow Ravishanker and Ray (1997)(Section 2.2) and construct the posterior distribution based on the conditional likelihood function, rather than the exact likelihood function. See Appendix 9.2.4 for details.

[14] As pointed out in Chiriac and Voev (2010), Ξ is irrelevant for constructing a point forecast. However, it's used in determining the forecast errors, and is also needed in simulation.

3.2.2 Wishart Autoregressive Model

Gourieroux et al. (2009) introduce the Wishart Autoregressive process (WAR) to model the dynamics of RCOV by a noncentral Wishart distribution. The Wishart Autoregressive process of order 1 (WAR(1)) is defined as

$$\begin{aligned} \Sigma_t|\nu, S, V_{t-1} &\sim NCW_k(\nu, S, V_{t-1}) \quad (22) \\ V_t &= M\Sigma_t M'. \quad (23) \end{aligned}$$

$NCW_k(\nu, S, V_{t-1})$ denotes a noncentral Wishart distribution over positive definite matrices of dimension k. ν is the real-valued degree of freedom and $\nu > k-1$. S is the scale matrix, which is symmetric positive definite. V_t is the noncentrality matrix, which is symmetric positive semi-definite. M is the $k \times k$ matrix of autoregressive parameters. The central Wishart previously discussed is a special case with $V_{t-1} = 0$.

Bonato et al. (2009) propose a block structure on the matrix M to reduce the number of parameters and also incorporate the HAR structure of Corsi (2009) to account for persistence in RCOV. In our paper we implement their diagonal-HAR-WAR specification extended to a joint model with returns as follows:

$$\begin{aligned} r_t|\Sigma_t &\sim N(0, \Sigma_t^{1/2}\Lambda(\Sigma_t^{1/2})') \quad (24) \\ \Sigma_t|\nu, S, V_{t-1} &\sim NCW_k(\nu, S, V_{t-1}) \\ V_t &= M_1\Gamma_{t,1}M_1' + M_2\Gamma_{t,5}M_2' + M_3\Gamma_{t,22}M_3' \quad (25) \\ \Gamma_{t,\ell} &= \frac{1}{\ell}\sum_{i=0}^{\ell-1}\Sigma_{t-i}, \qquad \ell = 1, 5, 22. \quad (26) \end{aligned}$$

where M_1, M_2, M_3 are diagonal matrices. Under this specification, the conditional mean of Σ_t becomes:

$$E(\Sigma_t|\Sigma_{1:t-1}) = M_1\Gamma_{t-1,1}M_1 + M_2\Gamma_{t-1,5}M_2 + M_3\Gamma_{t-1,22}M_3 + \nu S. \quad (27)$$

In estimation, we reparametrize S by its Cholesky factor L_S (i.e. $S = L_S L_S'$, L_S is lower triangular), and restrict the diagonal elements of L_S to be positive. For M_1, M_2 and M_3, we restrict the $(1,1)$ element of each matrix to be positive for identification purpose. For ν, we consider 2 cases. In the first case, we impose the condition that $\nu > k-1$. In the second case, in addition to the first condition, we also restrict it to be integer-valued for the purpose of simulation.[15] All forecasting and empirical applications (where simulation is needed) are based on the second case. See Appendix 9.2.5 for estimation details.

[15] To simulate a noncentral Wishart, we use the method proposed by Gleser (1976). In fact this is the only method we know of that is practically feasible and easy to implement. For this method to work, however, either ν needs to be greater than $2k-1$, or ν needs to be an integer. There is no easy way to simulate noncentral Wishart in all cases (Gauthier and Possamai 2009). In estimation we first allow ν to be any real number greater than $k-1$, which results in a posterior mean around 8.4. In our empirical work $k = 5$, this result does not satisfy the condition of $\nu > 2k-1$, in which case we cannot simulate Σ_t. To solve the problem, we estimate the model and restrict ν to be integer-valued, which results in a posterior mean of 8. See Table 10

3.3 GARCH Models of Daily Returns

3.3.1 Vector-diagonal GARCH Model

Ding and Engle (2001) introduce the vector-diagonal GARCH (VD-GARCH-t) model to which we add Student-t innovations as follows

$$
\begin{aligned}
r_t | r_{1:t-1} &\sim t(0, H_t, \zeta) & (28)\\
H_t &= CC' + aa' \odot r_{t-1} r_{t-1}' + bb' \odot H_{t-1}, & (29)
\end{aligned}
$$

where r_t is a k-dimensional daily return series, and $r_{1:t-1} = \{r_1, \ldots, r_{t-1}\}$. The parameters are C, a $k \times k$ lower triangular matrix; a and b are $k \times 1$ vectors, and ζ is the degree of freedom in the Student-t density. In estimation, covariance targeting is achieved by replacing CC' with $\text{Cov}(r) \odot \left(\frac{\zeta-2}{\zeta}\iota\iota' - aa' - \frac{\zeta-2}{\zeta}bb'\right)$, where $\text{Cov}(r)$ is the sample covariance matrix estimated from daily returns, and ι is a $k \times 1$ vector of 1. This model assumes that the conditional covariance $h_{ij,t}$ is only a function of the past shock $r_{i,t-1}r_{j,t-1}$, and the past conditional covariance $h_{ij,t-1}$. The conditional covariance of returns is $\frac{\zeta}{\zeta-2}H_t$ assuming $\zeta > 2$.

Besides the VD-GARCH-t model above, we also include a component VD-GARCH-t model that is analogous to the additive component Wishart Model:

$$
\begin{aligned}
r_t | r_{1:t-1} &\sim t(0, H_t, \zeta) & (30)\\
H_t &= CC' + \sum_{j=1}^{K} c_j c_j' \odot \tilde{\Gamma}_{t-1,\ell_j} & (31)\\
\tilde{\Gamma}_{t,\ell} &= \frac{1}{\ell}\sum_{i=0}^{\ell-1} r_{t-i} r_{t-i}' & (32)\\
1 = \ell_1 < &\cdots < \ell_K. & (33)
\end{aligned}
$$

c_j are $k \times 1$ vectors for $j = 1, \ldots, K$. $K \geq 1$ is the number of components computed from daily returns. In our empirical study, we let $K = 3$. Except for the first component, each $\tilde{\Gamma}_{t,\ell}$ is an average of past $r_t r_t'$ over ℓ observations. Each ℓ is estimated, as is the case in the component Wishart Models. In estimation, covariance targeting is achieved similarly by replacing CC' with $\text{Cov}(r) \odot \left(\frac{\zeta-2}{\zeta}\iota\iota' - c_1 c_1' - \ldots - c_K c_K'\right)$.

3.3.2 Dynamic Conditional Correlation Model

The second model is a dynamic conditional correlation (DCC-t) model of Engle (2002) with Student-t innovations,

$$\begin{aligned} r_t|r_{1:t-1} &\sim t(0, H_t, \zeta) && (34)\\ H_t &= D_t R_t D_t && (35)\\ D_t &= \text{diag}(\sigma_{i,t}) && (36)\\ \sigma^2_{i,t} &= \omega_i + \kappa_i r^2_{i,t-1} + \lambda_i \sigma^2_{i,t-1}, i = 1, \dots, k && (37)\\ \epsilon_t &= \left(\frac{\zeta-2}{\zeta}\right)^{1/2} D_t^{-1} r_t && (38)\\ Q_t &= \overline{Q}(1-\alpha-\beta) + \alpha\epsilon_{t-1}\epsilon'_{t-1} + \beta Q_{t-1} && (39)\\ R_t &= \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}. && (40) \end{aligned}$$

D_t, R_t, Q_t, $\overline{Q}$ are all $k \times k$ matrices. The parameters are $\omega_1, \dots, \omega_k$, $\kappa_1, \dots, \kappa_k$, $\lambda_1, \dots, \lambda_k$, α, β, ζ. Following Engle (2002), $\overline{Q}$ is replaced by $\text{Corr}(\epsilon_t)$, the sample correlation. In this way the number of parameters is greatly reduced from $\frac{k^2+5k}{2} + 2$ to $3k + 2$. Equation (37) governs the dynamics of the conditional variances of each individual return by a univariate GARCH process; equation (39) governs the dynamics of the time-varying conditional correlation of the whole return vector. Because $\text{Corr}(\epsilon_t)$ 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.

4 Model Estimation

We apply standard Bayesian estimation techniques to estimate the models using MCMC methods for posterior simulation. The posterior distribution is unknown for all the models considered, but a Markov Chain that has as its limiting distribution the posterior distribution of the parameters of interest can be sampled from using MCMC simulations. Features of the posterior density can then be estimated consistently based on the samples obtained from the posterior. 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).

In the following we outline estimation for the Wishart-RCOV-A(3) model and provide specific details for this model and others in the Appendix. To apply Bayesian inference, we need to first assign priors to the parameters. In general all priors are uninformative but proper. The priors on the elements of b_j's are all $N(0, 100)$, except the first element of each b_j is truncated to be positive for identification purposes. For the degree of freedom parameter $\nu \sim \exp(\lambda_0) I_{\nu > k-1}$, an exponential distribution with support truncated to be greater than $k-1$.[16] To make the prior flat, λ_0 is set to 100. In the empirical work focus is given to $K = 3$ components as this was found to produce good results. The priors for ℓ_2 and ℓ_3 are uniform discrete with support $\{2, 3, \dots, 200\}$, with the restriction that $\ell_2 < \ell_3$ for identification. We assume independence among the prior distributions of parameters.

[16]In posterior simulation only draws of $\nu > k - 1$ are accepted.

The joint density of returns and realized covariances is decomposed as

$$p(r_t, \Sigma_t|\Lambda, \Theta, r_{1:t-1}, \Sigma_{1:t-1}) = p(r_t|\Lambda, \Sigma_t)p(\Sigma_t|, \Theta, \Sigma_{1:t-1}) \tag{41}$$

where Θ is the parameters in the RCOV specification. $p(r_t|\Lambda, \Sigma_t)$ has a density in (5) while $p(\Sigma_t|\Theta, \Sigma_{1:t-1})$ has the density from (6). Equation (41) implies that estimation of Λ and Θ can be done separately.

Bayes' rule gives the posterior for Θ in the Wishart model as

$$p(\Theta|\Sigma_{1:T}) \propto \left[\prod_{t=1}^{T} p(\Sigma_t|\Theta, \Sigma_{1:t-1})\right] p(\Theta) \tag{42}$$

where $p(\Theta)$ is the prior discussed above. Conditional distributions used in posterior simulation are proportional to this density.

The Wishart-RCOV-A(3) model has parameters $\Theta = \{b_1, b_2, b_3, \ell, \nu\}$, with $\ell = \{\ell_2, \ell_3\}$. MCMC sampling iterates making parameter draws from the following conditional distributions.

- $\Theta_j|\Theta_{-j}, \Sigma_{1:T}$,

where Θ_j denotes one element of the parameter vector Θ and Θ_{-j} is Θ excluding Θ_j. For the parameters in b_1, b_2, b_3, ν a single-move Metropolis-Hastings step using a random walk proposal is employed. The conditional posterior of ℓ_2 and ℓ_3 has support on discrete points and the proposal density is a random walk with Poisson increments that are equally likely to be positive or negative.

Taking a draw from all of the conditional distributions constitutes one sweep of the sampler. After dropping an initial set of draws as burnin we collect N draws to obtain $\{\Theta^{(i)}\}_{i=1}^N$. 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 $N^{-1}\sum_{i=1}^N \Theta^{(i)}$.

Posterior simulation from $\Lambda|r_{1:T}, \Sigma_{1:T}$ is based on recognizing that $\tilde{r}_t = \Sigma_t^{-1/2} r_t \sim N(0, \Lambda)$. Setting the prior density of Λ^{-1} to $Wishart_k(k+1, I_k)$, results in a standard conjugate result for the multivariate normal model. This is done separately from the estimation for the RCOV models.

All of the details of the conditional distributions and proposal distributions along with details for the other models are collected in the Appendix.

4.1 Monte Carlo Study

To assess the estimation procedure for our preferred Wishart-RCOV-A model, we present a Monte Carlo study on a 3-dimensional system. We focus our study on a Wishart-RCOV-A(3) model with data generating process (DGP)

$$\Sigma_t|\nu, S_{t-1} \sim Wishart_3(\nu, S_{t-1}) \tag{43}$$

$$\nu S_t = B_0 + \sum_{j=1}^{3} b_j b_j' \odot \Gamma_{t,\ell_j}, \tag{44}$$

where parameter values are found in the second column of Table 3. We simulate 1000 samples from the model using the same set of DGP values. The sample size is 2000 each. We then apply the estimation procedure described earlier to each sample, where we calculate the posterior mean of a particular parameter as its estimator. Table 3 records the mean and the root mean squared error (RMSE) of the estimator. The results show that the estimation procedure accurately estimates the model parameters. Other DGP values yield similar results.

5 Density Forecasts of Returns

It is important to consider density forecasts of returns since this is the quantity that in principle enters into all financial decisions such as portfolio choice and risk measurement.[17] Another reason for comparing models this way is that the daily returns are common to both the GARCH and the joint return-RCOV models and provides a common metric to compare models that use high and low frequency data. In contrast to the value-at-risk measures that focus on the tails of a distribution the predictive likelihoods test the accuracy of the whole distribution. Finally, a term structure of forecasts is considered in order to assess model forecast strength at many different horizons.

From a Bayesian perspective the predictive likelihoods are a key input into model comparison through predictive Bayes factors (Geweke (2005)).[18] Following Maheu and McCurdy (2011) we evaluate a term structure of a model's density forecasts of returns. This is the cumulative log-predictive likelihood based on out-of-sample data for $h = 1, ..., H$ period ahead density forecasts of returns.

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

$$\hat{p}_h^{\mathcal{A}} = \sum_{t=T_0-h}^{T-h} \log(p(r_{t+h}|I_t, \mathcal{A})), \tag{45}$$

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 ahead forecasts to 10 day ahead forecasts using model $\mathcal{A}$. We are also interested in comparing $\hat{p}_h^{\mathcal{A}}$ for a fixed h with another specification $\mathcal{B}$, using its cumulative log-predictive likelihood $\hat{p}_h^{\mathcal{B}}$. Better models, in terms of more accurate predictive densities, will have larger (45).

For the joint return-RCOV models $I_t = \{r_{1:t}, \Sigma_{1:t}\}$ while for the MGARCH models $I_t = \{r_{1:t}\}$. 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},

$$p(r_{t+h}|I_t, \mathcal{A}) = \int p(r_{r+h}|\theta, \Omega_{t+h}, \mathcal{A}) p(\Omega_{t+h}|\theta, I_t, \mathcal{A}) p(\theta|I_t, \mathcal{A}) \mathrm{d}\theta \mathrm{d}\Omega_{t+h}. \tag{46}$$

[17]In general the covariance of returns is not a sufficient statistic for the future return distribution except with a Gaussian assumption.

[18]Classical approaches to comparison of density forecasts include Amisano and Giacomini (2007), Bao, Lee, and Saltoglu (2007) and Weigend and Shi (2000).

Parameter uncertainty from θ and the future latent covariance of returns Ω_{t+h} are both integrated out and will in general result in a highly non-Gaussian density on the left hand side of (46). In the DCC-t and VD-GARCH-t models $\Omega_t \equiv H_t$ while for each of the models that exploit RCOV information $\Omega_t \equiv \Sigma_t$, while θ is the respective parameter vector. The integration is approximated as

$$\int p(r_{t+h}|\theta, \Omega_{t+h}, \mathcal{A}) p(\Omega_{t+h}|\theta, I_t, \mathcal{A}) p(\theta|I_t, \mathcal{A}) \mathrm{d}\theta \mathrm{d}\Omega_{t+h} \approx \frac{1}{N} \sum_{i=1}^{N} p(r_{t+h}|\theta^{(i)}, \Omega_{t+h}^{(i)}, \mathcal{A}), \quad (47)$$

where $\Omega_{t+h}^{(i)} \sim p(\Omega_{t+h}|\theta^{(i)}, I_t, \mathcal{A})$, and $\theta^{(i)} \sim p(\theta|I_t, \mathcal{A})$. $\{\theta^{(i)}\}_{i=1}^{N}$ are the MCMC draws from the posterior distribution $p(\theta|I_t, \mathcal{A})$ for the model given the information I_t.

For the GARCH models, $p(r_{t+h}|\theta^{(i)}, \Omega_{t+h}^{(i)}, \mathcal{A})$ is the pdf of a multivariate Student-t density with mean 0, scale matrix $H_{t+h}^{(i)}$ and degree of freedom $\zeta^{(i)}$ evaluated at r_{t+h}. $H_{t+h}^{(i)}$ is simulated out from the last in-sample value $H_t^{(i)}$ which is computed using the GARCH recursion and the parameter draw $\theta^{(i)}$ from the posterior density given data $I_t = \{r_{1:t}\}$.

For the RCOV models, $p(r_{t+h}|\theta^{(i)}, \Omega_{t+h}^{(i)}, \mathcal{A})$ is the pdf of a multivariate Normal density with mean 0 and covariance $(\Sigma_{t+h}^{(i)})^{1/2} \Lambda^{(i)} ((\Sigma_{t+h}^{(i)})^{1/2})'$ evaluated at r_{t+h}. $\Sigma_{t+h}^{(i)}$ is simulated out using the Wishart, Cholesky-VARFIMA, or diagonal-HAR-WAR dynamics of the particular model and conditional on $\theta^{(i)}, \Lambda^{(i)}$ from the posterior density, given data $I_t = \{r_{1:t}, \Sigma_{1:t}\}$.

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* (47). In other words the full set of models is recursively estimated for $t = T_0 - H, \ldots, T - 1$.

Given a model $\mathcal{A}$ with log-predictive likelihood $\hat{p}^{\mathcal{A}}$, and model $\mathcal{B}$ with log-predictive likelihood $\hat{p}^{\mathcal{B}}$, based on the common data $\{r_{T_0}, \ldots, r_T\}$, the log predictive Bayes factor in favor of model $\mathcal{A}$ versus model $\mathcal{B}$ is $\log(BF_{\mathcal{AB}}) = \hat{p}^{\mathcal{A}} - \hat{p}^{\mathcal{B}}$. The log Bayes factor is a relative ranking of the ability of the models to account for the data. A value greater than 0 means that model $\mathcal{A}$ is better able to account for the data compared to model $\mathcal{B}$. Kass and Raftery (1995) suggest interpreting the evidence for $\mathcal{A}$ as: not worth more than a bare mention if $0 \leq \log(BF_{\mathcal{AB}}) < 1$; positive if $1 \leq \log(BF_{\mathcal{AB}}) < 3$; strong if $3 \leq \log(BF_{\mathcal{AB}}) < 5$; and very strong if $\log(BF_{\mathcal{AB}}) \geq 5$.

6 Results

6.1 Density Forecasts of Returns

In this section, we compare the joint return-RCOV models to the other benchmark models, focusing on their out-of-sample 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. The full set of models is recursively estimated for $t = T_0 - H, \ldots, T - 1$ with a burnin of 1000 iterations after which N=5000 draws are collected to compute the predictive likelihoods and other predictive quantities. Figures 3 and 4 present the full range of log-predictive likelihoods for the models while Table 4 presents specific values for selected h. The table can be used to compute log predictive Bayes factors by taking the difference in $\hat{p}_h$ for two models.

Figure 3 plots $\hat{p}_h$ for the MGARCH models against $h = 1, 2, \ldots, H = 60$, giving each model a cumulative log-predictive likelihood term structure. Included are the DCC model with Gaussian innovations and the DCC-t, VD-GARCH-t and component VD-GARCH-t all with Student-t innovations. For component VD-GARCH-t, we use 3 components. All specifications have a downward sloping term structure. Intuitively, forecasting further out is more difficult. The t-distribution provides significant improvements in density forecasts of returns at all forecast horizons. In general, the VD-GARCH-t model has the best performance among the MGARCH specifications. In particular, the fact that the component VD-GARCH-t is inferior to the original VD-GARCH-t model at almost each forecast horizon suggests that the component structure provides no gains in terms of forecasting compared to the standard VD-GARCH-t specification. In further discussion, we include the VD-GARCH-t as a benchmark that uses only daily return data.

Turning to Figure 4 the term structure of log-predictive likelihood for returns is presented for several of the joint return-RCOV models. Included are the following models: Wishart-RCOV-A(3), Wishart-RCOV-M(3), Cholesky-VARFIMA, diagonal-HAR-WAR as well as the VD-MGARCH-t and a new specification, diagonal-HAR-NCW, which we will discuss below.

First the VD-MGARCH-t model is competitive and is generally producing better density forecasts than the Cholesky-VARFIMA model that exploits high frequency information. For instance, the log predictive Bayes factor in favor of the VD-MGARCH-t is 7.04, $h = 5$, 6.78, $h = 10$ and 10.66 $h = 20$.[19] Recall that the log Bayes factor represents the improvement that the GARCH model gives in describing the data. Although Chiriac and Voev (2010) demonstrate that point forecasts of RCOV matrices are improved using their model as compared to DCC alternatives this does not translate into better density forecasts of returns. The dynamics and the distribution of the RCOV matrices is also important. The Wishart-RCOV-M(3) model provide further gains but the Wishart-RCOV-A(3) dominates all competitors across the forecast horizon.[20]

To further investigate the statistical significance of these results Figures 5 and 6 display log predictive Bayes factors over each forecast horizon h. The first plot shows that the Wishart-RCOV-M(3) does not always improve on the VD-MGARCH-t model. For $h = 17 - 35$ the MGARCH model has better density forecasts. The Wishart-RCOV-A(3) beats the MGARCH model at each h. It provides substantial improvements particularly for small and large h.

The second plot shows that the Wishart-RCOV-M(3) is often better than the Cholesky-VARFIMA specification but there are forecast horizons that the latter performs well, particularly for $h > 55$. On the other hand, the Wishart-RCOV-A(3) strongly dominates the Cholesky-VARFIMA model for all h. This translates into log predictive Bayes factors on the order of 10 to 20 in favor of the Wishart-RCOV-A(3).

Next we turn to the diagonal-HAR-WAR model which is shown in Figure 4 to have very poor performance compared to all other models. Why does this occur? After exploring

[19]For $h = 1$ the VD-MGARCH-t and Cholesky-VARFIMA models have essentially the same predictive power.

[20]Regulation NMS was established in 2007 by the SEC to foster efficient and fair price formation across securities markets. Splitting our sample into data before 2007 and after shows the improvements from Wishart-RCOV-A(3) return density forecasts to be robust.

other similar specifications we conjecture that the time variation in the diagonal-HAR-WAR model comes through the wrong channel. This model makes the noncentral parameter time-varying while fixing the scale matrix. Our Wishart models have a noncentral parameter of 0 but time variation in the scale matrix. To investigate the importance of where the time variation in the model should be we propose the following diagonal-HAR-noncentral Wishart (diagonal-HAR-NCW) specification as

$$\begin{aligned} \Sigma_t|\nu, S_{t-1}, V &\sim NCW_k(\nu, S_{t-1}, \nu V) \\ \nu S_t &= \tilde{M}_1 \Gamma_{t,1} \tilde{M}_1 + \tilde{M}_2 \Gamma_{t,5} \tilde{M}_2 + \tilde{M}_3 \Gamma_{t,22} \tilde{M}_3 \qquad (48) \\ \Gamma_{t,\ell} &= \frac{1}{\ell} \sum_{i=0}^{\ell-1} \Sigma_{t-i}, \qquad \ell = 1, 5, 22. \qquad (49) \end{aligned}$$

ν is the real-valued degree of freedom, S_{t-1} is the scale matrix, νV is the noncentralily matrix. $\tilde{M}_1, \tilde{M}_2, \tilde{M}_3$ are diagonal matrices. Under this specification, the conditional mean of Σ_t becomes:

$$E(\Sigma_t|\Sigma_{1:t-1}) = \tilde{M}_1 \Gamma_{t-1,1} \tilde{M}_1 + \tilde{M}_2 \Gamma_{t-1,5} \tilde{M}_2 + \tilde{M}_3 \Gamma_{t-1,22} \tilde{M}_3 + \nu V, \qquad (50)$$

and is exactly the same form as the conditional mean for the diagonal-HAR-WAR model in (27). The difference between the diagonal-HAR-WAR and the diagonal-HAR-NCW is that the roles of the scale matrix and the noncentrality matrix in the noncentral Wishart transition density are switched. In the diagonal-HAR-WAR model, the time series dependence in Σ_t is captured in the noncentrality matrix V_t, while the scale matrix S is set to a constant. In the diagonal-HAR-NCW model, however, the time dependence goes into the scale matrix S_t, while the noncentrality matrix V (up to a constant ν) is constant.

Figure 4 shows that switching the time variation from the noncentrality matrix to the scale matrix results in a huge improvement in density forecasts. Further improvements are possible for this model by estimating the lag length of the components $\Gamma_{t,\ell}$ (not reported). From Table 4 the log predictive Bayes factors in favor of the diagonal-HAR-NCW vs the diagonal-HAR-WAR model range from 75.75, $h = 1$ to 186.38, $h = 60$. We conclude that the existing WAR models, according to our results, are likely to be poor performers unless additional time variation in the conditional mean is incorporated into the scale matrix.

To see why 3 components are chosen for the Wishart RCOV models in the empirical study, Figure 7 plots $\hat{p}_h$ for Wishart-RCOV-A(3), Wishart-RCOV-A(2) and Wishart-RCOV-A(1). Wishart-RCOV-A(1) is completely dominated by the other two models. As a matter of fact, though not shown in the figure, the Wishart-RCOV-A(1) is even dominated by the DCC model with Gaussian innovations for all forecast horizons. With 2 components, Wishart-RCOV-A(2) substantially improves on Wishart-RCOV-A(1) for all forecast horizons, while Wishart-RCOV-A(3) improves the term structure even further.[21]

As mentioned before, instead of presetting the components to weekly and monthly terms (e.g., Chiriac and Voev (2010), Bonato et al. (2009)), we estimate the lag length of the components. To see the benefits of doing so, we compare the Wishart-RCOV-A(3) with ℓ_2 and ℓ_3 both estimated versus ℓ_2 and ℓ_3 fixed at 5 and 22, respectively. Figure 8 plots the

[21] We also tried a 4-component model Wishart-RCOV-A(4). It produces no significant gains in terms of density forecasts compared to Wishart-RCOV-A(3).

log-predictive Bayes factor, which strongly supports our approach of estimating as opposed to presetting the component lag lengths.

Finally, the log predictive Bayes factor found in Figure 9 compares the model of Wishart-RCOV-A(3) with Λ estimated versus $\Lambda = I$. The latter says that RCOV is perfectly synonymous with the covariance of daily returns. The evidence across the term structure of forecasts is in favor of Λ being a free estimated parameter.

In summary, the Wishart-RCOV-A(3) provides superior density forecasts for returns as compared to MGARCH models and existing RCOV models. The forecast improvements of the Wishart-RCOV-A(3) are due to: a better specification for the dynamics of RCOV compared to other RCOV models, modelling the dynamics of RCOV through the scale matrix rather than the noncentralily matrix of a Wishart distribution, including a sufficient number of components, estimation the lag length of the components and letting Λ being a free estimated parameter, as opposed to setting $\Lambda = I$.

6.2 Forecasts of RCOV

Figure 10 and Table 5 report the root mean squared error for predicting Σ_t based on the predictive mean from each model. This is reported for each of the forecast horizons h. As in the density forecasts, each model is re-estimated for each observation in the out-of-sample period to produce the predictive mean. Similar to the density forecasts, the Wishart-RCOV-A model has the best point forecasts of Σ_t amongst all the models. Both the Wishart-RCOV-M and Cholesky-VARFIMA are similar while the diagonal-HAR-WAR is the worst.[22] Compared to the diagonal-HAR-WAR, the alternative diagonal-HAR-NCW model shows great improvements, as it did in the density forecasts discussed above.

6.3 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

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

$\Omega_{t+h|t}$ is the predictive mean of the covariance matrix at time $t+h$ given time t information for a particular model. 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

$$w_{t+h|t} = \frac{\Omega^{-1}_{t+h|t}\iota}{\iota'\Omega^{-1}_{t+h|t}\iota}. \tag{51}$$

[22]Chiriac and Voev (2010) also find the diagonal-HAR-WAR model performs poorly in point forecasts.

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-A(3), Wishart-RCOV-M(3), Cholesky-VARFIMA, diagonal-HAR-WAR, and the VD-GARCH-t. As in the density forecasts the models are estimated using data up to and including time t and weights are computed from (51). For the MGARCH models $\Omega_{t+h|t} = E[\frac{\zeta}{\zeta-2} H_{t+h}|r_{1:t}]$ and for the RCOV models $\Omega_{t+h|t} = E[\Sigma_{t+h}^{1/2}\Lambda(\Sigma_{t+h}^{1/2})'|r_{1:t}, \Sigma_{1:t}]$. These terms are computed by simulation and have parameter uncertainty integrated out. Observation $t + 1$ is added and the models are re-estimated and the new weights computed, etc.

We report the sample variances[23] of the GMVPs across models in Figure 11. Table 6 reports the portfolio variance for selected values of h. 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 portfolio choice. With the exception of the diagonal-HAR-WAR model, all of the return-RCOV time-series models improve upon the VD-GARCH-t model for about 15 days out after which the portfolio variance is similar.[24] The Wishart-RCOV-M(3) model has the lowest portfolio variance but the Wishart-RCOV-A(3) model remains very competitive.

The maximum portfolio weight for the Wishart-RCOV-A(3) is fairly stable over time at 0.85. The weights from the VD-GARCH-t and Wishart-RCOV-A(3) models are often moving in the same direction but diverge significantly in 2007. The weight for the latter model tends to be a bit more volatile.

6.4 Full Sample Estimates

This section presents full sample estimates for selected models. Tables 7 and 8 report posterior moments for the new time-varying Wishart models.[25] Both models have a degree of freedom parameter of about 14 and components with windows of length 9 and 64. These components correspond to roughly 2 weeks and 3 months of data and the estimates are very precise. The component models deliver a dramatic improvement in capturing the time series autocorrelations of the eigenvalues of the RCOV matrices. These estimates and Figure 8 reinforce why setting $\ell_2 = 5, \ell_3 = 22$ is suboptimal. We found it critical to have 3 components and to estimate the window width of each component to obtain good out-of-sample results. All the 0.95 posterior density intervals show the parameters to be precisely estimated.

For the Wishart-RCOV-A model the second component with $\ell_2 = 9$ has the largest impact

[23]An alternative to comparing the sample variances is to compare the realized variance for each portfolio. However, as we discuss in the next section, Σ_t is biased for the covariance of returns and would not represent the true variance that investors face.

[24]Similar to our results, Fleming, Kirby and Ostdiek (2003) show that for a risk-averse investor using a volatility-timing strategy to allocate funds among assets, switching from a GARCH model to a covariance estimate based on intradaily returns, can provide substantial gains for one-day ahead investments.

[25]The *inefficiency factors* in the tables 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.

on the conditional mean of RCOV. The second component in the Wishart-RCOV-M(3) is also the most important with the largest d at 0.45.

The remaining joint return-RCOV model estimates are reported in Tables 9-11. The common long-memory parameter m has a posterior mean of 0.4295 in the VARFIMA specification. The parameter estimates of the diagonal-HAR-WAR model with a real valued degree of freedom parameter (not reported) are almost identical to the estimates in Table 10 which imposes an integer value of ν. The degree of freedom parameter is concentrated at 8. Recall, that for the noncentral Wishart it is necessary to impose an integer value on ν in order to simulate the model for $\nu < 2k-1$. For the diagonal-HAR-NCW model (Table 11) we estimate a $\nu = 14.55 > 2k-1 = 9$ which allows us to treat ν as real valued.[26] This latter estimate of ν is close to the estimation results for ν in the Wishart-RCOV models.

Finally, the posterior mean and standard deviation are reported for the lower triangular elements of Λ in Table 12. Λ is close to a diagonal matrix with four diagonal elements being significantly smaller than 1. The effect this matrix has is to reduce the conditional variance of returns. For instance, the determinant (generalized variance, Muirhead (1982)) is reduced since $|\Sigma_t^{1/2}\Lambda(\Sigma_t^{1/2})'|/|\Sigma_t| = |\Lambda| = 0.53$ and the eigenvalues of the covariance matrix of returns are reduced.[27] In other words, Λ serves to scale down RCOV as it enter the covariance of returns. The Bayes factor is strongly in favor of Λ being estimated versus it being set to an identity matrix (Figure 9).

7 Extensions

7.1 Asymmetric Wishart-RCOV Model

To capture the leverage effect that is common in stock returns, an asymmetric term can be included in the additive component Wishart model. More specifically, the asymmetric Wishart-RCOV-A(3) model is defined as follows:

$$
\begin{aligned}
r_t|\Sigma_t &\sim N(0, \Sigma_t^{1/2}\Lambda(\Sigma_t^{1/2})') && (52)\\
\Sigma_t|\nu, S_{t-1} &\sim Wishart_k(\nu, S_{t-1}) && (53)\\
\nu S_t &= B_0 + \sum_{j=1}^{3} B_j \odot \Gamma_{t,\ell_j} + B_4 \odot g_t g_t' \odot \Sigma_t && (54)
\end{aligned}
$$

$\Gamma_{t,\ell}$ is defined the same as before. $B_j = b_j b_j'$ for $j = 1, \ldots, 4$. g_t is $k \times 1$ and $g_{t,i} = 0$ if $r_{t,i} > 0$ and $g_{t,i} = 1$ if $r_{t,i} < 0$, $i = 1, \ldots, k$. The last term in the right-hand side of equation (54) allows the elements of RCOV (both realized variances and covariances) to respond asymmetrically to the signs of the lagged returns. In the special case where $b_4 = \mathbf{0}$, the leverage effect is precluded and we go back to the original Wishart-RCOV-A(3). Table 13 reports the full sample model estimates. Several of the b_4 parameters have density intervals that contain 0. The remaining parameters are similar to the symmetric Wishart-RCOV-A estimates.

[26] In fact all posterior draws of ν in the diagonal-HAR-NCW model are above $2k-1$.

[27] This is based on the posterior mean of Λ.

Does accounting for the leverage effect provide any gain for the Wishart-RCOV model in terms of forecasting? How does the asymmetric Wishart-RCOV perform compared to the asymmetric version of the benchmark MGARCH model? Figure 12 plots the term structure of log-predictive likelihood for returns for the asymmetric Wishart-RCOV-A(3) and asymmetric VD-GARCH-t. The symmetric Wishart-RCOV-A(3) model and VD-GARCH-t from Figure 4 are reproduced for comparison. The asymmetric VD-GARCH-t model used here was introduced by Hansson and Hordahl (1998), which has the following specification:

$$r_t | r_{1:t-1} \sim t(0, H_t, \zeta) \tag{55}$$

$$H_t = CC' + aa' \odot r_{t-1} r'_{t-1} + bb' \odot H_{t-1} + ee' \odot \eta_{t-1} \eta'_{t-1} \tag{56}$$

where $\eta_t = \max[\mathbf{0}, -r_t]$, and e is a $k \times 1$ vector. The results show that overall, for both the Wishart-RCOV model and the VD-GARCH-t model, accommodating the leverage effect provides marginal gains in terms of forecasting. The results also show that the superiority of the Wishart-RCOV model over the VD-GARCH-t model is robust under the presence of an asymmetric effect.

7.2 Overnight Returns

Up until now we have been using open-to-close returns in the empirical work. This exactly matches the realized covariance and corresponds to theory. In this section, we assess the performance of the models when dealing with close-to-close return data.

The overnight return could be included into the construction of RCOV but we found that excluding it provided better results. The Λ in the joint models allow for a general scaling up or down of RCOV which automatically corrects for the fact that the RCOV is only for part of the day.

Figure 13 plots the term structure of log-predictive likelihood using close-to-close returns for Wishart-RCOV-A(3) and Cholesky-VARFIMA models and the asymmetric VD-GARCH-t. Figure 14 shows the corresponding log-predictive Bayes factors for Wishart-RCOV-A vs the asymmetric VD-GARCH-t model. Table 14 presents specific values of the log-predictive likelihood for selected h. As in the previous results the Wishart-RCOV-A(3) is significantly better in density forecasts than the Cholesky-VARFIMA specification. Against the MGARCH model the Wishart-RCOV-A(3) still provides good density forecasts 1-5 days ahead and for forecast horizons longer than 7 weeks. The log-predictive Bayes factor in favor of Wishart-RCOV-A(3) is 16.58, $h = 1$, 4.53, $h = 40$ and 21.43 for $h = 60$. For forecast horizons between 1 week and 6 weeks, the asymmetric VD-GARCH-t is either equal to or better than the Wishart-RCOV-A(3) model. The VD-GARCH-t density forecasts are superior to the Cholesky-VARFIMA over the whole term structure.

For the GMVP portfolio selection exercise, we found similar results to Section 6.3. That is, the gains from using RCOV in the Wishart models with close-to-close returns is confined to short-term investment horizons.

The full sample estimation of the Wishart-RCOV model reveals a different value for Λ compared to the case with open-to-close return, see Table 15[28]. For instance, the diagonal elements are significantly larger, suggesting a set of larger eigenvalues and hence a larger

[28]For other parameters, the values are the same since the same set of RCOV data is used.

determinant. As a matter of fact, $|\Sigma_t^{1/2}\Lambda(\Sigma_t^{1/2})'|/|\Sigma_t| = |\Lambda| = 2.14$.[29] When using close-to-close returns Λ scales up the RCOV matrices; whereas in the case with open-to-close returns, the opposite (scaling down) is achieved.

Finally, our model could be extended to a larger number of assets. To maintain parsimony the parameter vectors b_j could be reduced to scalars. Another option is to consider a factor structure similar to Philipov and Glickman (2006b) or the independent component analysis of Matteson and Tsay (2011).

8 Conclusion

This paper provides a new approach to modelling multivariate returns that consists of joint models of returns and RCOV matrices. Our preferred specification is a Gaussian-Wishart model in which the scale matrix in the Wishart density has additive components. This specification allows the RCOV entering the Gaussian distribution for returns to be scaled up or down and thus can accommodate any systematic biases that RCOV may have. We find that it is critical to include and estimate the components to obtain improved performance relative to MGARCH models and other existing models of RCOV. We expand and empirically investigate several alternative models that have not been subjected to joint return-RCOV modelling. Model forecasts uniformly benefit from either covariance or realized covariance targeting. Models are compared based on the quality of multiple horizon density forecasts of returns, point forecasts of RCOV and portfolio selection. Our preferred model delivers improvements in all three areas of comparison.

9 Appendix

9.1 Stationarity Condition for Wishart-RCOV-A(K)

For the existence of the unconditional mean of Σ_t, let $E_{t-1}[\Sigma_t] = E(\Sigma_t|\Sigma_{1:t-1})$. Rewrite equation (11) as the following:

$$E_{t-1}[\Sigma_t] \quad = \quad B_0 + \sum_{j=1}^{K} B_j \odot \Gamma_{t-1,\ell(j)}, \tag{57}$$

[29] This is based on the posterior mean of Λ.

thus

$$
\begin{aligned}
\Sigma_t &= B_0 + \sum_{j=1}^{K}\left(B_j \odot \frac{1}{\ell(j)}\sum_{i=1}^{\ell(j)}\Sigma_{t-i}\right) + \Sigma_t - E_{t-1}[\Sigma_t] \\
&= B_0 + \sum_{j=1}^{K}\sum_{i=1}^{\ell(j)}\left(\frac{B_j}{\ell(j)} \odot \Sigma_{t-i}\right) + \Sigma_t - E_{t-1}[\Sigma_t] \\
&= B_0 + \sum_{i=1}^{\ell(K)}\sum_{j=\ell^{-1}(i)}^{K}\left(\frac{B_j}{\ell(j)} \odot \Sigma_{t-i}\right) + \Sigma_t - E_{t-1}[\Sigma_t] \\
&= B_0 + \sum_{i=1}^{\ell(K)}\left(\sum_{j=\ell^{-1}(i)}^{K}\frac{B_j}{\ell(j)}\right) \odot \Sigma_{t-i} + \Sigma_t - E_{t-1}[\Sigma_t] \\
&= B_0 + \sum_{i=1}^{\ell(K)}\tilde{B}_i \odot \Sigma_{t-i} + \Sigma_t - E_{t-1}[\Sigma_t] \qquad (58)
\end{aligned}
$$

where $\tilde{B}_i = \sum_{j=\ell^{-1}(i)}^{K}\frac{B_j}{\ell(j)}$, and $\ell^{-1}(i) := \min\{j:\ \ell(j) \geq i\}$. By stacking the lower triangular part of both sides using the operator vech(), we have

$$
\mathrm{vech}(\Sigma_t) = \mathrm{vech}(B_0) + \sum_{i=1}^{\ell(K)}\mathrm{diag}(\mathrm{vech}(\tilde{B}_i))\mathrm{vech}(\Sigma_{t-i}) + \tilde{v}_t \qquad (59)
$$

where $\tilde{v}_t = \mathrm{vech}(\Sigma_t - E_{t-1}[\Sigma_t])$ is a martingale difference. The unconditional mean of $\mathrm{vech}(\Sigma_t)$ exits if all the eigenvalues of the matrix sum $\sum_{i=1}^{\ell(K)}\mathrm{diag}(\mathrm{vech}(\tilde{B}_i))$ are less than 1 in modulus (Golosnoy et al. (2012)). More specifically,

$$
\begin{aligned}
\sum_{i=1}^{\ell(K)}\mathrm{diag}(\mathrm{vech}(\tilde{B}_i)) &= \sum_{i=1}^{\ell(K)}\mathrm{diag}(\mathrm{vech}(\sum_{j=\ell^{-1}(i)}^{K}\frac{B_j}{\ell(j)})) \\
&= \mathrm{diag}\left(\mathrm{vech}\left(\sum_{i=1}^{\ell(K)}\sum_{j=\ell^{-1}(i)}^{K}\frac{B_j}{\ell(j)}\right)\right) \\
&= \mathrm{diag}\left(\mathrm{vech}\left(\sum_{j=1}^{K}\sum_{i=1}^{\ell(j)}\frac{B_j}{\ell(j)}\right)\right) \\
&= \mathrm{diag}\left(\mathrm{vech}\left(\sum_{j=1}^{K}B_j\right)\right)
\end{aligned}
\qquad (60)
$$

So the condition becomes that all the elements of the matrix sum $\sum_{j=1}^{K}B_j$ are less than 1 in modulus.

For the existence of the unconditional second moment of Σ_t and hence the stationarity condition, define

$$\Delta_i = \sum_{j=1}^{i} \text{diag}(\text{vech}(\tilde{B}_j))\Delta_{i-j}, \quad i = 1, 2, \ldots, \quad \Delta_0 = I_{k(k+1)/2}. \tag{61}$$

and

$$\mathcal{O} = \frac{1}{\nu}(\mathcal{L}_k \otimes \mathcal{L}_k)[I_{k^2} \otimes (I_{k^2} + \mathcal{K}_{kk})](I_k \otimes \mathcal{K}_{kk} \otimes I_k)(\mathcal{D}_k \otimes \mathcal{D}_k) \tag{62}$$

where $\mathcal{L}_k$, $\mathcal{D}_k$ and $\mathcal{K}_{kk}$ denote the elimination matrix, the duplication matrix and the commutation matrix (as given in Lütkepohl, 1996, p.9-10 and p.115), respectively. $\otimes$ denotes the Kronecker product. By Golosnoy et al. (2012, Proposition 2), the unconditional second moment of Σ_t exists iff all eigenvalues of $\sum_{i=1}^{\infty}(\Delta_i \otimes \Delta_i)\mathcal{O}$ are less than 1 in modulus.

9.2 Estimation Details

9.2.1 Wishart-RCOV-A(K) Estimation

Parameters are $\Theta = \{\nu, b_1, \ldots, b_K, \ell_2, \ldots, \ell_K\}$. The likelihood function is the product of the Wishart densities,

$$p(\Sigma_{1:T}|\Theta) = \prod_{t=1}^{T} Wishart_k(\Sigma_t|\nu, S_{t-1}). \tag{63}$$

The joint posterior distribution of the parameters $p(\Theta|\Sigma_{1:T})$ then is the product of the data density and the individual priors for each parameter, with the priors given in Section 4. For a particular parameter $\Theta_i \in \Theta$, the conditional posterior distribution is:

$$\begin{aligned} p(\Theta_i|\Theta_{-i}, \Sigma_{1:T}) &\propto p(\Theta_i) \times \prod_{t=1}^{T} Wishart_k(\Sigma_t|\nu, S_{t-1}) \\ &= p(\Theta_i) \times p(\Sigma_{1:T}|\Theta) \\ &= p(\Theta_i) \prod_{t=1}^{T} \frac{|\Sigma_t|^{\frac{\nu-k-1}{2}} |S_{t-1}^{-1}|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})} \exp\left(-\frac{1}{2} Tr(\Sigma_t S_{t-1}^{-1})\right) \end{aligned} \tag{64}$$

We iteratively sample from the conditional posterior distribution of each parameter conditional on the other parameters by Metropolis-Hastings scheme. For each parameter, a random walk with normal proposal is applied, except for $\ell_i, i = 2, \ldots, K$, in which case 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

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

In the empirical work $\lambda = 2$. Given the value ℓ_i in the Markov chain, the new proposal $\ell_i^{'} \sim q(\ell)$ is accepted with probability

$$\min\left\{\frac{p(\ell_i^{'}|b_1,\ldots,b_K,\ell_{-i},\nu,\Sigma_{1:T})}{p(\ell_i|b_1,\ldots,b_K,\ell_{-i},\nu,\Sigma_{1:T})},1\right\}. \tag{65}$$

9.2.2 Wishart-RCOV-M(K) Estimation

The parameters are $\{A, d, \ell, \nu\} = \Theta$. For priors, $A^{-1} \sim Wishart_k(\gamma_0, Q_0)$, a Wishart distribution with $Q_0 = I_k$ and $\gamma_0 = k+1$ set to reflect a proper but relatively uninformative prior. Each d_j follows a uniform prior, $d_j \sim U(-1,1)$, and $\nu \sim \exp(\lambda_0) I_{\nu>k-1}$, an exponential distribution with support truncated to be greater than $k-1$. To make the prior flat, λ_0 is set to 100. Given the priors, the conditional posterior distributions for the parameters are as follows.

$$\begin{aligned}
p(A^{-1}|\nu,d,\ell,\Sigma_{t:T}) &\propto Wishart_k(A^{-1}|\gamma_0,Q_0)\times\prod_{t=1}^{T} Wishart_k(\Sigma_t|\nu,S_{t-1})\\
&\propto \frac{|A^{-1}|^{\frac{\gamma_0-k-1}{2}}|Q_0^{-1}|^{\frac{\gamma_0}{2}}}{2^{\frac{\gamma_0 k}{2}}\prod_{j=1}^{k}\Gamma(\frac{\gamma_0+1-j}{2})}\exp\left(-\frac{1}{2}Tr(A^{-1}Q_0^{-1})\right)\\
&\quad\times\prod_{t=1}^{T}\frac{|\Sigma_t|^{\frac{\nu-k-1}{2}}|S_{t-1}^{-1}|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}}\prod_{j=1}^{k}\Gamma(\frac{\nu+1-j}{2})}\exp\left(-\frac{1}{2}Tr(\Sigma_t S_{t-1}^{-1})\right)\\
&\propto |A^{-1}|^{\frac{T\nu+\gamma_0-k-1}{2}}\exp\left(-\frac{1}{2}Tr[A^{-1}(Q_0^{-1}+\nu\sum_{t=1}^{T}\left[\prod_{j=1}^{K}\Gamma_{t-1,\ell_j}^{\frac{-d_j}{2}}\right]\Sigma_t\left[\prod_{j=K}^{1}\Gamma_{t-1,\ell_j}^{\frac{-d_j}{2}}\right])]\right)\\
&\propto Wishart_k(A^{-1}|\widetilde{\gamma},\widetilde{Q})
\end{aligned} \tag{66}$$

Where $\widetilde{Q}^{-1} = \nu\sum_{t=1}^{T}\left[\prod_{j=1}^{K}\Gamma_{t-1,\ell_j}^{\frac{-d_j}{2}}\right]\Sigma_t\left[\prod_{j=K}^{1}\Gamma_{t-1,\ell_j}^{\frac{-d_j}{2}}\right]+Q_0^{-1}$, $\widetilde{\gamma} = T\nu+\gamma_0$.

For $d_i, i = 1,\ldots,K$ we have,

$$\begin{aligned}
p(d_i|A,d_{-i},\ell,\nu,\Sigma_{1:T}) &\propto p(d_i)\times\prod_{t=1}^{T} Wishart_k(\Sigma_t|\nu,S_{t-1})\\
&= p(d_i)\prod_{t=1}^{T}\frac{|\Sigma_t|^{\frac{\nu-k-1}{2}}|S_{t-1}^{-1}|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}}\prod_{j=1}^{k}\Gamma(\frac{\nu+1-j}{2})}\exp\left(-\frac{1}{2}Tr(\Sigma_t S_{t-1}^{-1})\right)\\
&\propto p(d_i)\exp\left(-\frac{d_i\nu\phi_i}{2}-\frac{1}{2}Tr(\nu A^{-1}Q^{-1})\right),
\end{aligned} \tag{67}$$

where $\phi_i = \sum_{t=1}^{T}\log|\Gamma_{t-1,\ell_i}|$, and $Q^{-1} = \sum_{t=1}^{T}\left[\prod_{j=1}^{K}\Gamma_{t-1,\ell_j}^{\frac{-d_j}{2}}\right]\Sigma_t\left[\prod_{j=K}^{1}\Gamma_{t-1,\ell_j}^{\frac{-d_j}{2}}\right]$. To sample from this density we do the following. If d_i is the previous value in the chain we propose $d_i^{'} = d_i + u$ where $u \sim N(0,\sigma^2)$ and accept $d_i^{'}$ with probability

$$\min\left\{\frac{p(d_i^{'}|A,d_{-i},\ell,\nu,\Sigma_{1:T})}{p(d_i|A,d_{-i},\ell,\nu,\Sigma_{1:T})},1\right\}, \tag{68}$$

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

For $\ell_i, i = 2, \ldots, K$ we have,

$$\begin{aligned} p(\ell_i|A, d, \ell_{-i}, \nu, \Sigma_{1:T}) &\propto p(\ell_i) \times \prod_{t=1}^{T} Wishart_k(\Sigma_t|\nu, S_{t-1}) \\ &\propto p(\ell_i) \exp\left(-\frac{d_i \nu \phi_i}{2} - \frac{1}{2} Tr(\nu A^{-1} Q^{-1})\right), \end{aligned} \tag{69}$$

where ϕ_i and Q^{-1} are defined the same way as in the previous case. To sample from the conditional posterior we use a simple random walk proposal. The proposal distribution is a "symmetric Poisson", as in Wishart-RCOV-A(K). Finally, ν has the conditional posterior density

$$\begin{aligned} p(\nu|A, d, \ell, \Sigma_{1:T}) &\propto p(\nu) \times p(\Sigma_{t:T}|A, d, \nu) \\ &= p(\nu) \prod_{t=1}^{T} \frac{|\Sigma_t|^{\frac{\nu-k-1}{2}} |S_{t-1}^{-1}|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})} \exp\left(-\frac{1}{2} Tr(\Sigma_t S_{t-1}^{-1})\right) \\ &= p(\nu) \frac{\prod_{t=1}^{T} |\Sigma_t|^{\frac{\nu}{2}} \times \nu^{\frac{T\nu k}{2}} |A^{-1}|^{\frac{T\nu}{2}} \times \prod_{t=1}^{T} \prod_{i=1}^{K} |\Gamma_{t-1,\ell_i}|^{-\frac{d_i \nu}{2}}}{2^{\frac{T\nu k}{2}} (\prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2}))^T} \\ &\quad \times \exp\left(-\frac{1}{2} Tr(\nu A^{-1} Q^{-1})\right) \\ &\propto \exp\left(-\lambda_0 \nu + \frac{T\nu}{2} \log|A|^{-1} + \frac{T\nu k}{2} \log\frac{\nu}{2} - T \sum_{j=1}^{k} \log\Gamma\left(\frac{\nu+1-j}{2}\right)\right) \\ &\quad \times \exp\left(\frac{\nu}{2} \sum_{t=1}^{T} \log|\Sigma_t| - \frac{\nu}{2} \sum_{i=1}^{K} d_i \phi_i - \frac{1}{2} Tr(\nu A^{-1} Q^{-1})\right) \end{aligned} \tag{70}$$

where Q^{-1} and ϕ_i are defined as in previous cases. 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.

9.2.3 Sampling from $\Lambda|r_{1:T}, \Sigma_{1:T}$

To estimate Λ, define $\tilde{r}_t = \Sigma_t^{-1/2} r_t$, then $\tilde{r}_t \sim N(0, \Lambda)$. We let the prior of Λ^{-1} be $Wishart_k(\gamma_1, Q_1)$, and we set $Q_1 = I$ and $\gamma_1 = k + 1$. The posterior distribution of Λ^{-1} is

$$\begin{aligned} p(\Lambda^{-1}|\tilde{r}_{1:T}) &\propto Wishart_k(\Lambda^{-1}|\gamma_1, Q_1) \times \prod_{t=1}^{T} N(\tilde{r}_t|0, \Lambda) \\ &\propto Wishart_k(\Lambda^{-1}|\widehat{\gamma}, \widehat{Q}) \end{aligned} \tag{71}$$

by the conjugacy of the Wishart prior of the precision matrix with respect to a multivariate Normal likelihood. Here $\widehat{\gamma} = T + \gamma_1$, and $\widehat{Q} = (\sum_{t=1}^{T} \tilde{r}_t \tilde{r}_t' + Q_1^{-1})^{-1}$

9.2.4 VARFIMA(1,m,1) Estimation

Pre-multiply both sides of equation (21) by $(1-\psi L)^{-1}$, we get the VAR representation of VARFIMA$(1,m,1)$:

$$\begin{aligned}\xi_t &= (1-\psi L)^{-1}(1-\delta L)(1-L)^m[Z_t-c]\\ &= \Psi(L)(Z_t-c)\\ &= \sum_{i=0}^{\infty}\Psi_i(Z_{t-i}-c)\end{aligned} \tag{72}$$

Follow Ravishanker and Ray (1997), let

$$\xi_t=\sum_{i=0}^{t-1}\Psi_i(Z_{t-i}-c) \qquad t=1,\ldots,T, \tag{73}$$

then the conditional likelihood function is proportional to

$$|\Xi|^{-\frac{T}{2}}\times\exp\left(\sum_{t=1}^{T}\xi_t'\Xi^{-1}\xi_t\right) \tag{74}$$

The parameters are $\Theta=\{\delta,m,\psi,\Xi\}$, where c is set at the sample mean of Z_t. For $\Theta_i=\delta$, m, or ψ, the conditional posterior distribution is:

$$p(\Theta_i|\Theta_{-i},\Sigma_{1:T})\propto p(\Theta_i)\times\exp\left(\sum_{t=1}^{T}\xi_t'\Xi^{-1}\xi_t\right), \tag{75}$$

with the all the priors being independent Normal of mean 0 and variance 100 truncated on the interval $(-1,1)$. To sample from the posterior distributions, we use Metropolis-Hastings scheme with a random walk proposal analogous to the sampling in the previous subsections. For Ξ, we use an inverse Wishart distribution $Wishart^{-1}(\gamma_2,Q_2)$ as the prior, where $\gamma_2=\frac{k*(k+1)}{2}+1$ and $Q_2=I$. By the conjugacy of inverse Wishart prior for the covariance matrix of a multivariate Normal, the conditional posterior distribution is:

$$\begin{aligned}p(\Xi|\delta,m,\psi,\Sigma_{1:T}) &\propto Wishart^{-1}(\Xi|\gamma_2,Q_2)\times|\Xi|^{-\frac{T}{2}}\times\exp\left(\sum_{t=1}^{T}\xi_t'\Xi^{-1}\xi_t\right)\\ &\propto Wishart^{-1}(\Xi|\widehat{\gamma}_2,\widehat{Q}_2),\end{aligned} \tag{76}$$

where $\widehat{\gamma}_2=T+\gamma_2$, $\widehat{Q}_2=\sum_{t=1}^{T}\xi_t\xi_t'+Q_2$.

9.2.5 diagonal-HAR-WAR Estimation

The parameters are $\Theta=\{\nu,M_1,M_2,M_3,L_S\}$. M_1, M_2, M_3 are diagonal matrices, L_S is a lower triangular matrix. The prior on ν is an exponential distribution with support truncated to be greater than $k-1$, $p(\nu)=\exp(100)I_{\nu>k-1}$.[30] All the other parameters are each

[30]In posterior simulation only draws of $\nu>k-1$ are accepted.

assigned an independent Normal prior $p(\Theta_i)$ with mean 0 and variance 100, with the following truncations:

$$M_1(1,1) > 0,\ M_2(1,1) > 0,\ M_3(1,1) > 0,\ L_S(i,i) > 0,\ i = 1, \dots, k \tag{77}$$

Given Θ the likelihood function is a product of the noncentral Wishart densities(see Muirhead (1982) p. 442),

$$\begin{aligned} p(\Sigma_{1:T}|\Theta) &= \prod_{t=1}^{T} NCW_k(\Sigma_t|\nu, S, V_{t-1}) \\ &= \prod_{t=1}^{T} \frac{|\Sigma_t|^{\frac{\nu-k-1}{2}} |S^{-1}|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})} \exp\left(-\frac{1}{2} Tr[S^{-1}(\Sigma_t + V_{t-1})]\right) \\ &\quad \times {}_0F_1(\nu;\ \frac{1}{4} S^{-1} V_{t-1} S^{-1} \Sigma_t) \end{aligned} \tag{78}$$

where ${}_0F_1$ is the hypergeometric function of matrix argument, which we evaluate using the Laplace approximation method developed in Butler and Wood (2003, 2005).
The conditional posterior distribution of the each parameter is proportional to the product of its prior and the likelihood function:

$$p(\Theta_i|\Theta_{-i}, \Sigma_{1:T}) \propto p(\Theta_i) \times \prod_{t=1}^{T} NCW_k(\Sigma_t|\nu, S, V_{t-1}) \tag{79}$$

To sample from the posterior distributions, we use Metropolis-Hastings scheme with a random walk proposal analogous to the sampling in the previous subsections. In the case where ν is real-valued, the proposal is normal. In the case where ν is integer-valued, the proposal distribution is a Poisson random variable multiplied by a random variable that takes on values 1 and -1 with equal probability.

9.2.6 VD-GARCH-t Estimation

The parameters are $\{a_1, \dots, a_k, b_1, \dots, b_k, \zeta\} = \Theta$. All parameters are assigned an independent Normal prior with mean 0 and variance 100, with a_1 and b_1 restricted to be positive for identification purpose. The joint prior $p(\Theta)$ is just the product of the individual priors. The likelihood function $p(r_{1:T}|\Theta)$ is:

$$p(r_{1:T}|\Theta) = \prod_{t=1}^{T} \frac{\Gamma[(\zeta+k)/2][1 + \frac{1}{\zeta} r_t' H_t^{-1} r_t]^{-(\zeta+k)/2}}{\Gamma(\zeta/2)(\zeta\pi)^{k/2} |H_t|^{1/2}} \tag{80}$$

The posterior of the parameters $p(\Theta|r_{1:T})$ is:

$$p(\Theta|r_{1:T}) \propto p(\Theta) \prod_{t=1}^{T} \frac{\Gamma[(\zeta+k)/2][1 + \frac{1}{\zeta} r_t' H_t^{-1} r_t]^{-(\zeta+k)/2}}{\Gamma(\zeta/2)(\zeta\pi)^{k/2} |H_t|^{1/2}} \tag{81}$$

To sample from the joint posterior distribution $p(\Theta|r_{1:T})$, 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 burnin, 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 recursively estimated 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.

9.2.7 DCC-t Estimation

The parameters are $\{\omega_1, \dots, \omega_k, \kappa_1, \dots, \kappa_k, \lambda_1, \dots, \lambda_k, \alpha, \beta, \zeta\} = \Theta$. All parameters are assigned an independent Normal prior with mean 0 and variance 100, with the following restrictions are imposed:

$$\omega_i > 0,\ \kappa_i \geq 0,\ \lambda_i \geq 0,\ \zeta > 2,\ \frac{\kappa_i \zeta}{\zeta - 2} + \lambda_i < 1,\ i = 1, ..., k,\ \alpha \geq 0,\ \beta \geq 0,\ \alpha + \beta < 1. \quad (82)$$

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

$$p(r_{1:T}|\Theta) = \prod_{t=1}^{T} \frac{\Gamma[(\zeta + k)/2][1 + \frac{1}{\zeta} r_t' H_t^{-1} r_t]^{-(\zeta+k)/2}}{\Gamma(\zeta/2)(\zeta\pi)^{k/2}|H_t|^{1/2}}. \quad (83)$$

The posterior of the parameters $p(\Theta|r_{1:T})$ is:

$$p(\Theta|r_{1:T}) \propto p(\Theta) \prod_{t=1}^{T} \frac{\Gamma[(\zeta + k)/2][1 + \frac{1}{\zeta} r_t' H_t^{-1} r_t]^{-(\zeta+k)/2}}{\Gamma(\zeta/2)(\zeta\pi)^{k/2}|H_t|^{1/2}}. \quad (84)$$

For the special case of DCC with Normal innovations, the parameters are

$$\{\omega_1, \dots, \omega_k, \kappa_1, \dots, \kappa_k, \lambda_1, \dots, \lambda_k, \alpha, \beta, \} = \Theta.$$

The restriction on the priors are similar:

$$\omega_i > 0,\ \kappa_i \geq 0,\ \lambda_i \geq 0,\ \kappa_i + \lambda_i < 1,\ i = 1, ..., k,\ \alpha \geq 0,\ \beta \geq 0,\ \alpha + \beta < 1. \quad (85)$$

The posterior of the parameters $p(\Theta|r_{1:T})$ is:

$$p(\Theta|r_{1:T}) \propto p(\Theta)(2\pi)^{\frac{Tk}{2}} \prod_{t=1}^{T} |D_t R_t D_t|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2} \sum_{t=1}^{T} r_t'(D_t R_t D_t)^{-1} r_t\right) \quad (86)$$

The sampling procedure is similar to that of the VD-GARCH-t.

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. Asai M., M.K.P. So (2010) "Stochastic Covariance Models" Available at SSRN: http:// ssrn.com/ abstract=1673764

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

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

10. 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", *Econometrica*, 76, 1481-1536.

11. 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.

12. Bauer, G. H., Vorkink, K. (2011), "Forecasting Multivariate Realized Stock Market Volatility", *Journal of Econometrics*, 160(1), 93-101.

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

14. Bonato M., M. Caporin, A. Ranaldo (2009), "Forecasting Realized (Co)Variances with a Block Structure Wishart Autoregressive Model", Swiss National Bank working paper 2009-3.

15. Butler, R.W., A.T.A. Wood, (2003), “Laplace approximation for Bessel functions of matrix argument”, *Journal of Computational and Applied Mathematics*, 155, 359-382.

16. Butler, R.W., A.T.A. Wood, (2005), “Laplace approximations to hypergeometric functions of two matrix arguments”, *Journal of Multivariate Analysis*, 94, 1-18.

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

18. Chiriac, R., Voev, V. (2010): “Modelling and Forecasting Multivariate Realized Volatility”, forthcoming *Journal of Applied Econometrics*

19. Chiriac, R. (2006), “Estimating Realized Volatility Wishart Autoregressive Model”, Working Paper, University of Konstanz.

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

21. Ding, Z. and R. Engle (2001), “Large scale conditional covariance matrix modeling, estimation and testing”,NYU Working Paper No. FIN-01-029.

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

23. 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.

24. Fleming, J., C. Kirby and B. Ostdiek (2003) ”The economic value of volatility timing using “realized” volatility,” *Journal of Financial Economics*, 67(3), 473-509.

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

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

27. Gleser, L.J., (1976), “A Canonical Representation for the Noncentral Wishart Distribution Useful for Simulation”, *Journal of the American Statistical Association*, 71(355), 690-695.

28. Golosnoy V., B. Gribisch, R. Liesenfeld (2012) “The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility”, *Journal of Econometrics*, 167, 211-223.

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

30. Hansson, B. and P. Hordahl (1998), “Testing the Conditional CAPM using Multivariate GARCH-M”, *Applied Financial Economics*, 8, 377-388.

31. Hautsch, N., L. M. Kyj and R. C. A. Oomen (2010) "A blocking and regularization approach to high-dimensional realized covariance estimation", *Journal of Applied Econometrics*, doi: 10.1002/jae.1218.

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

33. Kollo T. and D. von Rosen (2005) Advanced Multivariate Statistics with Matrices, Springer, The Netherlands.

34. Lütkepohl, H. (1996), *Handbook of Matrices*, Wiley: Chichester.

35. Maheu, J. M., T. H. McCurdy (2011), "Do high-frequency measures of volatility improve forecasts of return distributions?", *Journal of Econometrics* 160(1), 69-76.

36. Matteson D. S. and R. S. Tsay (2011), "Independent Component Analysis via Distance Covariance", Cornell University, Department of Statistical Science.

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

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

39. Noureldin D., N. Shephard, K. Sheppard (2011) "Multivariate high-frequency-based volatility (HEAVY) models", *Journal of Applied Econometrics*, forthcoming.

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

41. Philipov, A. and M. E. Glickman (2006b) "Factor Multivariate Stochastic Volatility via Wishart Processes", *Econometric Reviews*, 25(2-3), 311-334

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

43. Ravishanker, N., B. K. Ray (1997), "Bayesian Analysis of Vector ARFIMA Processes", *Australian Journal of Statistics*, 39(3), 295-311.

44. 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: Monte Carlo study results for 3-dimensional Wishart-RCOV-A(3)

Parameter	DGP Value	Mean	RMSE
b_{11}	0.30	0.2962	0.0144
b_{12}	0.35	0.3436	0.0159
b_{13}	0.25	0.2499	0.0179
b_{21}	0.65	0.6493	0.0138
b_{22}	0.55	0.5546	0.0142
b_{23}	0.60	0.5977	0.0135
b_{31}	0.60	0.5913	0.0167
b_{32}	0.70	0.6914	0.0144
b_{33}	0.55	0.5404	0.0194
ν	10.00	10.0213	0.0840
ℓ_2	5	5.0004	0.0004
ℓ_3	30	30.1051	0.6416

This table reports the mean and the root mean squared error (RMSE) for the posterior means. The number of replications is 1000, the sampe size is 2000.

Table 4: Cumulative log-predictive likelihoods $\hat{p}_h$ across models for various forecast horizon h

Model	$h = 1$	$h = 5$	$h = 10$	$h = 20$	$h = 60$
Wishart-RCOV-A(3)	-2737.53	-2762.59	-2768.55	-2791.22	-2812.09
Wishart-RCOV-M(3)	-2740.11	-2764.99	-2774.46	-2797.86	-2832.94
Cholesky-VARFIMA(1,m,1)	-2754.29	-2776.32	-2788.35	-2807.29	-2825.10
diagonal-HAR-WAR	-2816.14	-2897.86	-2939.19	-2983.14	-3034.82
diagonal-HAR-NCW	-2740.39	-2773.12	-2786.52	-2820.60	-2848.44
VD-GARCH-t	-2754.37	-2769.28	-2781.57	-2796.63	-2835.97

Table 5: Root mean squared error $RMSE_h$ of forecasts of Σ_t across models for various forecast horizon h

Model	$h = 1$	$h = 5$	$h = 10$	$h = 20$	$h = 60$
Wishart-RCOV-A(3)	2.969	3.349	3.508	3.741	3.751
Wishart-RCOV-M(3)	2.969	3.384	3.564	3.881	4.029
Cholesky-VARFIMA(1,m,1)	3.199	3.507	3.655	3.842	3.981
diagonal-HAR-WAR	3.678	4.619	5.062	5.504	5.989
diagonal-HAR-NCW	3.025	3.495	3.726	4.095	4.192

$RMSE_h = \frac{1}{T-T_0+1}\sum_{t=T_0-h}^{T-h}\|\Sigma_{t+h} - E[\Sigma_{t+h}|I_t]\|$, where $\|A\| = \sqrt{\sum_i \sum_j |a_{ij}|^2}$, and $E[\Sigma_{t+h}|I_t]$ denotes a model's predictive mean.

Table 6: Sample variances of GMVP across models for various forecast horizon h

Model	$h=1$	$h=5$	$h=10$	$h=20$	$h=60$
Wishart-RCOV-A(3)	0.419	0.438	0.452	0.476	0.477
Wishart-RCOV-M(3)	0.419	0.435	0.446	0.470	0.470
Cholesky-VARFIMA(1,m,1)	0.431	0.446	0.457	0.474	0.483
diagonal-HAR-WAR	0.452	0.480	0.495	0.511	0.526
VD-GARCH-t	0.452	0.462	0.464	0.473	0.484

Table 7: Estimation results for Wishart-RCOV-A(3)

Parameter	Mean	NSE	0.95 DI	Ineff
b_{11}	0.5744	0.0012	$(0.5556, 0.5905)$	55.6555
b_{12}	0.5579	0.0019	$(0.5354, 0.5783)$	90.5498
b_{13}	0.5995	0.0017	$(0.5835, 0.6155)$	138.5520
b_{14}	0.4888	0.0005	$(0.4628, 0.5127)$	3.5098
b_{15}	0.5878	0.0010	$(0.5668, 0.6077)$	25.4376
b_{21}	0.6732	0.0010	$(0.6518, 0.6932)$	28.8173
b_{22}	0.6536	0.0023	$(0.6314, 0.6821)$	104.4400
b_{23}	0.6536	0.0013	$(0.6381, 0.6691)$	73.1365
b_{24}	0.6918	0.0005	$(0.6649, 0.7174)$	3.7541
b_{25}	0.5623	0.0017	$(0.5290, 0.5963)$	33.6903
b_{31}	0.4242	0.0010	$(0.3992, 0.4519)$	16.6831
b_{32}	0.4854	0.0024	$(0.4544, 0.5111)$	76.9035
b_{33}	0.4410	0.0019	$(0.4187, 0.4644)$	85.4987
b_{34}	0.4475	0.0010	$(0.4066, 0.4833)$	7.9212
b_{35}	0.5384	0.0013	$(0.5064, 0.5709)$	17.9957
ν	14.6666	0.0037	$(14.4875, 14.8439)$	4.6488
ℓ_2	8.9967	0.0031	$(9.0000, 9.0000)$	8.7925
ℓ_3	63.8190	0.0379	$(62.0000, 66.0000)$	3.3739

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 8: Estimation results for Wishart-RCOV-M(3)

Parameter	Mean	NSE	0.95 DI	Ineff
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 9: Estimation results for VARFIMA$(1, m, 1)$

Parameter	Mean	NSE	0.95 DI	Ineff
δ	0.3477	0.0012	$(0.3157, 0.3769)$	18.7313
m	0.4295	0.0011	$(0.4058, 0.4501)$	32.6552
ψ	0.6121	0.0017	$(0.5808, 0.6430)$	35.9129

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

Table 10: Estimation results for diagonal-HAR-WAR with integer-valued ν

Parameter	Mean	NSE	0.95 DI	Ineff
$M_1(1,1)$	0.4425	0.0010	$(0.4167, 0.4677)$	17.8611
$M_1(2,2)$	0.4502	0.0012	$(0.4184, 0.4818)$	16.8463
$M_1(3,3)$	0.4968	0.0011	$(0.4727, 0.5191)$	28.6441
$M_1(4,4)$	0.2981	0.0022	$(0.2472, 0.3415)$	26.5445
$M_1(5,5)$	0.4770	0.0012	$(0.4431, 0.5056)$	14.6781
$M_2(1,1)$	0.4956	0.0022	$(0.4646, 0.5228)$	65.4310
$M_2(2,2)$	0.4589	0.0025	$(0.4121, 0.5089)$	29.8145
$M_2(3,3)$	0.4112	0.0018	$(0.3649, 0.4491)$	21.9140
$M_2(4,4)$	0.6201	0.0028	$(0.5792, 0.6528)$	75.6611
$M_2(5,5)$	0.4559	0.0015	$(0.4077, 0.5013)$	13.4569
$M_3(1,1)$	0.4159	0.0022	$(0.3879, 0.4416)$	73.9334
$M_3(2,2)$	0.4475	0.0017	$(0.4075, 0.4835)$	25.0029
$M_3(3,3)$	0.4488	0.0014	$(0.4176, 0.4802)$	23.4271
$M_3(4,4)$	0.3297	0.0034	$(0.2814, 0.3780)$	54.6142
$M_3(5,5)$	0.3591	0.0024	$(0.3091, 0.4024)$	33.4427
ν	8	0.0000	$(8, 8)$	1.0000

This table reports the posterior mean, its numerical standard error (NSE), a 0.95 density interval (DI) and the inefficiency factor for M_1,M_2,M_3 and ν

Table 11: Estimation results for diagonal-HAR-NCW

Parameter	Mean	NSE	0.95 DI	Ineff
$\tilde{M}_1(1,1)$	0.4827	0.0024	$(0.4500, 0.5145)$	65.6861
$\tilde{M}_1(2,2)$	0.4927	0.0025	$(0.4587, 0.5228)$	74.8778
$\tilde{M}_1(3,3)$	0.5418	0.0024	$(0.5143, 0.5753)$	75.9262
$\tilde{M}_1(4,4)$	0.2679	0.0018	$(0.2140, 0.3205)$	13.5440
$\tilde{M}_1(5,5)$	0.5488	0.0021	$(0.5106, 0.5835)$	40.0412
$\tilde{M}_2(1,1)$	0.6046	0.0027	$(0.5733, 0.6355)$	95.3429
$\tilde{M}_2(2,2)$	0.5626	0.0055	$(0.5197, 0.6038)$	191.0160
$\tilde{M}_2(3,3)$	0.5386	0.0039	$(0.4804, 0.5870)$	66.6818
$\tilde{M}_2(4,4)$	0.7407	0.0019	$(0.7077, 0.7695)$	43.8908
$\tilde{M}_2(5,5)$	0.4468	0.0046	$(0.3779, 0.5039)$	58.2833
$\tilde{M}_3(1,1)$	0.6123	0.0019	$(0.5889, 0.6405)$	65.8317
$\tilde{M}_3(2,2)$	0.6587	0.0031	$(0.6280, 0.6859)$	128.5340
$\tilde{M}_3(3,3)$	0.6438	0.0012	$(0.6127, 0.6761)$	19.1210
$\tilde{M}_3(4,4)$	0.5719	0.0025	$(0.5390, 0.6072)$	61.7975
$\tilde{M}_3(5,5)$	0.6903	0.0025	$(0.6602, 0.7208)$	86.7010
ν	14.5491	0.0021	$(14.3550, 14.7476)$	1.3168

This table reports the posterior mean, its numerical standard error (NSE), a 0.95 density interval (DI) and the inefficiency factor for $\tilde{M}_1$,$\tilde{M}_2$,$\tilde{M}_3$ and ν

Table 12: Estimation results for Λ

1.0783				
(0.0314)				
−0.0043	0.8923			
(0.0208)	(0.0264)			
−0.0188	0.0723	0.7905		
(0.0198)	(0.0176)	(0.0237)		
0.0050	0.0182	0.0057	0.8346	
(0.0199)	(0.0178)	(0.0168)	(0.0249)	
−0.0140	0.0173	0.0070	0.0171	0.8599
(0.0199)	(0.0181)	(0.0176)	(0.0180)	(0.0253)

This table reports the posterior mean, and the posterior standard deviation in parentheses for the lower triangle of Λ.

Table 13: Estimation results for asymmetric Wishart-RCOV-A(3)

Parameter	Mean	NSE	0.95 DI	Ineff
b_{11}	0.5471	0.0016	$(0.5276, 0.5691)$	82.2395
b_{12}	0.5384	0.0019	$(0.5171, 0.5625)$	84.7408
b_{13}	0.5922	0.0020	$(0.5708, 0.6108)$	129.3920
b_{14}	0.4694	0.0045	$(0.4301, 0.5126)$	120.5610
b_{15}	0.5513	0.0020	$(0.5265, 0.5771)$	74.7850
b_{21}	0.6638	0.0023	$(0.6412, 0.6848)$	145.0200
b_{22}	0.6426	0.0035	$(0.6167, 0.6685)$	205.7850
b_{23}	0.6515	0.0028	$(0.6254, 0.6673)$	195.7090
b_{24}	0.6875	0.0016	$(0.6583, 0.7165)$	38.2943
b_{25}	0.5633	0.0019	$(0.5331, 0.5959)$	47.1854
b_{31}	0.4338	0.0017	$(0.4099, 0.4562)$	61.5820
b_{32}	0.4857	0.0026	$(0.4560, 0.5099)$	123.2430
b_{33}	0.4415	0.0027	$(0.4130, 0.4680)$	114.6430
b_{34}	0.4484	0.0014	$(0.4075, 0.4857)$	175.0930
b_{35}	0.5352	0.0013	$(0.5014, 0.5680)$	20.5319
b_{41}	0.2187	0.0016	$(0.1732, 0.2541)$	12.5574
b_{42}	0.2447	0.0014	$(0.1978, 0.2871)$	196.8950
b_{43}	0.0256	0.0016	$(-0.0708, 0.1798)$	199.7410
b_{44}	0.1272	0.0036	$(-0.1463, 0.2645)$	45.7077
b_{45}	0.2905	0.0029	$(0.2410, 0.3355)$	14.4112
ν	14.7198	0.0022	$(14.4535, 14.9235)$	1.4796
ℓ_2	9.0000	0.0000	$(9.0000, 9.0000)$	1.0000
ℓ_3	63.4043	0.0490	$(62.0000, 65.0000)$	5.2721

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

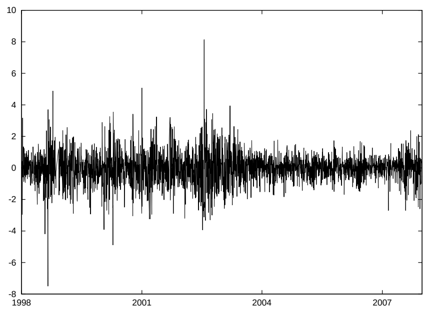
Table 14: Cumulative log-predictive likelihoods $\hat{p}_h$ using close-to-close returns for various forecast horizon

Model	$h = 1$	$h = 5$	$h = 10$	$h = 20$	$h = 60$
Wishart-RCOV-A(3)	-2972.49	-3008.04	-3021.44	-3047.23	-3062.13
Cholesky-VARFIMA(1,m,1)	-3007.97	-3036.96	-3045.08	-3066.65	-3087.77
VD-GARCH-t	-2989.07	-3008.07	-3013.54	-3038.39	-3083.56

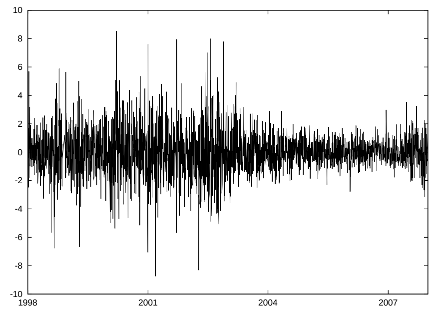
Table 15: Estimation results for Λ with close-to-close returns

1.5376				
(0.0455)				
0.0584	1.0942			
(0.0268)	(0.0328)			
0.0590	0.0973	0.9804		
(0.0257)	(0.0218)	(0.0292)		
0.0682	0.0352	0.0305	1.1071	
(0.0268)	(0.0226)	(0.0217)	(0.0323)	
0.0130	0.0205	−0.0106	0.0542	1.1959
(0.0282)	(0.0236)	(0.0227)	(0.0237)	(0.0363)

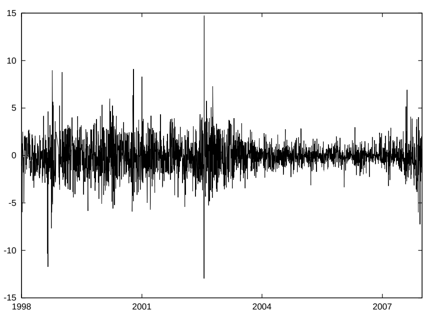
This table reports the posterior mean, and the posterior standard deviation in parentheses for the lower triangle of Λ.



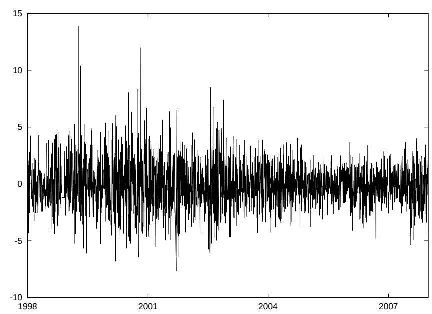
(a) SPYDER



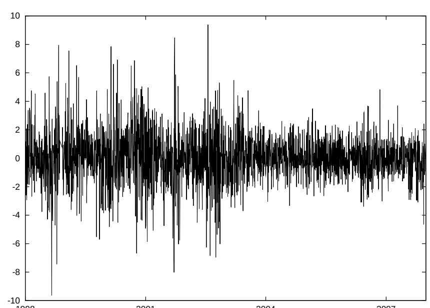
(b) GE



(c) Citigroup Inc.

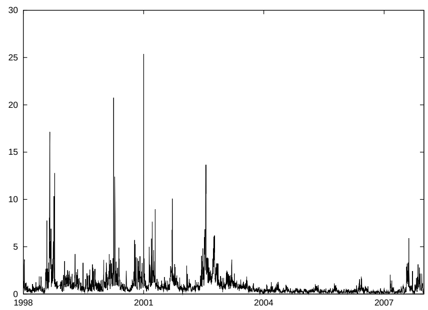


(d) Alcoa Inc.

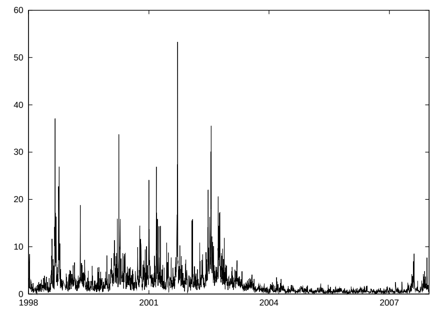


(e) Boeing Co.

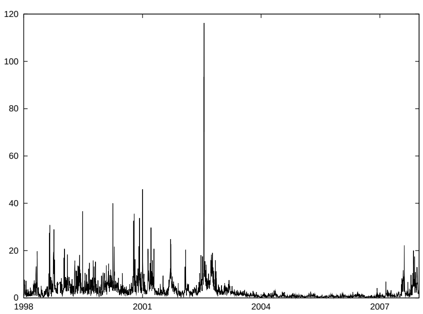
Figure 1: Daily returns



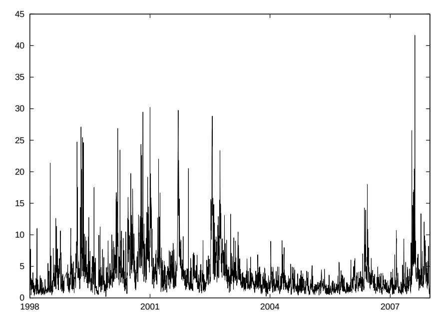
(a) SPYDER



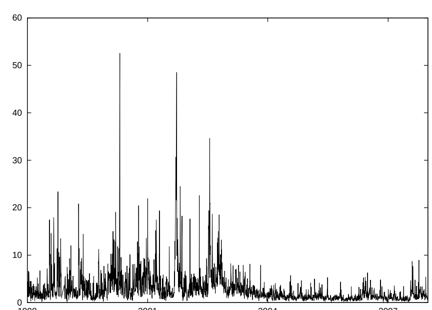
(b) GE



(c) Citigroup Inc.



(d) Alcoa Inc.



(e) Boeing Co.

Figure 2: RV for individual assets

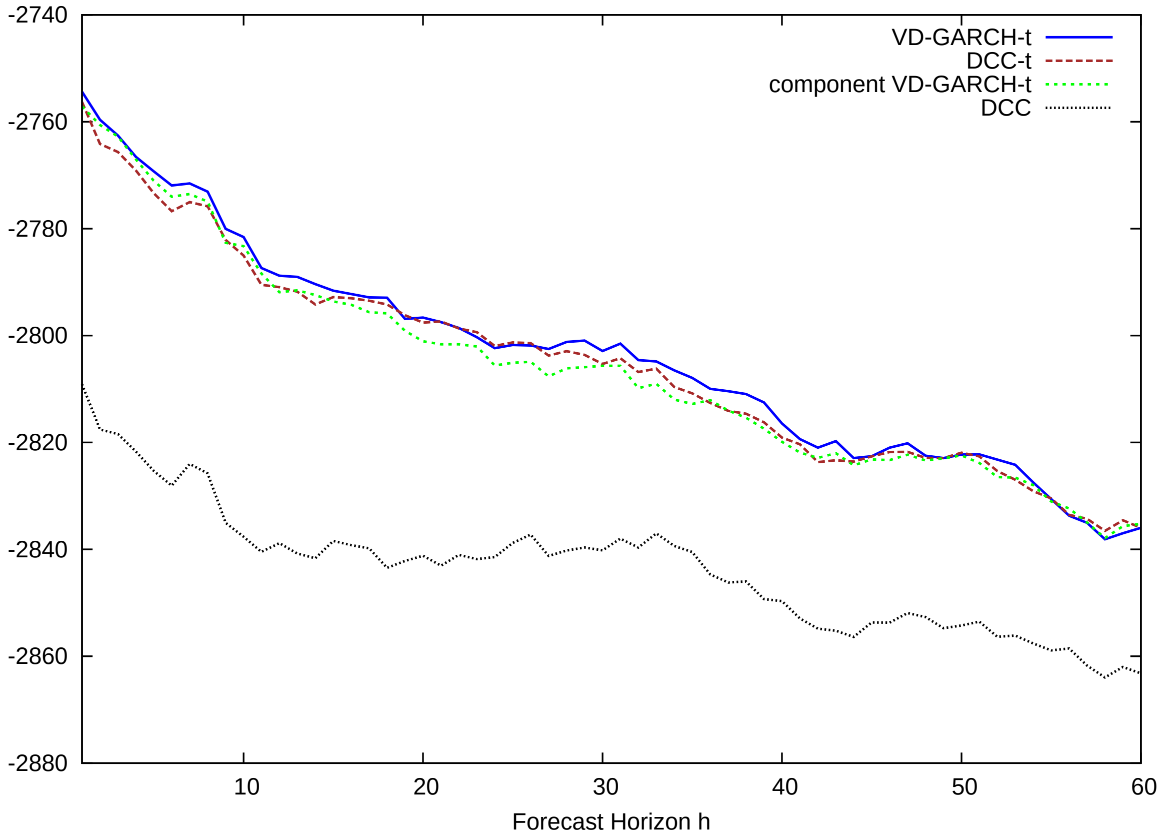


Figure 3: Term structure of cumulative log-predictive likelihoods for MGARCH models

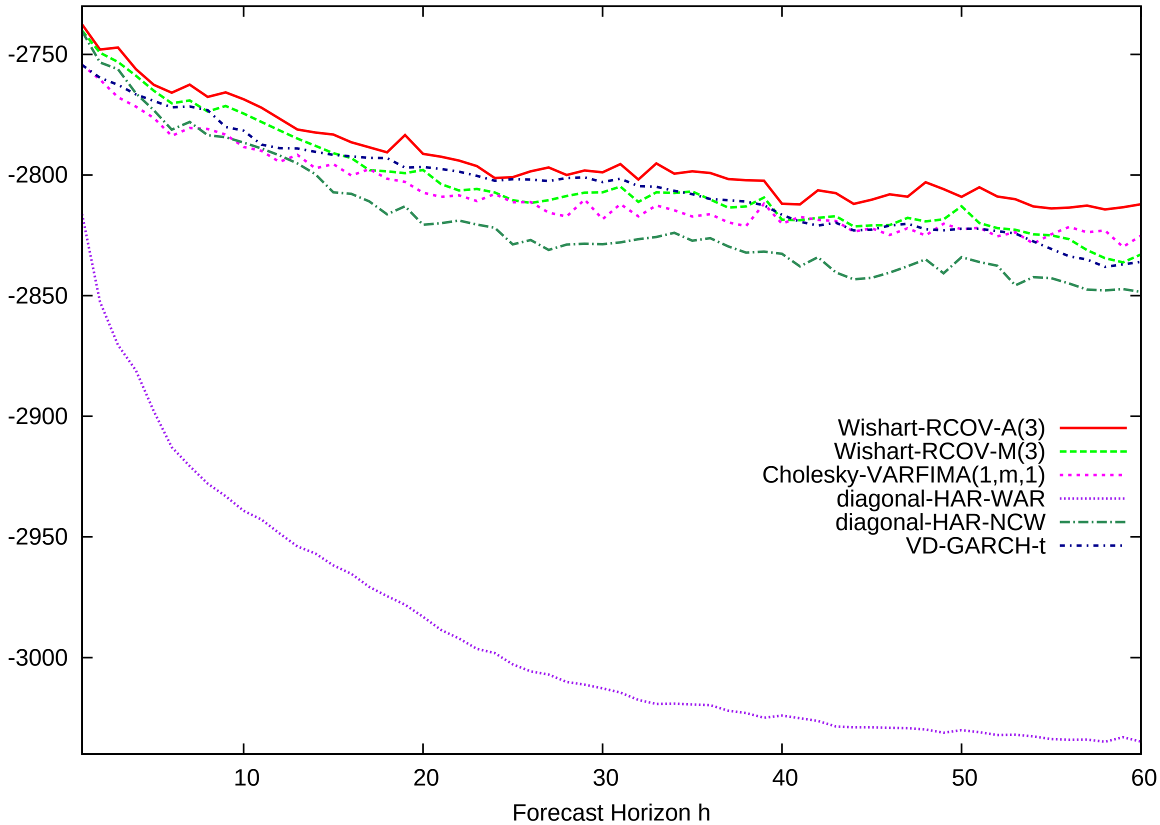


Figure 4: Term structure of cumulative log-predictive likelihoods for return-RCOV models

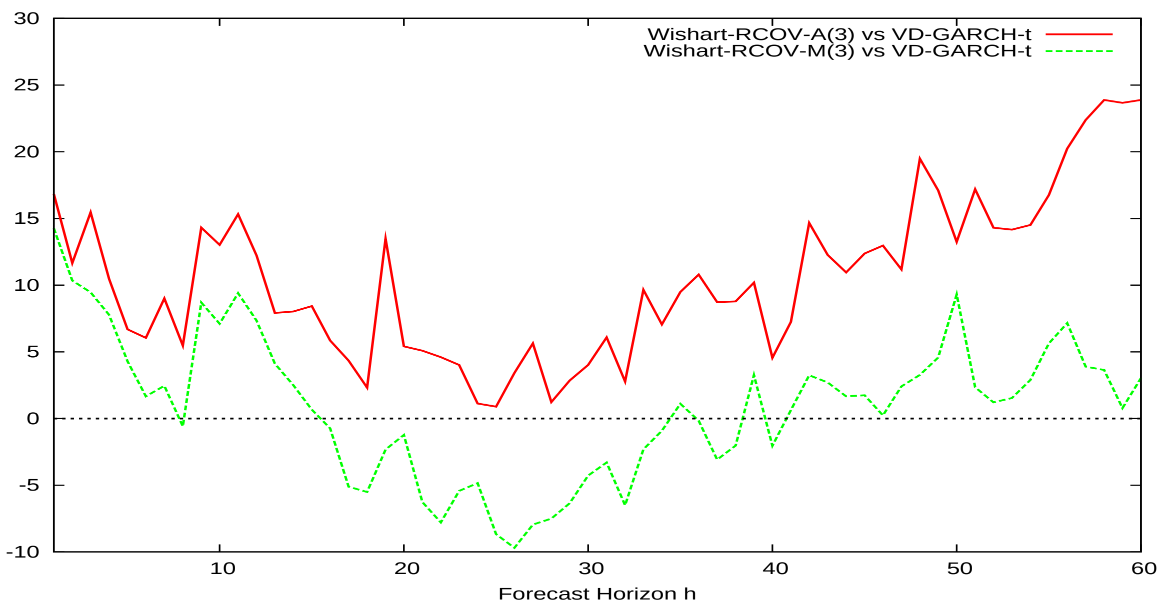


Figure 5: Log Predictive Bayes Factors: Wishart-RCOV vs VD-GARCH

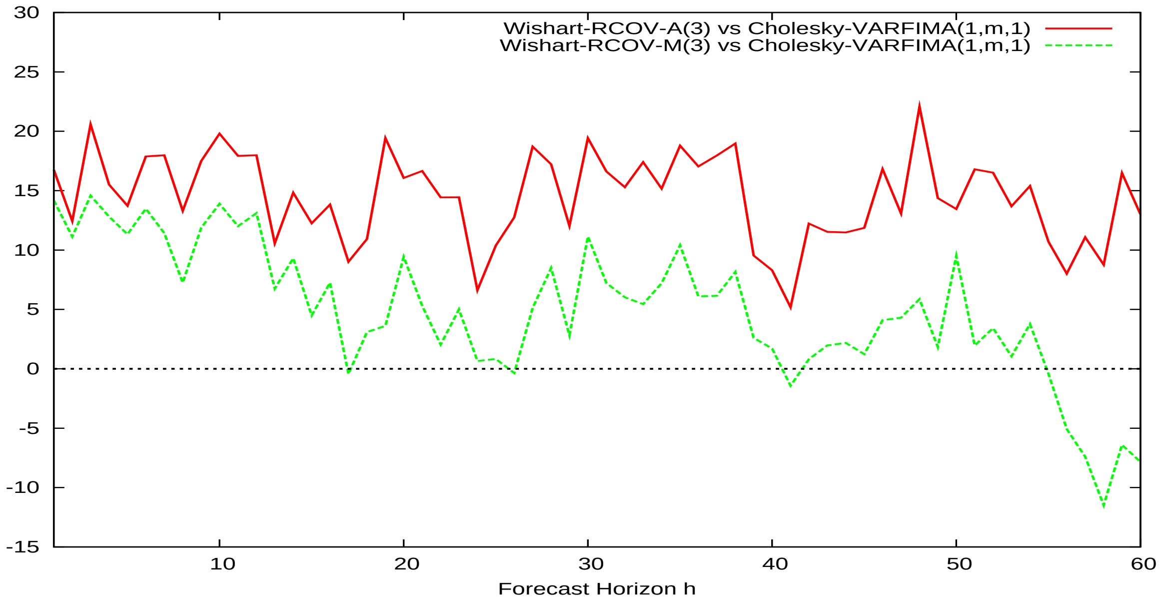


Figure 6: Log Predictive Bayes Factors: Wishart-RCOV vs Cholesky-VARFIMA

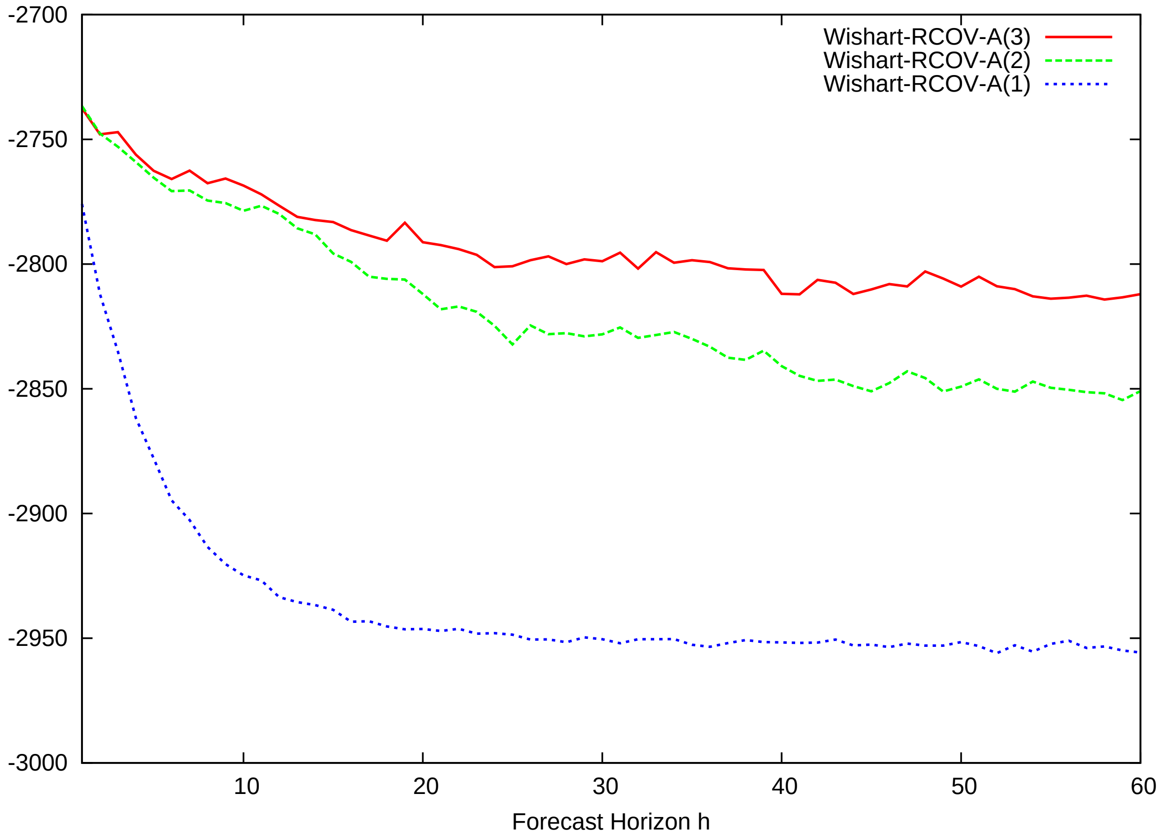


Figure 7: Term structure of cumulative log-predictive likelihoods for Wishart-RCOV-A(K), $K = 1, 2, 3$

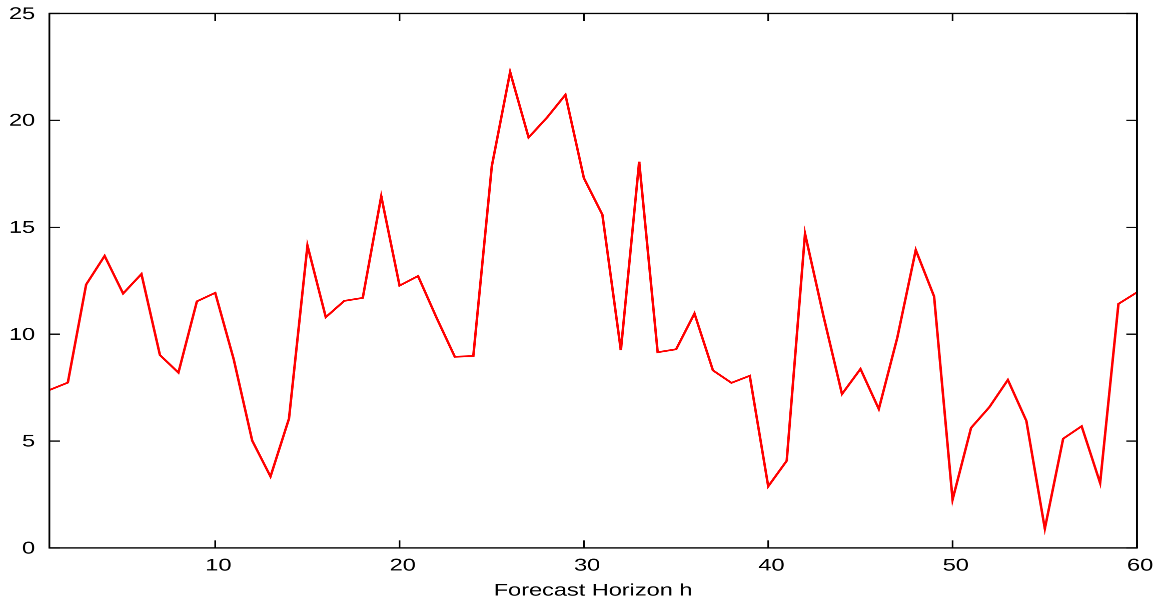


Figure 8: Log Predictive Bayes Factor: ℓ_2 and ℓ_3 estimated vs $\ell_2 = 5$ and $\ell_3 = 22$ for Wishart-RCOV-A(3)

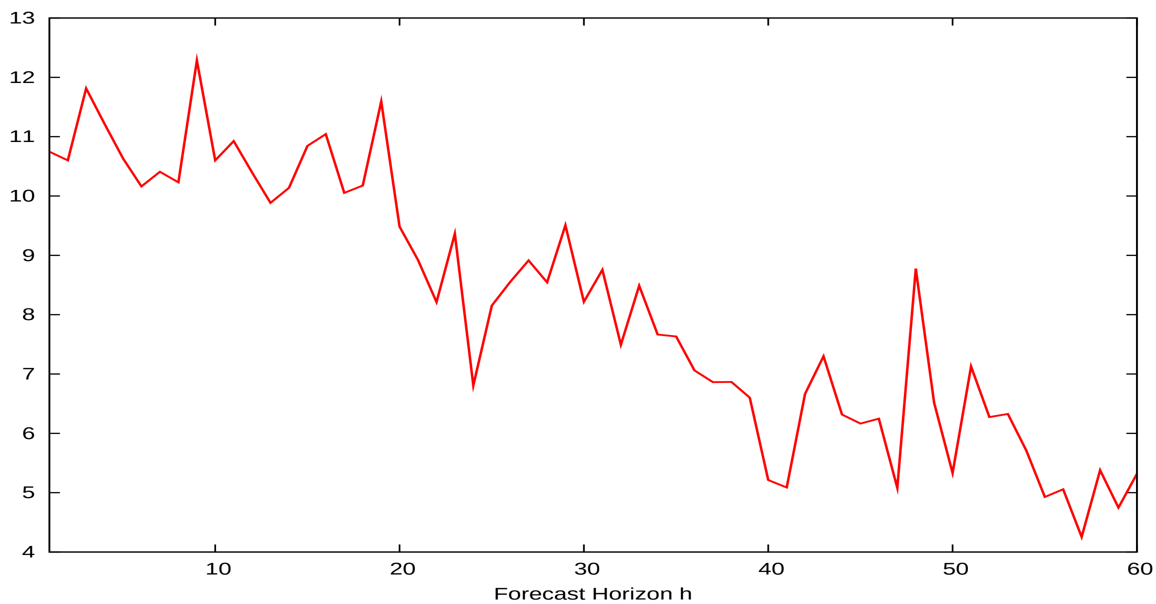


Figure 9: Log Predictive Bayes Factor: Λ estimated vs $\Lambda = I$

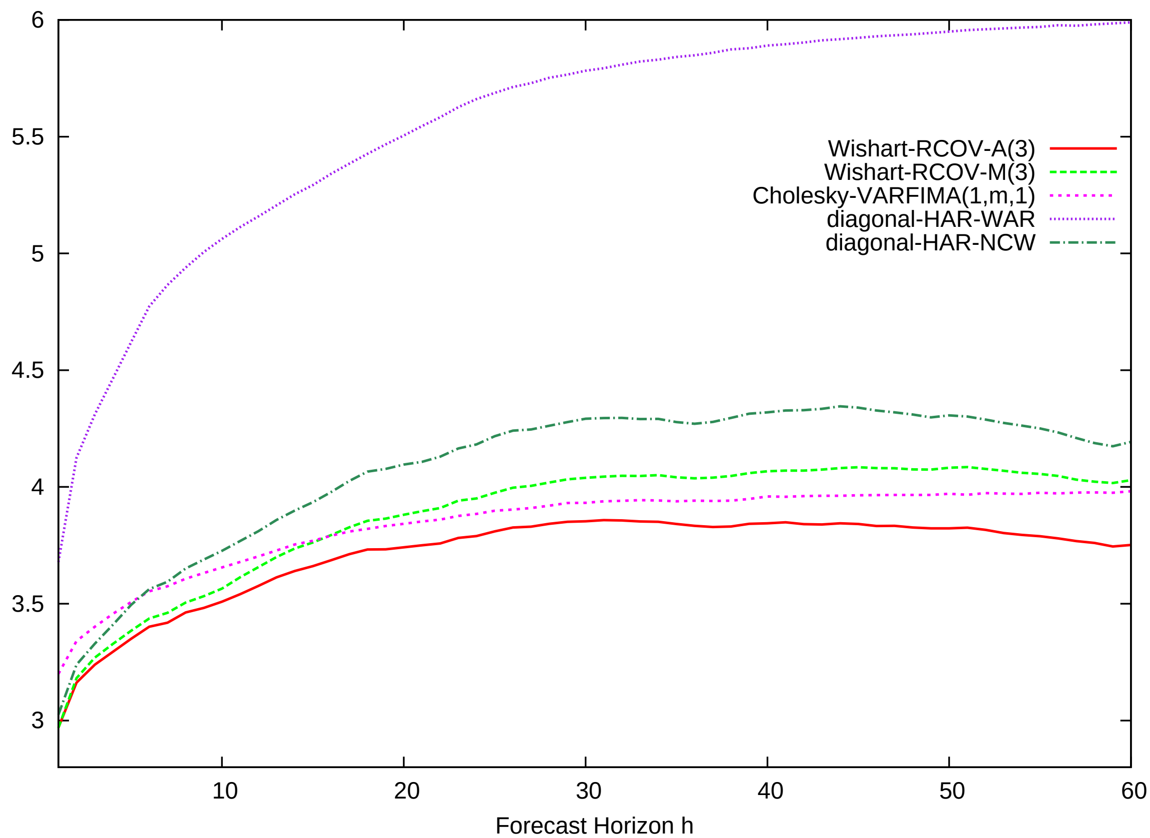


Figure 10: Root mean squared error of forecasts of Σ_t. $RMSE_h = \frac{1}{T-T_0+1}\sum_{t=T_0-h}^{T-h} \|\Sigma_{t+h} - E[\Sigma_{t+h}|I_t]\|$, where $\|A\| = \sqrt{\sum_i \sum_j |a_{ij}|^2}$, and $E[\Sigma_{t+h}|I_t]$ denotes a model's predictive mean.

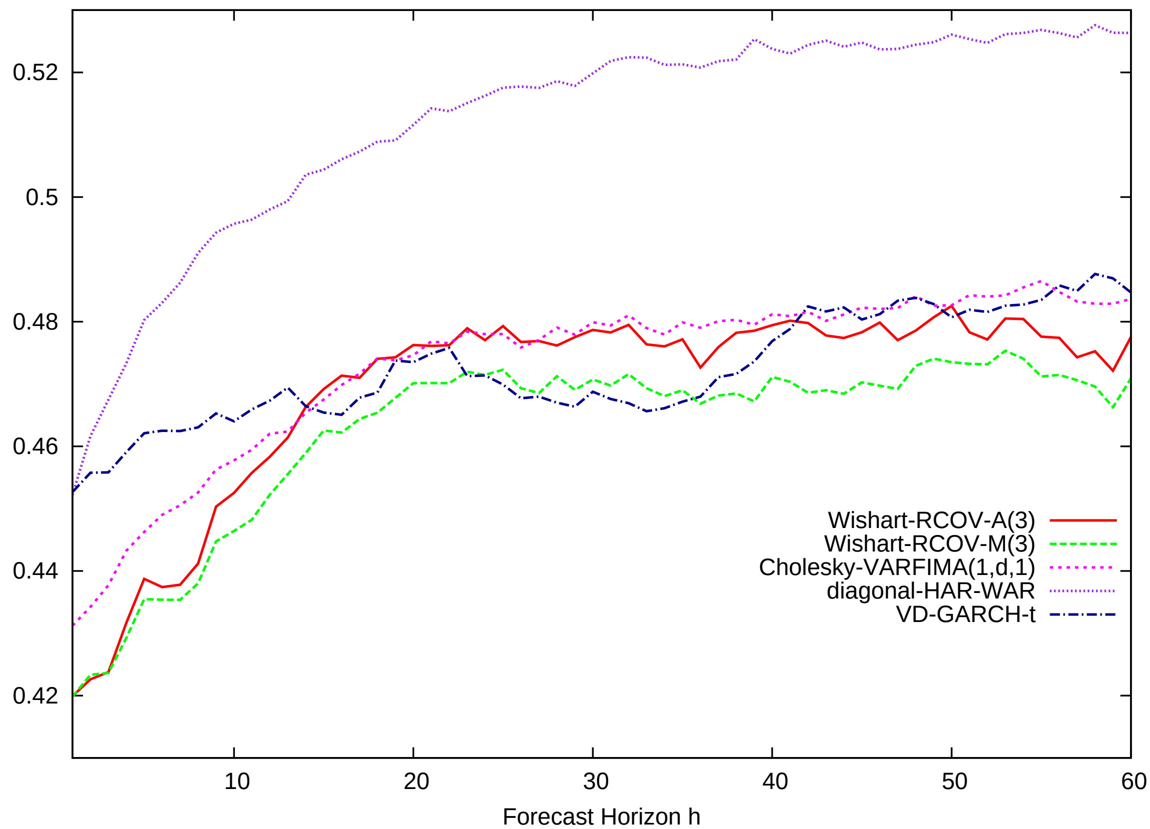


Figure 11: Sample variances of global minimum variance portfolios against forecast horizon for various models.

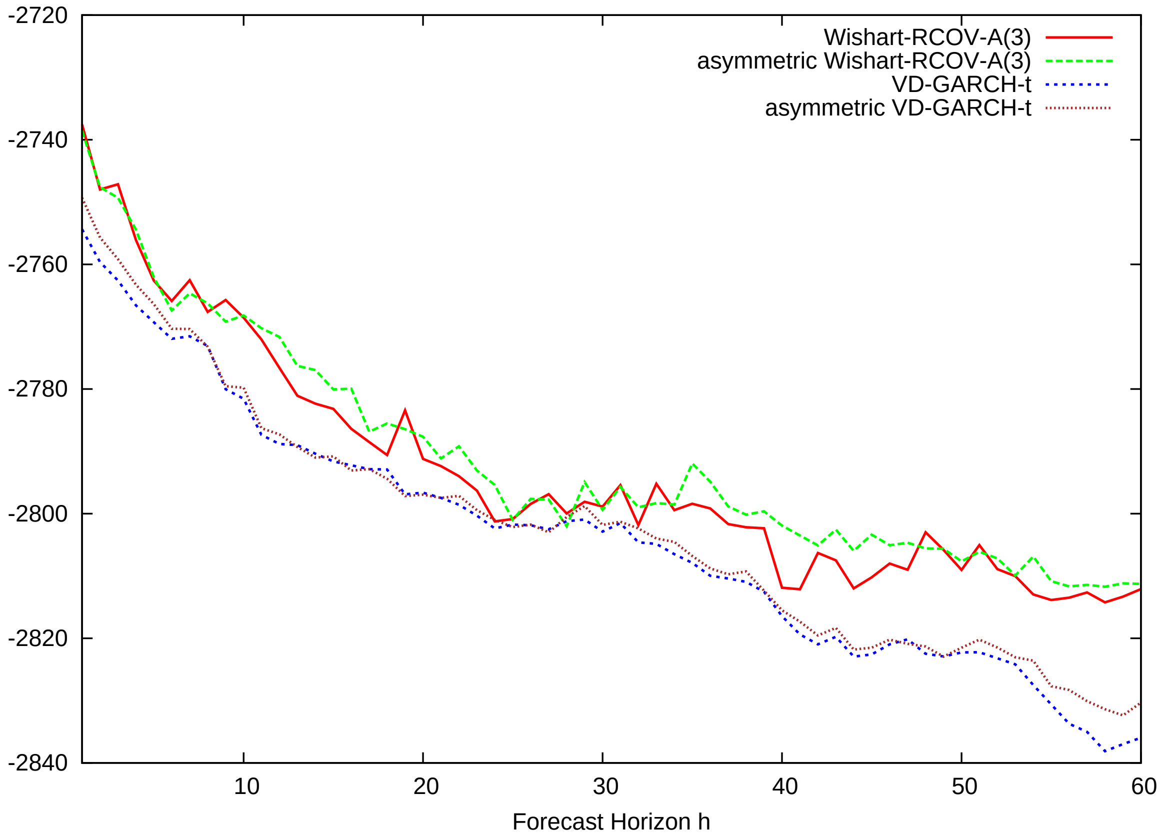


Figure 12: Term structure of cumulative log-predictive likelihoods for asymmetric Wishart-RCOV model and VD-GARCH-t model

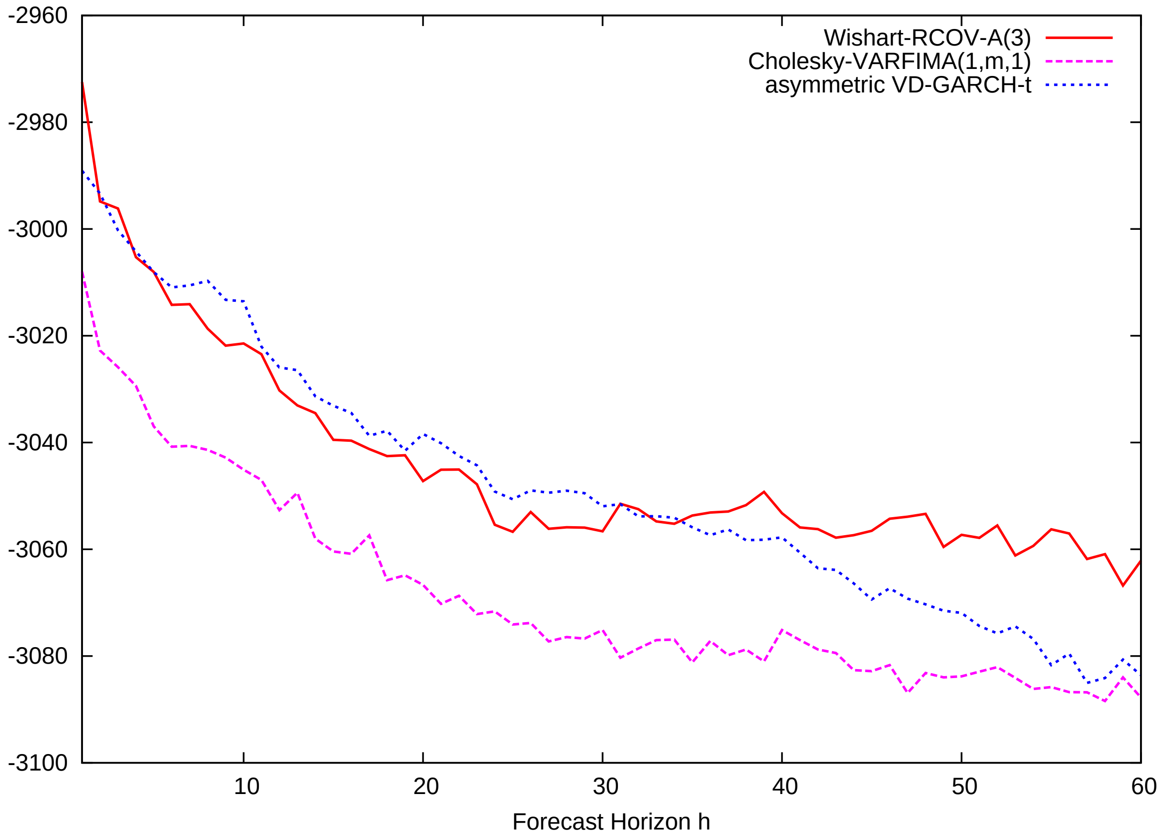


Figure 13: Term structure of cumulative log-predictive likelihoods: Wishart-RCOV-A(3), Cholesky-VARFIMA and asymmetric VD-GARCH-t model, using close-to-close returns

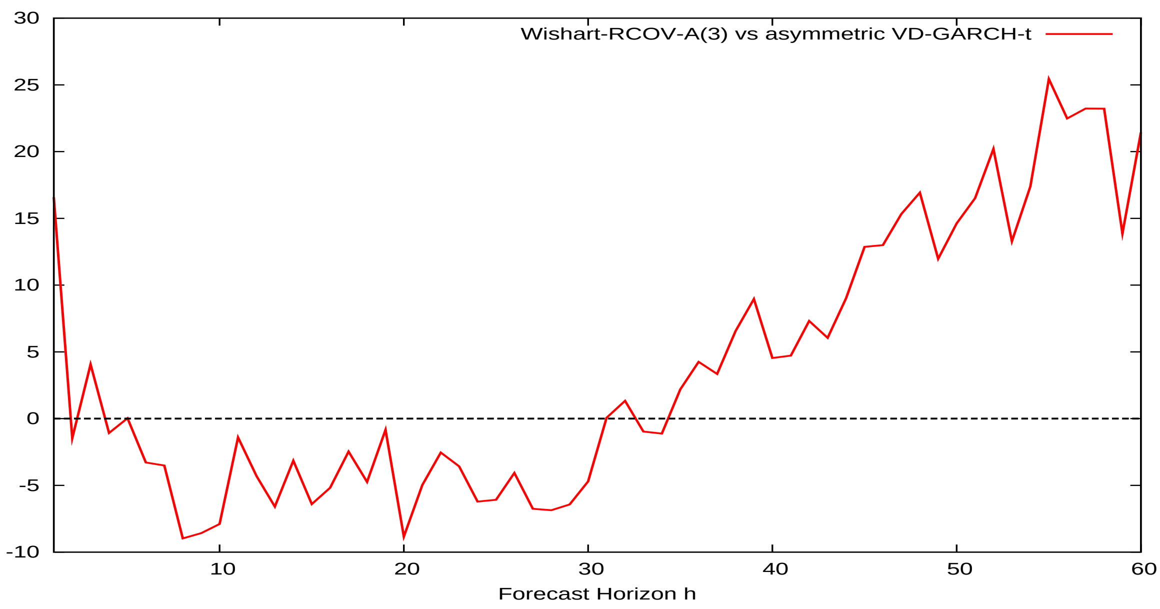


Figure 14: Log Predictive Bayes Factor using close-to-close return: Wishart-RCOV-A(3) vs asymmetric VD-GARCH-t