TI 2016-061/III
Tinbergen Institute Discussion Paper

Realized Wishart-GARCH: A Score-driven Multi-Asset Volatility Model

Peter Reinhard Hansen[a]
Pawel Janus[b]
Siem Jan Koopman[c]

[a] *University of North Carolina at Chapel Hill, United States;*
[b] *UBS Global Asset Management, Zürich, Switzerland;*
[c] *Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, the Netherlands.*

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Realized Wishart-GARCH: A Score-driven Multi-Asset Volatility Model

Peter Reinhard Hansen[a] Paweł Janus[b] Siem Jan Koopman[c*]

[a] *University of North Carolina, Chapel Hill & CREATES*

[b] *UBS Global Asset Management, Zürich*

[c] *Vrije Universiteit Amsterdam, Tinbergen Institute & CREATES*

August 6, 2016

Abstract

We propose a novel multivariate GARCH model that incorporates realized measures for the variance matrix of returns. The key novelty is the joint formulation of a multivariate dynamic model for outer-products of returns, realized variances and realized covariances. The updating of the variance matrix relies on the score function of the joint likelihood function based on Gaussian and Wishart densities. The dynamic model is parsimonious while each innovation still impacts all elements of the variance matrix. Monte Carlo evidence for parameter estimation based on different small sample sizes is provided. We illustrate the model with an empirical application to a portfolio of 15 U.S. financial assets.

Keywords: high-frequency data; multivariate GARCH; multivariate volatility; realized covariance; score; Wishart density.

JEL Classification: C32, C52, C58

---

*Corresponding author: Siem Jan Koopman, Vrije Universiteit Amsterdam, Department of Econometrics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, phone: +31 20 598 6019, fax: +31 20 598 6020, email: s.j.koopman@vu.nl. The first and third authors acknowledge financial support by the Center for Research in Econometric Analysis of Time Series (DNRF78), CREATES, funded by the Danish National Research Foundation. The views expressed herein are those of the authors and not necessarily those of UBS and UBS Global Asset Management, which does not accept any responsibility for the contents and opinions expressed in this paper. Finally we would like to thank Asger Lunde for his support and comments.

1 Introduction

Modeling conditional dependency structure of financial assets through time-varying variance matrices is typically based on multivariate extensions of generalized autoregressive conditional heteroskedasticity (GARCH) models and stochastic volatility (SV) models for daily returns. These classes of models aim to extract time-varying variance matrices from vector time series of financial returns. The dynamic process for multivariate volatility (variances and covariances) is typically specified as a vector autoregressive moving average process. Various multivariate GARCH and SV models have been developed and applied in recent years. For a comprehensive overview of multivariate GARCH models, we refer to Bauwens, Laurent and Rombouts (2006) and Silvennoinen and Teräsvirta (2009). Reviews of multivariate SV models are provided by Asai, McAleer and Yu (2006) and Jungbacker and Koopman (2006)

The main shortcoming of traditional multivariate GARCH and SV models is that they solely rely on daily returns to infer the current level of multivariate volatility. Given the increasing availability of high-frequency intraday data for a vast range of financial assets, the use of only low-frequency daily data appears inefficient for making statistical inference on time-varying multivariate volatility. One important consequence is that models based on daily data do not adapt quickly enough to changes in volatilities which is key to track the financial risk in a timely manner; see Andersen, Bollerslev, Diebold and Labys (2003) for a more detailed discussion. Various attempts have been made to use high-frequency intraday data into the modeling and analysis of volatility. For instance, information from high-frequency data can be incorporated by adding it in the form of an explanatory variable to the GARCH or SV volatility dynamics; see Engle (2002) and Koopman, Jungbacker and Hol (2005).

With the advent of high-frequency data, one can estimate ex-post daily return variation with so-called realized variance (or realized volatility) measures; see Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold and Labys (2001) and Barndorff-Nielsen and Shephard (2002). Inherent to high-frequency data is the microstructure noise (bid-ask bounce, decimal misplacement etc.) which leads to bias and inconsistency of standard measures. A number of related measures have been developed to restore the consistency; see Aït-Sahalia, Mykland and Zhang (2005), Barndorff-Nielsen, Hansen, Lunde and Shephard (2008), Jacod, Li, Mykland, Podolskij and Vetter (2009), Hansen and Horel (2009), and references therein. In the case of multiple assets, realized measures of asset covariance have also been proposed and considered; see Malliavin and Mancino (2002), Hayashi and Yoshida (2005), Christensen, Kinnebrock and

Podolskij (2010), Barndorff-Nielsen, Hansen, Lunde and Shephard (2011*a*), Griffin and Oomen (2011), and references therein. The analysis and forecasting of realized volatility series can be based on existing time series methods. Andersen et al. (2001) have explored the use of autoregressive models to analyze time series of realized volatilities. They have found considerable improvements in volatility forecasts over standard GARCH models. More recently, some new promising models have been proposed that rely on time series of realized measures. Gourieroux, Jasiak and Sufana (2009) have proposed (non-central) Wishart autoregressive model for realized covariance matrix. Asai and So (2013) and Golosnoy, Gribisch and Liesenfeld (2012) have proposed alternative dynamic formulation for covariance parameters with the underlying Wishart density. Chiriac and Voev (2011) and Bauer and Vorkink (2011) have proposed models for realized covariances using appropriate transformations to ensure the positive definiteness of the variance matrix. In our study we also rely on the Wishart density but we propose a novel conditional model formulation for the variance matrix. For the updating of the conditional variance matrix, daily as well as intra-daily financial returns are used.

An approach that combines possibly several measures of volatility based on low- and high-frequency data is recently proposed by Engle and Gallo (2006). They model jointly close-to-close returns, range and realized variance with the multiplicative error model (MEM) where each measure has its own dynamics for the update of latent volatility augmented with lagged values of other two measures. Engle and Gallo (2006) find that combination of these three noisy measures of volatility brings gains when making medium-run volatility forecasts. Shephard and Sheppard (2010) explore a similar model structure and refer to it as the HEAVY model, which was extended to the multivariate setting in Noureldin, Shephard and Sheppard (2012). Then a further extension based on the use of more heavy-tailed distributions is proposed by Janus, Lucas and Opschoor (2014). In the aforementioned models, a time-varying parameter is introduced for every realized measure that is included in the model. An alternative approach is the Realized GARCH framework by Hansen, Huang and Shek (2012) where daily returns and realized measures of volatility are both associated with the same latent volatility which circumvents the need for additional latent variables. The Realized GARCH framework has been developed further in Hansen and Huang (2012) and Hansen et al. (2014). A Realized SV model is proposed by Koopman and Scharth (2013). Our present work can be regarded as an extension of the Realized GARCH model to the multivariate case, but it is novel since we adopt a score-driven approach to the time-varying conditional variance matrix.

We contribute to the recent developments in the joint modeling of daily returns and realized measures. Our primary aim is to specify a model for the daily time-varying variance matrix and to extract it by using both low- and high-frequency data. We propose a specification for the unobserved daily variance matrix as a function of realized measures of daily covariance matrices and past outer-products of daily return vectors. The challenge is to suitably weight these different variance and covariance signals. For our purpose, we adopt the score-driven framework of Creal, Koopman and Lucas (2013). Our joint modeling framework relies on a Wishart density for realized variance matrices and on a Gaussian density for vectors of daily returns. The updating of the time-varying variance matrix is driven by the scaled score of the predictive joint likelihood function. The score function turns out to be a weighted combination of the outer-product of daily returns and the actual realized measures. The weighting relies on the number of degrees of freedom in the Wishart distribution. We refer to the resulting model as the Wishart-GARCH model. In our empirical illustration for a portfolio of 15 U.S. financial assets, the parameter estimates imply that the realized measures receive more weight than the outer-product of the vector of daily returns. It confirms that the realized measure is a more accurate measure of the variance matrix as it exploits intraday high-frequency data. We also present a model formulation that can accommodate several realized variance matrix measures.

A key feature of the Wishart-GARCH model is that the dynamics of the conditional variance matrix relies on the score function of the predictive likelihood. Blasques et al. (2015) argue that variance updating based on the score function is locally optimal in a Kullback-Leibler sense. It also offers a flexible structure that allows for complex interdependence between variances and covariances since they are all influenced by the score vector. These features distinguish our model from low frequency GARCH models that rely on outer-product of daily returns only. It also distinguishes our model from the MEM and HEAVY models which are driven by realized variances and covariances only. The aforementioned models potentially could be adapted to allow for cross-asset effects, at the expense of a more complex parameterization. In contrast, the Wishart-GARCH model achieves this flexibility in a simple parsimonious framework.

The structure of the paper is organized as follows. In Section 2, after we have set out notation and assumptions, we introduce the Wishart-GARCH model for multivariate volatility. In Section 3, we conduct a set of simulations to study to the likelihood-based estimation. Section 4 presents the results of our empirical analysis for some portfolio of NYSE equities, while Section 5 concludes the paper. The proofs of the main results in the paper are given in the Appendix.

2 The Wishart-GARCH Model

The development of our model starts with the assumption that at the end of each trading day we have a vector of daily returns and a measure (or possibly several measures) of daily realized covariance of assets under analysis. Our primary goal is to build a model for the vector of returns, while making use of both low- and high-frequency data. The proposed structure of the model permits the use of several realized measures, possibly computed with different sampling frequencies. We start this section by discussing our modeling assumptions. We then describe the modeling strategy and we provide technical details of new model for multivariate volatility. Some matrix notation and preliminary results are presented in Appendix A and proofs are collected in Appendix B.

2.1 Modeling assumptions

Let $r_t \in \mathbb{R}^k$ denote a $k \times 1$ vector of daily (demeaned) log returns for k assets and let the $X_t \in \mathbb{R}^{k\times k}$ denote a $k \times k$ realized covariance matrix of k assets on day t, with $t = 1, \dots, T$. Let $\mathcal{F}_{t-1}$ be the sigma field generated by the past values of r_t and X_t. We assume the following conditional densities

$$r_t|\mathcal{F}_{t-1} \sim N_k(0, H_t), \tag{1}$$

$$X_t|\mathcal{F}_{t-1} \sim W_k(V_t, \nu), \tag{2}$$

where H_t is the $k \times k$ variance matrix of the multivariate normal distribution $N_k(0, H_t)$ with mean zero and V_t is the mean of the k-th dimensional Wishart distribution $W_k(V_t, \nu)$ with degrees of freedom $\nu \geq k$. The variance matrices H_t and V_t are both measurable with respect to $\mathcal{F}_{t-1}$. The densities in (1) and (2) are conditionally independent. Only the dependence between H_t and V_t leads to the unconditional dependence of (1) and (2). The coefficient ν encapsulates the precision by which X_t measures V_t. A larger value of ν implies a more accurate measurement X_t for V_t.

The normal density function for $r_t|\mathcal{F}_{t-1}$ is given by

$$\frac{1}{(2\pi)^{\frac{k}{2}}|H_t|^{\frac{1}{2}}}\exp\Big\{-\frac{1}{2}\mathrm{tr}\big(H_t^{-1}r_t r_t'\big)\Big\}, \tag{3}$$

and the density function of the k-variate standard Wishart distribution for $X_t|\mathcal{F}_{t-1}$ is given by

$$\frac{|X_t|^{(\nu-k-1)/2}}{2^{(\nu k)/2}\nu^{-(\nu k)/2}|V_t|^{\nu/2}\Gamma_k\left(\frac{\nu}{2}\right)}\exp\left\{-\frac{\nu}{2}\mathrm{tr}\left(V_t^{-1}X_t\right)\right\}, \tag{4}$$

with Γ_k as the multivariate Gamma function $\Gamma_k(a) = \pi^{\frac{k(k-1)}{4}}\prod_{i=1}^{k}\Gamma\left(a+(1-i)/2\right)$ for any $a > 0$.

Remark 1. We assume that realized covariance X_t is available on each day t as it can be measured consistently by the multivariate realized kernel of Barndorff-Nielsen, Hansen, Lunde and Shephard (2011*a*) or related measures described by Griffin and Oomen (2011).

The distributional assumption (1) implies that the outer product of daily returns is distributed as

$$r_t r_t' \sim W_k(H_t, 1), \tag{5}$$

which is the singular Wishart distribution with one degree of freedom since matrix $r_t r_t'$ has rank one by construction; see Srivastava (2003). If r_t is the vector of close-to-close returns, then H_t measures overnight variation along with intraday variation, while V_t measures the intraday variation. If r_t is the vector of open-to-close returns, then both H_t and V_t measure the variation over a particular trading day. It is standard to exclude the overnight return for computation of realized measures, while modeling of daily returns is based on both open-to-close and close-to-close returns.

The measurement equations are given by

$$r_t = H_t^{1/2}\varepsilon_t, \qquad X_t = V_t^{1/2}\eta_t V_t^{1/2}, \tag{6}$$

where $A^{1/2}$ denotes the square root matrix of A and where measurement innovations are assumed to be *iid* distributed as

$$\varepsilon_t \sim N_k(0, I_k), \qquad \eta_t \sim W(\nu, I_k/\nu),$$

where ε_t is a vector and η_t is a matrix random variable.

Remark 2. Given the result in (5), we can redefine the conditional densities as

$$\begin{aligned} r_t r_t' &= H_t^{1/2}\zeta_t H_t^{1/2}, \qquad && \zeta_t \sim W_k(I_k, 1), \\ X_t &= V_t^{1/2}\eta_t V_t^{1/2}, && \eta_t \sim W_k(I_k, \nu), \end{aligned}$$

so that model measurement equations are expressed in terms of variances and covariances.

It is natural to assume that variation of conditional variance of realized returns and of conditional mean of realized covariance share the same source. For this reason, we impose the following structure,

$$H_t = \Lambda^{1/2} V_t \Lambda^{1/2}, \tag{7}$$

where we assume that $\Lambda^{1/2}$ is a $k \times k$ diagonal matrix with $\lambda_{ii} = \lambda_i > 0$, $i = 1, \ldots, k$. We could have proposed to have two separate models for H_t and V_t, and to link them through the lagged values of each other. This is the modeling strategy of Noureldin et al. (2012) who refer to it as the multivariate HEAVY model. This approach implies the explicit modeling of two latent matrix variables H_t and V_t. Instead, we aim to provide a single dynamic formulation for conditional multivariate volatility, based on one-step ahead predictions, denoted by $V_{t+1|t}$, which contains any information in the form of noisy measures of current level of volatilities. Finally, when r_t is defined as a vector of daily close-to-close returns, the individual elements of $\Lambda^{1/2}$ should be larger than the corresponding elements when r_t includes overnight variation.

Our set of distributional assumptions implies the following,

$$\mathrm{E}[r_t r_t' | \mathcal{F}_{t-1}] = \Lambda^{1/2} V_t \Lambda^{1/2}, \qquad \mathrm{E}[X_t | \mathcal{F}_{t-1}] = V_t, \tag{8}$$

and

$$\begin{aligned}
\mathrm{Var}[\mathrm{vec}(r_t r_t') | \mathcal{F}_{t-1}] &= (I_{k^2} + K_k)\big(\Lambda^{1/2} \otimes \Lambda^{1/2}\big)\big(V_t \otimes V_t\big)\big(\Lambda^{1/2} \otimes \Lambda^{1/2}\big), && (9)\\
\mathrm{Var}[\mathrm{vec}(\Lambda^{-1/2} r_t r_t' \Lambda^{-1/2}) | \mathcal{F}_{t-1}] &= (I_{k^2} + K_k)\big(V_t \otimes V_t\big), && (10)\\
\mathrm{Var}[\mathrm{vec}(X_t) | \mathcal{F}_{t-1}] &= \nu^{-1}(I_{k^2} + K_k)\big(V_t \otimes V_t\big). && (11)
\end{aligned}$$

The results of (9) and (11) follow directly from Magnus and Neudecker (1979). We notice that the result in (8) corresponds to the conditional second moment, while the results in (9) and (10) correspond to the conditional fourth moment (kurtosis) of returns, possibly adapted for overnight variation. It is a convenient feature of our modeling framework that conditional second moments of realized covariance (11) provides model-implied volatilities-of-volatilities and volatility cross, or spillover, effects.

To model the dynamic properties of V_t, we introduce the vector f_t which is assumed to represent V_t fully and uniquely. To ensure a positive definite variance matrix V_t in our analyses, we can employ the Cholesky decomposition as given by

$$V_t = C_t C_t', \tag{12}$$

and we model the dynamic properties of the lower-triangular matrix C_t with unique Cholesky factors. In this case we can consider the specification $f_t = \text{vech}(C_t)$ where $\text{vech}(A)$ stacks the diagonal and lower-triangular elements of some matrix A into a vector. Another example is the specification $f_t = \text{vech}(\text{logm}(V_t))$ where logm is the matrix-logarithm operator. In our empirical example we employ the first specification, whereas we detail the second specification for the univariate case in Section 2.4. It is also possible to impose structure on V_t with the purpose to lower the dimension of the time-varying parameter vector f_t.

2.2 Score-driven dynamics

In this section we discuss how the dynamic properties of the time-varying parameter f_t can be specified. We provide details of how the model formulation is derived taking into account the measurement densities that are introduced in the previous section. We adopt the score-driven approach to time-varying parameters as developed by Creal et al. (2013). They construct a general dynamic modeling framework in which the local (at time t) score function of the conditional likelihood function is used for updating time-varying parameters. Given that the conditional score function is a function of past observations, the model belongs to the class of observation-driven models; see Cox (1981).

Consider a series of m vector or matrix variables $Z_t^1, \ldots, Z_t^m$. The measurement density for the ith variable is given by

$$Z_t^i \sim \varphi_i(Z_t^i | f_t, \mathcal{F}_{t-1}; \psi), \qquad \text{for} \quad i = 1, \ldots, m, \quad \text{and} \quad t = 1, \ldots, T, \tag{13}$$

where f_t is $d \times 1$ vector of latent time-varying parameters, $\mathcal{F}_t = \{Z_1, \ldots, Z_t\}$ is the information set containing all observations up to time t, and ψ is a vector of (unknown) static model parameters. In this framework, the individual densities φ_i may correspond to different families of distributions. All distributions however depend partially on the same vector of the time-varying parameter vector f_t. In particular, consider (1) and (2), and decomposition (7), where return vector and variance matrix have different distributions but are assumed to be propelled by a common variance matrix, $V_t = V(f_t)$. Different mappings of f_t to V_t can be considered, see the discussion below equation (12). We assume that innovations of variables Z_t are independent conditional on f_t and on the information set $\mathcal{F}_{t-1}$. The log-likelihood is then given by

$$\mathcal{L}(\psi) = \sum_{t=1}^{T} \sum_{i=1}^{m} \log\ \varphi_i(Z_t^i | f_t, \mathcal{F}_{t-1}; \psi).$$

Our assumption rules out the possibility of correlated innovations, that is $\mathrm{E}[\varepsilon_{it}\eta_{jt}] = 0$, for any i, j in (6). However, any asymmetric effects can be easily introduced when specifying updating equation for the time-varying parameter f_t as we discuss below. The density in (13) may also depend on some exogenous variables, we however omit this extension for simplicity in notation.

The time-varying f_t is updated via the recursive equation

$$f_{t+1} = \omega + \sum_{i=1}^{p} B_i f_{t-i+1} + \sum_{j=1}^{q} A_j s_{t-j+1}, \tag{14}$$

where ω is an $d \times 1$ vector of constants, s_t is a mean-zero and finite variance martingale difference sequence, B_i and A_j are $d \times d$ matrices with loadings. The parameters ω, $B_1, \ldots, B_p$, $A_1, \ldots, A_q$ and some possible density specific unknown parameters, such as the number of degrees of freedom in the Wishart density, are all collected in the static parameter vector ψ. The vector autoregressive moving average representation (14) proves convenient for understanding the statistical dynamic properties of the f_t process. The specification (14) can be extended to incorporate some exogenous variables or other functions of lagged endogenous variables, or one could also consider long-memory specification of (14).

Given the linear updating in (14), the main challenge is to formulate the martingale innovation s_t. Here we adopt an observation-driven approach in which we formulate the innovation term s_t as a function of directly observable variables. Our modeling approach follows Creal et al. (2013) by setting the innovation s_t equal to the scaled score of the predictive likelihood function which under standard regularity conditions forms a martingale sequence. In particular, the score vector takes an additive form given by

$$\nabla_t = \sum_{i=1}^{m} \nabla_{i,t} = \sum_{i=1}^{m} \frac{\partial \log\ \varphi_i(Z_t^i | f_t, \mathcal{F}_{t-1}; \psi)}{\partial f_t}, \tag{15}$$

which corresponds to the sum of individual scores. The scaling term is based on the Fisher information matrix and can also be expressed in additive form,

$$\mathcal{I}_t = \sum_{i=1}^{m} \mathcal{I}_{i,t} = \sum_{i=1}^{m} \mathrm{E}[\nabla_{i,t}\nabla_{i,t}'|\mathcal{F}_{t-1}]. \tag{16}$$

The innovation term is now defined as

$$s_t = \mathcal{I}_t^{-1/2}\nabla_t, \tag{17}$$

such that $\mathrm{E}[s_t|\mathcal{F}_{t-1}] = 0$ and $\mathrm{E}[s_t s_t'|\mathcal{F}_{t-1}] = I_d$. In this approach, the one-step ahead prediction of latent parameters f_t is based on the scaled score that exploits the full likelihood function at time t. Along with the measurement densities (13), equations (14) and (17) are similarly formulated as in Creal et al. (2013). In the remainder we will take $p = q = 1$.

2.3 The main result

In this section we discuss the details of the score-driven model where the time-varying parameter f_t is based on the Cholesky decomposition of V_t. Given the score-driven model as formulated generally, we specify the dynamic specification for the one-step ahead prediction of the variance matrix $V_{t+1|t}$. To operate the model we have V_t as a function of f_t, that is $V_t = V(f_t)$ and we need expressions for the score function and the Fisher information matrix.

The log-likelihood function at time t can be decomposed into two contributions, that is

$$\mathcal{L}_t(\psi) = \mathcal{L}_{1,t} + \mathcal{L}_{2,t},$$

where the individual log-likelihoods are given by

$$\mathcal{L}_{1,t} = d(k) - \frac{1}{2}\log|\Lambda^{1/2}V_t\Lambda^{1/2}| - \frac{1}{2}\mathrm{tr}\big((\Lambda^{1/2}V_t\Lambda^{1/2})^{-1}r_t r_t'\big), \tag{18}$$

$$\mathcal{L}_{2,t} = d(k,\nu) + \frac{\nu-k-1}{2}\log|X_t| - \frac{\nu}{2}\log|V_t| - \frac{\nu}{2}\mathrm{tr}\big(V_t^{-1}X_t\big), \tag{19}$$

with $d(k) = -\frac{k}{2}\log(2\pi)$ and $d(k,\nu) = \frac{\nu k}{2}\log(\nu) - \frac{\nu k}{2}\log(2) - \log\left(\Gamma_k\left(\frac{\nu}{2}\right)\right)$.

The positive definiteness of the variance matrix V_t is ensured by employing the Cholesky decomposition of (12), that is $V_t = C_t C_t'$. It then suffices to specify a model for the diagonal and lower-triangular parts of matrix C_t and the time-varying parameter vector is defined as

$$f_t = \mathrm{vech}(C_t) \qquad \Leftrightarrow \qquad V_t = \mathrm{unvech}(f_t)\mathrm{unvech}(f_t)', \tag{20}$$

such that f_t is a $k^* \times 1$ vector with $k^* = k(k+1)/2$. For the updating equation (14), we require the score vector and Fisher information matrix that we obtain as described in Section 2.2.

Theorem 1. *For the measurements densities* (1) *and* (2)*, and for the decomposition (12), the score vector of dimension* $k^* \times 1$ *is given by*

$$\nabla_t = \frac{1}{2}\dot{V}_t' D_k' \big(V_t^{-1} \otimes V_t^{-1}\big)\Big(\nu\big[\mathrm{vec}(X_t) - \mathrm{vec}(V_t)\big] + \big[\mathrm{vec}(\Lambda^{-1/2} r_t r_t' \Lambda^{-1/2}) - \mathrm{vec}(V_t)\big]\Big),$$

where $\dot{V}_t = L_k(I_{k^2} + K_k)(C_t \otimes I_k)\tilde{L}_k'$.

It follows that $\mathrm{E}[\nabla_t|\mathcal{F}_{t-1}] = 0$ under standard regularity conditions implying that ∇_t forms martingale sequence. The expression for the score in Theorem 1 indicates that when making one-step ahead prediction from V_t to V_{t+1}, information from the deviations of realized covariance X_t from its mean V_t receives a weight ν, whereas information from deviations of $r_t r_t'$ from V_t (correcting for overnight variation Λ if r_t is vector of close-to-close returns) receives a weight of one. This model feature is pertinent as the outer-product of daily returns contains only a weak signal about the current covariance of assets as it does not exploit intraday information. It follows from Theorem 1 that score-based derivation of the model is not invariant to the decomposition applied to the covariance matrix V_t to ensure positive definiteness. The dependence enters through the term $\dot{V}_t$ which is unique for a selected decomposition. We find that $\dot{V}_t$ collects the first order derivatives of the full covariance matrix $V_t = V(f_t)$ with respect to f_t.

Theorem 2. *For the measurements densities (1) and (2), and for the decomposition (12), the Fisher information matrix of dimension* $k^* \times k^*$ *is given by*

$$\mathcal{I}_t = \mathrm{E}[\nabla_t \nabla_t' | \mathcal{F}_{t-1}] = \frac{1+\nu}{4}\dot{V}_t' D_k' \big(V_t^{-1} \otimes V_t^{-1}\big)\big(I_{k^2} + K_k\big) D_k \dot{V}_t.$$

The square root of the inverse of the conditional information matrix may be used to scale the score vector, such that $\mathrm{E}[s_t s_t' | \mathcal{F}_{t-1}] = I_{k^*}$ with s_t defined in (17). The scaling (17) implies the need to invert the Fisher information matrix $\mathcal{I}_t$ whose dimension grows at a rate proportional to $O(k^2)$. This step is therefore the most computationally demanding.

Remark 3. The results in Theorems 1 and 2 hold for two measurement equations defined as in (1) and (2). Applying the results presented in the Appendix, it is straightforward to extend the model setup to incorporate several noisy measures of daily equity covariance matrix. For instance, let

$$X_t^i = V_t^{1/2} \eta_t^i V_t^{1/2}, \qquad \eta_t^i \sim W_k(I_k, \nu^i), \qquad i = 1, \ldots, G,$$

where X_t^i is a noisy measure of daily realized covariance, for $i = 1, \ldots, G$, with $G \in \mathbb{N}$. Then

$$\nabla_t = \frac{1}{2} \dot{V}_t' D_k' \big(V_t^{-1} \otimes V_t^{-1}\big) \left(\sum_{i=1}^{G} \nu^i \big[\mathrm{vec}(X_t^i) - \mathrm{vec}(V_t)\big] \right),$$

and

$$\mathcal{I}_t = \mathrm{E}[\nabla_t \nabla_t' | \mathcal{F}_{t-1}] = \dot{V}_t' D_k' \big(V_t^{-1} \otimes V_t^{-1}\big) \big(I_{k^2} + K_k\big) D_k \dot{V}_t \, \frac{\sum_{i=1}^{G} \nu^i}{4},$$

where the numbers of degrees of freedom $\nu^1, \nu^2, \ldots, \nu^G$ are estimated along with other model static parameters, and where $\nu^i \equiv 1$ if $X_t^i = r_t r_t'$.

The key distinguishing feature of our model specification is that each element of the innovation vector s_t exploits the full likelihood information. This feature turns out to be relevant when avoiding the curse of dimensionality as we may consider diagonal specifications such as $B_i = \mathrm{diag}(\beta_1^i, \ldots, \beta_{k^*}^i)$ and $A_j = \mathrm{diag}(\alpha_1^j, \ldots, \alpha_{k^*}^j)$ or even a simple scalar version defined through $B_i = \beta^i I_{k^*}$ and $A_j = \alpha^j I_{k^*}$ in (14). In either case, the model dynamics allows for a complex interdependence between all variances and covariances such that the one-step update from V_t to $V_{t+1|t}$ is driven by own as well as cross-asset effects. This feature is model-specific and distinguishes our model from low frequency standard GARCH model that are driven by daily returns, typically the outer-product of daily returns $r_t r_t'$. It also distinguishes our model from the high frequency HEAVY model that is driven by a single realized measure and does not immediately lend itself to allow for cross-asset effects.

2.4 Special Case $k = 1$: The univariate model

We provide the details for the case of a single asset, the univariate case. We formulate an alternative volatility model for a single asset. Consider the case of $k = 1$, let r_t denote daily return, let X_t denote realized measure of variance and let V_t denote the unobserved daily variance. Our modeling assumptions from Section 2.1 apply in this specific case. Hence, the Wishart reduce to Gamma distributions. We obtain

$$r_t^2 | \mathcal{F}_{t-1} \sim \mathrm{Gamma}(1, V_t), \tag{21}$$

$$X_t | \mathcal{F}_{t-1} \sim \mathrm{Gamma}(\nu, V_t/\nu), \tag{22}$$

where Gamma(a, b) denotes the Gamma distribution with shape parameter $a > 0$ and scale parameter $b > 0$, which has the density

$$f(x) = \frac{1}{\Gamma(a/2)\,(2b)^{a/2}}\, x^{a/2-1} \exp\big(-\frac{x}{2b}\big).$$

In the univariate case, an alternative specification for V_t in terms of f_t can be considered. To guarantee a non-negative V_t, it is natural to take $V_t = \exp(f_t)$. Our score-driven model is then based on $f_t = \log V_t$ and we obtain

$$\nabla_t = \frac{1}{2V_t}\Big(\nu\big(X_t - V_t\big) + \big(r_t^2 - V_t\big)\Big) \qquad \text{and} \qquad \mathcal{I}_t = \frac{1+\nu}{2},$$

which can also be straightforwardly extended to incorporate several noisy measures of daily variance V_t in a similar manner as discussed in Remark 3. For instance, this result provides the possibility to define a new realized EGARCH model by

$$f_{t+1|t} = \omega + \beta f_t + \alpha \left\{ \nu\left(\frac{X_t}{V_t} - 1\right) + \left(\frac{r_t^2}{V_t} - 1\right) \right\},$$

such that daily log-variance is propelled by model-implied weighted sum of squared return and realized variance measure.

3 Estimation procedure and Monte Carlo study

We discuss the maximum likelihood estimation procedure and present simulation evidence for the statistical small-sample properties of the maximum likelihood estimation method for our model. We study estimation performance for varying sample size T and number of assets k.

3.1 Estimation procedure

The log-likelihood function is given by

$$\mathcal{L} = \sum_{t=1}^{T} \big(\mathcal{L}_{1,t} + \mathcal{L}_{2,t}\big), \tag{23}$$

where $\mathcal{L}_{1,t}$ and $\mathcal{L}_{2,t}$ are given in (18) and (19), respectively. The time-variation of V_t is determined by the score recursion (14), decomposition (12) and parametrization (20). The static parameter vector is given by

$$\psi = \left[\omega', \text{vec}(B_1)', \ldots, \text{vec}(B_p)', \text{vec}(A_1)', \ldots, \text{vec}(A_q)', \nu, (\lambda_1, \ldots, \lambda_k)\right]',$$

and contains $pk^2 + q(k(k+1)/2) + k + 1$ elements that need to be estimated; the number of parameters is therefore of order $O(k^2)$. The recursion (14) needs to be initialized and it is natural to set $s_0 = 0$ and f_0 either to the unconditional first moment estimated from the data or it can be added to the vector of parameters ψ which is then jointly estimated. In our empirical analysis we set f_0 to be the sample average of the matrices $R_1, \ldots, R_T$. For given parameter values ψ, the log-likelihood function can be evaluated in a straightforward way. In practice, ψ is unknown and estimation of all parameters is carried out by numerically maximizing (23) with respect to ψ. Maximization can be based on a standard quasi-Newton numerical optimization procedure and initial values of ψ can be determined through a grid search method. In the simulation study and the empirical application, the model parameters are estimated using numerical derivatives.

With increasing dimension k, the estimation of model parameters may become computationally demanding. A possible approach to reduce the number of parameters can be based on covariance targeting as proposed by Engle and Mezrich (1996) for GARCH models. Since the score recursion (14) admits a vector ARMA representation, the model intercept can be expressed, if stationarity conditions are satisfied. We can verify this by considering the unconditional moment and obtain

$$f_{t+1} = (I_{k^*} - \sum_{i=1}^{p} B_i)\text{E}[f_t] + \sum_{i=1}^{p} B_i f_{t-i+1} + \sum_{j=1}^{q} A_j s_{t-j+1},$$

where $\text{E}[f_t]$ is replaced by the moment estimator, $\hat{\text{E}}[f_t] = \text{vech}(\hat{C})$, with $T^{-1}\sum_{t=1}^{T} RK_t = \hat{C}\hat{C}'$ such that $\hat{C}$ is the lower-triangular matrix with Cholesky factors of a long run target as measured by the mean of realized measures series. The introduction of targeting leads to a two-step approach in estimation. We first remove the vector of constants by replacing it through some consistent estimator of the unconditional mean. Then maximize the log-likelihood function with respect to the remaining parameters. To avoid the curse of dimensionality further, parameter reduction can be achieved by setting $A_1, \ldots, A_q$ and $B_1, \ldots, B_p$ as diagonal matrices or to scalars. In either case, the diagonal and off-diagonal elements of covariance matrix are driven by own lagged values and importantly by cross-asset effects captured by individual score elements which all contain full likelihood information.

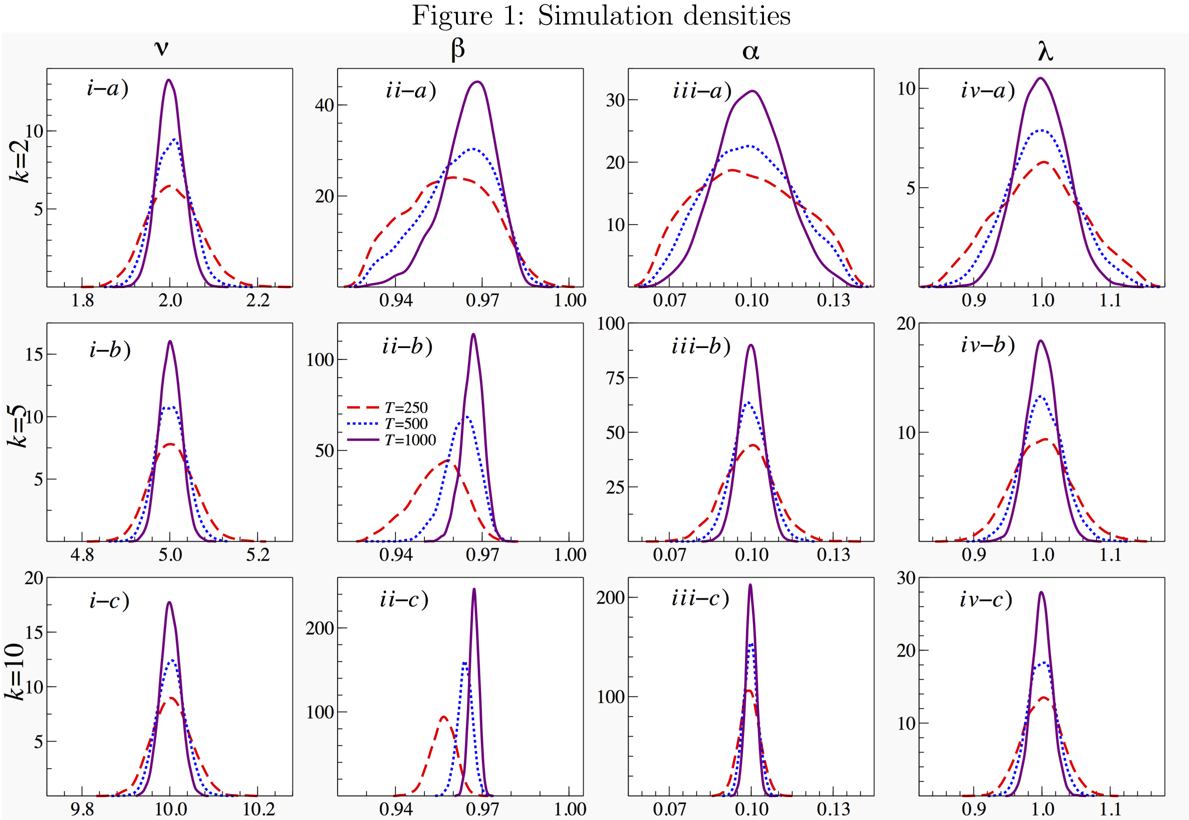
3.2 Monte Carlo study

We study properties of the likelihood-based estimation method by means of simulation exercises. We consider a dimension of $k \in \{2, 5, 10\}$ and we simulate a series of $T \in \{250, 500, 1000\}$ daily returns and daily variance matrices. For simplicity, we study the scalar specification. The Monte Carlo data generation process has taken the following parameter values

$$\nu = k, \qquad \omega = 0.10\ \mathbf{e}_{k^*}, \qquad \beta = 0.97, \qquad \alpha = 0.10, \qquad \text{and} \qquad \lambda = 1, \tag{24}$$

where $\mathbf{e}_{k^*}$ is the $k^* \times 1$ vector of ones. The parameter values are roughly in line with the empirical estimates that we present in Section 4. A high value of the autoregressive coefficient $\beta = 0.97$ is typically found in many volatility studies. We simulate 5000 datasets in our Monte Carlo study. For each generated dataset, we maximize the likelihood and we collect the estimates of parameters (24). We estimate the parameters without constraints and with covariance targeting.

Figure 1: Simulation densities



In Figure 1 we present the density kernel estimates of the histograms of the 5000 estimates for each parameter in ψ. Each graph contains three densities which are associated with the three time series dimensions 250, 500 and 1000. For increasing sample size T, the estimates concentrate more at their true values while the densities become more symmetric. We find some more skewness and heavy tails in the densities of the estimates obtained from the smaller sample size $T = 250$. In particular, the density for the memory parameter β is skewed to the left and the mode is shifted to the left near $\beta = 0.97$. This bias for β in small samples is somewhat expected since its estimation requires a relatively long time series. The number of degrees of freedom of the Wishart density ν can be estimated fairly robustly, even at moderate sample sizes. This finding is somewhat surprising given that ν is a highly nonlinear parameter.

If we increase k, the number of assets in our simulation study, the shapes of the densities become considerably more symmetric and more peaked around their respective true values, for example, compare panels a) to c). The improvement is particularly remarkable for parameters α and β. We may conclude from a practical viewpoint that the maximum likelihood method works well in terms of estimating the model parameters as long as the time series dimension is sufficiently large.

4 Empirical illustration

4.1 Dataset

In our empirical study for a portfolio of equities, we aim to measure the variation across firms and across market conditions. The equities consist of fifteen Dow Jones Industrial Average components with ticker symbols AA, AXP, BA, CAT, GE, HD, HON, IBM, JPM, KO, MCD, PFE, PG, WMT and XOM. The empirical study is based on consolidated trades (transaction prices) extracted from the Trade and Quote (TAQ) database through the Wharton Research Data Services (WRDS) system. The time stamp precision is one second. The sample period spans ten years, from January 2, 2001 to December 31, 2010, with a total of $T = 2515$ trading days for all equities.

Before we construct realized kernels, we carry out cleaning procedures to the raw transaction data. The importance of tick-by-tick data cleaning is highlighted by Hansen and Lunde (2006) and Barndorff-Nielsen et al. (2009) who provide a guideline on cleaning procedures based on the TAQ qualifiers that are included in the files (see TAQ User's Guide from WRDS). In particular, we carry out the following steps: (i) we delete entries with a time stamp outside the 9:30am-4:00pm window; (ii) we delete entries with transaction price equal to zero; (iii) we retain entries originating from a single exchange (NYSE in our application); (*iv*) we delete entries with corrected trades (trades with a correction indicator, "CORR" $\neq 0$); (*v*) we delete entries with abnormal sale condition (trades with "COND" has a letter code, except for "E" and "F"); (*vi*) we use the median price for multiple transactions with the same time stamp; (*vii*) we delete entries with prices that are above the ask plus the bid-ask spread.

We have in total 15 equities and will present results for a selection of $k \in \{2, 5, 15\}$. To conserve space, we will present results for the randomly selected ten bivariate systems and ten 5-variate models amongst the 15 equities. We also present results for our model of all 15 equities which require the modeling of a 15×15 conditional variance matrix.

In our empirical study we use the realized kernel based on the Parzen kernel and a subsampled realized covariance. Barndorff-Nielsen et al. (2011*b*) has shown that the subsampled realized covariance is equivalent to the Realized Kernel using the Bartlett kernel which we denote by RK_B. This estimator is based on returns with a sample frequency of 5 minutes. By shifting the starting time in 1-second increments, we obtain 300 different estimates and these are averaged to obtain RK_B.

Table 1 provides the number of observations and Table 2 provides the data fractions that we have retained in constructing the refresh sampling scheme. Given the dimension k, we record the resulting daily number of price observations. These statistics are averaged over particular years in our sample. We observe that for the 2-variate system we retain on average of around $60 - 65\%$ observations and this fraction is somewhat robust over time and across equities. The average number of refresh time observations is around 2800 and it moderately varies in time with higher volatility during crisis in 2007-2009. The average p statistics implies that we need, on average, to refresh observation every 8 seconds for $k = 2$.

For the 5×5 case the data loss is more pronounced. We retain around $35 - 40\%$ and we have 1800 refresh observations on average. Again, the p statistics is rather flat in time. In this case, we have an observation on average around every 13 seconds. For the 15×15 case, the overall average of fraction of retained observations equals around 22%, while the average number of observations is around 950 implying refresh observation every 25 seconds.

4.2 Capturing overnight variation

First we focus on the difference between modeling open-to-close and close-to-close returns. While the realized measures are defined over a period of a trading day i.e. from open-to-close,

Equities	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
2x2										
AA/CAT	803	1043	1340	1899	1919	2458	3538	3730	2810	2006
AXP/PFE	1805	2081	2486	2198	2372	2413	4007	4355	3527	2108
AXP/WMT	1508	1760	1865	2062	2323	2449	3994	4816	3900	2827
BA/HON	959	1248	1665	1719	2036	2171	3111	3069	2407	2154
CAT/KO	831	1144	1516	1934	2059	2382	3469	3809	3049	2585
GE/PFE	2064	2753	3061	3135	3156	3201	5105	5374	3514	1935
HD/JPM	1657	2022	2421	2329	2523	2817	4706	5454	3693	2906
IBM/PG	1566	1971	2390	2618	2659	3017	4252	4549	3493	2895
JPM/XOM	1476	1980	2516	2607	3044	3531	6187	7799	5747	4169
MCD/PG	1147	1516	1847	1969	2397	2517	3531	4330	3315	2442
5x5										
AA/AXP/IBM/JPM/WMT	827	940	1048	1304	1405	1553	2632	3074	2210	1526
AA/BA/CAT/GE/KO	570	736	933	1172	1247	1466	2340	2584	1790	1266
AXP/CAT/IBM/KO/XOM	671	885	1141	1272	1352	1520	2521	2787	2239	1924
BA/HD/JPM/PFE/PG	847	1060	1336	1332	1472	1639	2665	2920	2039	1395
BA/HD/MCD/PG/XOM	748	990	1232	1238	1462	1596	2483	2834	2009	1620
CAT/GE/KO/PFE/WMT	680	887	1055	1367	1481	1646	2625	2912	2070	1333
CAT/HON/IBM/MCD/WMT	626	783	951	1172	1332	1440	2186	2342	1857	1614
GE/IBM/JPM/PG/XOM	947	1256	1548	1586	1709	1915	3283	3773	2616	1863
HD/HON/KO/MCD/PG	662	868	1066	1136	1371	1414	2196	2443	1768	1432
HON/IBM/MCD/WMT/XOM	745	940	1079	1266	1537	1585	2408	2602	1994	1669
15x15										
AA/ ... /XOM	430	530	649	759	856	951	1613	1779	1267	894

Table 1: Summary statistics for the refresh sampling scheme. Note: Statistics are measured separately for each year in our sample. The averages over the daily number of high-frequency observations maintained by the refresh sampling scheme.

Equities	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
2x2										
AA/CAT	0.599	0.588	0.587	0.601	0.584	0.579	0.625	0.640	0.612	0.543
AXP/PFE	0.646	0.625	0.627	0.565	0.579	0.572	0.625	0.653	0.620	0.548
AXP/WMT	0.637	0.629	0.600	0.576	0.584	0.570	0.631	0.666	0.652	0.625
BA/HON	0.616	0.601	0.615	0.603	0.598	0.586	0.627	0.636	0.629	0.632
CAT/KO	0.583	0.573	0.577	0.595	0.584	0.585	0.636	0.651	0.643	0.626
GE/PFE	0.655	0.642	0.655	0.640	0.640	0.624	0.668	0.663	0.617	0.548
HD/JPM	0.644	0.625	0.635	0.615	0.621	0.607	0.652	0.636	0.575	0.582
IBM/PG	0.579	0.646	0.648	0.636	0.626	0.628	0.662	0.672	0.654	0.642
JPM/XOM	0.626	0.618	0.629	0.620	0.584	0.566	0.672	0.732	0.699	0.668
MCD/PG	0.643	0.628	0.624	0.610	0.621	0.597	0.634	0.662	0.643	0.637
5x5										
AA/AXP/IBM/JPM/WMT	0.338	0.338	0.322	0.347	0.354	0.348	0.396	0.407	0.371	0.329
AA/BA/CAT/GE/KO	0.314	0.288	0.308	0.336	0.334	0.339	0.385	0.394	0.375	0.334
AXP/CAT/IBM/KO/XOM	0.305	0.324	0.338	0.348	0.322	0.313	0.374	0.398	0.394	0.400
BA/HD/JPM/PFE/PG	0.357	0.348	0.363	0.345	0.354	0.354	0.400	0.395	0.360	0.328
BA/HD/MCD/PG/XOM	0.373	0.353	0.365	0.352	0.337	0.319	0.369	0.399	0.373	0.373
CAT/GE/KO/PFE/WMT	0.296	0.290	0.303	0.330	0.340	0.348	0.393	0.404	0.384	0.331
CAT/HON/IBM/MCD/WMT	0.305	0.326	0.330	0.333	0.339	0.336	0.382	0.396	0.385	0.388
GE/IBM/JPM/PG/XOM	0.358	0.366	0.384	0.371	0.361	0.352	0.416	0.426	0.392	0.362
HD/HON/KO/MCD/PG	0.359	0.347	0.354	0.353	0.362	0.348	0.393	0.405	0.385	0.389
HON/IBM/MCD/WMT/XOM	0.333	0.340	0.333	0.335	0.337	0.316	0.357	0.374	0.366	0.370
15x15										
AA/ ... /XOM	0.197	0.189	0.195	0.206	0.208	0.207	0.247	0.253	0.234	0.210

Table 2: Summary statistics for the refresh sampling scheme. Note: Statistics are measured separately for each year in our sample. The table presents the average ratio of the data maintained by the refresh sampling scheme.

the daily returns in the traditional GARCH modeling are defined as either open-to-close or close-to-close returns. Thus, for consistency with the realized measure one should use the open-to-close daily returns over the same interval. However, since it is commonly of interest to gauge and predict overnight variation, we provide a model specification that should be sufficiently flexible to distinguish the definition of daily returns. In case we use close-to-close returns, the individual elements of Λ in (7) need to capture the additional variation due to the overnight effects. Based on the univariate realized GARCH model of Hansen et al. (2012), we find that overnight variation may stand for around 25% of total daily variation. Hence we can expect that the individual elements of Λ are equal to approximately 1.25.

Table 3 presents the maximum likelihood estimates of the parameters in the Wishart-GARCH model based on the open-to-close returns data while Table 4 presents the estimates based on the close-to-close returns; both for a selection of equity portfolios of size $k = 2, 5, 15$. These estimates are based on a model specification with $A = \alpha I_{k^*}$ and $B = \beta I_{k^*}$. A key observation is that the parameter estimates of ν, α and β do not display large variation across different equity pairs. Nor are there large differences between the reported estimates in Tables 3 and 4. The exceptions however are the estimates of the elements in Λ that differ considerably in the two Tables. Table 5 reports the estimates of Λ for the full portfolio of 15 equities, and for both open-to-close and close-to-close returns. In the former case of open-to-close returns, the estimates of Λ (and for all portfolios), are close to unity and somewhat below. For the close-to-close returns, the estimates of Λ are between 1.20 and 1.40. These results suggest that if the interest is also to predict covariances of daily close-to-close returns, then the basic link function (7) is convenient and useful for modeling additional overnight variation. The estimates of other model parameters are virtually the same. In general, we find estimates of β being rather close to unity implying high persistence of the covariance matrix. We observe also that the dynamics of V_t put more weight on realized kernel measures as implied by the highly significant estimates of ν. Furthermore, we find that for a higher dimension k, more reliance is given to the realized measures as the estimates of the estimates of ν become higher and more significant. We notice that the degrees of freedom ν needs to grow with dimension k in order to ensure that the Wishart variance matrix does not become non-singular; also see the discussion Seber (1998, Section 2.3). However, when the dimension of k is fixed, a larger value for ν implies that the information coming from the realized measure is given more prominence in our Wishart model specification. The estimates of ν appear to be moderately high in relation to the dimension k

Equities	ν	β	α	λ_1	λ_2	λ_3	λ_4	λ_5	$\log L$
2 × 2									
AA/CAT	11.613 (0.176)	0.979 (0.002)	0.085 (0.003)	0.903 (0.025)	0.970 (0.027)	-	-	-	-20543.6
AXP/PFE	9.704 (0.145)	0.979 (0.002)	0.079 (0.003)	1.032 (0.029)	0.889 (0.025)	-	-	-	-17697.9
AXP/WMT	10.677 (0.161)	0.982 (0.001)	0.075 (0.003)	1.045 (0.029)	0.805 (0.023)	-	-	-	-15913.9
BA/HON	10.074 (0.151)	0.963 (0.003)	0.097 (0.003)	0.941 (0.026)	0.844 (0.023)	-	-	-	-18192.9
CAT/KO	11.981 (0.182)	0.972 (0.002)	0.068 (0.002)	1.012 (0.029)	0.849 (0.024)	-	-	-	-14579.3
GE/PFE	9.960 (0.149)	0.968 (0.002)	0.087 (0.003)	0.894 (0.025)	0.861 (0.024)	-	-	-	-16178.0
HD/JPM	11.289 (0.170)	0.973 (0.002)	0.112 (0.003)	0.891 (0.025)	0.955 (0.026)	-	-	-	-19021.1
IBM/PG	10.558 (0.159)	0.954 (0.003)	0.066 (0.002)	0.980 (0.028)	0.776 (0.022)	-	-	-	-11769.8
JPM/XOM	11.558 (0.174)	0.974 (0.002)	0.098 (0.003)	0.988 (0.027)	0.891 (0.025)	-	-	-	-16722.6
MCD/PG	9.521 (0.142)	0.959 (0.003)	0.061 (0.002)	0.840 (0.024)	0.771 (0.022)	-	-	-	-13099.5
5 × 5									
AA/AXP/IBM/JPM/WMT	16.842 (0.108)	0.983 (0.001)	0.054 (0.001)	0.924 (0.025)	0.964 (0.024)	0.960 (0.025)	0.916 (0.023)	0.794 (0.021)	-46209.0
AA/BA/CAT/GE/KO	16.858 (0.108)	0.981 (0.001)	0.051 (0.001)	0.906 (0.024)	0.963 (0.026)	0.972 (0.025)	0.875 (0.023)	0.865 (0.024)	-46256.4
AXP/CAT/IBM/KO/XOM	17.664 (0.114)	0.979 (0.001)	0.048 (0.001)	1.005 (0.026)	0.981 (0.026)	0.994 (0.026)	0.854 (0.023)	0.894 (0.024)	-35028.6
BA/HD/JPM/PFE/PG	16.425 (0.105)	0.975 (0.001)	0.053 (0.001)	0.965 (0.026)	0.904 (0.024)	0.948 (0.025)	0.882 (0.024)	0.794 (0.021)	-44412.5
BA/HD/MCD/PG/XOM	17.048 (0.110)	0.970 (0.001)	0.047 (0.001)	0.971 (0.026)	0.933 (0.025)	0.865 (0.024)	0.791 (0.021)	0.907 (0.024)	-38077.3
CAT/GE/KO/PFE/WMT	17.000 (0.109)	0.976 (0.001)	0.047 (0.001)	1.003 (0.027)	0.888 (0.023)	0.858 (0.023)	0.889 (0.024)	0.819 (0.022)	-35613.5
CAT/HON/IBM/MCD/WMT	16.195 (0.103)	0.972 (0.001)	0.048 (0.001)	0.994 (0.026)	0.856 (0.022)	0.986 (0.026)	0.893 (0.024)	0.812 (0.022)	-39998.9
GE/IBM/JPM/PG/XOM	17.427 (0.112)	0.973 (0.001)	0.056 (0.001)	0.848 (0.021)	0.976 (0.025)	0.922 (0.024)	0.775 (0.021)	0.880 (0.023)	-33329.3
HD/HON/KO/MCD/PG	15.971 (0.102)	0.968 (0.001)	0.048 (0.001)	0.922 (0.025)	0.884 (0.024)	0.847 (0.023)	0.872 (0.024)	0.795 (0.021)	-36269.8
HON/IBM/MCD/WMT/XOM	16.980 (0.109)	0.972 (0.001)	0.047 (0.001)	0.874 (0.023)	0.998 (0.026)	0.891 (0.024)	0.816 (0.022)	0.919 (0.024)	-35295.1

Table 3: Maximum likelihood estimates for open-to-close returns. Note: Standard errors are shown in parentheses.

Equities	ν	β	α	λ_1	λ_2	λ_3	λ_4	λ_5	$\log L$
2 × 2									
AA/CAT	11.644 (0.176)	0.979 (0.002)	0.079 (0.003)	1.351 (0.037)	1.462 (0.040)	-	-	-	-21556.2
AXP/PFE	9.750 (0.146)	0.978 (0.002)	0.073 (0.003)	1.531 (0.044)	1.297 (0.037)	-	-	-	-18649.3
AXP/WMT	10.732 (0.162)	0.983 (0.001)	0.068 (0.002)	1.545 (0.044)	1.141 (0.032)	-	-	-	-16821.3
BA/HON	10.095 (0.151)	0.963 (0.003)	0.096 (0.003)	1.327 (0.036)	1.156 (0.031)	-	-	-	-19013.1
CAT/KO	12.030 (0.182)	0.970 (0.002)	0.068 (0.002)	1.523 (0.043)	1.139 (0.032)	-	-	-	-15446.3
GE/PFE	9.991 (0.149)	0.966 (0.002)	0.087 (0.003)	1.318 (0.037)	1.249 (0.035)	-	-	-	-17120.7
HD/JPM	11.341 (0.170)	0.973 (0.002)	0.114 (0.003)	1.251 (0.034)	1.262 (0.035)	-	-	-	-19777.6
IBM/PG	10.579 (0.159)	0.953 (0.003)	0.068 (0.002)	1.397 (0.039)	1.082 (0.031)	-	-	-	-12627.7
JPM/XOM	11.593 (0.175)	0.974 (0.002)	0.100 (0.003)	1.295 (0.036)	1.175 (0.032)	-	-	-	-17398.1
MCD/PG	9.531 (0.142)	0.960 (0.003)	0.060 (0.002)	1.144 (0.033)	1.064 (0.030)	-	-	-	-13887.9
5 × 5									
AA/AXP/IBM/JPM/WMT	16.878 (0.108)	0.983 (0.001)	0.053 (0.001)	1.359 (0.036)	1.414 (0.036)	1.364 (0.035)	1.215 (0.031)	1.144 (0.030)	-48380.2
AA/BA/CAT/GE/KO	16.878 (0.108)	0.980 (0.001)	0.051 (0.001)	1.313 (0.035)	1.355 (0.036)	1.445 (0.037)	1.252 (0.032)	1.158 (0.031)	-48440.9
AXP/CAT/IBM/KO/XOM	17.702 (0.114)	0.978 (0.001)	0.048 (0.001)	1.450 (0.038)	1.479 (0.039)	1.408 (0.037)	1.161 (0.031)	1.190 (0.031)	-37140.7
BA/HD/JPM/PFE/PG	16.442 (0.105)	0.975 (0.001)	0.054 (0.001)	1.364 (0.037)	1.257 (0.033)	1.255 (0.033)	1.276 (0.034)	1.079 (0.029)	-46438.4
BA/HD/MCD/PG/XOM	17.066 (0.110)	0.970 (0.001)	0.047 (0.001)	1.368 (0.037)	1.292 (0.035)	1.176 (0.032)	1.081 (0.029)	1.204 (0.032)	-40026.1
CAT/GE/KO/PFE/WMT	17.031 (0.109)	0.975 (0.001)	0.048 (0.001)	1.493 (0.040)	1.294 (0.033)	1.172 (0.032)	1.286 (0.035)	1.167 (0.031)	-37847.6
CAT/HON/IBM/MCD/WMT	16.223 (0.104)	0.972 (0.001)	0.048 (0.001)	1.481 (0.039)	1.189 (0.031)	1.395 (0.037)	1.217 (0.033)	1.167 (0.031)	-42156.5
GE/IBM/JPM/PG/XOM	17.443 (0.112)	0.973 (0.001)	0.056 (0.001)	1.230 (0.031)	1.375 (0.036)	1.217 (0.032)	1.064 (0.028)	1.166 (0.030)	-35307.5
HD/HON/KO/MCD/PG	15.987 (0.102)	0.969 (0.001)	0.048 (0.001)	1.277 (0.034)	1.232 (0.033)	1.148 (0.031)	1.192 (0.032)	1.080 (0.029)	-38233.2
HON/IBM/MCD/WMT/XOM	17.001 (0.109)	0.971 (0.001)	0.048 (0.001)	1.213 (0.032)	1.398 (0.037)	1.209 (0.033)	1.172 (0.031)	1.219 (0.032)	-37297.0

Table 4: Maximum likelihood estimates for close-to-close returns. Note: Standard errors are shown in parentheses.

	15 × 15 AA/.../XOM	
	Open-to-Close	Close-to-Close
ν	28.3960 (0.0572)	28.4060 (0.0572)
β	0.9828 (0.0003)	0.9827 (0.0003)
α	0.0268 (0.0002)	0.0269 (0.0002)
λ_1	0.9196 (0.0235)	1.3078 (0.0331)
λ_2	0.9741 (0.0237)	1.3415 (0.0322)
λ_3	0.9755 (0.0251)	1.3511 (0.0342)
λ_4	0.9716 (0.0241)	1.4280 (0.0351)
λ_5	0.8328 (0.0188)	1.1993 (0.0270)
λ_6	0.8907 (0.0221)	1.2406 (0.0307)
λ_7	0.8603 (0.0200)	1.1426 (0.0262)
λ_8	0.9900 (0.0246)	1.4063 (0.0349)
λ_9	0.9101 (0.0220)	1.2030 (0.0291)
λ_{10}	0.8642 (0.0225)	1.1720 (0.0304)
λ_{11}	0.8939 (0.0237)	1.1958 (0.0316)
λ_{12}	0.8962 (0.0235)	1.2751 (0.0332)
λ_{13}	0.8031 (0.0209)	1.0857 (0.0280)
λ_{14}	0.8120 (0.0203)	1.1624 (0.0292)
λ_{15}	0.9032 (0.0227)	1.1677 (0.0290)
$\log L$	−69226.50	−75245.35

Table 5: Maximum likelihood estimates for open-to-close and close-to-close returns for the 15×15 model. Note: Standard errors are shown in parentheses.

and we may therefore conclude that realized covariance measures play a considerable role in our estimation framework.

4.3 Benchmarking against common alternatives

We compare performance of the Wishart-GARCH model against alternative models. We relate our model to the BEKK model of Engle and Kroner (1995). In the standard BEKK$(1,1)$ model it is assumed that $r_t|\mathcal{F}_{t-1} \sim N(0, V_t)$ and the covariance of assets is driven by the outer-products of daily returns,

$$V_{t+1} = CC' + BV_tB' + Ar_tr_t'A', \qquad t = 1, \dots, T, \tag{25}$$

where A, B, and C are $k \times k$ parameter matrices of which C is restricted to be lower-triangular. The scalar BEKK model with covariance targeting can be represented by

$$V_{t+1} = (1 - b - a)\bar{R} + bV_t + ar_tr_t', \qquad a, b \geq 0, \quad a + b < 1 \tag{26}$$

where $\bar{R} = T^{-1}\sum_{t=1}^{T} r_tr_t'$ is the sample covariance of daily returns with a and b being unknown coefficients. When contrasting the BEKK model to the result in Theorem 1, it is evident that BEKK does not exploit high-frequency data to infer about the current level of covariances.

We also consider a simple exponentially weighted moving average (EWMA) one-step ahead forecasting scheme for high-frequency based realized kernels. The EMWA filter is often used by practitioners and regulators, see, for example, in the RiskMetrics of J.P.Morgan (1996). The EMWA assumes that the conditional variance matrix is an integrated process given by the updating equation

$$V_{t+1} = c\, V_t + (1 - c)\, RK_t,$$

where c is a fixed smoothing parameter which is typically set equal to $c = 0.96$.

The considered models in our study are non-nested and log-likelihood ratio tests cannot be used. We evaluate the performance of our Wishart-GARCH model relative to the BEKK and EWMA using two loss functions. We use the quasi-likelihood loss function given by

$$Q(V_t, \Sigma_t) = \log |V_t| + \mathrm{tr}\big(V_t^{-1}\Sigma_t\big), \tag{27}$$

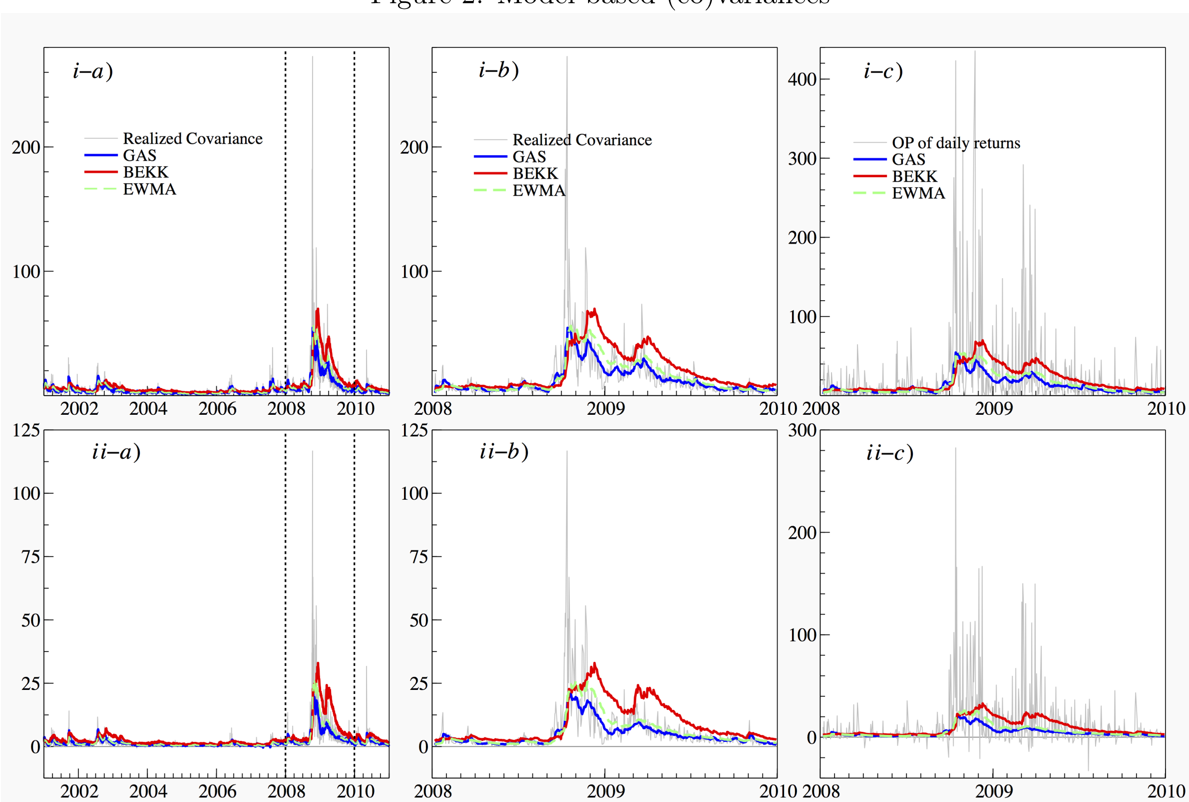
and we use the root mean squared error based on the matrix norm given by

$$F(V_t, \Sigma_t) = ||\Sigma_t - V_t||^{1/2} = \left[\sum_{i,j} (\Sigma_{ij,t} - V_{ij,t})^2 \right]^{1/2},$$

where Σ_t is true latent covariance matrix and V_t is the model-based covariance matrix. In our application, the proxy for the latent covariance matrix is based on the realized kernel and on the subsampled realized covariance using $5min$ returns. As we have done for our in-sample analysis, we report the average loss functions. Furthermore, we study the open-to-close returns and we use baseline $p = q = 1$ model specification.

Table 6 presents the results for the loss functions for ten randomly selected cases with $k = 2$ and $k = 5$, and for the case with all equities, $k = 15$. These results show that our Wishart-GARCH model delivers overall a substantially better fit than the BEKK specification. For two bivariate cases of IBM/PG and MCD/PG, the BEKK model provides better in-sample fits.

Figure 2: Model-based (co)variances



Note: Panels *i-a*) to *i*)-*c*) plot the case of variance of AA equity, while panel *ii-a*) to *ii*)-*c*) present the case of covariance of pair AA/CAT.

Figure 2 presents the model-based covariances for the case of the AA/CAT pair. When

	Wishart-GARCH Model				BEKK Model				EWMA			
	RK_P		RK_B		RK_P		RK_B		RK_P		RK_B	
	Q	F	Q	F	Q	F	Q	F	Q	F	Q	F
2×2												
AA/CAT	4.173	3.462	**4.149**	*3.387*	4.352	4.186	4.331	4.145	4.237	3.938	4.209	3.878
AXP/PFE	3.487	3.149	**3.453**	*3.042*	3.695	3.812	3.643	3.703	3.552	3.500	3.511	3.402
AXP/WMT	3.213	2.989	**3.173**	*2.884*	3.437	3.622	3.380	3.511	3.285	3.384	3.240	3.294
BA/HON	3.642	2.760	**3.631**	*2.672*	3.823	3.100	3.809	3.018	3.730	3.099	3.707	2.984
CAT/KO	3.012	1.893	**2.980**	*1.813*	3.147	2.168	3.105	2.089	3.085	2.174	3.046	2.101
GE/PFE	3.241	2.689	**3.205**	*2.644*	3.421	2.863	3.363	2.805	3.312	2.982	3.267	2.938
HD/JPM	3.864	3.976	**3.824**	*3.934*	4.118	5.078	4.054	5.028	3.963	4.597	3.919	4.560
IBM/PG	2.411	1.771	**2.360**	1.735	2.652	1.697	2.602	*1.634*	2.489	1.706	2.440	1.671
JPM/XOM	3.410	3.563	**3.383**	*3.547*	3.631	4.457	3.607	4.457	3.532	4.188	3.503	4.181
MCD/PG	2.645	1.755	**2.624**	1.710	2.892	1.717	2.890	*1.670*	2.708	1.762	2.689	1.711
5×5												
AA/BA/CAT/GE/KO	**8.117**	6.170	8.209	*5.936*	8.618	7.262	8.729	7.112	8.247	6.703	8.324	6.496
AXP/CAT/IBM/KO/XOM	**6.908**	5.294	6.976	*5.087*	7.540	6.462	7.613	6.303	7.075	5.726	7.134	5.548
BA/HD/JPM/PFE/PG	**7.878**	6.198	7.952	*5.991*	8.516	7.223	8.619	7.071	8.016	6.468	8.085	6.293
BA/HD/MCD/PG/XOM	**7.226**	4.616	7.300	*4.477*	7.775	4.908	7.887	4.807	7.374	4.793	7.447	4.674
CAT/GE/KO/PFE/WMT	**6.942**	4.761	7.037	*4.563*	7.520	5.329	7.603	5.154	7.066	5.148	7.145	4.959
CAT/HON/IBM/MCD/WMT	**7.346**	5.014	7.448	*4.765*	7.927	5.608	8.038	5.402	7.481	5.306	7.567	5.060
GE/IBM/JPM/PG/XOM	**6.740**	6.514	6.764	*6.374*	7.562	7.496	7.577	7.402	6.922	6.681	6.939	6.578
HD/HON/KO/MCD/PG	**6.931**	4.742	7.047	*4.519*	7.497	5.000	7.659	4.816	7.067	4.876	7.175	4.669
HON/IBM/MCD/WMT/XOM	**6.901**	4.594	6.991	*4.414*	7.482	5.077	7.579	4.915	7.062	4.915	7.139	4.731
15×15												
AA/.../XOM	**20.502**	16.249	21.506	*15.580*	23.246	19.334	24.467	18.904	20.778	16.480	21.768	15.853

Table 6: The sample Q-loss and F-loss for three specifications and two realized measures. Note: RK_P and RK_B, denote the realized kernel based on the Parzen kernel and Bartlett kernels, respectively. The best configuration in terms of Q-loss is identified by bold font, and the best configuration in terms of F-loss is identified by italic font. The Wishart-GARCH Model tend to have the smallest sample loss, with just two exceptions in the case of F-loss..

we compare the model-based estimates with the subsampled realized covariance, we find a clear improvement for the Wishart-GARCH model. Its estimated covariances quickly adapt to changes in the location of the covariances. In contrast, the traditional BEKK model smooths the outer-product of daily returns without exploiting high-frequency information. The panels *i-b*) and *ii-b*) zoom in on a period of high volatile markets, the years 2008-2009. The BEKK response to changes is remarkably longer in this period. Since the Wishart-GARCH model weighs the realized measure more than the outer-product of returns, the time-varying variance matrix remains more robust to large return innovations; see panels *i-c*) and *ii-c*).

5 Conclusions

We have proposed a new model for the modeling and predicting of daily time series of covariance matrices of financial assets: the Wishart-GARCH model. The challenge is to capture the complex temporal interdependencies among variances and covariances of assets. There are two distinguishing features of our model when compared to alternative frameworks. First, the model relies both on low- (daily) and on high-frequency (intraday) information. It turns out that the high-frequency measures receive most weight given that it exploits intraday data of financial assets to infer about current level of covariances. Several noisy and frequency-varying measures of current covariance levels can be adopted. A second feature of the Wishart-GARCH model is that the innovations driving the update of covariance matrix exploits full likelihood information. Consequently, even a simplified scalar specification allows for complex interdependences between variances and covariances of all assets. The model can therefore attain without loss of flexibility parsimonious formulation which is convenient property for multivariate volatility models. In an empirical study for a portfolio of fifteen NYSE equities, we have compared our Wishart-GARCH model with other multivariate GARCH models and with exponentially weighted moving avarage schemes. The in-sample fit of our model dominates the fit in the comparisons. This finding carries over to different subsets of equities but also to the full portfolio of equities. The proposed model is capable to track sudden changes in volatility and the dependence structure of the assets in a more efficient way than standard multivariate GARCH.

References

Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2005), 'A tale of two time scales: Determining integrated volatility with noisy high-frequency data', *Journal of the American Statistical Association* **100**(472), 1394–1411.

Andersen, T. G. and Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001), 'The distribution of realized exchange rate volatility', *Journal of the American Statistical Association* **96**(453), 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**, 579–625.

Asai, M., McAleer, M. and Yu, J. (2006), 'Multivariate stochastic volatility: a review', *Econometric Reviews* **25**(2-3), 145–175.

Asai, M. and So, M. K. P. (2013), 'Stochastic covariance models', *Journal of the Japan Statistical Society* **43**(2), 127–162.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008), 'Designing realized kernels to measure the ex post variation of equity prices in the presence of noise', *Econometrica* **76**(6), 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2009), 'Realized kernels in practice: trades and quotes', *Econometrics Journal* **12**, 1–32.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2011*a*), 'Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading', *Journal of Econometrics* **162**(2), 149–169.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2011*b*), 'Subsampling realised kernels', *Journal of Econometrics* **160**, 204–219.

Barndorff-Nielsen, O. E. and Shephard, N. (2002), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society, Series B* **64**(2), 253–280.

Bauer, G. H. and Vorkink, K. (2011), 'Forecasting multivariate realized stock market volatility', *Journal of Econometrics* **160**(1), 93–101.

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006), 'Multivariate GARCH models: a survey', *Journal of Applied Econometrics* **21**, 79–109.

Blasques, F., Koopman, S. J. and Lucas, A. (2015), 'Information theoretic optimality of observation driven time series models for continuous responses', *Biometrika* **102**(2), 325–343.

Chiriac, R. and Voev, V. (2011), 'Modelling and forecasting multivariate realized volatility', *Journal of Applied Econometrics* **26**(6), 922–947.

Christensen, K., Kinnebrock, S. and Podolskij, M. (2010), 'Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data', *Journal of Econometrics* **159**(1), 116–133.

Cox, D. R. (1981), 'Statistical analysis of time series: some recent developments', *Scandinavian Journal of Statistics* **8**, 93–115.

Creal, D. D., Koopman, S. J. and Lucas, A. (2013), 'Generalized autoregressive score models with applications', *Journal of Applied Econometrics* **28**, 777–795.

Engle, R. F. (2002), 'New frontiers for ARCH models', *Journal of Applied Econometrics* **17**, 425–446.

Engle, R. F. and Gallo, G. M. (2006), 'A multiple indicators model for volatility using intra-daily data', *Journal of Econometrics* **131**(3), 3–27.

Engle, R. F. and Kroner, K. F. (1995), 'Multivariate simultaneous generalized ARCH', *Econometric Theory* **11**(1), 122–150.

Engle, R. F. and Mezrich, J. (1996), 'GARCH for groups', *Risk* **9**, 36–40.

Golosnoy, V., Gribisch, B. and Liesenfeld, R. (2012), 'The conditional autoregressive wishart model for multivariate stock market volatility', *Journal of Econometrics* **167**(1), 211–223.

Gourieroux, C., Jasiak, J. and Sufana, R. (2009), 'The Wishart autoregressive process of multivariate stochastic volatility', *Journal of Econometrics* **150**(2), 167–181.

Griffin, J. E. and Oomen, R. C. A. (2011), 'Covariance measurement in the presence of nonsynchronous trading and market microstructure noise', *Journal of Econometrics* **160**, 58–68.

Hansen, P. R. and Horel, G. (2009), 'Quadratic variation by Markov chains', *working paper* .

Hansen, P. R. and Huang, Z. (2012), 'Exponential GARCH modeling with realized measures of volatility', *working paper* .

Hansen, P. R., Huang, Z. and Shek, H. (2012), 'Realized GARCH: A joint model of returns and realized measures of volatility', *Journal of Applied Econometrics* **27**, 877–906.

Hansen, P. R. and Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**, 127–161.

Hansen, P. R., Lunde, A. and Voev, V. (2014), 'Realized beta GARCH: A multivariate GARCH model with realized measures of volatility', *Journal of Applied Econometrics* **29**, 774–799.

Hayashi, T. and Yoshida, N. (2005), 'On covariance estimation of non-synchronously observed diffusion processes', *Bernoulli* **11**, 359–379.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. and Vetter, M. (2009), 'Microstructure noise in the continuous case: The pre-averaging approach', *Stochastic Processes and Their Applications* **119**, 2249–2276.

Janus, P., Lucas, A. and Opschoor, A. (2014), New HEAVY models for fat-tailed returns and realized covariance kernels, Tinbergen Institute Discussion Papers 2014-073/4, Tinbergen Institute.

J.P.Morgan (1996), *Riskmetrics Technical Document, 4th ed.*, J.P. Morgan, New York.

Jungbacker, B. and Koopman, S. J. (2006), 'Monte carlo likelihood estimation for three multivariate stochastic volatility models', *Econometric Reviews* **25**(2–3), 385–408.

Koopman, S. J., Jungbacker, B. and Hol, E. (2005), 'Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements', *Journal of Empirical Finance* **12**, 445–475.

Koopman, S. J. and Scharth, M. (2013), 'The analysis of stochastic volatility in the presence of daily realised measures', *Journal of Financial Econometrics* **11**, 76–115.

Magnus, J. R. and Neudecker, H. (1979), 'The commutation matrix: some properties and applications', *Annals of Statistics* **7**(2), 381–394.

Magnus, J. R. and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Economics*, Wiley, New York.

Malliavin, P. and Mancino, M. E. (2002), 'Fourier series method for measurement of multivariate volatilities', *Finance and Stochastics* **6**, 49–61.

Noureldin, D., Shephard, N. and Sheppard, K. (2012), 'Multivariate high-frequency-based volatility (HEAVY) models', *Journal of Applied Econometrics* **27**, 907–933.

Seber, G. (1998), *Multivariate Observations*, Wiley Interscience, New York.

Seber, G. A. F. (2007), *A matrix handbook for statisticians*, Wiley Series in Probability and Statistics, Wiley-Interscience.

Shephard, N. and Sheppard, K. (2010), 'Realising the future: forecasting with high frequency based volatility (HEAVY) models', *Journal of Applied Econometrics* **25**, 197–231.

Silvennoinen, A. and Teräsvirta, T. (2009), Multivariate GARCH models, *in* T. G. Andersen, R. A. Davis, J.-P. Kreißand T. Mikosch, eds, 'Handbook of Financial Time Series', Springer-Verlag, Berlin, pp. 201–229.

Srivastava, M. S. (2003), 'Singular wishart and multivariate beta distribution', *Annals of Statistics* **31**(5), 1537–1560.

APPENDICES

A Matrix Notation and Preliminary Results

The results in this paper make use of the following matrix notation and definitions. Let A and B be $k \times k$ matrices, then $A \otimes B$ denotes the Kronecker product, which is a $k^2 \times k^2$ block matrix $\{a_{ij}B\}$ where a_{ij} is the (i,j) element of matrix A. The vec(A) operator stacks the columns of matrix A consecutively into the $k^2 \times 1$ column vector, while vech(A) stacks the lower triangular part including diagonal into $k^* \times 1$ column vector, with $k^* = k(k+1)/2$. The reverse operators are denoted as unvec and unvech, respectively The $k \times k$ identity matrix is denoted by I_k. We define the $k^2 \times k^2$ commutation matrix K_k, the $k^2 \times k^*$ duplication matrix, D_k, and the $k^* \times k^2$ elimination matrix L_k, by the identities

$$K_k \text{vec}(B) = \text{vec}(B') \qquad D_k \text{vech}(A) = \text{vec}(A), \quad \text{and} \quad L_k \text{vec}(A) = \text{vech}(A),$$

where B is an arbitrary $k \times k$ matrix and A an arbitrary symmetric $k \times k$ matrix. Here $L_k = (D_k' D_k)^{-1} D_k'$ is the Moore-Penrose inverse of the duplication matrix D_k. Similarly, for a lower triangular $k \times k$ matrix, C, we let $\tilde{D}_k$ denote the $k^2 \times k^*$ duplication matrix and $\tilde{L}_k$ the $k^* \times k^2$ the elimination matrix defined by the identities

$$\tilde{D}_k \text{vech}(C) = \text{vec}(C) \qquad \text{and} \qquad \tilde{L}_k \text{vec}(C) = \text{vech}(C),$$

with $\tilde{D}_k' \tilde{D}_k = I_{k^*}$ and $\tilde{L}_k = \tilde{D}_k'$. The difference between L_k and $\tilde{L}_k$ arises because vec(C) contains zeros. Additional properties and results related to these matrices can be found in Magnus and Neudecker (1988) and Seber (2007).

The proofs below make use of the following results on matrix calculus. For a $p \times q$ matrix function $F(X)$ and a $m \times n$ matrix of variables X, the derivative of $F(X)$ with respect to $x = \text{vec}(X)$, denoted by the $pq \times mn$ matrix $DF(X)$, is given by

$$DF(X) = \frac{\partial \text{vec}(F(X))}{\partial x'}.$$

If $F(X)$ is $m \times m$ symmetric matrix, we have

$$\frac{\partial \text{vec}(F(X))}{\partial x'} = D_m \frac{\partial \text{vech}(F(X))}{\partial x'}.$$

The intermediate results for any $k \times k$ matrices A, X and B are

$$\begin{aligned} \frac{\partial \log |AXB|}{\partial X} &= \text{vec}[(X^{-1})']', \qquad (A, B \text{ nonsingular}), \\ \frac{\partial X^{-1}}{\partial X} &= -(X^{-1})' \otimes X^{-1}, \\ \frac{\partial \text{tr}(AXB)}{\partial X} &= A'B'. \end{aligned} \tag{28}$$

The above results are combined with the following result for any $k \times k$ matrices A, B and C with matrix B being symmetric,

$$\text{vec}(ABC) = (C' \otimes A)\text{vec}(B). \tag{29}$$

B Proofs

Lemma 1. *Suppose V is a symmetric positive definite $k \times k$ matrix and C is a lower triangular $k \times k$ matrix such that $V = CC'$. Then* vech(V) *is a differentiable function of* vech(C) *and*

$$\dot{V} = \frac{\partial \text{vech}(V)}{\partial \text{vech}(C)} = L_k(I_{k^2} + K_k)(C \otimes I_k)\tilde{L}_k'.$$

Proof. The proof is based on various matrix manipulations stated in Magnus and Neudecker (1988). Since $V = CC'$, we obtain

$$\text{d}V = (\text{d}C)C' + C(\text{d}C').$$

By vectorization and using the result that $\text{vec}(\text{d}X)A = (A' \otimes I)\text{dvec}X$, Magnus and Neudecker (1988, p. 182), we have

$$\begin{aligned} \text{dvec}(V) &= (C \otimes I_k)\text{dvec}(C) + (I_k \otimes C)\text{dvec}(C'), \\ &= [(C \otimes I_k) + (I_k \otimes C)K_k]\text{dvec}(C). \end{aligned}$$

Since for any $k \times k$ matrices A and B, we have $K_k(A \otimes B) = (B \otimes A)K_k$, see Magnus and Neudecker (1988, p. 46) for more general rules on commutation matrix. We have

$$\begin{aligned} \text{dvec}(V) &= \big((C \otimes I_k) + K_k(C \otimes I_k)\big)\text{dvec}(C), \\ &= \big((I_{k^2} + K_k)(C \otimes I_k)\big)\text{dvec}(C). \end{aligned}$$

Finally, after pre-multiplication by L_k, we obtain

$$\text{dvech}(V) = L_k\big((I_{k^2} + K_k)(C \otimes I_k)\big)\tilde{L}_k'\text{dvech}(C),$$

where

$$\text{dvec}(C) = \tilde{D}_k\text{dvech}(C) = \tilde{L}_k'\text{dvech}(C) \tag{30}$$

since matrix C is a lower triangular matrix. □

Proof of Theorem 1. We derive the score vector whose general form is given by (15). It follows from (18) and (19) that the relevant parts of log-likelihoods for the score vector derivation are given by

$$\begin{aligned}
\mathcal{L}_{1,t} &= -\frac{1}{2}\bigg(\log|\Lambda^{1/2}V_t\Lambda^{1/2}| + \text{tr}\big((\Lambda^{1/2}V_t\Lambda^{1/2})^{-1}r_tr_t'\big)\bigg), &(31)\\
\mathcal{L}_{2,t} &= -\frac{\nu}{2}\bigg(\log|V_t| + \text{tr}\big(V_t^{-1}X_t\big)\bigg). &(32)
\end{aligned}$$

We consider the Cholesky decomposition (12) of the covariance matrix V_t and parameter vector f_t, as given by (20). Using the chain rule for vector differentiation, the score for individual measurement densities (1) and (2) can be expressed by

$$\frac{\partial \log \varphi_i(Z_t^i|f_t, \mathcal{F}_{t-1};\psi)}{\partial f_t'} = \frac{\partial \log \varphi_i(Z_t^i|f_t, \mathcal{F}_{t-1};\psi)}{\partial(\text{vec}(V_t))'}\frac{\partial\text{vec}(V_t)}{\partial\text{vech}(V_t)'}\frac{\partial\text{vech}(V_t)}{\partial f_t'}.$$

We first differentiate the measurement density of returns (31). Using (28) and (29), together with noting that V_t is symmetric and $V_t^{-1} = V_t^{-1}V_tV_t^{-1}$, we obtain

$$\begin{aligned}
\frac{\partial\mathcal{L}_{1,t}}{\partial\text{vec}(V_t)'} &= -\frac{1}{2}[\text{vec}(V_t^{-1})' - \text{vec}(\Lambda^{-1/2}r_tr_t'\Lambda^{-1/2})'(V_t^{-1}\otimes V_t^{-1})]\\
&= -\frac{1}{2}[\text{vec}(V_t)'(V_t^{-1}\otimes V_t^{-1}) - \text{vec}(\Lambda^{-1/2}r_tr_t'\Lambda^{-1/2})'(V_t^{-1}\otimes V_t^{-1})]\\
&= \frac{1}{2}[\text{vec}(\Lambda^{-1/2}r_tr_t'\Lambda^{-1/2})' - \text{vec}(V_t)'](V_t^{-1}\otimes V_t^{-1}), &(33)
\end{aligned}$$

and similarly for the measurement density of covariance (32), we have

$$\begin{aligned}
\frac{\partial\mathcal{L}_{2,t}}{\partial\text{vec}(V_t)'} &= -\frac{\nu}{2}[\text{vec}(V_t^{-1})' - (\text{vec}(X_t)'(V_t^{-1}\otimes V_t^{-1})]\\
&= -\frac{\nu}{2}[(\text{vec}(V_t)'(V_t^{-1}\otimes V_t^{-1}) - \text{vec}(X_t)'(V_t^{-1}\otimes V_t^{-1})]\\
&= \frac{\nu}{2}[\text{vec}(X_t) - \text{vec}(V_t)]'(V_t^{-1}\otimes V_t^{-1}). &(34)
\end{aligned}$$

The results (33) and (34), combined with the Lemma in Appendix 1 and with the score defined in (15), the proof of Theorem 1 is completed. □

Proof of Theorem 2: We derive the Fisher information matrix whose general form is given by (16). Using the results from the proof of Theorem 1, the individual score functions are given by

$$\begin{aligned}\nabla_{1,t} &= \frac{1}{2}\dot{V}_t' D_k' \big(V_t^{-1} \otimes V_t^{-1}\big)\big[\mathrm{vec}(\Lambda^{-1/2} r_t r_t' \Lambda^{-1/2}) - \mathrm{vec}(V_t)\big],\\ \nabla_{2,t} &= \frac{\nu}{2}\dot{V}_t' D_k' \big(V_t^{-1} \otimes V_t^{-1}\big)\big[\mathrm{vec}(X_t) - \mathrm{vec}(V_t)\big],\end{aligned}$$

for the measurement densities of returns and of covariance, respectively. Taking $\mathrm{E}[\nabla_{i,t}\nabla_{i,t}'|\mathcal{F}_{t-1}]$, we obtain

$$\mathcal{I}_{1,t} = \frac{1}{4}\dot{V}_t' D_k' \big(V_t^{-1} \otimes V_t^{-1}\big)\mathrm{var}\big[\mathrm{vec}(\Lambda^{-1/2} r_t r_t' \Lambda^{-1/2}) - \mathrm{vec}(V_t)|\mathcal{F}_{t-1}\big]\big(V_t^{-1} \otimes V_t^{-1}\big) D_k \dot{V}_t,$$

$$\mathcal{I}_{2,t} = \frac{\nu^2}{4}\dot{V}_t' D_k' \big(V_t^{-1} \otimes V_t^{-1}\big)\mathrm{var}\big[\mathrm{vec}(X_t) - \mathrm{vec}(V_t)|\mathcal{F}_{t-1}\big]\big(V_t^{-1} \otimes V_t^{-1}\big) D_k \dot{V}_t.$$

Using the results (10) and (11), and given that $(V_t^{-1} \otimes V_t^{-1})(V_t \otimes V_t) = I_{k^2}$, we have

$$\mathcal{I}_{1,t} = \frac{1}{4}\dot{V}_t' D_k' \big(V_t^{-1} \otimes V_t^{-1}\big)(I_{k^2} + K_k) D_k \dot{V}_t,$$

$$\mathcal{I}_{2,t} = \frac{\nu^2}{4}\dot{V}_t' D_k' \big(V_t^{-1} \otimes V_t^{-1}\big)(I_{k^2} + K_k) D_k \dot{V}_t = \frac{\nu^2}{4}\mathcal{I}_{1,t},$$

which combined with (16) completes the proof. □