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

P. Gorgi[a] P.R. Hansen[b,c] P. Janus[d] S.J. Koopman[a,c*]

[a] *Vrije Universiteit Amsterdam, Tinbergen Institute*

[b] *University of North Carolina, Chapel Hill*

[c] *CREATES, Aarhus University*

[d] *UBS Asset Management, Zürich*

November 25, 2017

Abstract

We propose a novel multivariate GARCH model that incorporates realized measures for the covariance matrix of returns. The joint formulation of a multivariate dynamic model for outer-products of returns, realized variances and realized covariances leads to a feasible approach for analysis and forecasting. The updating of the covariance matrix relies on the score function of the joint likelihood function based on Gaussian and Wishart densities. The dynamic model is parsimonious while the analysis relies on straightforward computations. In a Monte Carlo study we show that parameters are estimated accurately for different small sample sizes. We illustrate the model with an empirical in-sample and out-of-sample analysis for a portfolio of 15 U.S. financial assets.

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

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 Asset Management, which does not accept any responsibility for the contents and opinions expressed in this paper. We would like to thank Asger Lunde for his support and comments. Finally, we are indebted to the editor and two referees who have made many insightful comments that have improved the paper considerably. Additional material and the R code are made available at `http://www.pgorgi.com`.

1 Introduction

Modeling conditional dependency structure of financial assets through time-varying covariance 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 covariance matrices from vector time series of financial returns. The dynamic process for multivariate volatility (variances, covariances and correlations) 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 et al. (2006), Silvennoinen and Teräsvirta (2009) and Anudrino and Trojani (2011). Reviews of multivariate SV models are provided by Asai et al. (2006) and Jungbacker and Koopman (2006). These developments in financial econometrics are also related with the theoretical developments in finance and in particular with the literature on option pricing, optimal portfolio modeling, and term structure modeling. For example, Driessen et al. (2009) investigate individual volatility risk premia differences (typically in relation to a portfolio or index) and they explain them by a high correlation risk premium. More recently, the study of Buraschi et al. (2017) focuses on a priced disagreement risk that explains returns of option volatility and correlation in trading strategies. In all such studies, the multivariate GARCH and SV models for volatilities and correlations in multiple asset returns (possibly within a portfolio) are of key importance.

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 intra-day 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 et al. (2003) for a more detailed discussion. The relevance of these issues in the context of discrete volatility models, possibly with leverage effects, and their relations to option pricing models have been discussed and reviewed recently in Khrapov and Renault (2016). Various attempts have been made to use high-frequency intra-day 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 (2002b) and Koopman et al. (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 et al. (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 et al. (2005), Barndorff-Nielsen et al. (2008), Jacod et al. (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 Christensen et al. (2010), Barndorff-Nielsen et al. (2011a), Griffin and Oomen (2011), and references therein. 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 et al. (2009) have proposed (non-central) Wishart autoregressive model for realized covariance matrix. Asai and So (2013) and Golosnoy et al. (2012) have proposed alternative dynamic formulation for covariance parameters with the underlying Wishart distribution. 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 covariance matrix. In our study we also rely on the Wishart distribution but we propose a novel conditional model formulation for the covariance matrix. For the updating of the conditional covariance matrix, we use daily as well as intra-daily financial returns.

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 et al. (2012). Then a further extension based on the use of more heavy-tailed distributions is proposed by Opschoor et al. (2017). 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 et al. (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 et al. (2014). A Realized SV model is proposed by Koopman and Scharth (2013). Our present work introduces an extension of the Realized GARCH model to the multivariate case and the use of a score-driven framework for the time-varying conditional covariance matrix.

Our primary aim is to specify a model for the daily time-varying covariance matrix and to extract it by using both low- and high-frequency data. For this purpose we propose a specification for the unobserved daily covariance 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 this purpose, we adopt the score-driven framework of Creal et al. (2013). Our joint modeling framework relies on a Wishart distribution for realized covariance matrices and on a Gaussian distribution for vectors of daily returns. The updating of the time-varying covariance matrix is driven by the scaled score of the predictive joint likelihood function; Blasques et al. (2015) have argued that such updating is locally optimal in a Kullback-Leibler sense. 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 our resulting model as the Realized-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. We confirm that the realized measure is a more accurate measure of the covariance matrix as it exploits intra-day high-frequency data. In an out-of-sample study we show that our modeling framework can lead to accuracy improvements in forecasting, especially those for the density in daily returns.

The remainder of the paper is organized as follows. In Section 2, we introduce the Realized-Wishart-GARCH model for multivariate conditional volatility. In Section 3, we conduct a Monte Carlo study to verify the performance of likelihood-based estimation. Section 4 presents the results of our in-sample and out-of-sample empirical study for a portfolio of fifteen NYSE equities. It includes a thorough forecasting comparison of our model against several other competitive models and methods. Section 5 concludes. The Appendices provide some matrix algebra results, proofs of the main results and additional estimation results.

2 The Realized-Wishart-GARCH Model

The development of our model for the time-varying conditional covariance matrix starts with the assumption that for each trading day and for a selection of assets, we have a data vector of daily returns and a measure (or possibly several measures) of the daily realized covariance matrix. We build a model for these data sources and implicitly use both low- and high-frequency data. The proposed structure of the model permits the use of several realized measures that are based on different sampling frequencies. In this section we discuss our modeling assumptions. We then describe the modeling strategy and we provide technical details of our new model for multivariate conditional 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 $t = 1, \ldots, T$. Let $\mathcal{F}_{t-1}$ be the sigma field generated by the past values of r_t and X_t, that is $\mathcal{F}_{t-1} = \sigma(r_s, X_s; s = 1, \ldots, t-1)$. We assume the following conditional densities

$$\begin{aligned} r_t|\mathcal{F}_{t-1} &\sim N_k(0, H_t), && (1)\\ X_t|\mathcal{F}_{t-1} &\sim W_k(V_t/\nu, \nu), && (2) \end{aligned}$$

with non-singular $k \times k$ covariance matrix H_t of the zero-mean multivariate normal distribution $N_k(0, H_t)$ and non-singular $k \times k$ covariance matrix V_t as the mean of the k-th dimensional Wishart distribution $W_k(V_t/\nu, \nu)$ with degrees of freedom $\nu \geq k$. The covariance matrices H_t and V_t are both measurable with respect to $\mathcal{F}_{t-1}$. The variables r_t and X_t in (1) and (2) are conditionally independent of each other. The (unconditional) dependence between r_t and X_t is assumed to rely only on the dependence between H_t and V_t. 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

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

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 multivariate Gamma function $\Gamma_k(a) = \pi^{\frac{k(k-1)}{4}}\prod_{i=1}^{k}\Gamma(a+(1-i)/2)$ for any $a>0$. The measurement equations can be formally given by

$$r_t|\mathcal{F}_{t-1} = H_t^{1/2}\varepsilon_t, \qquad X_t|\mathcal{F}_{t-1} = V_t^{1/2}\eta_t V_t^{1/2}, \tag{5}$$

where $A^{1/2}$ denotes the square root matrix of A and where the measurement innovations are assumed to be, mutually and serially, identically and independently distributed (iid) random variables, that is

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

with $k\times 1$ random vector ε_t and $k\times k$ random matrix η_t with property $\mathrm{E}(\varepsilon_t\,\eta_s') = 0$, for $t,s = 1,\ldots,T$.

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 et al. (2011a) or related measures described by Griffin and Oomen (2011). The distributional assumption (1) implies that the outer product of the daily returns vector is distributed as

$$r_t r_t'|\mathcal{F}_{t-1} \sim W_k^s(H_t, 1), \tag{6}$$

where $W_k^s(H_t,1)$ is the Singular Wishart distribution with mean H_t and one degree of freedom, see Uhlig (1994) and Srivastava (2003). We notice that the covariance matrix H_t is non-singular; the distinctive feature of the Singular Wishart is that $\nu<k$ and in (6) we have $\nu=1$ while for the Wishart we have $\nu>k$. Given the specification in (6), we can formulate the measurement equations alternatively as

$$r_t r_t'|\mathcal{F}_{t-1} \;=\; H_t^{1/2}\zeta_t H_t^{1/2}, \qquad X_t|\mathcal{F}_{t-1} \;=\; V_t^{1/2}\eta_t V_t^{1/2},$$

with $\zeta_t \sim W_k^s(I_k,1)$, and where ζ_t and η_t are, serially and mutually, iid processes of $k\times k$ stochastic matrices. In this representation, the measurement equations are expressed in terms of variances and covariances.

The developments in our study are based on the assumption that the conditional covariance matrix of (daily) returns and the conditional mean of the realized covariance matrix share the same dynamic processes. Specifically, we let the covariance matrix H_t to be fully dependent on V_t, and vice-versa, that is

$$H_t = \Lambda V_t \Lambda', \tag{7}$$

where $\Lambda = (\lambda_{ij})$ is a $k \times k$ non-singular matrix. Due to the quadratic form in (7), a sign restriction on Λ needs to be imposed to ensure identifiability. For this purpose, we impose the sign restriction $\lambda_{11} > 0$. The specific role and economic interpretation of Λ depends on whether daily returns are computed as close-to-close or open-to-close; we refer to the empirical study for a discussion. Our model specification implies that the conditional statistical properties of the measurements can be expressed in terms of V_t and Λ, that is

$$\begin{aligned}
\mathrm{E}[r_t r_t' | \mathcal{F}_{t-1}] &= \Lambda V_t \Lambda', \qquad \mathrm{E}[X_t | \mathcal{F}_{t-1}] = V_t, & (8)\\
\mathrm{Var}[\mathrm{vec}(r_t r_t') | \mathcal{F}_{t-1}] &= (I_{k^2} + K_k)\big(\Lambda \otimes \Lambda\big)\big(V_t \otimes V_t\big)\big(\Lambda' \otimes \Lambda'\big), & (9)\\
\mathrm{Var}[\mathrm{vec}(\Lambda^{-1} r_t r_t' (\Lambda')^{-1}) | \mathcal{F}_{t-1}] &= (I_{k^2} + K_k)\big(V_t \otimes V_t\big), & (10)\\
\mathrm{Var}[\mathrm{vec}(X_t) | \mathcal{F}_{t-1}] &= \nu^{-1}(I_{k^2} + K_k)\big(V_t \otimes V_t\big), & (11)
\end{aligned}$$

where K_k is the $k^2 \times k^2$ commutation matrix as discussed in detail in Magnus and Neudecker (1979) from which also the results of (9) and (11) follow directly. 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. 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-asset effects (also known as spill-over effects).

We introduce the time-varying vector process f_t for which the details of its dynamic model specification are given below. We assume that V_t is a function of f_t, that is $V_t = V_t(f_t)$ for $t = 1, \ldots, T$. This flexible specification can accommodate a covariance matrix V_t that is only partly time-varying. But it can also allow for specifications that lead to a fully time-varying matrix V_t. In our study we consider the specification $f_t = \mathrm{vech}(V_t)$ where the operator $\mathrm{vech}(V_t)$ stacks the diagonal and lower-triangular elements of the covariance matrix V_t into a vector.

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 score function (at time t) of the conditional or predictive 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 the set Z_t consisting of m vector or matrix variables, we have $Z_t = \{Z_t^1, \ldots, Z_t^m\}$, for which observations or measurements are available for $t = 1, \ldots, T$. For our Realized-Wishart-GARCH model, we have $m = 2$, $Z_t^1 = r_t r_t'$ and $Z_t^2 = X_t$. It is a straightforward extension to include more variables into Z_t, such as other realized measures that can possibly provide more information on $V_t = V_t(f_t)$. The measurement distribution for the ith variable in Z_t is given by

$$Z_t^i \sim \varphi_i(Z_t^i | f_t, \mathcal{F}_{t-1}; \psi), \qquad i = 1, \ldots, m, \qquad t = 1, \ldots, T, \tag{12}$$

where f_t is the $d \times 1$ vector of time-varying parameters, $\mathcal{F}_{t-1} = \sigma(Z_s; s = 1, \ldots, t-1)$ is the sigma field generated by all observations up to time $t-1$, and ψ is a vector of (unknown) static model parameters. In this framework, the individual distribution φ_i may correspond to different families of distributions. All distributions however depend partially on the same time-varying parameter vector f_t. For our Realized-Wishart-GARCH model with conditional distributions (1) and (2), and with specification (7), return vector r_t and realized covariance matrix X_t have different distributions but are assumed to be propelled by the common covariance matrix $V_t = V_t(f_t)$. Finally, the distribution φ_i in (12) may depend on exogenous variables; we omit this extension for simplicity in notation.

We assume that the m variables in Z_t are conditionally independent, conditional on both f_t and the information set $\mathcal{F}_{t-1}$. We further assume that the distributional functions φ_i are at least differentiable up to the first order with respect to f_t. The log-likelihood function is then given by

$$\mathcal{L}(\psi) = \sum_{t=1}^{T} \sum_{i=1}^{m} \log \, \varphi_i(Z_t^i | f_t, \mathcal{F}_{t-1}; \psi). \tag{13}$$

The time-varying parameter 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 of coefficients. The unknown parameters in ω, $B_1, \ldots, B_p$, $A_1, \ldots, A_q$ and those associated with the measurement equations, such as the number of degrees of freedom in the Wishart distribution, are 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 but also for parameter estimation. 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. Under the assumption of correct model specification, the score has the convenient property that it forms a martingale difference 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 existence of ∇_t relies on the assumption of differentiability of φ_i with respect to f_t up to the first order. 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 existence of $\mathcal{I}_t$ relies on the assumption of differentiability of φ_i with respect to f_t up to the second order. The innovation term is now defined as

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

where the invertibility of $\mathcal{I}_t$ is assumed but is often simply implied by the choices of distribution φ_i, for $i = 1, \dots, m$. Further, the martingale property of ∇_t implies that $\mathrm{E}[s_t|\mathcal{F}_{t-1}] = 0$. In this approach, the one-step ahead prediction of the time-varying parameter vector, f_{t+1}, is primarily based on the scaled score that exploits the full likelihood contribution at time t. The score-driven time-varying parameter equations (14) and (17) are formulated as in Creal et al. (2013), for the case of the measurement distributions in (12). The details for the Realized-Wishart-GARCH model are given next. In the remainder of this treatment, we consider the updating equation (14) with $p = q = 1$ to obtain

$$f_{t+1} = \omega + B f_t + A s_t, \tag{18}$$

with $A = A_1$ and $B = B_1$.

2.3 The details for the Realized-Wishart-GARCH model

We provide the details of the score-driven model as introduced above for the Realized-Wishart-GARCH model with the time-varying covariance matrix $V_t = V_t(f_t)$ for the specification that f_t simply represents all elements of V_t. In particular, we require expressions for the score function and the Fisher information matrix. Given the conditional independence assumption for the variables in Z_t, in our case $Z_t^1 = r_t r_t'$ and $Z_t^2 = X_t$, we can decompose the contribution of the log-likelihood function (22) at time t in two parts, that is

$$\mathcal{L}(\psi) = \sum_{t=1}^{T} \mathcal{L}_t(\psi), \qquad \mathcal{L}_t(\psi) = \mathcal{L}_{r,t} + \mathcal{L}_{X,t},$$

with the log-likelihood parts given by

$$\mathcal{L}_{r,t} = \frac{1}{2} d_r(k) - \frac{1}{2} \log |\Lambda V_t \Lambda'| - \frac{1}{2} \mathrm{tr}\big((\Lambda V_t \Lambda')^{-1} r_t r_t'\big), \tag{19}$$

$$\mathcal{L}_{X,t} = \frac{1}{2} d_X(k, \nu) + \frac{\nu - k - 1}{2} \log |X_t| - \frac{\nu}{2} \log |V_t| - \frac{\nu}{2} \mathrm{tr}\big(V_t^{-1} X_t\big), \tag{20}$$

where $d_r(k) = -k \log(2\pi)$, $d_X(k, \nu) = \nu k \log(\nu/2) - 2 \log \Gamma_k(\nu/2)$ and $\Gamma_k()$ is the multivariate Gamma function for dimension k. In case of the Realized-Wishart-GARCH model, the two log-likelihood expressions follow immediately since the distribution $\varphi_1 = W_k^s(H_t, 1)$ is the singular Wishart distribution and $\varphi_2 = W_k(V_t/\nu, \nu)$ is the k-th dimensional Wishart distribution.

Our aim is to specify a dynamic model for the matrix V_t and the time-varying parameter vector f_t is therefore simply defined as

$$f_t = \text{vech}(V_t), \tag{21}$$

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)*, the score vector of dimension* $k^* \times 1$ *is given by*

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

where D_k is the duplication matrix as discussed in detail by Magnus and Neudecker (1979). □

Given the statistical properties in (8), it follows that $\text{E}[\nabla_t | \mathcal{F}_{t-1}] = 0$ under correct model specification; it implies that ∇_t forms a martingale difference sequence. The expression for the score shows that for the updating of f_t, and hence V_t, 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 receives a weight of one. This model feature is pertinent as the outer-product of daily returns typically contains a weak signal about the current covariance of assets as it does not exploit intra-day information.

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

$$\mathcal{I}_t = \text{E}[\nabla_t \nabla_t' | \mathcal{F}_{t-1}] = \frac{1+\nu}{2} D_k' \big(V_t^{-1} \otimes V_t^{-1}\big) D_k. \qquad \square$$

The inverse of the conditional information matrix exists since we have assumed that V_t is nonsingular. This inverse matrix will be used to scale the score vector.

Theorem 3. *For the measurements densities (1) and (2), the scaled score vector $s_t = \mathcal{I}_t^{-1} \nabla_t$ is given by*

$$s_t = \frac{1}{\nu+1} \Big(\nu \text{vech}(X_t) + \text{vech}(\Lambda^{-1} r_t r_t' (\Lambda')^{-1})\Big) - \text{vech}(V_t). \qquad \square$$

The proofs of Theorems 1, 2 and 3 are given in Appendix A.

For the updating of the time-varying parameter vector f_t in (18), and to avoid the curse of dimensionality, we can consider specifications with diagonal matrices for $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_{k^*})$ and $B = \mathrm{diag}(\beta_1, \ldots, \beta_{k^*})$, or with even more simpler scalar versions that have $A = \alpha I_{k^*}$ and $B = \beta I_{k^*}$. We need to impose some constraints on the parameters to guarantee that the covariance matrix V_t is positive definite with probability 1. For the scalar specification, the conditions $\alpha \geq 0$ and $\beta - \alpha \geq 0$ are sufficient to ensure that V_t is positive definite. Other constraints are needed for the diagonal specification which are discussed in more detail in Appendix C.

2.4 The Realized-Wishart-GARCH model with multiple measures

The results in Theorems 1-3 hold for our model with the two measurement equations (1) and (2). However, it is straightforward to extend our Realized-Wishart-GARCH modeling framework to incorporate several noisy measures of the daily equity covariance matrix V_t. For example, 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 the daily realized covariance matrix, for $i = 1, \ldots, G$, with $G \in \mathbb{N}$. We define $\nu^* = \sum_{i=1}^G \nu^i$ and we have

$$\nabla_t = \frac{1}{2} D_k' \big(V_t^{-1} \otimes V_t^{-1}\big) \sum_{i=1}^{G} \nu^i \big[\mathrm{vec}(X_t^i) - \mathrm{vec}(V_t)\big], \quad \mathcal{I}_t = \mathrm{E}[\nabla_t \nabla_t' | \mathcal{F}_{t-1}] = D_k' \big(V_t^{-1} \otimes V_t^{-1}\big) D_k \frac{\nu^*}{2},$$

and

$$s_t = \left(\sum_{i=1}^{G} \frac{\nu^i}{\nu^*} \mathrm{vech}(X_t^i) \right) - \mathrm{vech}(V_t),$$

where the numbers of degrees of freedom $\nu^1, \nu^2, \ldots, \nu^G$ are estimated along with other model static parameters. We notice that $\nu^i \equiv 1$ if $X_t^i = r_t r_t'$ or for any matrix X_t^i that has rank one.

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

$$\mathcal{L}(\psi) = \sum_{t=1}^{T} \big(\mathcal{L}_{r,t} + \mathcal{L}_{X,t}\big), \tag{22}$$

where $\mathcal{L}_{r,t}$ and $\mathcal{L}_{X,t}$ are given in (19) and (20), respectively. The time-variation of V_t is determined by the score recursion (14) and parameterization (21). The static parameter vector is given by

$$\psi = \big(\text{vec}(\Lambda)', \omega', \text{vec}(A)', \text{vec}(B)'\big)',$$

and contains at least $k^2 + k(k+1)/2$ elements for ω and Λ and more elements depending on the specification of A and B; the number of parameters is therefore of order $O(k^2)$. The computation of the log-likelihood function (22) requires the updating equations (18) that needs to be initialized. 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. In our empirical analysis we set f_0 to be (the vec of) the sample average of the realized covariance matrices $X_1, \ldots, X_T$. For a given parameter vector ψ, the log-likelihood function can be evaluated in a straightforward manner. In practice, ψ is unknown and estimation of all parameters is carried out via the numerical maximization of (22) with respect to ψ. The maximization relies typically on a standard quasi-Newton numerical optimization procedure; the initial values for ψ can be determined through a grid search method. For both the simulation study and the empirical application, the model parameters are estimated using numerical derivatives.

As the dimension k increases, parameter estimation can 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 updating equation (18) admits a vector autoregressive moving average (VARMA) representation, an analytical expression for the intercept can be provided, if stationarity conditions are satisfied. When we replace ω in (18) by its unconditional mean, we obtain

$$f_{t+1} = (I_{k^*} - B)\text{E}[f_t] + Bf_t + As_t,$$

where $\text{E}[f_t]$ is replaced by $\text{vech}(T^{-1}\sum_{t=1}^{T} X_t)$. 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 reductions can be achieved by setting A and B as diagonal matrices or to scalars.

3.2 Monte Carlo study

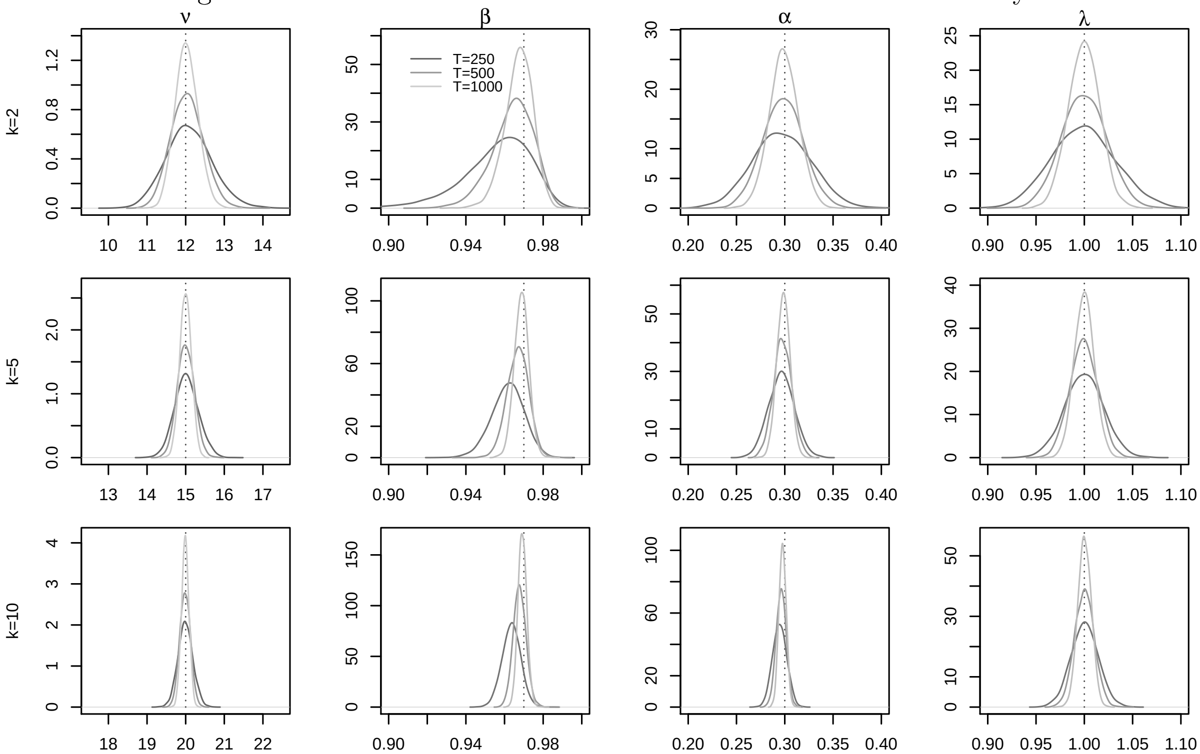
We study properties of the likelihood-based estimation method by means of simulation exercises. We consider a dimension of $k \in \{2, 5, 10\}$ and we simulate a series of $T \in \{250, 500, 1000\}$ daily returns and daily covariance matrices. For simplicity, we study the scalar specification for the time-varying parameter (18) with $A = \alpha I_{k^*}$ and $B = \beta I_{k^*}$. We further consider that all elements of Λ are the same, that is $\lambda_{i,j} = \lambda$, for $i, j = 1, \ldots, k$. The Monte Carlo data generation process has adopted the following parameter values

$$\nu = k + 10, \qquad \omega = 0.10 \text{ vech}(I_k), \qquad \beta = 0.97, \qquad \alpha = 0.30, \qquad \lambda = 1, \tag{23}$$

These parameter values are roughly in line with the empirical estimates that we present in Section 4. A close-to-unity value for 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 (23). We estimate the parameters without constraints but with covariance targeting. We emphasize that we do not simulate intra-day prices as we do not analyze the properties of high-frequency realized measures but we only aim to validate the estimation procedures for our model.

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 an 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 autoregressive coefficients require generally a relatively long time series for its estimation. Moreover, it is likely that the ad-hoc treatment of the initial value f_0 will require some strong adjustments for f_t in the first part of the sample. This will cause a (negative) bias in the estimation of β for relatively small samples. For an increasing sample size, this initial estimation bias will vanish. The number of degrees of freedom of the Wishart distribution ν

Figure 1: Parameter estimate densities from the Monte Carlo study



is estimated rather accurately, even for moderate sample sizes. This finding is promising but somewhat surprising given that ν is a highly nonlinear parameter.

By increasing k, this is the number of assets in our simulation study, the shapes of the densities become considerably more symmetric and more peaked around their true values; in particular, compare the panels for $k = 2$ and $k = 10$. We notice that in the Monte Carlo study our parameterization is parsimonious and therefore increasing k will lead to more pooling for the estimation of the parameters. Also, the data size increases with k^2 while the number of parameters increases with k. The improvement is however remarkable for parameters α and β. We may conclude overall that the maximum likelihood method is successful in the accurate estimation of model parameters.

4 Empirical illustration

4.1 Dataset: open-to-close daily returns and realized covariance matrices

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.

We analyze these 15 equities using the Realized-Wishart-GARCH model for different dimensions of $k \in \{2, 5, 15\}$. To conserve space, we will present results for a randomly selected set of ten bivariate models and ten 5-variate models amongst the 15 equities; the random selection is justified as our primary aim is to verify estimation results, to understand their implications and to detect similarities. We also present results for our model with all 15 equities included which requires the modeling of a 15×15 conditional covariance matrix. The sample period 2001-2010 represents two characteristic periods: first a period of low volatility and then a period of high or even extreme volatility due to the "financial crises". The length of a ten-year period is rather standard in the GARCH literature.

In our study we have followed the standard practice of excluding the overnight return for the computation of realized measures while daily asset returns can be based on both open-to-close and close-to-close returns. The vector of daily asset returns r_t is taken as open-to-close returns in our study. The conditional covariance matrix H_t therefore measures the intra-day variations and co-variations. Hence the covariance matrices H_t and V_t contain similar information. Given the specification $H_t = \Lambda V_t \Lambda'$ in (7), we may expect matrix Λ to be close to an identity matrix. However, the diagonal elements may be close to unity, the off-diagonal elements may reveal some interesting information on cross-asset or spillover effects. When we would have considered close-to-close returns, the overnight market risk, specific for each individual stock, would have been accounted for by the parameter matrix Λ; this overnight effect is of key interest to many market players such as liquidity providers or market makers who generally want to minimize this risk and hedge it effectively.

Before we compute the realized measures, 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.

For the computation of the realized covariance matrices, we adopt a kernel that is based on a subsampling scheme. We use an overall sample frequency of 5 minutes and adopt the refresh sampling scheme of Barndorff-Nielsen et al. (2011b). The refresh sampling scheme refers to the irregular sampling over time: a time interval ends when at least one realization is recorded for all considered k stocks. By shifting the starting time by 1-second increments, we obtain 300 different estimates in a 5 minutes interval; the average is our subsampled realized covariance measure. 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 for each year in our sample. We observe that for the 2×2 datasets we retain on average of around $60 - 65\%$ observations; 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 the financial crisis period of 2007-2009. For the 5×5 case the data loss is more pronounced. We retain around $35 - 40\%$ and we have 1800 refresh observations on average. 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.

4.2 Estimation results

We present the parameter estimation results from the Realized-Wishart-GARCH model when applied to the datasets as described. The dynamic specification for the covariance matrix V_t is based on the updating equation (18) for $f_t = \text{vech}(V_t)$ with $A = \alpha I_{k^*}$ and $B = \beta I_{k^*}$. In Appendix C, we consider the estimation results for a less parsimonious specification that allows for different dynamics for the variances (α_v and β_v) and covariances (α_c and β_c). The additional results do not suggest that a more flexible specification provides better results compared to those for the basic specification. We also investigate the presence of cross-effects by having Λ as a diagonal matrix and as a full matrix. When off-diagonal elements of Λ are estimated to be

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: Average daily number of high-frequency observations maintained by the refresh sampling scheme of Barndorff-Nielsen et al. (2011b). The averages are over the days in each year of our sample.

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: Average ratio of the data maintained by the refresh sampling scheme of Barndorff-Nielsen et al. (2011b). The averages are over the days in each year of our sample.

significantly different from zero, it implies that cross-effects are present. Table 3 presents the maximum likelihood estimation results for the parameters in the Realized-Wishart-GARCH model for $k = 2$. We report the estimates of the models for a full matrix Λ (first panel) and for a diagonal matrix Λ (second panel). The estimates of the diagonal elements of Λ tend to be close-to-unity but most have estimated values just below unity, the smallest estimate is 0.88 and the largest is 1.03. Many off-diagonal elements are estimated as not being significantly different from zero, only five out of twenty appear to have some statistical impact. The significantly estimated off-diagonal elements of Λ are all positive and range from 0.03 to 0.23. Although in most cases, the Akaike information criterion (AIC) points weakly towards a model specification with a full Λ matrix, other aspects of our analyses, including the estimates of ν, β and α, are not affected when we restrict Λ to be diagonal. Table 4 presents the results for the model with $k = 5$ and Table 5 presents those for the model with $k = 15$, both with a diagonal matrix Λ.

Taking all results together, the estimates of the parameters amongst the different stock combinations are very similar. In general, we find that the estimates of β are close-to-unity from which we can infer that the time-varying process of the covariance matrix is highly persistent. We also observe that the dynamics of V_t rely more on the realized kernel measures given the highly significant estimates of ν. Furthermore, we find that for a higher dimension k, the estimates of ν become higher and more significant. It implies that for models with more stocks, more reliance is given to the realized measures. We emphasize that the degrees of freedom ν needs to grow with the dimension k in order to ensure that the Wishart covariance matrix does not become non-singular; see 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 Realized-Wishart-GARCH model. The estimates of ν appear to be higher in relation to the dimension k and we therefore conclude that the realized measures play a considerable role in our analysis.

4.3 Forecasting study: other forecasting models and methods

In our forecasting study, we compare the out-of-sample performance of the Realized-Wishart-GARCH (RWG) model against four alternative forecasting models and methods. Our model allows for a joint analysis of daily returns and realized variance variables. In our comparisons, we consider two forecasting approaches for daily returns and two for realized measures. The two models for the vector of daily returns are the dynamic conditional correlation (DCC) model

Equities	ν	β	α	λ_{11}	λ_{22}	λ_{12}	λ_{21}	$\log L$	AIC
2×2									
AA/CAT	12.428 (0.189)	0.977 (0.002)	0.331 (0.011)	0.893 (0.034)	1.022 (0.026)	0.226 (0.073)	-0.032 (0.050)	-20171.5	40356.9
AXP/PFE	10.876 (0.164)	0.991 (0.001)	0.378 (0.012)	1.032 (0.016)	0.918 (0.016)	-0.018 (0.030)	0.078 (0.021)	-17107.3	34228.5
AXP/WMT	11.907 (0.180)	0.993 (0.001)	0.360 (0.012)	1.018 (0.017)	0.887 (0.015)	0.033 (0.033)	0.025 (0.017)	-15347.7	30709.4
BA/HON	10.681 (0.161)	0.975 (0.002)	0.354 (0.011)	0.986 (0.028)	0.894 (0.029)	0.026 (0.053)	0.104 (0.055)	-17860.8	35735.5
CAT/KO	12.829 (0.195)	0.977 (0.002)	0.354 (0.011)	0.986 (0.022)	0.928 (0.017)	0.095 (0.074)	-0.037 (0.030)	-14227.6	28469.1
GE/PFE	11.015 (0.166)	0.984 (0.001)	0.405 (0.013)	0.943 (0.017)	0.911 (0.018)	0.016 (0.030)	0.072 (0.029)	-15622.3	31258.7
HD/JPM	12.458 (0.189)	0.988 (0.001)	0.447 (0.013)	0.953 (0.018)	0.944 (0.020)	0.020 (0.031)	0.125 (0.036)	-18481.1	36976.2
IBM/PG	12.409 (0.189)	0.977 (0.002)	0.383 (0.012)	0.984 (0.020)	0.866 (0.020)	-0.025 (0.052)	0.030 (0.035)	-10961.1	21936.2
JPM/XOM	13.086 (0.199)	0.989 (0.001)	0.441 (0.012)	0.987 (0.016)	0.930 (0.016)	0.012 (0.032)	0.034 (0.018)	-16082.0	32178.0
MCD/PG	10.427 (0.157)	0.979 (0.002)	0.310 (0.011)	0.919 (0.018)	0.880 (0.017)	0.039 (0.049)	0.015 (0.027)	-12645.9	25305.8
AA/CAT	12.424 (0.189)	0.977 (0.002)	0.333 (0.011)	0.952 (0.013)	0.978 (0.013)	-	-	-20201.6	40413.3
AXP/PFE	10.876 (0.164)	0.991 (0.001)	0.377 (0.012)	1.013 (0.014)	0.940 (0.013)	-	-	-17118.6	34247.2
AXP/WMT	11.908 (0.180)	0.993 (0.001)	0.360 (0.012)	1.016 (0.014)	0.890 (0.012)	-	-	-15353.2	30716.3
BA/HON	10.680 (0.161)	0.975 (0.002)	0.354 (0.011)	0.969 (0.013)	0.915 (0.012)	-	-	-17883.1	35776.3
CAT/KO	12.829 (0.195)	0.977 (0.002)	0.354 (0.011)	1.007 (0.014)	0.913 (0.013)	-	-	-14228.4	28466.8
GE/PFE	11.013 (0.166)	0.984 (0.001)	0.405 (0.013)	0.931 (0.013)	0.926 (0.013)	-	-	-15634.8	31279.6
HD/JPM	12.455 (0.189)	0.988 (0.001)	0.445 (0.013)	0.937 (0.013)	0.968 (0.013)	-	-	-18509.2	37028.4
IBM/PG	12.409 (0.189)	0.977 (0.002)	0.383 (0.012)	0.974 (0.014)	0.877 (0.012)	-	-	-10961.7	21933.3
JPM/XOM	13.088 (0.199)	0.989 (0.001)	0.441 (0.012)	0.979 (0.014)	0.939 (0.013)	-	-	-16087.1	32184.1
MCD/PG	10.426 (0.157)	0.979 (0.002)	0.310 (0.011)	0.921 (0.013)	0.879 (0.012)	-	-	-12649.5	25309.0

Table 3: Maximum likelihood estimates for the 2×2 models. Standard errors are shown in parentheses.

Equities	ν	β	α	λ_{11}	λ_{22}	λ_{33}	λ_{44}	λ_{55}	$\log L$
5 × 5									
AA/AXP/IBM/JPM/WMT	18.515	0.991	0.302	0.968	0.979	0.970	0.955	0.891	-43769.0
	(0.120)	(0.000)	(0.005)	(0.013)	(0.012)	(0.013)	(0.012)	(0.012)	
AA/BA/CAT/GE/KO	17.727	0.986	0.266	0.958	0.984	0.985	0.925	0.923	-44903.2
	(0.115)	(0.001)	(0.005)	(0.013)	(0.013)	(0.013)	(0.012)	(0.013)	
AXP/CAT/IBM/KO/XOM	19.185	0.990	0.296	0.998	0.991	0.977	0.919	0.946	-32878.6
	(0.125)	(0.001)	(0.004)	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)	
BA/HD/JPM/PFE/PG	17.856	0.987	0.300	0.986	0.949	0.968	0.941	0.889	-42213.9
	(0.116)	(0.001)	(0.005)	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)	
BA/HD/MCD/PG/XOM	18.178	0.981	0.273	0.987	0.962	0.932	0.888	0.954	-36404.4
	(0.118)	(0.001)	(0.005)	(0.013)	(0.013)	(0.013)	(0.012)	(0.013)	
CAT/GE/KO/PFE/WMT	18.084	0.985	0.283	1.004	0.933	0.921	0.944	0.902	-33962.4
	(0.117)	(0.001)	(0.005)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)	
CAT/HON/IBM/MCD/WMT	17.252	0.981	0.268	0.997	0.928	0.982	0.948	0.901	-38326.4
	(0.111)	(0.001)	(0.004)	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)	
GE/IBM/JPM/PG/XOM	19.650	0.989	0.342	0.912	0.969	0.958	0.879	0.940	-30238.6
	(0.129)	(0.001)	(0.005)	(0.011)	(0.013)	(0.012)	(0.012)	(0.012)	
HD/HON/KO/MCD/PG	17.018	0.980	0.276	0.956	0.937	0.915	0.939	0.890	-34545.2
	(0.109)	(0.001)	(0.005)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)	
HON/IBM/MCD/WMT/XOM	18.139	0.982	0.279	0.933	0.984	0.946	0.903	0.956	-33552.8
	(0.118)	(0.001)	(0.004)	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	

Table 4: Maximum likelihood estimates for the 5×5 models. Standard errors are shown in parentheses.

$\mathbf{15 \times 15}$	AA/.../XOM
ν	29.260 (0.060)
β	0.990 (0.000)
α	0.187 (0.001)
λ_{11}	0.966 (0.012)
λ_{22}	0.978 (0.012)
λ_{33}	0.992 (0.013)
λ_{44}	0.990 (0.012)
λ_{55}	0.908 (0.010)
λ_{66}	0.943 (0.012)
λ_{77}	0.922 (0.011)
λ_{88}	0.985 (0.012)
λ_{99}	0.951 (0.012)
λ_{1010}	0.928 (0.012)
λ_{1111}	0.954 (0.013)
λ_{1212}	0.954 (0.013)
λ_{1313}	0.900 (0.012)
λ_{1414}	0.903 (0.011)
λ_{1515}	0.952 (0.012)
$\log L$	-61418.1

Table 5: Maximum likelihood estimates for the 15×15 model. Standard errors are shown in parentheses.

of Engle (2002a) and the so-called BEKK model of Engle and Kroner (1995). The model-based forecasting framework for the realized covariance matrix is the conditional autoregressive Wishart (CAW) model of Golosnoy et al. (2012) while the non-parametric forecasting method is based on the exponentially weighted moving average (EWMA) scheme. In the forecasting study, we consider the scalar specifications for the updating of the conditional covariance matrix in the RWG model but also, where applicable, for the DCC, BEKK and CAW models. Finally, we assume matrix Λ to be diagonal in the RWG model. A short practical introduction to each model is provided next.

The CAW model assumes that the conditional distribution of the realized variance is Wishart with scale matrix V_t^c and degrees of freedom ν^c, we simply have $X_t|\mathcal{F}_{t-1} \sim W_k(V_t^c/\nu^c, \nu^c)$. The updating of the conditional covariance matrix is also subject to covariance targeting and to the scalar specification, that is

$$V_{t+1}^c = (1 - \beta^c - \alpha^c)\bar{X} + \beta^c V_t^c + \alpha^c X_t, \qquad \beta^c \geq 0,\ \alpha^c > 0,\ \alpha^c + \beta^c < 1,$$

for $t = 1, \ldots, T$ and with $\bar{X} = (1/T)\sum_{t=1}^{T} X_t$. *The EWMA method* is the one-step ahead forecasting scheme applied to the realized variance series; it is the default method used by practitioners and regulators; see, for example, RiskMetrics as described by J.P.Morgan (1996). The updating equation also has a scalar specification and is given by

$$V_{t+1}^e = \beta^e \cdot V_t^e + (1 - \beta^e) \cdot X_t, \qquad 0 < \beta^e < 1,$$

where we treat β^e as a fixed smoothing constant that we set equal to $\beta^e = 0.96$. In our implementation, we can regard EWMA as a special or limiting case of CAW with $\alpha^c = 0.04$ and $\beta^c = \beta^e = 0.96$. *The DCC model* assumes that the daily returns vector is conditionally normally distributed as $r_t|\mathcal{F}_{t-1} \sim N(0, V_t^d)$ with its covariance matrix given by $V_{t+1}^d = D_t R_t D_t$ where D_t is a diagonal matrix with its i-th diagonal element given by $\sqrt{h_{i,t}}$ and where R_t is the conditional correlation matrix with $R_t = \text{diag}[Q_t]^{-1/2} Q_t \text{diag}[Q_t]^{-1/2}$, for $t = 1, \ldots, T$. The updating of $h_{i,t}$ and Q_t takes place in two different steps. It is assumed that $h_{i,t}$ follows the GARCH(1,1) process as given by

$$h_{i,t+1} = \omega_i^d + \beta_i^d h_{i,t} + \alpha_i^d r_{i,t}^2, \qquad \omega_i^d > 0,\ \beta_i^d \geq 0,\ \alpha_i^d > 0,\ \alpha_i^d + \beta_i^d < 1,$$

for $i = 1, \ldots, k$ and where $r_{i,t}$ is the ith element of daily return vector r_t. The scalar updating equation with covariance targeting for Q_t is given by

$$Q_t = (1 - \beta^+ - \alpha^+)\bar{Q} + \beta^+ Q_t + \alpha^+ \epsilon_t \epsilon_t', \qquad \beta^+ \geq 0,\ \alpha^+ > 0,\ \alpha^+ + \beta^+ < 1,$$

where ϵ_t is the GARCH residual vector with its ith element given by $\epsilon_{i,t} = r_{i,t} \,/\, \sqrt{h_{i,t}}$, for $i = 1, \ldots, k$, and $\bar{Q} = T^{-1} \sum_{t=1}^{T} \epsilon_t \epsilon_t'$. The BEKK model assumes that $r_t | \mathcal{F}_{t-1} \sim N(0, V_t^b)$ and the covariance matrix of the vector of asset returns is driven by the outer-products of daily returns. The scalar updating equation with covariance targeting is given by

$$V_{t+1}^b = (1 - \beta^b - \alpha^b)\bar{V} + \beta^b V_t^b + \alpha^b r_t r_t', \qquad \beta^b \geq 0,\ \alpha^b > 0,\ \alpha^b + \beta^b < 1,$$

where $\bar{V} = T^{-1} \sum_{t=1}^{T} r_t r_t'$ is the sample covariance matrix of daily returns, and a^b and b^b are unknown coefficients.

4.4 Forecasting study: design and forecast loss functions

We split our original dataset in two subsamples: the in-sample data consists of the years 2001-2008 and the out-of-sample consists of the years 2009-2010. We consider these last two years as our forecasting evaluation period. The years 2009-2010 are somewhat representative of financial markets. In 2009 many large equity recovery operations have taken place in the U.S. while 2010 has shown a return to a modest market risk.

The estimation of the static parameter vector, for all model specifications, is done only once for the in-sample data. The one-step ahead forecasts are generated for the out-of-sample data (without the re-estimation of static parameters), for all model specifications. The evaluation of the out-of-sample forecasts is based on the Diebold-Mariano (DM) test to assess the statistical significance of the superiority of the forecasting performance of a specific model; see Diebold and Mariano (1995). In our study, we test whether our Realized-Wishart-GARCH (RWG) model has a significantly smaller out-of-sample loss compared to the loss of the other considered models in our forecasting study. For this purpose, we measure the performance of the models by means of two loss functions: the root mean squared error (RMSE) based on the matrix norm given by

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

and the quasi-likelihood (QL) loss function as given by

$$QL(V_t, S_t) = \log |V_t| + \mathrm{tr}\big(V_t^{-1} S_t\big),$$

where S_t is an observed measure of the covariance matrix and V_t is the covariance matrix as predicted by the model or method. Given that we jointly analyze r_t and X_t with our RWG model, we evaluate the performances of all models in forecasting the daily returns density and the realized variances and covariances. Therefore $S_t = X_t$ for the forecasting of the realized covariance matrix and $S_t = r_t r_t'$ for the forecasting of the density in daily returns. We notice that in case of daily returns with $S_t = r_t r_t'$, the quasi-likelihood loss is equivalent to the log-score criterion for a Gaussian distribution. The log-score criterion is widely used in density forecast comparisons between different models; see Geweke and Amisano (2011).

4.5 Forecasting study: empirical results

The results of our forecasting study are summarized in Tables 6 and 7: in Table 6 we report the forecasting results for the realized covariance matrix and in Table 7 for the density in daily returns. Both tables display the relative value of the loss function for our RWG model against the other models. We measure the relative performance by the ratio between the loss for a given model and the loss for the RWG model. When a model has a relative performance larger than unity, the implication is that it underperforms the RWG model. The opposite is also true. When the relative performance is smaller than unity, the model outperforms the RWG model.

We learn from Table 6 that the RWG and CAW models are the best performing models in forecasting the realized measures. Their performances are very similar in relative terms and, except for a few cases, there is not a statistically significant difference. This finding is to be expected given that the daily returns are not very informative to forecast the realized measures. Therefore the RWG model is not expected to outperform the CAW model by a large amount. However, from Table 7 we can conclude that the RWG model is by far the best performing model in forecasting the density in daily returns. The outperformance is in relative terms as well as in statistical terms because the reported DM tests are clearly significant in most cases. Here the RWG is able to outperform the DCC and BEKK convincingly. The reason is obvious since it exploits additional information as provided by the realized measures. In a similar fashion, the RWG model outperforms the CAW model and the EWMA method since our preferred model analyzes the daily returns jointly with the realized measures. On the other hand, the CAW model and the EWMA method only consider the realized measures. We can therefore conclude that the factor structure of the RWG model is particularly useful in exploiting the realized measures for the forecasting of the density in daily returns.

5 Conclusions

We have proposed a new model for the joint modeling and forecasting of daily time series of returns and realized covariance matrices of financial assets: the Realized-Wishart-GARCH model. There are many distinguishing features of our model when compared to alternative frameworks. First, the model relies both on low- (daily) and on high-frequency (intra-day) information. It turns out that the high-frequency measures are given most weight since they exploit intra-day data of financial assets to infer about the underlying covariance structures. Several noisy measures that are based on different sampling frequencies can be considered in the analysis. Second, the time-varying features of the Realized-Wishart-GARCH model are driven by updates of the covariance matrix that exploit full-likelihood information. The model relies on standard parsimonious formulations which is a convenient property for multivariate conditional volatility models. In particular, the model is closely connected with the multivariate GARCH literature and the dynamics are related with vector autoregressive moving average models. Third, the model parameters can be interpreted straightforwardly. An example is that overnight market risk can be measured directly via the parameter matrix Λ when daily close-to-close returns are considered in the analysis. Fourth, the modeling framework is flexible: it can be extended easily when more realized measures are considered. The multivariate model can also be used to simulate realistic dynamic paths for portfolios in order to facilitate the validation of investment strategies. Fifth, the likelihood function is available analytically and hence estimation is easy; nonetheless computer code is made available for its use. Finally, in an empirical study for a portfolio of fifteen NYSE equities, we have studied the Realized-Wishart-GARCH model and its different specifications. We have provided in-sample evidence that our basic specification can be effective in extracting the salient features in the data. In an out-of-sample forecasting study we compare our model performance against four competitive models and methods. The ability of our model to jointly capture the daily returns vector and the realized covariance matrix appears in particular to benefit the accuracy in forecasting the density of daily returns.

	RMSE loss					QL loss				
	RWG	CAW	EWMA	BEKK	DCC	RWG	CAW	EWMA	BEKK	DCC
2x2										
AA/CAT	**6.08**	1.01	1.38***	1.32***	1.19***	**4.74**	1.00	1.02***	1.03***	1.02***
AXP/PFE	3.86	**0.99**	1.55***	1.84***	1.76***	4.14	**1.00***	1.03***	1.06***	1.04***
AXP/WMT	**3.41**	1.00	1.56***	1.86***	1.71***	**3.40**	1.00	1.03***	1.05***	1.04***
BA/HON	3.20	**1.00**	1.40***	1.23***	1.17**	**3.41**	1.00	1.03***	1.03***	1.03***
CAT/KO	**2.51**	1.01	1.48***	1.15***	1.10*	3.18	**1.00**	1.03***	1.03***	1.03***
GE/PFE	**4.11**	1.02	1.57***	1.25***	1.23***	3.98	**1.00**	1.04***	1.05***	1.04***
HD/JPM	4.75	**1.00**	1.60***	1.78***	1.55***	4.12	**1.00**	1.03***	1.06***	1.04***
IBM/PG	**1.62**	1.01	1.54***	1.13	1.06	2.09	**1.00**	1.09**	1.24	1.21
JPM/XOM	4.19	**1.00**	1.63***	1.79***	1.59***	3.53	**1.00**	1.05***	1.09**	1.05**
MCD/PG	**1.33**	1.00	1.49***	1.14**	1.18***	2.03	**1.00**	1.07*	1.17	1.24
5x5										
AA/AXP/IBM/JPM/WMT	**20.77**	1.00	1.54***	2.16***	1.57***	8.59	**1.00**	1.03***	1.07***	1.05***
AA/BA/CAT/GE/KO	**20.04**	1.01*	1.46***	2.31***	1.16***	9.05	**1.00**	1.02***	1.07***	1.02***
AXP/CAT/IBM/KO/XOM	**13.29**	1.01	1.59***	1.90***	1.43***	6.91	**1.00**	1.04***	1.09***	1.05**
BA/HD/JPM/PFE/PG	13.34	**1.00**	1.53***	1.77***	1.34***	8.26	**1.00**	1.03***	1.11*	1.08*
BA/HD/MCD/PG/XOM	**9.73**	1.01	1.54***	1.33***	1.09	6.16	**1.00**	1.04*	1.10*	1.09
CAT/GE/KO/PFE/WMT	**12.40**	1.01	1.54***	1.42***	1.18***	7.38	**1.00**	1.03***	1.06***	1.03***
CAT/HON/IBM/MCD/WMT	**10.61**	1.01*	1.53***	1.38***	1.12*	5.97	**1.00**	1.03***	1.06***	1.05***
GE/IBM/JPM/PG/XOM	**15.79**	1.00	1.65***	1.98***	1.45***	7.10	**1.00**	1.06***	1.16*	1.11
HD/HON/KO/MCD/PG	**9.03**	1.01	1.57***	1.42***	1.12*	5.63	**1.00**	1.04**	1.11*	1.11
HON/IBM/MCD/WMT/XOM	**8.68**	1.01	1.59***	1.41***	1.10	5.00	**1.00**	1.04**	1.07***	1.06***
15x15										
AA/.../XOM	119.91	**1.00**	1.47***	2.06***	1.21***	20.09	**1.00**	1.03***	1.10***	1.06**

Table 6: Out-of-sample $RMSE$ loss and QL loss for the realized covariance matrix. The out-of-sample window is two years. The best configurations are identified in bold. The Realized-Wishart-GARCH (RWG) is the benchmark model. The average loss is reported for the benchmark model while the relative loss is reported for the other models. The relative loss is the ratio between the loss of a model and the loss of the benchmark.

	RMSE loss					QL loss				
	RWG	CAW	EWMA	BEKK	DCC	RWG	CAW	EWMA	BEKK	DCC
2x2										
AA/CAT	**19.31**	1.03***	1.11***	1.12***	1.08***	4.67	**1.00**	1.01	1.00	1.01
AXP/PFE	**13.76**	1.01	1.12***	1.15***	1.14***	**4.10**	1.00	1.02*	1.02	1.03*
AXP/WMT	**11.73**	1.01**	1.13***	1.13***	1.10***	**3.35**	1.01	1.03*	1.03	1.03
BA/HON	**8.91**	1.05***	1.15***	1.06**	1.03	**3.17**	1.01	1.03**	1.02	1.03*
CAT/KO	**7.20**	1.01***	1.13***	1.03	1.01	3.02	**1.00**	1.03	1.02	1.02
GE/PFE	**9.68**	1.07***	1.23***	1.08*	1.08*	**3.54**	1.01	1.05***	1.02*	1.02*
HD/JPM	**14.92**	1.02**	1.15***	1.19***	1.16***	**4.01**	1.00	1.03**	1.02*	1.03**
IBM/PG	**3.48**	1.08***	1.26***	1.06	1.03	**1.52**	1.04***	1.15***	1.09**	1.08**
JPM/XOM	**13.34**	1.01*	1.14***	1.18***	1.14*	**3.32**	1.00	1.05***	1.02	1.03**
MCD/PG	**3.11**	1.10***	1.25***	1.06*	1.07***	**1.27**	1.09***	1.17***	1.06*	1.05
5x5										
AA/AXP/IBM/JPM/WMT	**73.86**	1.02***	1.13***	1.26***	1.15***	8.48	**1.00**	1.02*	1.03**	1.03**
AA/BA/CAT/GE/KO	**57.63**	1.03***	1.16***	1.38***	1.06***	8.62	**1.00**	1.02**	1.05***	1.02*
AXP/CAT/IBM/KO/XOM	**44.77**	1.01***	1.12***	1.19***	1.08***	**6.44**	1.00	1.04***	1.05***	1.04**
BA/HD/JPM/PFE/PG	**39.74**	1.03***	1.15***	1.20***	1.09***	**7.45**	1.01***	1.04***	1.03**	1.02**
BA/HD/MCD/PG/XOM	**23.44**	1.04***	1.24***	1.15***	1.00	**4.95**	1.03***	1.07***	1.05***	1.04***
CAT/GE/KO/PFE/WMT	**32.58**	1.04***	1.18***	1.14***	1.04*	**6.55**	1.01*	1.05***	1.03**	1.03**
CAT/HON/IBM/MCD/WMT	**28.56**	1.04***	1.19***	1.14***	1.01	**5.20**	1.02***	1.05***	1.06***	1.05***
GE/IBM/JPM/PG/XOM	**44.24**	1.03***	1.18***	1.27***	1.13***	**5.93**	1.02***	1.08***	1.06***	1.04***
HD/HON/KO/MCD/PG	**21.33**	1.07***	1.27***	1.20***	1.02	**4.61**	1.02***	1.06***	1.04*	1.04**
HON/IBM/MCD/WMT/XOM	**21.11**	1.05***	1.24***	1.17***	1.02	**4.12**	1.02***	1.05***	1.07***	1.06***
15x15										
AA/.../XOM	**333.91**	1.03***	1.16***	1.31***	1.08***	**18.18**	1.01**	1.03***	1.06***	1.03***

Table 7: Out-of-sample $RMSE$ loss and QL loss for the density in daily returns. The out-of-sample window is two years. The best configurations are identified by bold font. The Realized-Wishart-GARCH (RWG) is the benchmark model. The average loss is reported for the benchmark model while the relative loss is reported for the other models. The relative loss is the ratio between the loss of a model and the loss of the benchmark.

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.

Anudrino, F. and Trojani, F. (2011). A general multivariate threshold GARCH model for dynamic correlations. *Journal of Business and Economic Statistics*, 29(1):138–149.

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. (2011a). 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. (2011b). 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.

Buraschi, A., Trojani, F., and Vedolin, A. (2017). When uncertainty blows in the orchard: Comovement and equilibrium volatility risk premia. *Journal of Finance*, 72:forthcoming.

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.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Driessen, J., Maenhout, P., and Vilkov, G. (2009). The price of correlation risk: Evidence from equity options. *Journal of Finance*, 64:1377–1406.

Engle, R. (2002a). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Engle, R. F. (2002b). 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.

Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164:130–141.

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., 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.

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.

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.

Khrapov, S. and Renault, E. (2016). Affine option pricing model in discrete time. Mimeo.

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. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, third edition.

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

Opschoor, A., Janus, P., Lucas, A., and van Dijk, D. (2017). New HEAVY models for fat-tailed returns and realized covariance kernels. *Journal of Business and Economic Statistics*, 35:forthcoming.

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

Seber, G. A. F. (2007). *A Matrix Handbook for Statisticians*. Wiley Interscience, New York.

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 Andersen, T. G., Davis, R. A., Kreiß, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*, pages 201–229. Springer-Verlag, Berlin.

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

Uhlig, H. (1994). On singular wishart and singular multivariate beta distributions. *The Annals of Statistics*, 22:395–405.

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 $\text{vec}(A)$ operator stacks the columns of matrix A consecutively into the $k^2 \times 1$ column vector, while $\text{vech}(A)$ stacks the lower triangular part including diagonal into $k^* \times 1$ column vector, with $k^* = k(k+1)/2$. 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

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

where B is an arbitrary $k \times k$ matrix and A is 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. Additional properties and results related to these matrices can be found in Magnus and Neudecker (2007) and Seber (2007).

The proofs in the next appendix make use of the following results in matrix calculus. For a $k \times k$ symmetric matrix X, the derivative of $\text{vec}(X)$ with respect to $\text{vech}(X)$ is given by

$$\frac{\partial \text{vec}(X)}{\partial \text{vech}(X)'} = D_k,$$

where the duplication matrix D_k is defined above. For all $k \times k$ nonsingular matrices A, X and B, we have

$$\begin{aligned} \frac{\partial \log |AXB|}{\partial \text{vec}(X)'} &= \text{vec}[(X^{-1})']', \\ \frac{\partial \text{vec}(X^{-1})}{\partial \text{vec}(X)} &= -(X^{-1})' \otimes X^{-1}, \\ \frac{\partial \text{tr}(AXB)}{\partial \text{vec}(X)} &= \text{vec}(A'B')'. \end{aligned} \tag{24}$$

Finally, for all $k \times k$ matrices A, B and C, we have

$$\text{vec}(ABC) = (C' \otimes A)\text{vec}(B). \tag{25}$$

B Proofs

Proof of Theorem 1. We derive the score vector of which the general form is given by (15). From the equations (19) and (20), the relevant parts of log-likelihoods for the score vector derivation can be explicitly given as in

$$\mathcal{L}_{r,t} = c_r - \frac{1}{2}\Big(\log |\Lambda V_t \Lambda'| + \mathrm{tr}\big((\Lambda V_t \Lambda')^{-1} r_t r_t'\big)\Big), \tag{26}$$

$$\mathcal{L}_{X,t} = c_X - \frac{\nu}{2}\Big(\log |V_t| + \mathrm{tr}\big(V_t^{-1} X_t\big)\Big), \tag{27}$$

where c_r and c_X are non-relevant constants. We consider the covariance matrix V_t and parameter vector f_t, given by (21), as two unknown, non-random variables. Using the chain rule for vector differentiation, the score functions for the individual measurements associated with (1) and (2) can be expressed by

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

We first differentiate the measurement density for returns (26). Using (24) and (25), together with noting that V_t is symmetric and $V_t^{-1} = V_t^{-1} V_t V_t^{-1}$, we obtain

$$\begin{aligned} \frac{\partial \mathcal{L}_{r,t}}{\partial \mathrm{vec}(V_t)'} &= -\frac{1}{2}\left[\mathrm{vec}(V_t^{-1})' - \mathrm{vec}(\Lambda^{-1} r_t r_t' (\Lambda')^{-1})' (V_t^{-1} \otimes V_t^{-1})\right] \\ &= -\frac{1}{2}\Big[\mathrm{vec}(V_t)'(V_t^{-1} \otimes V_t^{-1}) - \mathrm{vec}\left(\Lambda^{-1} r_t r_t' (\Lambda')^{-1}\right)' (V_t^{-1} \otimes V_t^{-1})\Big] \\ &= \frac{1}{2}\left[\mathrm{vec}(\Lambda^{-1} r_t r_t' (\Lambda')^{-1})' - \mathrm{vec}(V_t)'\right](V_t^{-1} \otimes V_t^{-1}), \end{aligned} \tag{28}$$

and similarly we differentiate the measurement density for the realized covariance (27), we have

$$\begin{aligned} \frac{\partial \mathcal{L}_{X,t}}{\partial \mathrm{vec}(V_t)'} &= -\frac{\nu}{2}[\mathrm{vec}(V_t^{-1})' - \mathrm{vec}(X_t)'(V_t^{-1} \otimes V_t^{-1})] \\ &= -\frac{\nu}{2}\left[\mathrm{vec}(V_t)'(V_t^{-1} \otimes V_t^{-1}) - \mathrm{vec}(X_t)'(V_t^{-1} \otimes V_t^{-1})\right] \\ &= \frac{\nu}{2}[\mathrm{vec}(X_t) - \mathrm{vec}(V_t)]'(V_t^{-1} \otimes V_t^{-1}). \end{aligned} \tag{29}$$

Therefore, given the results (28) and (29), combined with the fact that $\partial \mathrm{vec}(V_t)/\partial f_t' = D_k$ and with the score defined in (15), we conclude that 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

$$\nabla_{r,t} = \frac{1}{2}D_k'\big(V_t^{-1}\otimes V_t^{-1}\big)\big[\mathrm{vec}(\Lambda^{-1}r_t r_t'(\Lambda')^{-1}) - \mathrm{vec}(V_t)\big],$$
$$\nabla_{X,t} = \frac{\nu}{2}D_k'\big(V_t^{-1}\otimes V_t^{-1}\big)\big[\mathrm{vec}(X_t) - \mathrm{vec}(V_t)\big],$$

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

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

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

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

Finally, considering that $I_{k^2}+K_k = 2D_kL_k$ (see Theorem 12 in Chapter 3 of Magnus and Neudecker (2007)) and that $L_kD_k = I_{k^*}$, we obtain

$$\mathcal{I}_{r,t} = \frac{1}{2}D_k'\big(V_t^{-1}\otimes V_t^{-1}\big)D_k,$$
$$\mathcal{I}_{X,t} = \frac{\nu}{2}D_k'\big(V_t^{-1}\otimes V_t^{-1}\big)D_k,$$

which combined with (16) completes the proof. □

Proof of Theorem 3: The score ∇_t can be written as

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

since $D_k'\big(V_t^{-1}\otimes V_t^{-1}\big)D_kL_k = D_k'\big(V_t^{-1}\otimes V_t^{-1}\big)$; see Theorem 13 in Chapter 3 of Magnus and Neudecker (2007). Together with the expression of the conditional Fisher information $\mathcal{I}_t = \frac{\nu+1}{2}D_k'\big(V_t^{-1}\otimes V_t^{-1}\big)D_k$ and the equality $\big(D_k'\big(V_t^{-1}\otimes V_t^{-1}\big)D_k\big)^{-1}D_k'\big(V_t^{-1}\otimes V_t^{-1}\big)D_k = I_{k^*}$, we have completed the proof for Theorem 3. □

C Additional estimation results

In this Appendix we consider a less parsimonious dynamic specification for the covariance matrix V_t: we allow variances and covariances to have different persistency levels. The empirical results do not suggest that the more general specification leads to improvements in terms of in-sample goodness-of-fit.

We consider matrices A and B in (18) to be diagonal matrices where the coefficients α_i and β_i corresponding to a conditional variance are set equal to α_v and β_v, respectively. The coefficients α_i and β_i corresponding to a conditional covariance are set equal to α_c and β_c, respectively. The matrices A and B can also be defined as $A = \text{diag}(\text{vech}(\tilde{A}))$ and $B = \text{diag}(\text{vech}(\tilde{B}))$. The matrix $\tilde{A}$ is a $k \times k$ matrix with diagonal elements equal to α_v and outer diagonal elements equal to α_c. Similarly, the matrix $\tilde{B}$ is a $k \times k$ matrix with diagonal elements equal to β_v and outer diagonal elements equal to β_c. This specification allows us to explore whether the variances and covariances have different dynamic properties.

We impose the additional parameter constraints $\alpha_v \geq \alpha_c \geq 0$ and $\beta_v - \alpha_v \geq \beta_c - \alpha_c \geq 0$ to ensure that V_t is positive definite with probability 1. These constraints can be easily obtained when we notice that the covariance matrix V_t can be expressed as

$$V_{t+1} = \text{E}[V_t](I_k - \tilde{B}) + (\tilde{B} - \tilde{A}) \odot V_t + \tilde{A} \odot \left(\frac{1}{v+1} \left(vX_t + \Lambda^{-1} r_t r_t' (\Lambda')^{-1} \right) \right),$$

where $\odot$ denotes the Hadamard product. Therefore we impose that $\tilde{B} - \tilde{A}$ and $\tilde{A}$ are positive definite, which leads to the parameter constraints as stated above. Imposing $\tilde{B} - \tilde{A}$ and $\tilde{A}$ to be positive definite also guarantees that V_t is positive definite by an application of the Schur product theorem.

We estimate the parameters for the 2×2 models of Table 3 and consider both the case where Λ is a full matrix and the case where Λ is a diagonal matrix. The results are reported in Table 8. The results suggest that the variances and covariances have the same dynamics, that is, $\alpha_v = \alpha_c$ and $\beta_v = \beta_c$. This can be concluded since the estimates of α_v and α_c, as well as β_v and β_c, are not significantly different from each other. Finally we notice that imposing $\alpha_v = \alpha_c$ and $\beta_v = \beta_c$ leads to the scalar models that are estimated in Table 3.

Equities	ν	β_v	β_c	α_v	α_c	λ_{11}	λ_{22}	λ_{12}	λ_{21}	$\log L$	AIC
2 × 2											
AA/CAT	12.428 (0.190)	0.977 (0.002)	0.977 (0.004)	0.331 (0.011)	0.331 (0.017)	0.893 (0.035)	1.022 (0.026)	0.226 (0.073)	-0.032 (0.051)	-20171.5	40360.9
AXP/PFE	10.876 (0.164)	0.991 (0.001)	0.991 (0.003)	0.378 (0.012)	0.378 (0.023)	1.032 (0.016)	0.918 (0.016)	-0.018 (0.030)	0.078 (0.021)	-17107.3	34232.5
AXP/WMT	11.909 (0.181)	0.993 (0.001)	0.993 (0.002)	0.360 (0.012)	0.360 (0.020)	1.018 (0.017)	0.887 (0.015)	0.032 (0.033)	0.025 (0.017)	-15347.7	30713.4
BA/HON	10.684 (0.161)	0.975 (0.002)	0.970 (0.005)	0.355 (0.011)	0.351 (0.019)	0.985 (0.028)	0.894 (0.029)	0.026 (0.053)	0.102 (0.055)	-17859.4	35736.9
CAT/KO	12.826 (0.196)	0.977 (0.002)	0.973 (0.006)	0.355 (0.011)	0.351 (0.022)	0.986 (0.022)	0.928 (0.017)	0.093 (0.074)	-0.037 (0.030)	-14226.8	28471.6
GE/PFE	11.016 (0.166)	0.984 (0.001)	0.984 (0.004)	0.405 (0.013)	0.405 (0.021)	0.943 (0.017)	0.911 (0.018)	0.016 (0.030)	0.072 (0.029)	-15622.7	31263.5
HD/JPM	12.458 (0.190)	0.988 (0.001)	0.988 (0.003)	0.447 (0.013)	0.447 (0.021)	0.953 (0.018)	0.944 (0.020)	0.020 (0.031)	0.125 (0.036)	-18481.1	36980.2
IBM/PG	12.407 (0.189)	0.977 (0.002)	0.974 (0.005)	0.383 (0.012)	0.381 (0.020)	0.985 (0.020)	0.866 (0.020)	-0.026 (0.052)	0.030 (0.036)	-10960.8	21939.7
JPM/XOM	13.067 (0.199)	0.989 (0.001)	0.984 (0.003)	0.444 (0.012)	0.440 (0.020)	0.988 (0.016)	0.928 (0.016)	0.006 (0.032)	0.036 (0.018)	-16081.5	32181.0
MCD/PG	10.432 (0.157)	0.978 (0.002)	0.972 (0.006)	0.311 (0.011)	0.305 (0.021)	0.919 (0.018)	0.880 (0.017)	0.037 (0.049)	0.015 (0.027)	-12645.5	25308.9
AA/CAT	12.424 (0.190)	0.977 (0.002)	0.977 (0.004)	0.333 (0.011)	0.333 (0.017)	0.952 (0.013)	0.978 (0.013)	-	-	-20201.6	40417.3
AXP/PFE	10.876 (0.164)	0.991 (0.001)	0.991 (0.003)	0.377 (0.012)	0.377 (0.022)	1.014 (0.014)	0.940 (0.013)	-	-	-17118.6	34251.2
AXP/WMT	11.908 (0.181)	0.993 (0.001)	0.992 (0.002)	0.360 (0.012)	0.360 (0.020)	1.016 (0.014)	0.890 (0.012)	-	-	-15353.2	30720.3
BA/HON	10.681 (0.161)	0.974 (0.002)	0.969 (0.005)	0.355 (0.011)	0.351 (0.019)	0.968 (0.013)	0.914 (0.012)	-	-	-17881.1	35776.3
CAT/KO	12.838 (0.196)	0.977 (0.002)	0.972 (0.006)	0.355 (0.011)	0.350 (0.022)	1.007 (0.014)	0.913 (0.013)	-	-	-14227.6	28469.3
GE/PFE	11.013 (0.166)	0.984 (0.002)	0.982 (0.004)	0.406 (0.013)	0.404 (0.020)	0.931 (0.013)	0.926 (0.013)	-	-	-15635.9	31285.9
HD/JPM	12.454 (0.190)	0.988 (0.001)	0.987 (0.003)	0.446 (0.013)	0.445 (0.021)	0.937 (0.013)	0.968 (0.013)	-	-	-18509.2	37032.3
IBM/PG	12.412 (0.189)	0.977 (0.002)	0.973 (0.005)	0.383 (0.012)	0.380 (0.020)	0.974 (0.014)	0.877 (0.012)	-	-	-10961.6	21937.0
JPM/XOM	13.085 (0.200)	0.989 (0.001)	0.985 (0.003)	0.442 (0.012)	0.439 (0.020)	0.979 (0.014)	0.939 (0.013)	-	-	-16085.9	32185.8
MCD/PG	10.426 (0.157)	0.979 (0.002)	0.974 (0.006)	0.311 (0.011)	0.307 (0.021)	0.921 (0.013)	0.879 (0.012)	-	-	-12648.8	25311.5

Table 8: Maximum likelihood estimates for the 2×2 models. Standard errors are shown in parentheses.