A DCC-type approach for realized covariance modeling with score-driven dynamics*

Danilo Vassallo[1], Giuseppe Buccheri[2], and Fulvio Corsi[3,4]

[1]Scuola Normale Superiore, Italy.

[2]University of Rome Tor Vergata, Italy.

[3]University of Pisa, Italy

[4]City University of London, UK

Abstract

We propose a class of score-driven realized covariance models where volatilities and correlations are separately estimated. We can thus combine univariate realized volatility models with a recently introduced class of score-driven realized covariance models based on Wishart and matrix-F distributions. Compared to the latter, the proposed models remain computationally simple at high dimensions and allow for higher flexibility in parameter estimation. Through a Monte-Carlo study, we show that the two-step maximum likelihood procedure provides accurate parameter estimates in small samples. Empirically, we find that the proposed models outperform those based on joint estimation, with forecasting gains that become more significant as the cross-section dimension increases.

Keywords: Realized Covariance; Dynamic Dependencies; Covariance forecasting; Score-driven models; Estimation errors

JEL codes: C58; D53; D81

*We are particularly grateful to participants to the 12th CFE conference in Pisa, the 2019 INET conference on score-driven and nonlinear time series models in Cambridge and the 2019 SoFiE conference in Shanghai.

1 Introduction

Covariance modeling and forecasting is a prominent topic in many financial applications. Since the seminal work of Andersen and Bollerslev (1998), the use of realized measures computed from high-frequency data has emerged as a preferential tool to forecast volatilities and correlations. Several univariate specifications for realized volatility have appeared in the econometric literature; see, among others, Engle and Gallo 2006, Corsi 2009, Shephard and Sheppard 2010, Hansen et al. 2012. More recently, a parallel interest arose towards multivariate models. Notably examples are given by Chiriac and Voev 2011, Bauer and Vorkink 2011, Bonato et al. 2012, Bauwens et al. 2012, Callot et al. 2017, Gorgi et al. 2018, Opschoor et al. 2017. The dynamic modeling of realized covariance poses a number of challenges that are absent in an univariate framework. The curse of dimensionality, i.e. the fact that the number of parameters rapidly increases with the cross-section dimension, might significantly affect the statistical efficiency of the estimation. Another complication is the need to guarantee positive-definite estimates, which is relevant in financial applications. Furthermore, estimation errors on realized measures play a crucial role in a multivariate framework, as they can lead to unstable solutions in portfolio optimization (Jagannathan and Ma 2003).

In this paper, we introduce a methodology aimed at addressing the above mentioned difficulties. Inspired by the DCC model of Engle (2002), we apply a two-step estimation procedure to a class of score-driven realized covariance models based on Wishart and matrix-F distributions. In the new models, the dynamics of *correlations* are driven by the score of the conditional density. Volatilities are instead estimated in a previous step through a sequence of univariate models. Such approach inherits the main advantages of the DCC. In particular, it allows to model separately volatilities and correlations, which possess different dynamics, and leaves more flexibility in the estimation of the model. Each univariate specification is estimated independently from the others, and it is therefore characterized by different parameters. In a similar fashion to the DCC, in the second step of the estimation, a set of restrictions on the parameter space can be imposed in such a way that the number of parameters scales well with the cross-section dimension. The new models are thus less affected by the typical curse of dimensionality problem of multivariate realized covariance models. Conditions guaranteeing positive-definite covariance estimates are easily derived.

Bauwens et al. (2012)[1] introduce a related DCC approach to realized covariance modeling based on the Wishart density. The proposed approach differs in the method used to update the time-varying parameters. In our framework, the update rule is determined by the score of the conditional density (Creal et al. 2013, Harvey 2013). In the case of the Wishart density, if the score is scaled through the inverse of the Fisher information matrix, our method leads to the same update rule of Bauwens et al. (2012). If a matrix-F density is instead used, we obtain a different update rule, which is robust to outliers. In our empirical application, we show that the model based on matrix-F density provides a better fit to the data and is superior in terms of out-of-sample forecast performance. It is worth noticing that the score of the conditional likelihood is by construction a martingale difference. For this reason, the score-driven update automatically provides a correction to the DCC model analogous to that introduced by Aielli (2013). In particular, it turns out that the specific correction of Aielli (2013) is recovered in the case of the Wishart density, whereas if the matrix-F

[1] See also their extensions in Bauwens et al. (2016) and Bauwens et al. (2017).

density is used, the correction assumes a different form.

Score-driven models are a large class of observation-driven models (Cox 1981) where the time-varying parameters are driven by the score of the conditional density. They have been successfully applied in the recent econometric literature. Notably examples are given by Creal et al. (2011), Creal et al. (2014) and Oh and Patton (2018). Gorgi et al. (2018) and Opschoor et al. (2017) introduce a score-driven specification for realized covariances based on Wishart and matrix-F distributions, respectively. In both cases, variances and correlations are estimated in a single step. In contrast, we estimate such models using the above mentioned two-step procedure. We show that this leads to significant forecast gains in common empirical applications as a result of the higher level of flexibility in parameter estimation. The choice of the score-driven approach as a starting point of our work is motivated by three main reasons. First, score-driven models provide a general methodology to update the time-varying parameters based on the full shape of the conditional density function. We thus obtain an update mechanism for volatilities and correlations which is consistent with the choice of the observation density. Second, such approach acknowledges the existence of measurement errors on realized covariance estimates. The latter are indeed modeled as noisy observations of the true latent covariances. The relevance of measurement errors on forecasting with realized measures has been underlined by Bollerslev et al. (2016), Bollerslev et al. (2018), Bekierman and Manner (2018) and Buccheri and Corsi (2019). In the empirical application, we compare our methodology to alternative methods that ignore estimation errors and show the resulting advantages in forecasting. Finally, in a score-driven framework, the likelihood can be written in closed form and can be optimized numerically. Estimation is thus feasible even when dealing with high dimensional matrices. General state-space models with latent covariances can only be estimated by simulation-based techniques, which become infeasible at high dimensions.

Our Monte-Carlo study has three main objectives. First, we investigate the finite sample properties of the maximum likelihood estimator that is employed in the second step of our procedure. We find that, as the number of observations increases, the estimator becomes unbiased and its distribution concentrates around the true parameters. Second, we simulate the covariances through a misspecified DGP and, according to the empirical evidence that realized covariances have fat-tails, we generate observations from a matrix-F distribution. We thus misspecify the measurement density of the score-driven Wishart model. We find that both models can capture the misspecified dynamics of the covariances. However, the relative performance of the matrix-F model is superior, as it is robust to the outliers generated by the fat-tailed measurement density. As a final experiment, we compare the two-step estimated covariances with those resulting from joint estimation. To mimic a realistic setting, variances are generated based on realized measures computed from real data. We clearly find that, both in-sample and out-of-sample, the two-step procedure provides significantly lower loss measures.

We illustrate the advantages of the proposed method in an empirical study involving two different datasets. The first dataset includes 1-second transactions of 100 NYSE stocks. The second includes 1-minute transaction data of 2767 stocks belonging to the Russell 3000 index. Score-driven realized covariance models are typically estimated by restricting the volatility persistences to be the same across assets. We first motivate the newly introduced class of models by performing several empirical tests which show that

such assumption is too restrictive on real data. We then examine the in-sample and out-of-sample forecast performance of the two-step models and compare them to models based on joint estimation and to other benchmarks. We select groups of 5, 10, 25, 50, 100 assets. The proposed two-step procedure provides significant in-sample and out-of-sample forecast gains compared to models based on joint estimation. In particular, the Model Confidence Set of Hansen et al. (2011) is effective in selecting the two-step Wishart and matrix-F models and excluding the remaining alternatives. Notably, forecast gains over joint estimation based models become more significant as the cross-section dimension increases, given the higher level of heterogeneity in the persistences of realized variances. Such results confirm that the additional flexibility provided by the two-step procedure translates into better forecasts. We also find that, in the vast majority of the scenarios examined in the analysis, the model based on the matrix-F density performs better than that based on the Wishart density. We thus confirm the result of Opschoor et al. (2017) that the matrix-F density is more suited to model time-series of realized covariances. Two additional empirical experiments are performed in order to assess the advantages of the two-step approach. In the first experiment, we show the behavior of the methodology as estimation errors on realized covariance measures become more severe. We find that the model forecasts are significantly less affected by the noise compared to alternative methods which ignore estimation errors. As a second test, we assess the *economic* gains of switching from joint estimation based models to the proposed two-step models. By adopting a utility based framework similar to that of Fleming et al. (2001), Fleming et al. (2003) and Bollerslev et al. (2018), we show that a risk-averse investor is willing to pay a positive annual amount to employ the forecasts of the two-step models in constructing her portfolio. Essentially, these economic gains are imputable to the lower ex-post risk featured by the portfolios constructed with the proposed methodology.

The rest of this paper is organized as follows: Section (2) introduces the two-step estimation procedure and reports the main results; Section (3) shows the results obtained through the Monte-Carlo analysis; Section (4) describes the empirical application and reports the main results; Section (5) concludes.

2 Framework

2.1 Score-driven models for realized covariance

Let $\{X_t\}_{t=1}^T$ denote a sequence of $k \times k$ positive-definite realized covariance matrices and let $\mathcal{F}_t = \sigma\left(X_s : s \leqslant t\right)$ be the σ-field generated by past observations of X_t. We assume that the conditional distribution function of X_t is given by:

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{G}_k\left(V_t\right) \tag{2.1}$$

where $\mathcal{G}_k$ is a matrix-variate distribution and $V_t \in \mathbb{R}^{k \times k}$ is a latent covariance matrix. Let us define $q = \frac{k(k+1)}{2}$ and let $f_t = \mathsf{vech}\left(V_t\right) \in \mathbb{R}^q$ collect the diagonal and upper diagonal elements of V_t. In the score-driven framework of Creal et al. (2013) and Harvey (2013), f_t is modeled as:

$$f_{t+1} = \omega + A s_t + B f_t, \tag{2.2}$$

where:

$$s_t = (\mathcal{I}_{t|t-1})^{-1}\nabla_t, \quad \nabla_t = \frac{\partial \log p_{\mathcal{G}_k}(X_t, f_t)}{\partial f_t}, \quad \mathcal{I}_{t|t-1} = \mathrm{E}_{t-1}[\nabla_t \nabla_t'] \tag{2.3}$$

Here, $p_{\mathcal{G}_k}(X_t, f_t)$ denotes the conditional density function associated with $\mathcal{G}_k$, ∇_t is the score computed with respect to f_t and $\mathcal{I}_{t|t-1}$ is the Fisher information matrix. The score-driven framework provides two main advantages. First, the parameters in f_t evolve based on the full shape of the conditional density function. Different choices of $p_{\mathcal{G}_k}(X_t, f_t)$ thus determine different update rules. Second, the likelihood can be written in closed form and estimation can be performed by standard numerical optimization.

Two different specifications for the conditional density $\mathcal{G}_k$ have been proposed in the literature. Gorgi et al. (2018) set:

$$X_t|\mathcal{F}_{t-1} \sim W_k\left(V_t/\nu, \nu\right) \tag{2.4}$$

where $W_k\left(V_t/\nu, \nu\right)$ is the k-variate Wishart distribution with $\nu \geq k$ degrees of freedom. The corresponding density function is given by:

$$p_{W_k}\left(X_t; V_t/\nu, \nu\right) = \frac{|X_t|^{\frac{\nu-k-1}{2}}}{2^{\nu k/2}\nu^{-(\nu k)/2}|V_t|^{\nu/2}\Gamma_k\left(\frac{\nu}{2}\right)} \exp\left[-\frac{\nu}{2}\mathrm{tr}\left(V_t^{-1}X_t\right)\right], \tag{2.5}$$

where $\Gamma_k(\cdot)$ denotes the k-variate Gamma function. In this parameterization, the conditional mean is given by $\mathrm{E}_{t-1}[X_t] = V_t$. A similar density is employed by Golosnoy et al. (2012), Gourieroux et al. (2009) and Bonato et al. (2012), who introduce autoregressive processes for realized covariances based on the Wishart distribution.

The dynamics of realized covariances are often characterized by fat-tails and jumps. The Wishart distribution is not able to reproduce such features and, as a consequence, covariance estimates might be too sensitive to large movements. To obtain robust estimates, Opschoor et al. (2017) propose to model X_t through the matrix-F distribution:

$$X_t|\mathcal{F}_{t-1} \sim F_k\left(V_t, \nu_1, \nu_2\right), \tag{2.6}$$

where $\nu_1, \nu_2 \geq k$ are degrees of freedom. The density function of the matrix-F distribution is given by:

$$p_{F_k}(X_t; V_t, \nu_1, \nu_2) = K(\nu_1, \nu_2)\frac{\left|\frac{\nu_1}{\nu_2-k-1}V_t^{-1}\right|^{\nu_1/2}|X_t|^{\frac{\nu_1-k-1}{2}}}{\left|\mathbb{1}_k + \frac{\nu_1}{\nu_2-k-1}V_t^{-1}X_t\right|^{\frac{\nu_1+\nu_2}{2}}} \tag{2.7}$$

where $\mathbb{1}_k$ denotes the $k \times k$ identity matrix and:

$$K(\nu_1, \nu_2) = \frac{\Gamma_k\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma_k\left(\frac{\nu_1}{2}\right)\Gamma_k\left(\frac{\nu_2}{2}\right)} \tag{2.8}$$

The $k \times k$ positive-definite matrix V_t turns out to be the conditional mean of the matrix-F distribution. The Wishart distribution can be obtained as a limit from the matrix-F when ν_2 goes to infinity. If ν_2 is finite, the matrix-F distribution exhibits fat-tail behavior and it is thus more suited to describe the dynamics of realized measures. Indeed, the score computed from the matrix-F density exhibits a nonlinear

structure which underweights the impact of outliers on the dynamics of the covariances (see also discussions in Opschoor et al. 2017).

The main shortcoming of such modeling approach is that the dimension of f_t grows quadratically with k. Estimation is thus feasible only by imposing strong restrictions on the structure of the matrices A, B. For instance, Gorgi et al. (2018) and Opschoor et al. (2017) impose a scalar structure, namely they set $A = \alpha \mathbb{1}_q$ and $B = \beta \mathbb{1}_q$, with $\alpha, \beta \in \mathbb{R}$. Another drawback is that long-memory effects are not directly taken into account. To this end, Opschoor et al. (2017) impose a HAR specification in the dynamic in Eq. (2.2). However, the additional parameters are estimated under the same scalar restrictions. Our objective is to relax such restrictions while maintaining the model computationally simple to estimate.

2.2 DCC-type score-driven models for realized covariance

Inspired by Engle (2002), we write the covariance matrix V_t as:

$$V_t = D_t R_t D_t \tag{2.9}$$

where D_t is a diagonal matrix of standard deviations and R_t is a correlation matrix. We propose to estimate D_t and R_t in two steps. In the first step, the individual standard deviations are separately estimated by a sequence of k realized volatility models. We allow for complete flexibility in the choice of the univariate specification. In our applications, we choose the univariate density by setting $k = 1$ in Eq. (2.5) and (2.7). In the case of the Wishart density, we obtain, for $i = 1, \dots, k$:

$$x_t^{(i)} | \mathcal{F}_{t-1} \sim W_1 \left(\frac{v_t^{(i)}}{\nu^{(i)}}, \nu^{(i)} \right) \sim \frac{v_t^{(i)}}{\nu^{(i)}} \chi^2_{\nu^{(i)}} \tag{2.10}$$

where $x_t^{(i)}$ and $v_t^{(i)}$ denote the i-th diagonal element of X_t and V_t, respectively, and $\chi^2_{\nu^{(i)}}$ is a chi-squared distribution with $\nu^{(i)}$ degrees of freedom. In the case of the matrix-F density, we obtain, for $i = 1, \dots, k$:

$$x_t^{(i)} | \mathcal{F}_{t-1} \sim F_1 \left(v_t^{(i)}, \nu_1^{(i)}, \nu_2^{(i)} \right) \sim \frac{\nu_2^{(i)} - 2}{\nu_2^{(i)}} v_t^{(i)} F_{\nu_1^{(i)}, \nu_2^{(i)}} \tag{2.11}$$

where $F_{\nu_1^{(i)}, \nu_2^{(i)}}$ denotes the scalar F density with degrees of freedom $\nu_1^{(i)}$ and $\nu_2^{(i)}$. To guarantee positive variance estimates, we set:

$$v_t^{(i)} = e^{\lambda_t^{(i)}} \tag{2.12}$$

and model the log-variance $\lambda_t^{(i)}$ through the score of the conditional likelihood. Specifically, the dynamic of $\lambda_t^{(i)}$ is given by:

$$\lambda_{t+1}^{(i)} = \omega_i + \alpha_i s_t^{(i)} + \beta_i^{(d)} \lambda_t^{(i)} + \beta_i^{(w)} \lambda_{t-1|t-5}^{(i)} + \beta_i^{(m)} \lambda_{t-6|t-22}^{(i)} \tag{2.13}$$

where $\lambda_{t-m|t-n}^{(i)} = \frac{1}{n-m+1} \sum_{j=t-n}^{t-m} \lambda_j^{(i)}$ and $s_t^{(i)}$ denotes the scaled score of the i-th univariate density. The expression of $s_t^{(i)}$ depends on the choice of the probability density function and is computed in Appendix A. The HAR-like structure allows to parsimoniously capture the strong persistence observed on realized

variance. After estimating the k univariate models, we compute the diagonal elements of the matrix D_t as:

$$D_t^{(i)} = \exp\left[\frac{\lambda_t^{(i)}}{2}\right] \tag{2.14}$$

Thus, in the first step, k univariate score-driven models are independently estimated and their forecasts are used to build the matrix D_t of standard deviations. This operation is extremely simple from a computational viewpoint, and at the same time allows to have different parameters $\omega_i, \beta_i^{(d)}, \beta_i^{(w)}, \beta_i^{(m)}, \nu^{(i)}, \nu_1^{(i)}, \nu_2^{(i)}$ for different assets.

To model the correlations, as in Engle (2002), we introduce the matrix Q_t such that:

$$R_t = \Delta_t^{-1} Q_t \Delta_t^{-1} \tag{2.15}$$

where $\Delta_t = \mathsf{diag}\left(Q_t\right)^{1/2}$. We then model dynamically $f_t = \mathsf{vech}(Q_t)$ based on the score of the conditional density:

$$f_{t+1} = \omega + A s_t + B f_t \tag{2.16}$$

where the scaled score s_t is computed by assuming D_t known and given by Eq. (2.14). In the next two subsections, we compute the expression of s_t for both the Wishart and the matrix-F densities.

2.2.1 Wishart

The log-density function associated with the density in Eq. (2.5) is:

$$\log p_{W_k}\left(X_t; V_t, \nu\right) = \frac{1}{2} d_X\left(k, \nu\right) + \frac{\nu - k - 1}{2} \log\left|X_t\right| - \frac{\nu}{2} \log\left|V_t\right| - \frac{\nu}{2} \mathsf{tr}\left(V_t^{-1} X_t\right) \tag{2.17}$$

where $d_X\left(k, \nu\right) = \nu k \log\left(\nu/2\right) - 2\log\Gamma_k\left(\nu/2\right)$. In Appendix B, C we prove the following two results:

Proposition 2.1. *For the density in Eq. (2.17), the score* $\nabla_t^W = \frac{\partial \log p_{W_k}(X_t; f_t, \nu)}{\partial f_t}$ *is given by:*

$$\nabla_t^W = \frac{\nu}{2} \mathcal{D}_k' \Psi_t' \left(D_t^{-1} \Delta_t Q_t^{-1} \otimes D_t^{-1} \Delta_t Q_t^{-1}\right) \left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right] \tag{2.18}$$

where $\mathcal{D}_k$ *denotes the duplication matrix,* $\Psi_t = \mathbb{1}_{k^2} - \left(\Delta_t^{-1} Q_t \otimes \mathbb{1}_k + \mathbb{1}_k \otimes \Delta_t^{-1} Q_t\right) W_Q$ *and* W_Q *is a sparse* $k^2 \times k^2$ *diagonal matrix defined in Appendix B.*

Proposition 2.2. *For the density in Eq. (2.17), the Fisher information matrix* $\mathcal{I}_{t|t-1}^W = \mathrm{E}_{t-1}[\nabla_t^W \nabla_t^{W\prime}]$ *is given by:*

$$\mathcal{I}_{t|t-1}^W = \frac{\nu}{2} \mathcal{D}_k{}' \Psi_t' \left(H_t^{-1} Q_t^{-1} H_t \otimes H_t^{-1} Q_t^{-1} H_t\right) \mathcal{D}_k \mathcal{D}_k{}^+ \Psi_t \mathcal{D}_k \tag{2.19}$$

where $\mathcal{D}_t^+$ *denotes the elimination matrix and* $H_t = D_t \Delta_t^{-1}$.

Taking the inverse of $\mathcal{I}_{t|t-1}^W$, as required by Eq. (2.2), poses two problems. First, $\mathcal{I}_{t|t-1}^W$ is singular, as f_t includes $k(k+1)/2$ time-varying parameters whereas the number of time-varying parameters in R_t is $k(k-1)/2$. Second, for $k \gg 1$, matrix pseudo-inversion is computationally cumbersome and can lead to

numerical instabilities. We solve both problems by setting $\mathcal{I}_t^W = \frac{\nu}{2}\mathcal{D}_k{}'\Psi_t'\left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k$. This is equivalent to approximate the second Ψ_t factor appearing in Eq. (2.19) as $\Psi_t \approx \mathbb{1}_{k^2}$ (note that $\mathcal{D}_k^+\mathcal{D}_k = \mathbb{1}_k$), thus neglecting the term $\left(\Delta_t^{-1}Q_t \otimes \mathbb{1}_k + \mathbb{1}_k \otimes \Delta_t^{-1}Q_t\right)W_Q$. The latter has indeed a tiny influence on Ψ_t, as W_Q is a very sparse $k^2 \times k^2$ matrix with only k nonzero elements on the main diagonal. By assuming[2] such expression for $\mathcal{I}_{t|t-1}^W$, not only the latter becomes non-singular, but one can also compute in closed form the product $(\mathcal{I}_{t|t-1}^W)^{-1}\nabla_t^W$ appearing in Eq. (2.2). In particular, in Appendix D we prove the following:

Proposition 2.3. *For the density in Eq. (2.17), the scaled score vector $s_t^W = (\mathcal{I}_{t|t-1}^W)^{-1}\nabla_t^W$ is given by:*

$$s_t^W = \mathsf{vech}(H_t^{-1}X_tH_t^{-1}) - \mathsf{vech}\,(Q_t) \tag{2.20}$$

This means that, in order to update the time-varying parameter f_t, one only needs to compute the scaled score s_t, which is a q-dimensional vector having the simple expression given by Eq. (2.20). Correlations are obtained by updating the time-varying parameter f_t using Eq. (2.16) and then constructing the matrix R_t through Eq. (2.15). Note that imposing a scalar structure on parameters A, B, namely $A = \alpha\mathbb{1}_q$, $B = \beta\mathbb{1}_q$ is now less restrictive, as variances are estimated independently and have different persistences. The correlation matrix R_t is positive-definite if and only if Q_t is positive-definite. If a scalar structure is imposed, one has:

$$\begin{aligned}\mathsf{vech}\,(Q_{t+1}) =& \omega + \beta\mathsf{vech}\,(Q_t) + \alpha\mathsf{vech}(H_t^{-1}X_tH_t^{-1}) - \alpha\mathsf{vech}\,(Q_t) &(2.21)\\ =& \omega + (\beta-\alpha)\,\mathsf{vech}\,(Q_t) + \alpha\mathsf{vech}(H_t^{-1}X_tH_t^{-1}), &(2.22)\end{aligned}$$

which is positive-definite for all t's if ω is positive-definite and if $\alpha \geqslant 0$, $\beta - \alpha \geqslant 0$. Note that the update rule in Eq. (2.22) coincides with the Re-cDCC specification of Bauwens et al. (2012) which adopt the Aielli (2013) correction in the context of a DCC model with realized covariance.

2.2.2 Matrix-F

The log-density function in Eq. (2.7) is given by:

$$\log p_{F_k}\left(X_t; V_t, \nu_1, \nu_2\right) = d(\nu_1,\nu_2) + \frac{\nu_1-k-1}{2}\log|X_t| - \frac{\nu_1}{2}\log|V_t| - \frac{\nu_1+\nu_2}{2}\log|\tilde{W}_t| \tag{2.23}$$

where:

$$\tilde{W}_t = \mathbb{1}_k + \frac{\nu_1}{\nu_2-k-1}V_t^{-1}X_t \tag{2.24}$$

$$d(\nu_1,\nu_2) = \frac{\nu_1}{2}\log\left(\frac{\nu_1}{\nu_2-k-1}\right) + \log\Gamma_k\left(\frac{\nu_1+\nu_2}{2}\right) - \log\Gamma_k\left(\frac{\nu_1}{2}\right) - \log\Gamma_k\left(\frac{\nu_2}{2}\right) \tag{2.25}$$

[2]Such approximation is harmless from a dynamic modeling perspective. Indeed, the scaling matrix in score-driven models is typically related to the inverse of the Fisher information matrix but is not necessarily equal to the latter. For instance, different powers of the inverse can be chosen. In this case, our approximating expression turns out to be very close to the full Information matrix and very similar results are obtained by explicitly computing the pseudo-inverse. However, the latter operation is computationally unstable and becomes infeasible at large dimensions.

For more details on the Matrix-F distribution see Gupta and Nagar (1999). In Appendix E, we prove the following:

Proposition 2.4. *For the density in Eq. (2.23), the score* $\nabla_t^F = \frac{\partial \log p_{F_k}(X_t; f_t, \nu_1, \nu_2)}{\partial f_t}$ *is given by*

$$\nabla_t^F = \frac{\nu_1}{2} \mathcal{D}_k' \Psi_t' \left(H_t^{-1} Q_t^{-1} \otimes H_t^{-1} Q_t^{-1} \right) \left[\frac{\nu_1 + \nu_2}{\nu_2 - k - 1} \mathsf{vec} \left(X_t \tilde{W}_t^{-1} \right) - \mathsf{vec} \left(V_t \right) \right] \tag{2.26}$$

As in Opschoor et al. (2017), to take into account the curvature of the log-density, we scale the score through the inverse of the Fisher information matrix of the Wishart density. This has two main advantages. First, it avoids the numerical computation of the Fisher information matrix of the matrix-F distribution, which is not available in closed form. Second, as in the Wishart density, we can compute in closed form the product $(\mathcal{I}_{t|t-1}^W)^{-1} \nabla_t^F$. Indeed, in Appendix F we prove the following:

Proposition 2.5. *The scaled score vector* $s_t^F = (\mathcal{I}_{t|t-1}^W)^{-1} \nabla_t^F$ *is given by:*

$$s_t^F = \frac{\nu_1 + \nu_2}{\nu_2 - k - 1} \mathsf{vech} \left(H_t^{-1} X_t \tilde{W}_t^{-1} H_t^{-1} \right) - \mathsf{vech} \left(Q_t \right) \tag{2.27}$$

As in the Wishart density, the dynamic modeling of f_t solely requires the computation of the scaled score s_t, which is a q-dimensional vector given by Eq. (2.27). By imposing a scalar structure on parameters A, B, we have:

$$\begin{aligned}
\mathsf{vech} \left(Q_{t+1} \right) =& \omega + \beta \mathsf{vech} \left(Q_t \right) + \alpha s_t && (2.28) \\
=& \omega + \beta \mathsf{vech} \left(Q_t \right) + \alpha \frac{\nu_1 + \nu_2}{\nu_2 - k - 1} \mathsf{vech} \left(H_t^{-1} X_t \tilde{W}_t^{-1} H_t^{-1} \right) - \alpha \mathsf{vech} \left(Q_t \right) && (2.29) \\
=& \omega + (\beta - \alpha) \mathsf{vech} \left(Q_t \right) + \alpha \frac{\nu_1 + \nu_2}{\nu_2 - k - 1} \mathsf{vech} \left(H_t^{-1} X_t \tilde{W}_t^{-1} H_t^{-1} \right), && (2.30)
\end{aligned}$$

which is positive-definite for all t's if ω is positive-definite and if $\alpha \geqslant 0$, $\beta - \alpha \geqslant 0$. As in the previous case, the score-driven update mechanism includes by construction a correction based on the matrix H_t. Note that, with this choice for the scaling matrix, the scaled score of the matrix-F density converges to the scaled score of the Wishart density as $\nu_2 \to \infty$:

Proposition 2.6. *When* $\nu_2 \to \infty$, *we get* $s_t^F = s_t^W$

The proof is in Appendix G.

In principle, a HAR structure similar to that in Eq. (2.13) might be imposed even in the update rule for the correlations. As it will be extensively discussed in the empirical application in Section (4), a similar structure does not lead to any out-of-sample forecast gain. We thus maintain the current specification with a HAR structure in the variances and a simple AR(1) score-driven update for the correlations.

We conclude this section by noticing that both Gorgi et al. (2018) and Opschoor et al. (2017) consider a larger filtration also including daily returns. Under the assumption of conditional independence between realized covariances and daily returns, this translates into an additional term in the score of the two models. Our DCC-type approach is readily generalizable to include the effect of daily returns. However, as the latter

have less explanatory power, and given that the aim of the present work is to highlight the advantages of the two-step approach, we neglect them for simplicity.

2.3 Estimation

The two-step procedure can be summarized as follows:

1. Maximum likelihood estimation of the k univariate models described in Section (2.2) (or alternative univariate specifications), from which the matrices D_t are built
2. Maximum likelihood estimation of the Wishart log-density in Eq. (2.17) or the matrix-F log-density in Eq. (2.23) assuming D_t known and f_t evolving according to Eq. (2.16)

In a similar fashion to the DCC model, the static parameters A, B governing the dynamics of correlations are estimated under the scalar restrictions $A = \alpha \mathbb{1}_q$, $B = \beta \mathbb{1}_q$. In contrast, ω is a q-dimensional vector with different elements. We estimate it through (co)variance targeting, i.e. we set:

$$\omega = (\mathbb{1}_q - B)\, \mathsf{vech}(\overline{R}) \tag{2.31}$$

where $\overline{R}$ is the correlation matrix computed from the sample mean of realized covariances, namely $\overline{Q} = \frac{1}{T}\sum_{t=1}^{T} X_t$.

3 Monte-Carlo analysis

3.1 Finite sample properties

In this section, we study the finite sample properties of the two-step estimator via Monte-Carlo simulations. We consider both the two-step Wishart and the two-step matrix-F models. For easy of notation, we indicate the two models by "2-step-W" and "2-step-F". We set $D_t = \mathbb{1}_k$ for every t, i.e. we assume constant variances. Similar results are recovered when employing dynamic variances. We choose different values of k and T to test the properties of the estimator on a variety of different scenarios. Specifically, we set $k = \{2, 5, 10\}$ and $T = \{250, 500, 1000\}$. For each scenario, the model is simulated and estimated 1000 times. The two-step procedure is applied as described in Section (2.3). The initial parameter f_1 is set equal the correlation matrix computed from the sample mean of the simulated covariances.

We consider a specification for the Wishart model given by:

$$\alpha = 0.3, \quad \beta = 0.97, \quad \omega = (1-\beta)\mathsf{vech}\left[C(\rho)\right], \quad \nu = k + 10.$$

where $C(\rho)$ denotes the equi-correlation matrix with correlation parameter equal to ρ. We consider two different values for the correlation, namely $\rho = 0.5$ and $\rho = 0.9$. Figures (1), (2) show kernel density estimates of the probability density function of parameters in each scenario. The distribution concentrates around the true values as the time-series length T and/or the cross-section dimension k increase. This is due

to the scalar specification adopted for the correlations. The distribution of α and ν is symmetric, whereas the distribution of β is slightly skewed at $T = 250$ and exhibits a bias. However, note that for $T = 500$ and $T = 1000$ all parameters are estimated accurately.

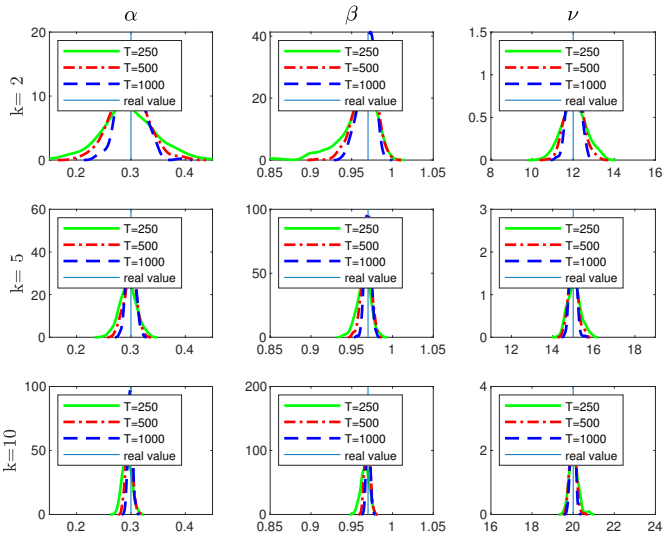

Figure 1: Kernel density estimates of maximum-likelihood estimates based on $N = 1000$ simulations of the two-step Wishart model with equi-correlation parameter $\rho = 0.5$.

We perform the same analysis for the two-step model based on matrix-F distribution. Parameters are set as follows:

$$\alpha = 0.3, \quad \beta = 0.97, \quad \omega = (1-\beta)\mathsf{vech}\left[C(\rho)\right], \quad \nu_1 = 22, \quad \nu_2 = 35.$$

where $\rho = \{0.5, 0.9\}$. Figures (3), (4) show the results. We find again that parameters are accurately estimated in small samples and tend to concentrate around the true values as the time-series length T and/or the cross-section dimension k increase.

3.2 Analysis based on a misspecified DGP

We test the performance of the two-step procedure under a misspecified data generating process for volatilities and correlations. For simplicity, we consider a bivariate model, in a similar fashion to Engle (2002) and Creal et al. (2011). In Section (4), we show that the matrix-F distribution provides a better fit to empirical data.

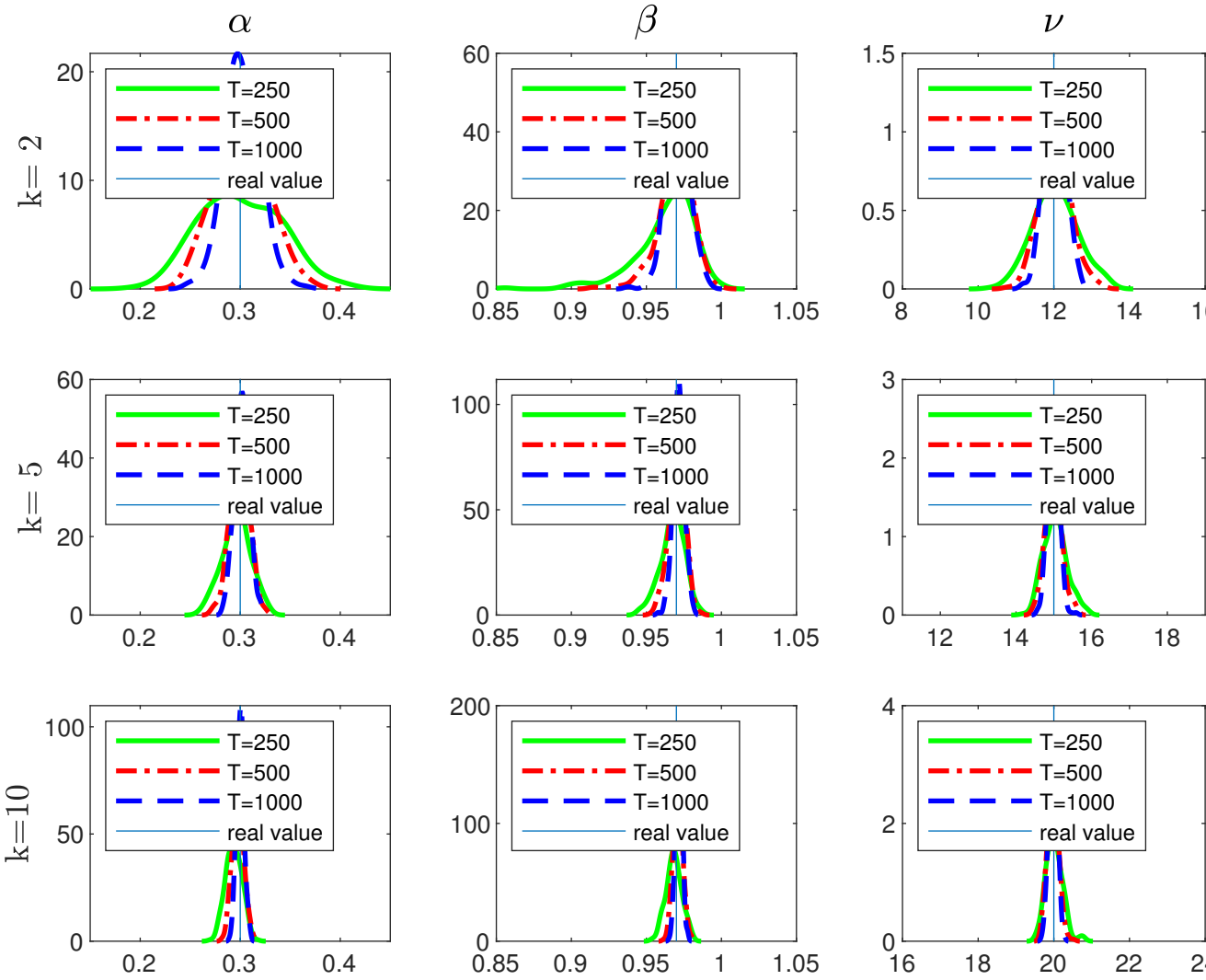

Figure 2: Kernel density estimates of maximum-likelihood estimates based on $N = 1000$ simulations of the two-step Wishart model with equi-correlation parameter $\rho = 0.9$.

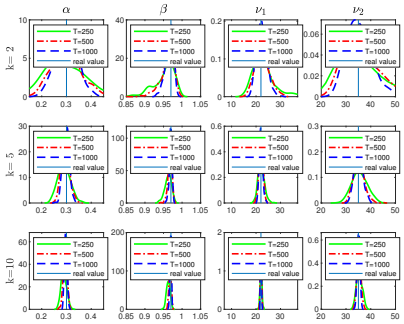

Figure 3: Kernel density estimates of maximum-likelihood estimates based on $N = 1000$ simulations of the two-step matrix-F model with equi-correlation parameter $\rho = 0.5$.

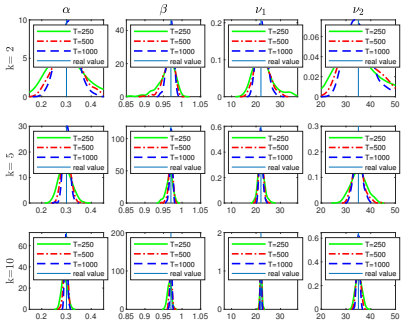

Figure 4: Kernel density estimates of maximum-likelihood estimates based on $N = 1000$ simulations of the two-step matrix-F model with equi-correlation parameter $\rho = 0.9$.

We therefore employ it as a conditional density to generate observations:

$$X_t|\mathcal{F}_{t-1} \sim F_2(V_t, \nu_1, \nu_2) \tag{3.1}$$

where we set $\nu_1 = 20$ and $\nu_2 = 8$. The latent covariance matrix V_t is written as:

$$V_t = \begin{pmatrix} c_1^2 & c_1 c_2 \rho_t \\ c_1 c_2 \rho_t & c_2^2 \end{pmatrix}$$

The misspecified volatilities c_1 and c_2 are modeled as:

$$c_{i,t} = \exp[0.97 \log(c_{i,t-1}) + \epsilon_t^{(i)}]$$

where $\epsilon_t^{(i)} \sim \mathcal{N}(0, 0.1)$, $i = 1, 2$. The correlation ρ_t evolves over time based on different dynamic patterns. They are given by:

$$\text{Sine:} \qquad \rho_t^{(1)} = 0.5 + 0.4 \sin(2\pi t/500) \tag{3.2}$$

$$\text{Fast sine:} \qquad \rho_t^{(2)} = 0.5 + 0.4 \sin(2\pi t/125) \tag{3.3}$$

$$\text{Step:} \qquad \rho_t^{(3)} = 0.9 - 0.4H\left(t - \frac{T}{4}\right) + 0.4H\left(t - \frac{T}{2}\right) - 0.4H\left(t - \frac{3T}{4}\right) \tag{3.4}$$

$$\text{Ramp:} \qquad \rho_t^{(4)} = \sum_{i=0}^{3} \left(\frac{4 \times 0.9}{T} t - 0.9i\right) \left[H\left(t - \frac{iT}{4}\right) - H\left(t - \frac{(i+1)\,T}{4}\right)\right] \tag{3.5}$$

where $T = 2000$ and H is the Heaviside step function. For each dynamic pattern, we generate $N = 1000$ time-series of realized covariances through the observation density in Eq. (3.1). We estimate both the 2-step-W and 2-step-F on the subsample comprising the first 1000 observations. In-sample loss measures are computed in this subsample. We then use the recovered parameter estimates to filter the covariances in the subsample comprising the last 1000 observations. Figures (8)-(11), reported in Appendix H, show the simulated patterns and the average filtered correlations of 2-step-W and 2-step-F. Both models are able to capture the pattern of correlations accurately. Note that the confidence bands of 2-step-W filtered estimates are slightly larger than those of the 2-step-F model. The motivation is that the 2-step-W is not robust to the outliers generated by the matrix-F density.

Pattern	In-sample		Out-of-sample	
	2-step-W	2-step-F	2-step-W	2-step-F
Sine	1.000	0.736	1.000	0.718
Fast sine	1.000	0.846	1.000	0.807
Step	1.000	0.933	1.000	0.891
Ramp	1.000	0.937	1.000	0.937

Table 1: Relative in-sample and out-of-sample RMSE of two-step-W and two-step-F for the all the correlation patterns.

Table (1) reports relative in-sample and out-of-sample average Frobenius norms, defined as:

$$\text{Frob}\left(\hat{V}_t, V_t\right) = \sqrt{\mathsf{Tr}[(V_t - \hat{V}_t)(V_t - \hat{V}_t)']} \tag{3.6}$$

Hereafter, we refer to the Frobenius norm as RMSE, as it coincides with the square root of the sum of the MSE's of the entries of $\hat{V}_t$. The 2-step-F provides much lower in-sample and out-of-sample RMSE in patterns "sine" and "fast sine". In patterns "step" and "ramp", the 2-step-F performs better, though with a lower relative difference. This is due to the discontinuity in the correlation pattern, which is treated as an outlier by the 2-step-F model. As a consequence, it adapts more slowly to the abrupt change of correlation.

3.3 Comparison with joint estimation

We compare now the performance of the two-step procedure to that of standard joint estimation. In order to generate variance paths that have the same dynamic behavior of real data, we construct the matrices D_t based on a dataset of 100 realized variance time-series computed from stocks traded on the NYSE. The dataset is described in more detail in Section (4.1). In each simulation, the matrices D_t are generated by randomly selecting k realized variance series from the dataset and replacing each diagonal element[3] $D_t^{(i)}$ with the true realized volatility $(X_t^{(i)})^{1/2}$:

$$D_t = \begin{pmatrix} (X_t^{(1)})^{1/2} & & 0 \\ & \ddots & \\ 0 & & (X_t^{(k)})^{1/2} \end{pmatrix}$$

To simulate the correlation matrix R_t, we use the equicorrelation model of Engle and Kelly (2012). It has the form:

$$R_t = \rho_t I_k + [1 - \rho_t] J_{k,k} \tag{3.7}$$

where $J_{k,k} \in \mathbb{R}^{k \times k}$ is a matrix of ones. The correlation parameter ρ_t evolves based on the following process:

$$\rho_t = 1 - \frac{1}{k-1}\left[1 - \tanh\left(\theta_t\right)\right] \tag{3.8}$$

$$\theta_{t+1} = \phi\theta_t + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}\left(0, \sigma^2\right) \tag{3.9}$$

where we set $\sigma^2 = 0.25$. We then compute the covariance matrix as $V_t = D_t R_t D_t$ and generate realized covariances through both the Wishart and the matrix-F density:

$$X_t^W \sim W_k\left(V_t/\nu, \nu\right) \tag{3.10}$$

$$X_t^F \sim F_k\left(V_t, \nu_1, \nu_2\right) \tag{3.11}$$

with $\nu = k + 10$, $\nu_1 = 140$ and $\nu_2 = 120$. These values are similar to those estimated on real data.

[3] As D_t is the matrix of latent standard deviations, we apply an EWMA smoothing scheme to clean the real time-series from measurement error effects.

We set $k = 10$ and perform $N = 500$ simulations of realized covariance series $\{X_t^W\}$, $\{X_t^F\}$ of $T = 2000$ observations. As in the previous simulation study, in-sample loss measures are computed in the subsample comprising the first 1000 observations. Out-of-sample loss-measures are instead computed in the remaining subsample of 1000 observations based on the parameters estimated in the first subsample. As loss measures, we consider the RMSE and the Quasi-likelihood (Qlike). The latter is defined as:

$$\text{Qlike}\left(\hat{V}_t, V_t\right) = -\log|\hat{V}_t^{-1}V_t| + \mathsf{Tr}\left(\hat{V}_t^{-1}V_t\right) - k \tag{3.12}$$

Table (2) compares in-sample and out-of-sample loss measures of two-step and joint estimation for both Wishart and matrix-F models. The joint estimation approach coincides with the models of Gorgi et al. (2018) and Opschoor et al. (2017). Both in-sample and out-of-sample, the two-step estimator performs

Loss	Model	Wishart		matrix-F	
		In-sample	Out-of-sample	In-sample	Out-of-sample
RMSE	two-step	1.000	1.000	1.000	1.000
		(420)	(395)	(475)	(500)
	joint	1.118	1.092	1.154	1.189
		(15)	(20)	(0)	(0)
Qlike	two-step	1.000	1.000	1.000	1.000
		(460)	(430)	(400)	(400)
	joint	1.096	1.086	1.213	1.141
		(15)	(20)	(0)	(0)

Table 2: Relative in-sample and out-of-sample RMSE and Qlike of two-step models compared with joint estimation models. We report in parenthesis the number of times the corresponding loss measure is judged as significantly smaller by the Diebold-Mariano test.

significantly better than the joint estimator, according to both loss measures. To test the significance of our results, we perform the Diebold-Mariano test (Diebold and Mariano 2002) at 5% significance level for each simulation. We report in the table the number of times each loss measure is judged as significantly smaller by the Diebold-Mariano test. It clearly emerges that, as a result of the higher flexibility provided by the two-step models, the latter provide significantly smaller loss measures in most of the $N = 500$ simulations.

4 Empirical evidence

4.1 Dataset

The empirical analysis is performed on two different datasets. The first dataset consists of transactions of 100 assets traded on the NYSE. The time resolution is 1-second. Data are available from 03-01-2006 to 31-12-2014, corresponding to 2265 business days. The second dataset consists of transactions of 2767 assets belonging to the Russell 3000 index. The time resolution is 1-minute and data are available from

03-01-2006 to 27-09-2013, corresponding to 1948 business days. As a realized covariance estimator, we use the multivariate Realized Kernel of Barndorff-Nielsen et al. (2011). Before computing it, we perform the standard cleaning procedures described by Barndorff-Nielsen et al. (2009).

4.2 Preliminary analysis

The main advantage of the proposed models is the higher flexibility in the estimation of the static parameters. Gorgi et al. (2018) and Opschoor et al. (2017) use the scalar restrictions $A = \alpha \mathbb{1}_q$ and $B = \beta \mathbb{1}_q$, i.e. they assume the same persistence for the volatilities and correlations of all the assets. In contrast, we find empirical evidences supporting the hypothesis of a high level of heterogeneity among the persistences of volatilities. Such hypothesis is tested by means of two different experiments.

The first experiment is performed using the dataset of 2767 assets of the Russell 3000 index. The results, presented in Figures (5), (6), (7), provide a first evidence that the persistences of volatilities can vary a lot among assets. The blue lines represent the (sorted) 2767 maximum likelihood estimates of the coefficients $\beta^{(d)}$, $\beta^{(w)}$ and $\beta^{(m)}$ appearing in Eq. (2.13). The red lines represent 2767 (sorted) realizations of $\beta^{(d)}$, $\beta^{(w)}$ and $\beta^{(m)}$ given the null hypothesis that the different assets have the same HAR coefficients. They are built using the following method: we simulate 2767×1000 random variables distributed as $\mathcal{N}\left(\bar{\beta}^{(j)}, \bar{\sigma^2}^{(j)}\right)$, $j = d, w, m$, where $\bar{\beta}^{(j)}$ represents the sample mean of the estimated $\beta^{(j)}$ coefficient across the 2767 assets, and $\bar{\sigma^2}^{(j)}$ represents the sample mean of the estimation error variances of the $\beta^{(j)}$ coefficient (computed as the inverse of the hessian at the maximum). We compute the blue line by averaging the sorted 2767 realizations over the 1000 replications. The 95% confidence bands are computed as empirical quantiles of the simulated distribution. They show that the real estimated coefficients are incompatible with the null assumption that volatility persistences are equal.

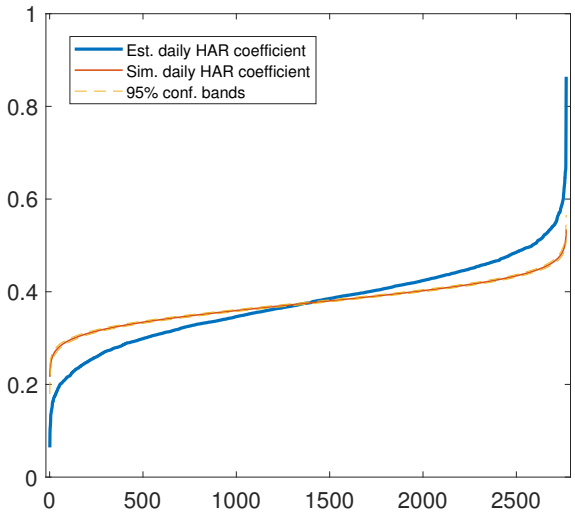

Figure 5: We report in blue the sorted maximum likelihood estimates of the $\beta^{(d)}$ coefficient corresponding to the 2767 assets in the Russel 3000 dataset. The red line and the 95% confidence bands are computed under the null hypothesis that the 2767 assets have the same $\beta^{(d)}$ coefficient.

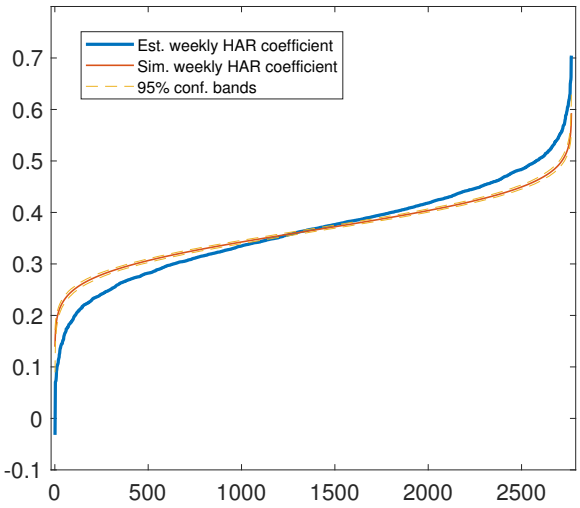

Figure 6: We report in blue the sorted maximum likelihood estimates of the $\beta^{(w)}$ coefficient corresponding to the 2767 assets in the Russel 3000 dataset. The red line and the 95% confidence bands are computed under the null hypothesis that the 2767 assets have the same $\beta^{(w)}$ coefficient.

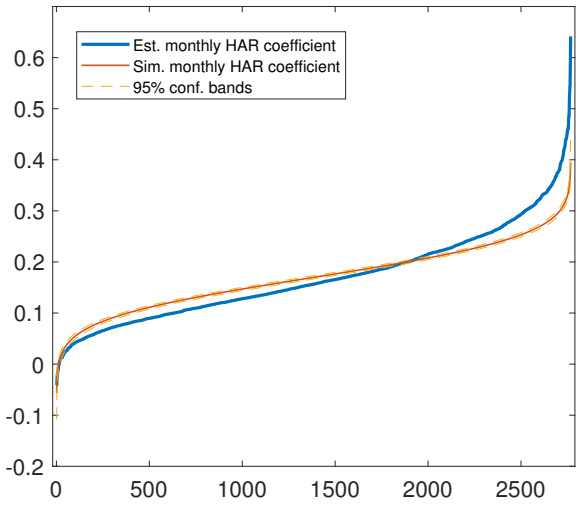

Figure 7: We report in blue the sorted maximum likelihood estimates of the $\beta^{(m)}$ coefficient corresponding to the 2767 assets in the Russel 3000 dataset. The red line and the 95% confidence bands are computed under the null hypothesis that the 2767 assets have the same $\beta^{(m)}$ coefficient.

In the second experiment, we formally test on real data the scalar restriction assumed by Gorgi et al. (2018). We consider 100 portfolios each composed by k randomly selected assets from the NYSE dataset. We estimate the Wishart model of Gorgi et al. (2018) by imposing two different restrictions on the matrix B. In the first case, we assume a scalar restriction, whereas in the second case we assume a diagonal restriction (i.e. the diagonal elements are not required to be equal among each other). In both cases, the matrix A is assumed to have a scalar structure. Under the diagonal restriction, the number of parameters scales as k^2. We thus set $k = 5$ in order to avoid the curse of dimensionality.

We compare the fitting ability of the two different restrictions using the likelihood ratio test. In particular, we test the null hypothesis

$$H_0 : B = \beta I \tag{4.1}$$

against the alternative hypothesis

$$H_1 : B = \begin{pmatrix} \beta_1 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \beta_{15} \end{pmatrix}, \quad \text{where } \beta_1, \ldots, \beta_{15} \text{ are not all equal.} \tag{4.2}$$

To compute the p-values, we use a χ^2 distribution for the test statistics, since the model with a scalar restriction is nested into the model with a diagonal restriction. The test rejects the null hypothesis with p-values smaller than 0.01 for all the 100 portfolios. We obtain the same results when testing the model based on matrix-F density.

The results of these two tests suggest that the scalar assumption of Gorgi et al. (2018) and Opschoor et al. (2017) might be too restrictive and warrant for a more flexible specification. In the next sections, we show the advantages of the proposed two-step estimation procedure over standard joint estimation.

4.3 Analysis settings and benchmark models

We perform the analysis at different cross-section dimensions and with different portfolios. In the NYSE dataset, which contains 1-second data, we set $k = 5, 10, 25, 50, 100$. The assets belonging to each group are randomly selected, except for the group of $k = 100$ stocks, which includes all the assets in the dataset. In the Russell 3000 dataset, the time resolution is 1-minute and the number of time-stamps per day is much smaller compared to the NYSE dataset. To avoid ill-conditioned realized covariance estimates, we limit the maximum cross-section to 50 assets, i.e. we set $k = 5, 10, 25, 50$. Even in this case, the assets in each portfolio are randomly selected.

In the in-sample analysis, we estimate the models in the entire NYSE (Russel 3000) dataset of 2265 (1948) business days. In the out-of-sample analysis, the models are estimated in the subsample comprising the first 1000 business days. The estimated parameters are then used to forecast the covariances of the last 1265 (948) observations, from 22-12-2009 to 31-12-2014 (27-09-2013). As a loss measure, we follow Patton (2011) and use the RMSE and the Qlike. We assess whether loss differences are significant through the Model Confidence Set (MCS) of Hansen et al. (2011).

The 2-step-W and 2-step-F are compared with the following benchmark models:

1. RWG: the Realized Wishart GARCH model of Gorgi et al. (2018) based on the Wishart density in Eq. (2.5)

2. RWG-HAR: the RWG model equipped with a HAR-like structure in the dynamic equation for the covariances:

$$f_{t+1} = \omega + B_1 f_t + B_2 f_{t-1|t-5} + B_3 f_{t-6|t-22} + A s_t$$

   where $f_t = \mathsf{vech}(V_t)$.

3. GAS-F: the model of Opschoor et al. (2017) based on the matrix-F density in Eq. (2.7)

4. GAS-F-HAR: the GAS-F model equipped with a HAR-like structure on static parameters similar to that of the RWG-HAR

5. HAR-DRD: the model of Oh and Patton (2018), combining univariate HAR models with a panel HAR specification for the correlations. Similarly to our approach, the estimation is made in two steps. In the first step, the volatilities are estimated using k distinct HAR models. In the second step, a multivariate HAR-like specification is fitted on the time-series of realized correlations. Covariance forecasts are then computed by combining volatility and correlation forecasts through the DRD decomposition in Eq. (2.9).

6. DCC: the standard Dynamic Conditional Correlation model Engle (2002) for daily returns estimated through the two-step procedure.

We include the RWG-HAR and the GAS-F-HAR in the set of benchmark models in order to exclude that forecast gains provided by our two-step procedure are merely due to long memory effects captured by the univariate models, which have a similar HAR-like structure, see Eq. (2.13). In principle, we might impose the same structure in our correlation model. We verified that such assumption does not provide significant forecast gains. We thus maintain a parsimonious specification with a HAR-like structure in the univariate models and a standard AR(1) score-driven specification for the correlations.

All the realized covariance models are estimated by assuming a scalar specification for the matrix parameters appearing in the score-driven update equation. These include the two matrices A, B in the RWG and GAS-F models, which are constrained as $A = \alpha \mathbb{1}_q$, $B = \beta \mathbb{1}_q$, and the matrices B_1, B_2, B_3 in the RWG-HAR and GAS-F-HAR, which are constrained as $B_1 = \beta_1 \mathbb{1}_q$, $B_2 = \beta_2 \mathbb{1}_q$, $B_3 = \beta_3 \mathbb{1}_q$.

4.4 In-Sample results

Table (3) reports the maximum likelihood estimates, obtained in the NYSE dataset, of the parameters of the models described in the previous section. They are estimated in the entire dataset of 2265 business days and for each cross-section dimension $k = 5, 10, 25, 50, 100$. Note that the parameter ν in the RWG model and the parameters ν_1, ν_2 in the GAS-F model are significantly different from the corresponding estimates in the 2-step-W and 2-step-F models. In particular, ν is lower in the RWG model, ν_1 is generally lower in

the GAS-F model and ν_2 is lower in the 2-step-F model for $k = 5, 10, 25$ and larger for $k = 50, 100$. Such differences indicate that the fat-tail behavior of covariances is significantly different from that of correlations, and it is a further motivation for the introduction of the proposed two-step approach.

Table (4) shows average in-sample loss measures computed in both datasets. We note that score-driven models based on the two-step procedure (2-step-W and 2-step-F) fit the data better than those based on joint estimation (RWG, RWG-HAR, GAS-F, GAS-F-HAR). The relative performance between the two classes of models tends to increase with the cross-section dimension, as can be seen from the RMSE and Qlike gains at $k = 50, 100$, which are both large and highly significant. This result is due to the fact that large portfolios are more likely to include assets with different volatility persistences, which are better captured by the two-step models. We also see that the 2-step-F often provides lower loss measures compared to the 2-step-W, as a result of being robust to the outliers that are typically observed on real time-series. In the vast majority of the cases, the 2-step-F has significantly lower loss measures, whereas in few cases the performances of the two models are statistically indistinguishable. The RWG-HAR and GAS-F-HAR tend to perform better than the RWG and GAS-F because of the long-memory captured by the HAR specification. We finally note that the HAR-DRD often provides a lower RMSE. This is natural when looking at in-sample results, as the HAR-DRD is estimated based on the same RMSE criterion.

We report in Table (5) the AIC and BIC statistics computed in the NYSE dataset. As can be seen, the 2-step-F has the lowest AIC and BIC for all the cross-section dimensions. We also note that the 2-step-W has lower AIC and BIC compared to both the RWG and the RWG-HAR. In order to further examine the in-sample performance of the proposed methodology, we carry out the likelihood based model selection test of Rivers and Vuong (2002), which is applicably to non-nested, possibly misspecified nonlinear dynamic models. In particular, we test 2-step and joint estimation based models (2-step-W vs RWG-HAR and 2-step-F vs GAS-F-HAR) on the one hand, and matrix-F and Wishart based models (2-step-F vs 2-step-W and GAS-F-HAR vs RWG-HAR) on the other hand. The first test compares the likelihood of 2-step models to that of joint estimation based models, whereas the second test compares the likelihood of matrix-F models to that of Wishart models. Table (6) shows the test statistics of Rivers and Vuong (2002), computed as the difference between the log-likelihoods of the two models indicated in the first line, divided by the square root of its asymptotic variance. The latter is computed through the Newey-West estimator. Under the null that the likelihoods of the two models are equal, the test statistics are asymptotically distributed as a standard normal. We see that the null is strongly rejected in all the cases, implying that 2-step models outperform joint estimation based models and matrix-F models outperform Wishart models.

k = 5

Model	α	$\beta(\beta_1)$	β_2	β_3	$\nu(\nu_1)$	ν_2
2-step-W	0.293	0.980			33.2	
	(0.012)	(0.000)			(0.242)	
2-step-W(1)	0.542	0.805	0.130	0.047	50.5	
	(0.020)	(0.029)	(0.033)	(0.015)	(1.376)	
2-step-F	0.900	0.986			151.1	55.3
	(0.060)	(0.003)			(9.018)	(0.754)
2-step-F(1)	2.247	0.820	0.117	0.049	201.6	78.6
	(0.087)	(0.024)	(0.031)	(0.016)	(7.841)	(1.750)
RWG	0.568	0.992			33.1	
	(0.007)	(0.000)			(0.253)	
GAS-F	0.905	0.995			75.4	72.3
	(0.015)	(0.012)			(1.438)	(1.326)
RWG-HAR	0.615	0.832	0.111	0.050	33.8	
	(0.007)	(0.007)	(0.007)	(0.004)	(0.272)	
GAS-F-HAR	0.987	0.870	0.080	0.046	72.7	77.4
	(0.016)	(0.007)	(0.008)	(0.004)	(1.484)	(1.357)
DCC	0.016	0.962				
	(0.002)	(0.005)				
HAR-DRD		0.399	0.437	0.107		
		(0.063)	(0.087)	(0.075)		
HAR-DRD(1)		0.609	0.268	0.086		
		(0.019)	(0.025)	(0.019)		

k = 10

Model	α	$\beta(\beta_1)$	β_2	β_3	$\nu(\nu_1)$	ν_2
2-step-W	0.221	0.999			57.7	
	(0.004)	(0.000)			(0.218)	
2-step-W(1)	0.528	0.799	0.141	0.040	51.9	
	(0.020)	(0.029)	(0.033)	(0.014)	(1.393)	
2-step-F	0.832	0.999			245.5	97.7
	(0.023)	(0.001)			(6.964)	(0.707)
2-step-F(1)	2.040	0.814	0.127	0.043	195.7	82.8
	(0.084)	(0.028)	(0.032)	(0.016)	(2.774)	(1.609)
RWG	0.481	0.992			56.8	
	(0.003)	(0.000)			(0.222)	
GAS-F	0.978	0.995			144.9	117.4
	(0.008)	(0.008)			(1.532)	(1.033)
RWG-HAR	0.546	0.813	0.131	0.049	58.0	
	(0.004)	(0.005)	(0.005)	(0.002)	(0.210)	
GAS-F-HAR	0.999	0.860	0.088	0.047	139.6	125.1
	(0.009)	(0.004)	(0.005)	(0.002)	(1.523)	(1.100)
DCC	0.011	0.970				
	(0.001)	(0.004)				
HAR-DRD		0.563	0.328	0.070		
		(0.028)	(0.039)	(0.036)		
HAR-DRD(1)		0.563	0.328	0.070		
		(0.020)	(0.025)	(0.020)		

k = 25

Model	α	$\beta(\beta_1)$	β_2	β_3	$\nu(\nu_1)$	ν_2
2-step-W	0.119	0.990			98.8	
	(0.001)	(0.000)			(0.149)	
2-step-W(1)	0.480	0.781	0.161	0.035	47.1	
	(0.017)	(0.029)	(0.033)	(0.015)	(1.152)	
2-step-F	0.282	0.999			190.4	152.9
	(0.004)	(0.000)			(2.890)	(0.565)
2-step-F(1)	2.422	0.807	0.134	0.041	229.1	68.1
	(0.087)	(0.030)	(0.035)	(0.016)	(2.576)	(1.413)
RWG	0.346	0.996			95.2	
	(0.001)	(0.000)			(0.139)	
GAS-F	0.824	0.999			161.2	182.0
	(0.000)	(0.000)			(0.000)	(0.000)
RWG-HAR	0.432	0.785	0.148	0.063	97.2	
	(0.001)	(0.002)	(0.002)	(0.001)	(0.140)	
GAS-F-HAR	0.9169	0.8646	0.0807	0.0534	284.9	182.1
	(0.003)	(0.002)	(0.002)	(0.001)	(1.027)	(0.921)
DCC	0.005	0.977				
	(0.000)	(0.007)				
HAR-DRD		0.357	0.406	0.154		
		(0.010)	(0.014)	(0.012)		
HAR-DRD(1)		0.531	0.328	0.097		
		(0.020)	(0.026)	(0.021)		

k = 50

Model	α	$\beta(\beta_1)$	β_2	β_3	$\nu(\nu_1)$	ν_2
2-step-W	0.064	0.992			133.3	
	(0.000)	(0.000)			(0.096)	
2-step-W(1)	0.466	0.769	0.170	0.039	46.0	
	(0.020)	(0.032)	(0.036)	(0.016)	(1.177)	
2-step-F	0.056	0.999			162.7	180.5
	(0.001)	(0.000)			(0.381)	(0.692)
2-step-F(1)	2.429	0.789	0.152	0.042	234.8	68.2
	(0.085)	(0.030)	(0.035)	(0.016)	(3.295)	(1.573)
RWG	0.242	0.995			127.5	
	(0.000)	(0.000)			(0.088)	
GAS-F	0.479	0.999			186.3	157.0
	(0.001)	(0.000)			(0.351)	(0.653)
RWG-HAR	0.320	0.777	0.138	0.080	129.4	
	(0.001)	(0.001)	(0.002)	(0.001)	(0.093)	
GAS-F-HAR	0.3812	0.9500	0.0000	0.0485	205.9	237.7
	(0.001)	(0.001)	(0.002)	(0.001)	(0.350)	(0.699)
DCC	0.002	0.978				
	(0.000)	(0.001)				
HAR-DRD		0.339	0.398	0.176		
		(0.005)	(0.007)	(0.006)		
HAR-DRD(1)		0.512	0.343	0.103		
		(0.020)	(0.027)	(0.022)		

k = 100

Model	α	$\beta(\beta_1)$	β_2	β_3	$\nu(\nu_1)$	ν_2
2-step-W	0.031	0.993			147.4	
	(0.000)	(0.000)			(0.044)	
2-step-W(1)	0.449	0.765	0.168	0.045	42.8	
	(0.004)	(0.037)	(0.033)	(0.019)	(0.459)	
2-step-F	0.059	0.993			193.1	507.4
	(0.000)	(0.000)			(0.091)	(0.635)
2-step-F(1)	2.312	0.790	0.149	0.045	219.1	66.0
	(0.078)	(0.031)	(0.036)	(0.017)	(2.635)	(1.132)
RWG	0.147	0.994			142.6	
	(0.000)	(0.000)			(0.042)	
GAS-F	0.256	0.992			140.7	202.5
	(0.000)	(0.000)			(0.000)	(0.001)
RWG-HAR	0.217	0.763	0.135	0.096	143.7	
	(0.000)	(0.001)	(0.001)	(0.000)	(0.042)	
GAS-F-HAR	0.3095	0.9062	0.0860	0.0000	107.9	222.0
	(0.000)	(0.001)	(0.001)	(0.000)	(0.094)	(0.538)
DCC	0.001	0.959				
	(0.000)	(0.006)				
HAR-DRD		0.300	0.392	0.206		
		(0.002)	(0.003)	(0.003)		
HAR-DRD(1)		0.508	0.337	0.113		
		(0.020)	(0.027)	(0.022)		

Table 3: Maximum likelihood estimates of the parameters of 2-step-W, 2-step-F and of benchmark models obtained in the NYSE dataset. Standard errors are reported in parenthesis. For the k univariate models, denoted by 2-step-W(1) and 2-step-F(1), we report the average of the maximum likelihood estimates and of the standard errors. Similarly, we report the averages of the parameters of the univariate models in the HAR-DRD.

Model	MSE					Qlike				
NYSE	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$
2-step-W	0.9859*	1.0143	1.0255	1.0308	1.0478	**0.9995***	0.9964	**0.9869***	**0.9962***	1.0051
2-step-F	1.0000	1.0000*	1.0000	1.0000	1.0000*	1.0000*	1.0000	1.0000	1.0000*	**1.0000***
RWG	1.0320	1.0285	1.0521	1.0494	1.0818	1.0476	1.0470	1.0458	1.0575	1.0548
RWG-HAR	1.0184	1.0240	1.0457	1.0407	1.0913	1.0203	1.0126	1.0137	1.0320	1.0356
GAS-F	1.0352	1.0254	1.0396	1.1230	1.3608	1.0522	1.0417	1.0393	1.1114	1.1925
GAS-F-HAR	1.0151	1.0169	1.0162	1.0620	1.3223	1.0287	1.0135	1.0189	1.0530	1.1504
DCC	1.8458	1.7460	1.5017	1.4146	1.3819	1.5974	1.6370	1.7858	1.6352	1.4153
HAR-DRD	**0.9858***	**0.9975***	**0.9946***	**0.9812***	**0.9957***	1.0024*	**0.9898***	0.9922	1.0250	1.0562
Russell	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$
2-step-W	1.0133	1.0208	1.0256	1.0469	-	1.0029*	1.0145	1.0081	1.0055	-
2-step-F	**1.0000***	**1.0000***	1.0000*	1.0000*	-	1.0000*	**1.0000***	**1.0000***	**1.0000***	-
RWG	1.0683	1.0573	1.0735	1.0941	-	1.0322	1.0490	1.0524	1.0583	-
RWG-HAR	1.0525	1.0528	1.0587	1.0878	-	1.0100	1.0190	1.0276	1.0347	-
GAS-F	1.0408	1.0201	1.0210	1.0730	-	1.0329	1.0346	1.0340	1.0744	-
GAS-F-HAR	1.0281	1.0116	1.0131	1.0633	-	1.0073*	1.0144	1.0278	1.0597	-
DCC	1.4863	1.5754	1.5065	1.8216	-	1.5392	1.6012	1.5702	1.4774	-
HAR-DRD	1.0035*	1.0140	**0.9988***	**0.9907***	-	**0.9976***	1.0107	1.0196	1.0365	-

Table 4: Relative in-sample average loss measures. All measures are reported with respect to the 2-step-F model. A value lower than one indicates that the corresponding model outperforms the 2-step-F model. Bold numbers denote the models having lowest loss measures. Asterisks denote the models that are part of the 90% model confidence set (MCS).

4.5 Out-of-Sample results

We now check whether the in-sample gains found in the previous analysis translate into out-of-sample forecast gains. In Table (7), we report the average out-of-sample loss measures of all the models considered above. Forecasts are computed in the NYSE (Russel 3000) subsample including the last 1265 (948) business days and are based on parameter estimates obtained in the subsample comprising the first 1000 business days. We immediately notice that, with only few exceptions, the 2-step-F model is the best performing model and it is included in the model confidence set. Similarly to the in-sample analysis, the relative performance of the 2-step-W and 2-step-F over joint estimation based models increases as the cross-section dimension increases. This confirms the intuition that the flexibility provided by the two-step procedure leads to better covariance forecasts. The results in the table also confirm that the matrix-F density is more suited than the Wishart density to capture the fat-tails of covariances and correlations.

The performance of the HAR-DRD is comparable to that of the 2-step-F model when we look at the RMSE loss. When we instead consider the Qlike loss, the relative performance of the HAR-DRD rapidly deteriorates when k increases. The comparison of our methodology with the HAR-DRD is examined further in the next section, where we study the behavior of both models in relation to measurement errors on covariance estimates.

One may argue that additional forecast gains might be recovered by imposing a HAR structure on the score-driven dynamic equation for the correlations in the 2-step-W and 2-step-F models. As outlined in Section (4.3), this is not the case, as we verified by performing similar experiments. Most of the benefits provided by modeling long-memory come from the univariate HAR models, whereas a similar HAR structure on correlations does not lead to significant out-of-sample improvements. This is confirmed by the fact that

Model	AIC				
	$k = 5$	$k = 10$	$k = 25$	$k = 50$	$k = 100$
	$(\times 10^5)$	$(\times 10^6)$	$(\times 10^7)$	$(\times 10^7)$	$(\times 10^8)$
2-step-W	−6.8400	−2.4546	−1.4677	−5.7651	−2.3129
2-step-F	**−7.1643**	**−2.6124**	**−1.5617**	**−6.0493**	**−2.4197**
RWG	−5.8542	−2.2686	−1.4190	−5.6213	−2.2535
GAS-F	−6.1604	−2.4064	−1.5091	−5.9459	−2.3749
RWG-HAR	−5.7970	−2.2475	−1.4059	−5.5688	−2.2318
GAS-F-HAR	−6.0981	−2.3833	−0.0028	−5.8736	−2.3432
Model	BIC				
	$k = 5$	$k = 10$	$k = 25$	$k = 50$	$k = 100$
	$(\times 10^5)$	$(\times 10^6)$	$(\times 10^7)$	$(\times 10^7)$	$(\times 10^8)$
2-step-W	−6.8381	−2.4542	−1.4676	−5.7649	−2.3129
2-step-F	**−7.1621**	**−2.6119**	**−1.5616**	**−6.0491**	**−2.4196**
RWG	−5.8541	−2.2686	−1.4190	−5.6213	−2.2535
GAS-F	−6.1602	−2.4064	−1.5091	−5.9459	−2.3749
RWG-HAR	−5.7967	−2.2475	−1.4059	−5.5688	−2.2318
GAS-F-HAR	−6.0978	−2.3832	−0.0028	−5.8736	−2.3432

Table 5: AIC and BIC statistics of all the considered models computed in the NYSE dataset.

	2-step vs joint		Matrix-F vs Wishart	
Cross-section	2-step-W vs RWG-HAR	2-step-F vs GAS-F-HAR	2-step-F vs 2-step-W	GAS-F-HAR vs RWG-HAR
$k = 5$	69.79	85.25	23.14	31.64
$k = 10$	67.18	94.15	28.24	30.73
$k = 25$	49.34	98.24	28.27	33.29
$k = 50$	37.49	57.35	39.69	43.84
$k = 100$	16.31	16.56	42.51	47.88

Table 6: Test statistics of the Rivers and Vuong (2002) test. In each column, we report the results of the test for the two models indicated in the first line. The test statistics are computed as the difference between the log-likelihoods of the two models divided by its asymptotic standard deviation. The latter is estimated through the Newey-West estimator.

the RWG-HAR and GAS-F-HAR, which have a HAR structure in both volatilities and correlations, have a performance which in some cases is sub-optimal with respect to the RWG and the GAS-F. In this sense, our proposed two-step approach provides a further advantage, as it easily allows for different modeling strategies for volatilities and correlations.

Model			MSE					Qlike		
NYSE	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$
2-step-W	0.9882*	1.0087*	1.0413	1.0814	1.0927	**0.9987***	1.0044*	1.0107	1.0141	1.0121
2-step-F	1.0000	**1.0000***	**1.0000***	**1.0000***	**1.0000***	1.0000*	**1.0000***	**1.0000***	**1.0000***	**1.0000***
RWG	1.0248	1.0278	1.0479	1.0523	1.0531	1.0261	1.0293	1.0659	1.0740	1.0588
RWG-HAR	1.0422	1.0471	1.0530	1.0602	1.0683	1.0507	1.0279	1.0555	1.0767	1.0708
GAS-F	1.0209	1.0122	1.0312	1.0778	1.0444	1.0248	1.0130	1.0731	1.0644	1.1180
GAS-F-HAR	1.0294	1.1061	1.0232	1.1489	1.1649	1.0360	1.0768	1.0467	1.1099	1.1275
DCC	2.0354	1.8998	1.6256	1.4739	1.3265	5.2720	3.5086	1.9520	1.4058	1.3014
HAR-DRD	**0.9840***	1.0007*	1.0051*	1.0091	1.0105	1.0091*	1.0196	1.0271	1.0433	1.0591
Russell	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$	$k=5$	$k=10$	$k=25$	$k=50$	$k=100$
2-step-W	1.0341	1.0315	1.0483	1.0284	-	1.0246	1.0266	1.0141	1.0107	-
2-step-F	**1.0000***	**1.0000***	**1.0000***	**1.0000***	-	**1.0000***	**1.0000***	**1.0000***	**1.0000***	-
RWG	1.0538	1.0542	1.0711	1.0435	-	1.0434	1.0586	1.0612	1.0554	-
RWG-HAR	1.0605	1.0513	1.0686	1.0561	-	1.1015	1.0751	1.0697	1.0699	-
GAS-F	1.0199	1.0172	1.0148	1.0496	-	1.0206	1.0169	1.0311	1.1345	-
GAS-F-HAR	1.0276	1.0190	1.0284	1.0231	-	1.0961	1.0554	1.0605	1.0877	-
DCC	1.1081	1.2586	1.2739	1.2929	-	1.2654	1.5282	1.4649	1.4237	-
HAR-DRD	1.0084	1.0099*	1.0083	1.0027*	-	1.0020*	1.0183	1.0263	1.0277	-

Table 7: Relative out-of-sample average loss measures. All measures are reported with respect to the 2-step-F model. A value lower than one indicates that the corresponding model outperforms the 2-step-F model. Bold numbers denote the models having lowest loss measures. Asterisks denote the models that are part of the 90% model confidence set (MCS).

4.6 Robustness to noise

As underlined in the introduction, one of the advantages of the proposed approach is that it allows to take into account the unavoidable estimation error that affects realized covariance estimates. The latter are indeed modeled as noisy observations of a latent matrix process representing the true integrated covariance of the intraday log-prices. In this sense, our methodology is in sharp contrast with the HAR-DRD, where realized covariances are modeled as if they were the true covariances. We refer the reader to Bollerslev et al. (2016), Bollerslev et al. (2018) and Buccheri and Corsi (2019) for a more detailed discussion on the impact of estimation errors on forecasting with realized measures.

In this section, we assess the effect of different levels of noise on covariance forecasts by comparing our methodology, which accounts for noise, to the HAR-DRD, which does not possess such ability. To increase the level of noise, we sample the log-prices at the frequency of 5-minutes. The number of sampled log-prices at this frequency is considerably smaller than that at 1-second (NYSE dataset) and 1-minute (Russell 3000 dataset). As a consequence, the multivariate realized kernel estimator of Barndorff-Nielsen et al. (2011) used in our analysis will be subject to larger estimation errors.

Table (8) shows the out-of-sample average loss measures of the HAR-DRD with respect to the 2-step-F

model at different sampling frequencies. Specifically, on the first line of each box, we show the result of Table (7) corresponding to the original frequencies Δ = 1-second (NYSE dataset) and Δ = 1-minute (Russell 3000 dataset). On the second line, we report the results obtained by sampling the log-prices at Δ = 5-minutes. With 5-minutes data, the maximum number of time-stamps per day is 78, and thus for $k = 100$ the realized covariance estimator is ill-conditioned. We thus report the results for $k = 5, 10, 25, 50$. We immediately note that the forecast gains of the 2-step-F are larger and more significant for Δ = 5-minutes. Of course, such effect is more pronounced in the NYSE dataset, since the new sampling at 5-minutes implies huge data reduction compared to the original 1-second sampling frequency. We also see that the statistical significance of the forecast gains improves considerably at 5-minutes, as shown by the model confidence set test, which always excludes the HAR-DRD from the 90% MCS.

The above results are a consequence of the score-driven filtering mechanism, which smooths out the noise and guarantees more robust estimates compared to the HAR-DRD. The fact that we can successfully test for such effect in the absence of observations of the true latent covariance is due to the use of loss measures which are robust to estimation errors (see Patton and Sheppard (2009) and Patton 2011). Finally, note that the use of lower frequencies is common in financial applications, as it may be dictated by liquidity or data source restrictions.

Model	MSE				Qlike			
NYSE	$k=5$	$k=10$	$k=25$	$k=50$	$k=5$	$k=10$	$k=25$	$k=50$
				Δ = 1-second				
2-step-F	1.0000	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***
HAR-DRD	**0.9840***	1.0007*	1.0051*	1.0091	1.0091*	1.0196	1.0271	1.0433
				Δ = 5-minutes				
2-step-F	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***
HAR-DRD	1.0833	1.0807	1.0915	1.0963	1.2760	1.2196	1.1189	1.0875
Russell	$k=5$	$k=10$	$k=25$	$k=50$	$k=5$	$k=10$	$k=25$	$k=50$
				Δ = 1-minute				
2-step-F	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***
HAR-DRD	1.0084	1.0099	1.0083	1.0027*	1.0020*	1.0183	1.0263	1.0277
				Δ = 5-minutes				
2-step-F	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***	**1.0000***
HAR-DRD	1.0247	1.0205	1.0131	1.0103	1.0192	1.0208	1.0256	1.0251

Table 8: Relative Out-of-sample average loss measures of HAR-DRD with respect to the 2-step-F model at different sampling frequencies. Bold numbers denote the models having lowest loss measures. Asterisks denote the models that are part of the 90% model confidence set (MCS).

4.7 Portfolio analysis

In this section, we use a framework similar to that of Fleming et al. (2001), Fleming et al. (2003) and Bollerslev et al. (2018) to provide a quantitative assessment of the *economic* gains of switching from joint estimation based models (RWG, RWG-HAR, GAS-F, GAS-F-HAR) to the proposed two-step estimation based models (2-step-W, 2-step-F). Let us consider an investor who allocates her funds into k risky assets by pursuing a volatility timing strategy in the period of time from day 1 to day T. On day $t-1$, $1 < t \leq T$,

the investor solves the Global Minimum Variance (GMV) problem:

$$\begin{aligned} &\min \omega_t' \Sigma_{t|t-1} \omega_t, \\ &\text{subject to } \omega_t' \mathbf{1} = 1 \end{aligned} \tag{4.3}$$

where ω_t is a $k \times 1$ vector of portfolio weights and $\Sigma_{t|t-1}$ is the covariance matrix of asset returns at time t computed based on information available up to time $t-1$. We assume that the investor faces transaction costs proportional to the portfolio turnover. The latter is defined as in DeMiguel et al. (2014), namely:

$$TO_{t-1} = \sum_{i=1}^{k} \left| \omega_t^{(i)} - \omega_{t-1}^{(i)} \frac{1 + r_{t-1}^{(i)}}{1 + \omega_{t-1}' r_{t-1}} \right| \tag{4.4}$$

where $\omega_t^{(i)}$, $r_t^{(i)}$ denote the i-th component of the vector of weights ω_t and the returns vector r_t, respectively. The portfolio return in excess of transaction costs is computed as:

$$r_{p,t} = \omega_t' r_t - cTO_t \tag{4.5}$$

where c is constant. We finally assume that the investor has a quadratic utility function:

$$U(r_{p,t}, \gamma) = (1 + r_{p,t}) - \frac{\gamma}{2(1+\gamma)} (1 + r_{p,t})^2 \tag{4.6}$$

where $\gamma > 0$ is the coefficient of risk aversion.

The economic gain of switching from a covariance model "s" to another covariance model "l" is defined as the quantity Δ_γ such that:

$$\sum_{t=1}^{T} U(r_{p,t}^{(s)}, \gamma) = \sum_{t=1}^{T} U(r_{p,t}^{(l)} - \Delta_\gamma, \gamma) \tag{4.7}$$

where $r_{p,t}^{(s)}$ and $r_{p,t}^{(l)}$ are the returns of the portfolios constructed with the covariance forecasts of model s and l, respectively. In other words, Δ_γ represents the return an investor with risk aversion γ would be willing to sacrifice in order to use the forecasts of model l in place of those of model s in solving the GMV problem[4]. Table (9) reports the values of Δ_γ computed for $\gamma = 1$ and $\gamma = 10$. For sake of clarity, we report in the table the results obtained with the group of $k = 25$ assets belonging to the NYSE dataset. However, similar results are obtained when considering other portfolio dimensions. The setting of the analysis is the same as in the out-of-sample test of Section (4.5): the models are estimated in the sub-sample comprising the first 1000 business days, and the out-of-sample portfolios are computed for the remaining 1265 business days.

The values of Δ_γ reported under the columns entitled RWG and RWG-HAR represent the economic gains in annual basis points of switching from the latter two models to the 2-step-W. Similarly, those reported under the columns entitled GAS-F and GAS-F-HAR represent the economic gains of switching from the

[4]In principle, the investor might solve a general mean-variance problem instead of the GMV problem considered here. As discussed by Jagannathan and Ma (2003) and DeMiguel et al. (2009), mean-variance problems rely on forecasts of expected returns, which are notoriously subject to large estimation errors and tend to distort the optimal portfolio solution. As we are interested in examining covariance forecasts, we focus on the GMV problem, in a similar fashion to Bollerslev et al. (2018).

	2-step-W	2-step-F	RWG	RWG-HAR	GAS-F	GAS-F-HAR
Turnover	0.239	0.263	0.205	0.241	0.246	0.323
Average return (%)	9.847	10.427	9.371	9.273	9.644	8.473
Ex-post volatility (%)	7.887	**7.875***	7.887	7.912	7.893	8.202
				$c = 0\%$		
Sharpe ratio	1.138	**1.209**	1.084	1.069	1.116	0.941
Δ_1	59.3		47.6	57.4	78.6*	198.7*
Δ_{10}	70.9		46.9	58.0	81.3*	228.1*
				$c = 1\%$		
Sharpe ratio	1.069	**1.132**	1.024	0.999	1.045	0.851
Δ_1	53.3		39.0	57.9	74.2*	213.8*
Δ_{10}	65.0		38.4	58.4	76.8*	243.2*
				$c = 2\%$		
Sharpe ratio	0.999	**1.055**	0.964	0.929	0.973	0.760
Δ_1	47.2		30.5	58.3	69.8*	228.9*
Δ_{10}	59.2		29.9	58.9	72.3*	258.3*

Table 9: For each covariance model, we report the portfolio transaction costs computed as in Eq. (4.4), the average portfolio return, the ex-post portfolio volatility and the Sharpe ratio. All quantities are annualized. We also report, for $c = 0\%, 1\%, 2\%$, the economic gains Δ_γ in annual basis points of switching from joint estimation based models with Wishart (matrix-F) density to the 2-step-W (2-step-F), for $\gamma = 1, 10$. Similarly, we report the economic gains of switching from the 2-step-W to the 2-step-F. The asterisks in portfolio volatilities indicate the models that belong to the 90% MCS of lowest ex-post volatilities. Significance levels on economic gains are instead computed based on the Reality Check of White (2000). The asterisks indicate the economic gains that are significantly different from zero at the 5% confidence level. Bold quantities denote the best performing model according to the measure specified in the first column.

latter two models to the 2-step-F. We also show in the first column the economic gains of switching from 2-step-W to 2-step-F. Significance levels on Δ_γ are assessed through the Reality Check of White (2000), using the stationary bootstrap of Politis and Romano (1994). In addition, we report in the table the turnover, computed as in Eq. (4.4), the ex-post portfolio volatility, the average portfolio return and the Sharpe ratio. The constant c multiplying the turnover is set as $c = 0\%, 1\%, 2\%$, as in Fleming et al. (2003) and Bollerslev et al. (2018).

We first note that all the economic gains are positive, meaning that the investor would be willing to pay a positive annual amount in order to switch to the proposed two-step estimation based models. In particular, the economic gains of switching from GAS-F and GAS-F-HAR to 2-step-F are highly significant. These gains are mainly imputable to the variance reduction featured by the 2-step-F model. On the one side, the flexibility provided by the proposed approach leads to better out-of-sample forecasts, which translate into a lower ex-post portfolio volatility (see Engle and Colacito 2006 and Patton and Sheppard 2009). On the other side, the two-step procedure naturally leads to a loss of statistical efficiency, which in turn determines more erratic portfolio weights and a larger turnover. Our results clearly indicate that the first effect, namely the superior forecasting ability, significantly dominates the loss of statistical efficiency in the context of a risk-averse investor maximizing a quadratic utility function. In the case of models based on the Wishart density, the ex-post volatility of the 2-step-W portfolio is statistically indistinguishable from that of the RWG portfolio. Accordingly, we find that the corresponding economic gains are not statistically significant. However, they are still positive due to the larger average return featured by the 2-step-W portfolio. We finally note that the economic gains of switching from the 2-step-W to the 2-step-F are also positive, confirming the result that the matrix-F density provides a better description of realized covariance time-series.

5 Conclusions

We have introduced a two-step estimation procedure for score-driven realized covariance models based on Wishart and matrix-F densities. By employing the score, which is a martingale difference by construction, the proposed models automatically include a correction in the spirit of Aielli (2013). In the case of the Wishart density, our class of models reduces to the model proposed by Bauwens et al. (2012). In the case of the matrix-F density, we obtain a new realized covariance model that can be estimated in two-steps and that accounts for the fat-tails of realized covariance time-series. More specifically, the main advantage of the method is that it has a higher degree of flexibility in the estimation of volatilities. In the first step, the latter are indeed separately estimated by univariate realized volatility models with different parameters. The model is therefore easy to estimate in large dimensions and, compared to joint estimation, is less affected by the curse of dimensionality.

Through a Monte-Carlo study, we have examined the finite sample properties of the two-step estimator and shown that it recovers accurate estimates of the parameters. We have also examined the ability of the model to capture misspecified correlation dynamics. The comparison between the two-step procedure and standard joint estimation has been studied extensively through both Monte-Carlo simulations and empirical data. We have found that the additional flexibility of the two-step procedure translates into in-sample and out-of-sample forecast gains. The latter are found to be more significant in portfolios characterized by different levels of heterogeneity in the persistence of the volatilities. In particular, we have found statistically significant evidences of superior forecast ability in portfolios of high-dimensions, where volatilities are more likely to exhibit different persistences.

We have also examined the performance of the methodology under very noisy estimates of the underlying matrix-variate process. It is obtained that the score-driven filtering mechanism leads to increasingly better forecasts compared to methods not accounting for estimation errors. As a final experiment, we have assessed the economic gains of switching from standard joint estimation to the proposed two-step approach. It is found that a risk averse investor would be willing to pay a positive amount to adopt the forecasts of the two-step models in constructing her portfolio.

References

Abadir, K. M. and J. R. Magnus (2005). *Matrix Algebra*. Econometric Exercises. Cambridge University Press.

Aielli, G. P. (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business & Economic Statistics 31*(3), 282–299.

Andersen, T. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*(4), 885–905.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal 12*(3), C1–C32.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2011). 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.

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

Bauwens, L., M. Braione, and G. Storti (2016). Forecasting Comparison of Long Term Component Dynamic Models for Realized Covariance Matrices. *Annals of Economics and Statistics* (123-124), 103–134.

Bauwens, L., M. Braione, and G. Storti (2017). A dynamic component model for forecasting high-dimensional realized covariance matrices. *Econometrics and Statistics 1*(C), 40–61.

Bauwens, L., G. Storti, F. Violante, et al. (2012). Dynamic conditional correlation models for realized covariance matrices. *CORE DP 60*, 104–108.

Bekierman, J. and H. Manner (2018). Forecasting realized variance measures using time-varying coefficient models. *International Journal of Forecasting 34*(2), 276 – 287.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics 192*(1), 1 – 18.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2018). Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. *Journal of Econometrics 207*(1), 71 – 91.

Bonato, M., M. Caporin, and A. Ranaldo (2012). A forecast-based comparison of restricted wishart autoregressive models for realized covariance matrices. *The European Journal of Finance 18*(9), 761–774.

Buccheri, G. and F. Corsi (2019, 08). HARK the SHARK: Realized Volatility Modeling with Measurement Errors and Nonlinear Dependencies*. *Journal of Financial Econometrics*.

Callot, L. A. F., A. B. Kock, and M. C. Medeiros (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics 32*(1), 140–158.

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

Corsi, F. (2009, Spring). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Cox, D. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics 8*(2), 93–115.

Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics 29*(4), 552–563.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *The Review of Economics and Statistics 96*(5), 898–915.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science 55*(5), 798–812.

DeMiguel, V., F. J. Nogales, and R. Uppal (2014, 02). Stock Return Serial Dependence and Out-of-Sample Portfolio Performance. *The Review of Financial Studies 27*(4), 1031–1073.

Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics 20*(1), 134–144.

Engle, R. (2002). Dynamic conditional correlation. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. and R. Colacito (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics 24*, 238–253.

Engle, R. and B. Kelly (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics 30*(2), 212–228.

Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*(1), 3 – 27.

Fleming, J., C. Kirby, and B. Ostdiek (2001). The economic value of volatility timing. *The Journal of Finance 56*(1), 329–352.

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

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

Gorgi, P., P. R. Hansen, P. Janus, and S. J. Koopman (2018, 04). Realized Wishart-GARCH: A Score-driven Multi-Asset Volatility Model*. *Journal of Financial Econometrics 17*(1), 1–32.

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

Gupta, A. and D. Nagar (1999). *Matrix Variate Distributions*. PMS Series. Addison-Wesley Longman, Limited.

Hansen, P. R., Z. Huang, and H. H. Shek (2012). Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics 27*(6), 877–906.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge University Press.

Jagannathan, R. and T. Ma (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance 58*(4), 1651–1683.

Magnus, J. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. PROBABILISTICS AND STATISTICS. Wiley.

Oh, D. H. and A. J. Patton (2018). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Journal of Business & Economic Statistics 36*(2), 181–195.

Opschoor, A., P. Janus, A. Lucas, and D. V. Dijk (2017). New heavy models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics 0*(0), 1–15.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics 160*(1), 246 – 256.

Patton, A. J. and K. Sheppard (2009). *Evaluating Volatility and Correlation Forecasts*, pp. 801–838. Berlin, Heidelberg: Springer Berlin Heidelberg.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*(428), 1303–1313.

Rivers, D. and Q. Vuong (2002). Model selection tests for nonlinear dynamic models. *The Econometrics Journal 5*(1), 1–39.

Shephard, N. and K. Sheppard (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics 25*(2), 197–231.

White, H. (2000). A reality check for data snooping. *Econometrica 68*(5), 1097–1126.

A Computation of the scaled score in the univariate models

A.1 χ^2 density

We compute the scaled score appearing in the dynamic of the log-variance $\lambda_t^{(i)}$ in Eq. (2.13) in the case of the χ^2 density. To simplify the notation, we suppress the subscript i. The conditional log-likelihood is:

$$\log p_{W_1}(x_t; v_t, \nu) = \frac{1}{2}c(\nu) + \left(\frac{\nu}{2} - 1\right)\log(x_t) - \frac{\nu}{2}\log(v_t) - \frac{\nu}{2}\left(\frac{x_t}{v_t}\right) \tag{A.1}$$

where $c(\nu) = \nu\log(\nu/2) - 2\log\Gamma(\nu/2)$. We now prove the following result (recall that $v_t = e^{\lambda_t}$):

Proposition A.1. *For the density in Eq. (A.1), the score* $\nabla_t^{W_1} = \frac{\partial \log p_{W_1}(x_t;\lambda_t,\nu)}{\partial \lambda_t}$ *is given by:*

$$\nabla_t^{W_1} = \frac{\nu}{2e^{\lambda_t}}\left[x_t - e^{\lambda_t}\right] \tag{A.2}$$

Proof.

$$\begin{aligned}
\nabla_t^{W_1} &= \frac{\partial \log p_{W_1}(x_t;\lambda_t,\nu)}{\partial \lambda_t} = \frac{\partial \log p_{W_1}(x_t;\lambda_t,\nu)}{\partial v_t} \times \frac{\partial v_t}{\partial \lambda_t} \\
&= \frac{\nu}{2v_t^2}\left[x_t - v_t\right] \times v_t \\
&= \frac{\nu}{2e^{\lambda_t}}\left[x_t - e^{\lambda_t}\right]
\end{aligned}$$

□

Then, we compute the information quantity:

Proposition A.2. *For the density in Eq. (A.1), the Fisher information* $\mathcal{I}_{t|t-1}^{W_1} = \mathrm{E}_{t|t-1}[\nabla_t^2]$ *is given by:*

$$\mathcal{I}_{t|t-1}^{W_1} = \frac{\nu}{2} \tag{A.3}$$

Proof.

$$\begin{aligned}
\mathcal{I}_{t|t-1}^{W_1} &= E_{t-1}\left[(\nabla_t^{W_1})^2\right] = E_{t-1}\left[\frac{\nu^2}{4e^{2\lambda_t}}(x_t - e^{\lambda_t})^2\right] \\
&= \frac{\nu^2}{4e^{2\lambda_t}} Var\left[x_t\right] =^* \frac{\nu^2}{4e^{2\lambda_t}}\frac{2e^{2\lambda_t}}{\nu} = \frac{\nu}{2}
\end{aligned}$$

where $(^*)$ follows from $Var[x] = Var[(v_t/\nu)k_\nu] = 2\frac{(v_t)^2}{\nu}$ if k_ν is distributed as a χ^2_ν. □

Finally, it is immediate to compute the scaled score:

Proposition A.3. *For the density in Eq. (A.1), the scaled score* $s_t^{W_1} = (\mathcal{I}_{t|t-1}^{W_1})^{-1}\nabla_t^{W_1}$ *is given by:*

$$s_t^{W_1} = \frac{1}{e^{\lambda_t}}\left[x_t - e^{\lambda_t}\right] \tag{A.4}$$

A.2 F density

In the case of the univariate F density, the conditional log-likelihood is:

$$\log p_F\left(x_t; v_t, \nu_1, \nu_2\right) = d(\nu_1, \nu_2) - \frac{\nu_1}{2}\log(v_t) + \left(\frac{\nu_1}{2} - 1\right)\log(x_t) - \frac{\nu_1 + \nu_2}{2}\log\left(\tilde{w}_t\right) \tag{A.5}$$

where:

$$\tilde{w}_t = 1 + \frac{\nu_1 x_t}{(\nu_2 - 2)v_t} \tag{A.6}$$

$$d(\nu_1, \nu_2) = \frac{\nu_1}{2}\log\left(\frac{\nu_1}{\nu_2 - 2}\right) + \log\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right) - \log\Gamma\left(\frac{\nu_1}{2}\right) - \log\Gamma\left(\frac{\nu_2}{2}\right) \tag{A.7}$$

We compute now the score of the conditional log-likelihood.

Proposition A.4. *For the density in Eq. (A.5), the score* $\nabla_t^F = \frac{\partial \log p_F(x_t; v_t, \nu_1, \nu_2)}{\partial \lambda_t}$ *is given by:*

$$\nabla_t^F = \frac{\nu_1}{2e^{\lambda_t}}\left[\frac{\nu_1 + \nu_2}{\nu_2 - 2}\frac{x_t}{\tilde{w}_t} - e^{\lambda_t}\right] \tag{A.8}$$

Proof.

$$\begin{aligned}
\nabla_t^F &= \frac{\partial \log p_F\left(x_t; v_t, \nu_1, \nu_2\right)}{\partial \lambda_t} = \frac{\partial \log p_F\left(x_t; v_t, \nu_1, \nu_2\right)}{\partial v_t} \times \frac{\partial v_t}{\partial \lambda_t} \\
&= -\frac{\nu_1}{2v_t} + \frac{\nu_1 + \nu_2}{2}\left[\tilde{w}_t \frac{\nu_1 x_t}{(\nu_2 - 2)(v_t)^2}\right] \times v_t \\
&= \frac{\nu_1}{2e^{2\lambda_t}}\left[\frac{\nu_1 + \nu_2}{\nu_2 - 2}\frac{x_t}{\tilde{w}_t} - e^{\lambda_t}\right] \times e^{\lambda_t} \\
&= \frac{\nu_1}{2e^{\lambda_t}}\left[\frac{\nu_1 + \nu_2}{\nu_2 - 2}\frac{x_t}{\tilde{w}_t} - e^{\lambda_t}\right]
\end{aligned}$$

□

As in the correlation model (cf. Section 2.2.2), we scale the score by the inverse of the Fisher information of the χ^2 density. Thus, we get:

Proposition A.5. *For the density in Eq. (A.5), the scaled score* $s_t^F = (\mathcal{I}_{t|t-1}^{W_1})^{-1}\nabla_t^F$ *is given by:*

$$s_t^F = \frac{1}{e^{\lambda_t}}\left[\frac{\nu_1 + \nu_2}{\nu_2 - 2}\frac{x_t}{\tilde{w}_t} - e^{\lambda_t}\right] \tag{A.9}$$

B Proposition 2.1

Proof. We need to compute $\frac{\partial l(X_t)}{\partial f_t'} = \frac{\partial l(X_t)}{\partial \mathsf{vech}(Q_t)^\intercal}$, where l is given by Eq. (2.17)

$$l\left(X_t\right) = \frac{1}{2}d_X(k, \nu) + \frac{\nu - k - 1}{2}\log|X_t| - \frac{\nu}{2}\log|V_t| - \frac{\nu}{2}tr\left(V_t^{-1}X_t\right).$$

Thanks to the chain rule, we can split our equation as:

$$\frac{\partial l\left(X_t\right)}{\partial \mathsf{vech}\left(Q_t\right)'} = \frac{\partial l\left(X_t\right)}{\partial \mathsf{vec}\left(V_t\right)'}\frac{\partial \mathsf{vec}\left(V_t\right)}{\partial \mathsf{vec}\left(R_t\right)'}\frac{\partial \mathsf{vec}\left(R_t\right)}{\partial \mathsf{vec}\left(Q_t\right)'}\frac{\partial \mathsf{vec}\left(Q_t\right)}{\partial \mathsf{vech}\left(Q_t\right)'}.$$

Then, starting with the first term and considering $d\log|X| = tr\left(X^{-1}\right)dX$ and $d\left(X^{-1}\right) = -X^{-1}\left(dX\right)X^{-1}$, see Magnus and Neudecker (1999),

$$\begin{aligned}\frac{\partial l\left(X_t\right)}{\partial \mathsf{vec}\left(V_t\right)'} &= -\frac{\nu}{2}\left[\mathsf{vec}\left(V_t^{-1}\right)' - \mathsf{vec}\left(X_t\right)'\left(V_t^{-1}\otimes V_t^{-1}\right)\right] \\ &= \frac{\nu}{2}\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right]'\left(V_t^{-1}\otimes V_t^{-1}\right).\end{aligned}$$

The second term is, thanks to the fact that $\mathsf{vec}\left(AXB\right) = \left(B'\otimes A\right)\mathsf{vec}\left(X\right)$

$$\frac{\partial \mathsf{vec}\left(V_t\right)}{\partial \mathsf{vec}\left(R_t\right)'} = \frac{\partial \mathsf{vec}\left(D_tR_tD_t\right)}{\partial \mathsf{vec}\left(R_t\right)'} = \left(D_t\otimes D_t\right).$$

By definition of duplication matrix, see Abadir and Magnus (2005), we have that

$$\frac{\partial \mathsf{vec}\left(Q_t\right)}{\partial \mathsf{vech}\left(Q_t\right)'} = \mathcal{D}_k.$$

The third term is a little bit more complicated, indeed defining $\Delta_t = \left(diag(Q_t)^{1/2}\right)$,

$$\begin{aligned}d\mathsf{vec}\left(R_t\right) =& d\mathsf{vec}\left(\Delta_t^{-1}Q_t\Delta_t^{-1}\right) \\ =& \Delta_t^{-1}\otimes\Delta_t^{-1}d\mathsf{vec}\left(Q_t\right) + \mathsf{vec}\left(d\left(\Delta_t^{-1}\right)Q_t\Delta_t^{-1}\right) + \mathsf{vec}\left(\Delta_t^{-1}Q_t d\left(\Delta_t^{-1}\right)\right) \\ =& \Delta_t^{-1}\otimes\Delta_t^{-1}d\mathsf{vec}\left(Q_t\right) + \left[\left(\Delta_t^{-1}Q_t\otimes I\right) + \left(I\otimes\Delta_t^{-1}Q_t\right)\right]d\mathsf{vec}\left(\Delta_t^{-1}\right) \\ =& \Delta_t^{-1}\otimes\Delta_t^{-1}d\mathsf{vec}\left(Q_t\right) - \left[\left(\Delta_t^{-1}Q_t\otimes I\right) + \left(I\otimes\Delta_t^{-1}Q_t\right)\right]\Delta_t^{-1}\otimes\Delta_t^{-1}d\mathsf{vec}\left(\Delta_t\right) \\ =& \Delta_t^{-1}\otimes\Delta_t^{-1}d\mathsf{vec}\left(Q_t\right) - \left[\left(\Delta_t^{-1}Q_t\otimes I\right) + \left(I\otimes\Delta_t^{-1}Q_t\right)\right]\Delta_t^{-1}\otimes\Delta_t^{-1}W_Q d\mathsf{vec}\left(Q_t\right),\end{aligned}$$

where $q_t = \mathsf{vec}\left(\Delta_t\right)$ and W_Q is a diagonal matrix with its ith diagonal elements equal to $1/2\sqrt{q_t^{(i)}}$ if $q_t^{(i)} \neq 0$ and zero otherwise.

Given these four results combined with the fact that $\left(D_t\otimes D_t\right)$ and $\left(V_t^{-1}\otimes V_t^{-1}\right)$ are symmetric , we get

$$\begin{aligned}\frac{\partial l\left(X_t\right)}{\partial \mathsf{vec}\left(V_t\right)'} =& \frac{\nu}{2}\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right]'\left(V_t^{-1}\otimes V_t^{-1}\right)\left(D_t\otimes D_t\right)\left(\Delta_t^{-1}\otimes\Delta_t^{-1}\right) \\ &\times\left[I - \left(Q_t\Delta_t^{-1}\otimes I\right) + \left(I\otimes Q_t\Delta_t^{-1}\right)\right]W_Q\mathcal{D}_k \\ =& \frac{\nu}{2}\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right]'\left(D_t^{-1}\Delta_tQ_t^{-1}\otimes D_t^{-1}\Delta_tQ_t^{-1}\right)\Psi_t\mathcal{D}_k,\end{aligned}$$

where $\Psi_t = \left[I - \left[\left(\Delta_t^{-1}Q_t\otimes I\right) + \left(I\otimes\Delta_t^{-1}Q_t\right)\right]W_Q\right]$. □

C Proposition 2.2

Proof. Starting with the definition of Fisher information matrix:

$$\begin{aligned}
E\left[\nabla_t^W \nabla_t^{W'} | \mathcal{F}_{t-1}\right] =& E\big[\frac{\nu^2}{4}\mathcal{D}_k'\Psi_t'\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\left[\mathsf{vec}\left(X_t\right)-\mathsf{vec}\left(V_t\right)\right] \\
&\times\left[\mathsf{vec}(X_t)-\mathsf{vec}(V_t)\right]'\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\Psi_t\mathcal{D}_k|\mathcal{F}_{t-1}\big] \\
=& \frac{\nu^2}{4}\mathcal{D}_k'\Psi_t'\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\times \\
&\times Var\left(\mathsf{vec}(X_t)-\mathsf{vec}(V_t)|\mathcal{F}_{t-1}\right)\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\Psi_t\mathcal{D}_k \\
=^{*}& \frac{\nu^2}{4}\mathcal{D}_k'\Psi_t'\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\frac{1}{\nu}\left(I_{k^2}+K_k\right)\times \\
&\times\left(V_t\otimes V_t\right)\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\Psi_t\mathcal{D}_k \\
=^{**}& \frac{\nu}{2}\mathcal{D}_k'\Psi_t'\left(D_t^{-1}\Delta_t Q_t^{-1}D_t\Delta_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}D_t\Delta_t^{-1}\right)\mathcal{D}_k\mathcal{D}_k^{+}\Psi_t\mathcal{D}_k
\end{aligned}$$

where * is thank to the vech formulation of the Wishart variance in Abadir and Magnus (2005) and ** is given by $2\mathcal{D}_k\mathcal{D}_k^{+}=\left(I_k+K_k\right)$ (recall that $V_t=\left(D_t\Delta_t^{-1}Q_t\Delta_t^{-1}D_t\right)$)

$$\begin{aligned}
A &= \left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\left(V_t\otimes V_t\right)\left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right) \\
&= \left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\left(D_t\Delta_t^{-1}Q_t\Delta_t^{-1}D_tD_t^{-1}\Delta_t Q_t^{-1}\otimes D_t\Delta_t^{-1}Q_t\Delta_t^{-1}D_tD_t^{-1}\Delta_t Q_t^{-1}\right) \\
&= \left(D_t^{-1}\Delta_t Q_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}\right)\left(D_t\Delta_t^{-1}\otimes D_t\Delta_t^{-1}\right) \\
&= \left(D_t^{-1}\Delta_t Q_t^{-1}D_t\Delta_t^{-1}\otimes D_t^{-1}\Delta_t Q_t^{-1}D_t\Delta_t^{-1}\right) \\
&= \left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)
\end{aligned}$$

□

D Proposition 2.3

Proof. Multiplying the inverse of 2.19 with the 2.18 we get

$$\begin{aligned}
s_t^W &= \left(\mathcal{D}_k{}'\Psi_t'\left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\Psi_t\mathcal{D}_k\right)^{-1}\mathcal{D}_k'\Psi_t' \\
&\quad\times\left(H_t^{-1}Q_t^{-1}\otimes H_t^{-1}Q_t^{-1}\right)\left[\mathsf{vec}\left(X_t\right)-\mathsf{vec}\left(V_t\right)\right] \\
&= \left(\mathcal{D}_k{}'\Psi_t'\left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\Psi_t\mathcal{D}_k\right)^{-1}\mathcal{D}_k'\Psi_t' \\
&\quad\times\left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)\left(H_t^{-1}\otimes H_t^{-1}\right)\left[\mathsf{vec}\left(X_t\right)-\mathsf{vec}\left(V_t\right)\right] \\
&= \left(\mathcal{D}_k{}'\Psi_t'\left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\Psi_t\mathcal{D}_k\right)^{-1}\mathcal{D}_k'\Psi_t' \\
&\quad\times\left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)\left(H_t^{-1}\otimes H_t^{-1}\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\left[\mathsf{vec}\left(X_t\right)-\mathsf{vec}\left(V_t\right)\right] \\
&= \left(\mathcal{D}_k{}'\Psi_t'\left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\Psi_t\mathcal{D}_k\right)^{-1}\mathcal{D}_k'\Psi_t' \\
&\quad\times\left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\mathcal{D}_k{}^{+}\left(H_t^{-1}\otimes H_t^{-1}\right)\left[\mathsf{vec}\left(X_t\right)-\mathsf{vec}\left(V_t\right)\right]
\end{aligned}$$

Now let us simplify $\mathcal{I}_t^W = \frac{\nu}{2}\mathcal{D}_k'\Psi_t'\left(H_t^{-1}Q_t^{-1}H_t \otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k$. This simplification is not restrictive because the matrix Ψ_t is very sparse, moreover using the approximation $\Psi_t = I$ allows us to define the inverse of $\mathcal{I}_t$ which is not full-rank with the original representation.
We get

$$\begin{aligned}
s_t =& \mathcal{D}_k{}^{+}\left(H_t^{-1}\otimes H_t^{-1}\right)\left[\mathsf{vec}\left(X_t\right) - \mathsf{vec}\left(V_t\right)\right] \\
=& \mathcal{D}_k{}^{+}\left(H_t^{-1}\otimes H_t^{-1}\right)\mathsf{vec}(X_t) - \mathcal{D}_k{}^{+}\left(H_t^{-1}\otimes H_t^{-1}\right)\left(H_t\otimes H_t\right)\mathsf{vec}(Q_t) \\
=& \mathcal{D}_k{}^{+}\left(H_t^{-1}\otimes H_t^{-1}\right)\mathsf{vec}(X_t) - \mathsf{vech}\left(Q_t\right)
\end{aligned}$$

□

E Proposition 2.4

Proof. As in the Wishart case, we can split

$$\frac{\partial l\left(X_t\right)}{\partial \mathsf{vech}\left(Q_t\right)'} = \frac{\partial l\left(X_t\right)}{\partial \mathsf{vec}\left(V_t\right)'}\frac{\partial \mathsf{vec}\left(V_t\right)}{\partial \mathsf{vec}\left(R_t\right)'}\frac{\partial \mathsf{vec}\left(R_t\right)}{\partial \mathsf{vec}\left(Q_t\right)'}\frac{\partial \mathsf{vec}\left(Q_t\right)}{\partial \mathsf{vech}\left(Q_t\right)'}.$$

The only terms which is different from the previous model is the first one. The log-likelihood function is and get, using again $d\log|X| = tr\left(X^{-1}\right)dX$ and $d\left(X^{-1}\right) = -X^{-1}\left(dX\right)X^{-1}$

$$\begin{aligned}
dl_{X_t} =& -\frac{\nu_1}{2}tr\left(V_t^{-1}dV_t\right) - \frac{\nu_1+\nu_2}{2}tr\left(\tilde{W}_t^{-1}d\tilde{W}_t\right) \\
=^{*}& -\frac{\nu_1}{2}\left(\mathsf{vec}V_t^{-1}\right)'d\mathsf{vec}V_t + \frac{\nu_1+\nu_2}{2}tr\left(\tilde{W}_t^{-1}\frac{\nu_1}{\nu_2-k-1}V_t^{-1}dV_tV_t^{-1}X_t\right) \\
=& -\frac{\nu_1}{2}\left(\mathsf{vec}V_t^{-1}\right)'d\mathsf{vec}V_t + \frac{\nu_1+\nu_2}{2}tr\left(\frac{\nu_1}{\nu_2-k-1}V_t^{-1}X_t\tilde{W}_t^{-1}V_t^{-1}dV_t\right) \\
=& -\frac{\nu_1}{2}\left(\mathsf{vec}V_t^{-1}\right)'d\mathsf{vec}V_t + \frac{\nu_1+\nu_2}{2}\mathsf{vec}\left(\frac{\nu_1}{\nu_2-k-1}V_t^{-1}X_t\tilde{W}_t^{-1}V_t^{-1}\right)'d\mathsf{vec}V_t
\end{aligned}$$

hence we obtain

$$\begin{aligned}
\frac{\partial l\left(X_t\right)}{\partial \mathsf{vec}\left(V_t\right)'} =& -\frac{\nu_1}{2}\left(\mathsf{vec}V_t^{-1}\right)' + \frac{\nu_1+\nu_2}{2}\mathsf{vec}\left(\frac{\nu_1}{\nu_2-k-1}V_t^{-1}X_t\tilde{W}_t^{-1}V_t^{-1}\right)' \\
=& \frac{\nu_1}{2}\left[\left(V_t^{-1}\otimes V_t^{-1}\right)\left(\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right) - \mathsf{vec}\left(V_t\right)\right)\right]'.
\end{aligned}$$

Combining all the formulas together we get

$$\begin{aligned}\frac{\partial l\left(X_t\right)}{\partial \mathsf{vech}\left(Q_t\right)'} =& \frac{\nu_1}{2}\left(\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right)-\mathsf{vec}\left(V_t\right)\right)'\left(V_t^{-1}\otimes V_t^{-1}\right)\left(D_t\otimes D_t\right)\\ &\times\left(\Delta_t^{-1}\otimes\Delta_t^{-1}\right)\Psi_t\mathcal{D}_k\\ =& \frac{\nu_1}{2}\left(\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right)-\mathsf{vec}\left(V_t\right)\right)'\\ &\times\left(D_t^{-1}\Delta_tQ_t^{-1}\Delta_tD_t^{-1}\otimes D_t^{-1}\Delta_tQ_t^{-1}\Delta_tD_t^{-1}\right)\left(D_t\Delta_t^{-1}\otimes D_t\Delta_t^{-1}\right)\Psi_t'\mathcal{D}_k'\\ =& \frac{\nu_1}{2}\left(\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right)-\mathsf{vec}\left(V_t\right)\right)'\left(H_t^{-1}Q_t^{-1}\otimes H_t^{-1}Q_t^{-1}\right)\end{aligned}$$

□

F Proposition 2.5

Proof. Consider the multiplication $\mathcal{I}_t^W\nabla_t^F$, considering the same approximation of $\mathcal{I}_t^W$ we used in D,

$$\begin{aligned}s_t^F =& \left(\mathcal{D}_k'\left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\right)^{-1}\mathcal{D}_k'\\ &\times\left(H_t^{-1}Q_t^{-1}\otimes H_t^{-1}Q_t^{-1}\right)\left[\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right)-\mathsf{vec}\left(V_t\right)\right]\\ =& \left(\mathcal{D}_k'\left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)\mathcal{D}_k\right)^{-1}\mathcal{D}_k'\left(H_t^{-1}Q_t^{-1}H_t\otimes H_t^{-1}Q_t^{-1}H_t\right)\\ &\times\mathcal{D}_k\mathcal{D}_k{}^{+}\left(H_t^{-1}\otimes H_t^{-1}\right)\left[\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vec}\left(X_t\tilde{W}_t^{-1}\right)-\mathsf{vec}\left(V_t\right)\right]\\ =& \frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vech}\left(H_t^{-1}X_t\tilde{W}_t^{-1}H_t^{-1}\right)-\mathsf{vech}\left(Q_t\right)\end{aligned}$$

□

G Proposition 2.6

Proof. The proof is straightforward:

$$\begin{aligned}\lim_{\nu_2\to\infty}s_t^F =& \lim_{\nu_2\to\infty}\frac{\nu_1+\nu_2}{\nu_2-k-1}\mathsf{vech}\left(H_t^{-1}X_t\tilde{W}_t^{-1}H_t^{-1}\right)-\mathsf{vech}\left(Q_t\right)\\ =& \mathsf{vech}\left(H_t^{-1}X_t\tilde{W}_t^{-1}H_t^{-1}\right)-\mathsf{vech}\left(V_t\right)=s_t^W\end{aligned}$$

since $\tilde{W}_t=I_k+\frac{\nu_1}{\nu_2-k-1}V_t^{-1}X_t$ □

H Figures of Section (3.2)

In this section, we report the figures of the experiment based on the misspecified DGP's presented in Section (3.2)

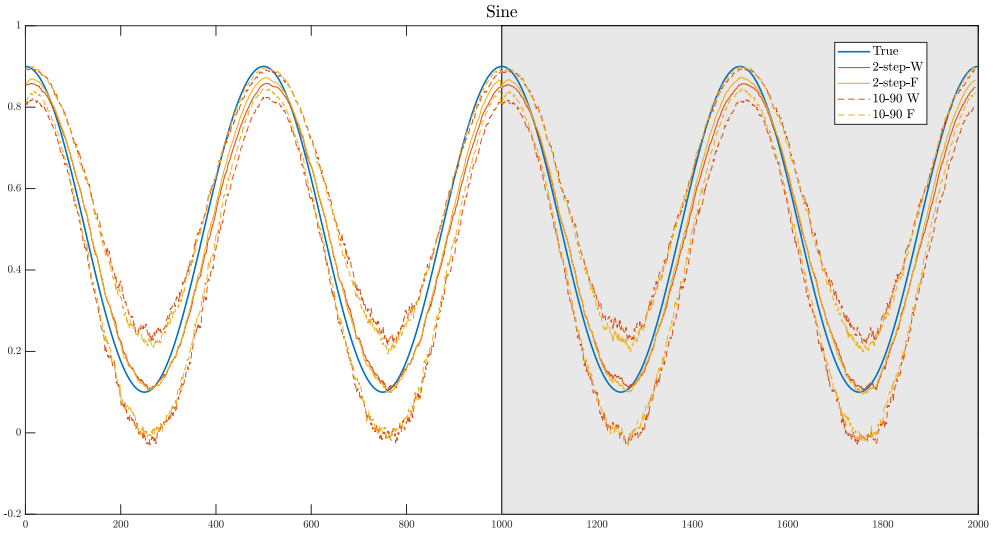

Figure 8: In-sample and out-of-sample filtered estimates of $\rho_t^{(1)}$ from both the 2-step-W and 2-step-F models. Estimates are averaged over the 1000 simulations. Confidence bands are constructed by computing the 10% and 90% empirical quantiles.

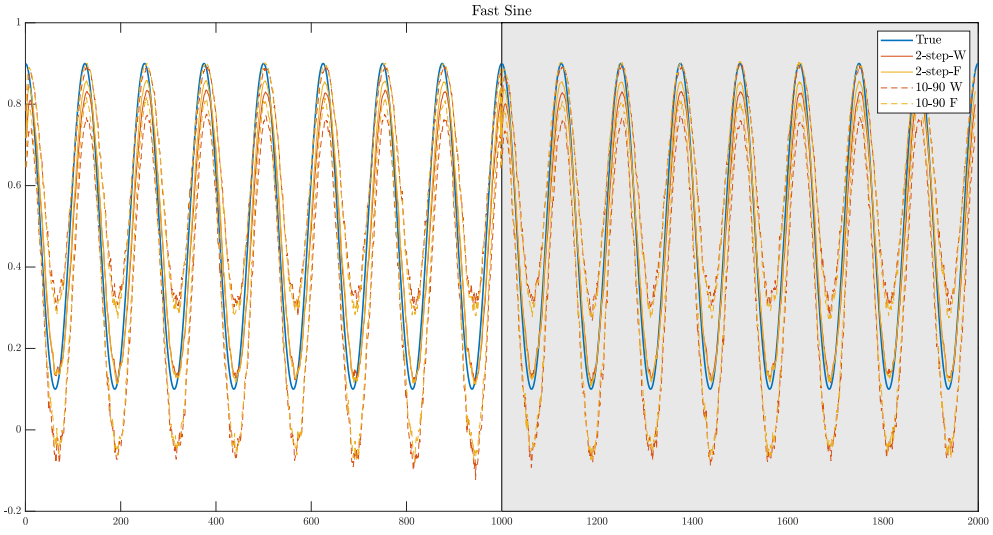

Figure 9: In-sample and out-of-sample filtered estimates of $\rho_t^{(2)}$ from both the 2-step-W and 2-step-F models. Estimates are averaged over the 1000 simulations. Confidence bands are constructed by computing the 10% and 90% empirical quantiles.

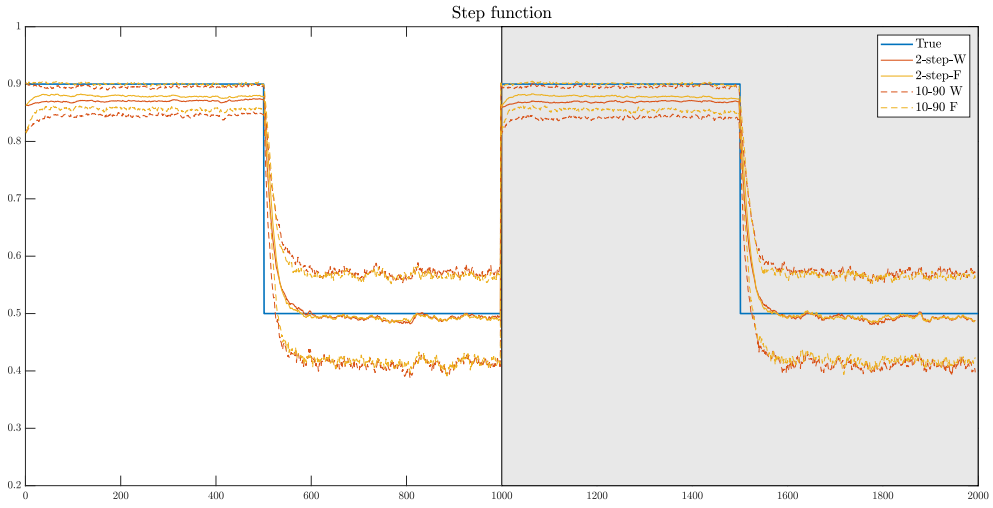

Figure 10: In-sample and out-of-sample filtered estimates of $\rho_t^{(3)}$ from both the 2-step-W and 2-step-F models. Estimates are averaged over the 1000 simulations. Confidence bands are constructed by computing the 10% and 90% empirical quantiles.

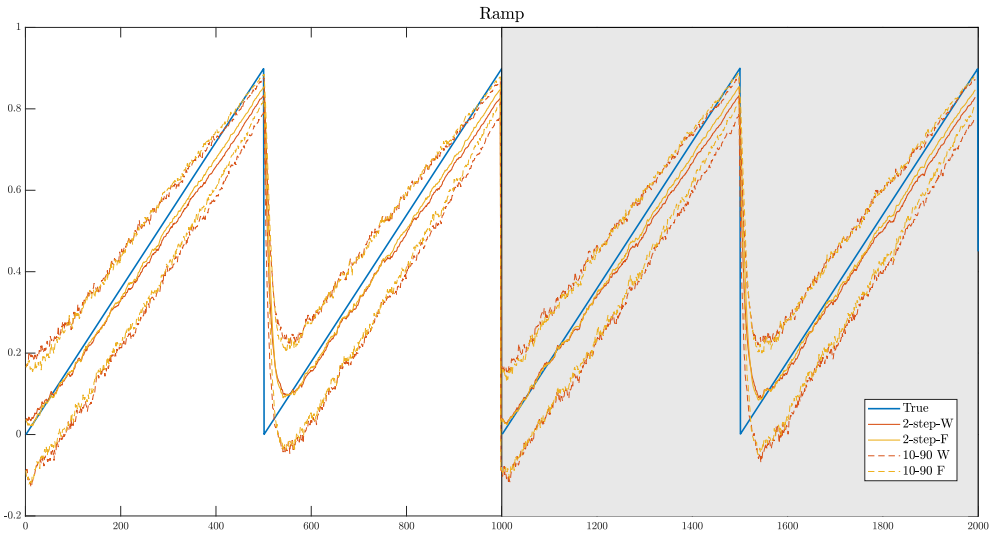

Figure 11: In-sample and out-of-sample filtered estimates of $\rho_t^{(4)}$ from both the 2-step-W and 2-step-F models. Estimates are averaged over the 1000 simulations. Confidence bands are constructed by computing the 10% and 90% empirical quantiles.