

TI 2014-073/IV
Tinbergen Institute Discussion Paper



New HEAVY Models for Fat-Tailed Returns and Realized Covariance Kernels

Pawel Janus¹

André Lucas²

Anne Opschoor²

Dick J.C. van Dijk³

¹ *UBS Global Asset Management;*

² *Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute;*

³ *Erasmus School of Economics, Erasmus University Rotterdam, Tinbergen Institute, the Netherlands..*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

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

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

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

New HEAVY Models for Fat-Tailed Realized Covariances and Returns*

Pawel Janus^a, André Lucas^b, Anne Opschoor^b, Dick J. van Dijk^c

^a *UBS Global Asset Management*

^b *VU University Amsterdam and Tinbergen Institute*

^c *Erasmus University Rotterdam and Tinbergen Institute*

This version: August 19, 2015

Abstract

We develop a new dynamic model for fat-tailed realized covariance matrix and daily return observations. The score dynamics for the unobserved true covariance matrix in our model are robust to outliers and incidentally large observations in both types of data by assuming a matrix- F distribution for the realized covariance measures and a multivariate Student's t distribution for the daily returns. The filter for the unknown covariance matrix has a convenient matrix formulation, which makes the model computationally highly efficient. We formulate convenient parameter restrictions for stationarity and positive definiteness. A simulation exercise shows that the model is able to capture unknown volatility dynamics even if the model is mis-specified. The model is applied to daily equity returns and realized covariance matrices of up to 30 dimensions. The new model both statistically and economically outperforms competing multivariate volatility models out-of-sample.

Keywords: matrix- F distribution; multivariate volatility; realized covariance matrices; heavy tails; generalized autoregressive score (GAS) dynamics.

Classification codes: C32, C58.

*We appreciate the comments of participants at the 8th Annual SoFiE Conference (Aarhus, June 2015), the 9th meeting of the NESG (Tilburg, June 2014) and seminar participants at the Econometrics Brown Bag Seminar Series at VU University Amsterdam. Lucas and Opschoor thank the Dutch National Science Foundation (NWO, grant VICI453-09-005) for financial support. The views expressed herein are those of the authors and not necessarily those of UBS, which does not accept any responsibility for the contents and opinions expressed in this paper. Email correspondence: pawel.janus@ubs.com, a.lucas@vu.nl, a.opschoor@vu.nl, djvandijk@ese.eur.nl.

1 Introduction

A substantial body of literature focusses on modeling volatilities and correlations of financial asset returns.; see Bauwens *et al.* (2006) and Asai *et al.* (2006) for surveys. More recently, the increasing availability of intraday data has led to the introduction of new types of volatility models that include so-called ‘realized measures’ of variances and covariances. These new models lead to more accurate measurements and forecasts of the conditional variance of daily financial returns. Examples of such models in the univariate case are the Multiplicative Error Model (MEM) model (Engle and Gallo, 2006), the HEAVY (High-frEQUENCY-bAsed VolatilitY) model (Shephard and Sheppard, 2010), and the Realized GARCH model (Hansen *et al.*, 2012). These models consist of dynamic specifications for both returns and realized (variance) measures. In the multivariate context, Noureldin *et al.* (2012) extend the univariate version of the HEAVY model, while Gouriéroux *et al.* (2009) develop a Wishart autoregressive (WAR) model for the covariance matrix and apply it to realized covariance matrices, discarding the daily return observations. Likewise, the multivariate volatility models of Chiriac and Voev (2011) and Bauer and Vorkink (2011) also focus on (multivariate) realized measures only, as does the Conditional Autoregressive Wishart (CAW) model of Golosnoy *et al.* (2012).

The asset prices used to estimate the above models are typically subject to the presence of fat-tails and jumps. These may not only affect the daily return observations, but also the realized measures. In particular, depending on the chosen estimator, the realized measure estimates either integrated variance, or both integrated variance and variation due to jumps. The latter may obviously substantially inflate realized measures occasionally whenever jumps occur; see for example Lee and Mykland (2008) on the estimation of spot variances in the presence of jumps. Huang and Tauchen (2005) show the importance of jumps and argue that they account for up to 7% of S&P 500 index variation.

None of the methods described earlier shows how to deal with fat tails in the realized covariance measures and returns simultaneously. In fact, most of the earlier methods center around the use of a Wishart distribution for the realized covariance matrix. The Wishart distribution is rather ill-suited to handle outliers and incidental large observations. In this paper we therefore develop a new model for the covariance matrix dynamics based on joint

measurements of possibly fat-tailed intraday-based realized covariance matrices and daily returns. Our set-up is particularly suitable for cases where no explicit robustification methods are applied while estimating realized measures. The new model is observation driven, thus allowing for easy likelihood evaluation, estimation, and inference. We describe the dynamics of the unobserved true daily return covariance matrix by adopting the generalized autoregressive score framework (GAS) of Creal *et al.* (2011, 2013); see also Harvey (2013). The GAS framework uses the score of the conditional density function to drive the dynamics of the time-varying parameters, which in our case is the unknown covariance matrix. Score driven dynamics possess information theoretic optimality properties even if the model is mis-specified; see Blasques *et al.* (2015). The framework has been successfully applied in the recent literature to a variety of different areas.¹

The key ingredient in our dynamic modeling framework for realized covariance matrices is the matrix- F distribution. For an introduction to the matrix- F distribution, see for example Konno (1991). Though the matrix- F distribution has been around for some time, we have not found any applications to economic or financial data. This is the more surprising given the typical fat-tailed nature of such data. Incidental large observations may easily corrupt the estimated dynamic pattern of the underlying covariance matrix if modeling distributions with relatively thin tails are used; see Creal *et al.* (2011), Janus *et al.* (2014), Harvey (2013), and Lucas *et al.* (2014). The matrix- F distribution provides a coherent approach to address such sensitivities.

The use of the matrix- F distribution together with the GAS dynamics of Creal *et al.* (2013) automatically yields a robust recursive method for filtering the covariance matrix dynamics. The form of this recursion is new, and a direct generalization of both Wishart (thin-tailed) dynamics, and multivariate Student's t (vector rather than matrix) dynamics. By a suitable choice of scaling, our recursion retains a convenient matrix format, rendering our approach numerically highly efficient, also in higher dimensional settings. The matrix format of our recursion contrasts sharply with the approaches of Lucas *et al.* (2014)

¹For example, Creal *et al.* (2011) use the GAS framework to model volatilities and correlations in stock returns; Lucas *et al.* (2014) develop new dynamic copula models under skewness and fat tails and apply this to systemic risk measurement; Harvey and Luati (2014) describe a new framework for dynamic local level models and state filtering based on scores; Creal *et al.* (2014) introduce observation driven mixed measurement dynamic factor models to describe default and loss-given-default dynamics; Andres (2014) studies score driven models for positive random variables; and Oh and Patton (2013) study high dimensional factor copula models based on GAS dynamics for systemic risk measurement.

or Hansen *et al.* (2014), which become infeasible in higher dimensions due to the use of vectorization and subsequent scaling operations.

We establish intuitive parameter restrictions for the stationarity and ergodicity properties of our new model. In particular, we show that our stochastic recurrence relation can be seen as a special case of the semi-polynomial Markov chains studied by Boussama (2006). The stationarity of the model then hinges on the simple and intuitive condition that the autoregressive roots in the GAS recursion lie outside the unit circle. In addition we show that the positivity of the filtered covariance matrices can easily be ensured.

We illustrate the performance of the new method in both a controlled simulation environment and in an empirical application. The simulation results indicate that the model is able to capture deterministic volatility patterns. In addition, when the (co)variance dynamics are generated by a stochastic volatility (SV) type process, the model outperforms existing multivariate models such as the CAW model of Golosnoy *et al.* (2012) and the normal-Wishart model of Hansen *et al.* (2014).

In our empirical application, we use the new model to describe daily returns and daily realized measures of 30 equities from the S&P 500 index over the period January 2001 to July 2014. We show that the volatilities and correlations estimated by our model produce fewer spikes due to incidental large tail observations. The differences follow directly from the fat-tailed nature of the observation densities we assume, and the GAS transition dynamics used to drive the time variation in the true daily covariance matrices. We compare density forecasts based on our matrix- F distribution and the familiar Wishart distribution used in competing model specifications and show that the matrix- F provides a much better fit to the data. In addition, we assess the economic significance by considering mean-variance efficient portfolios. The results indicate that our proposed model significantly produces lower ex-post conditional portfolio standard deviations than competing models.

The rest of this paper is set up as follows. In Section 2, we introduce the new GAS model for multivariate returns and realized covariance matrices under fat-tails. In Section 3, we study the performance of the model in a simulation setting. In Section 4, we apply the model to a high-dimensional panel of 30 daily equity returns and daily realized measures from the S&P 500 index. We conclude in Section 5. The appendix gathers the proofs.

2 Modeling Framework

2.1 The HEAVY GAS tF model

Let $y_t \in \mathbb{R}^k$ denote the vector of (demeaned) asset returns over day t , $t = 1, \dots, T$, and let $RC_t \in \mathbb{R}^{k \times k}$ denote the realized covariance matrix over period t , where RC_t is computed using high-frequency data, e.g., a standard realized covariance matrix estimator based on 5 minute return intervals.² We assume that y_t is fat-tailed and follows a standardized Student's t distribution with ν_0 degrees of freedom and positive definite time varying covariance matrix $V_t \in \mathbb{R}^{k \times k}$. The conditional observation density for y_t is

$$p_y(y_t|V_t, \mathcal{F}_{t-1}; \nu_0) = \frac{\Gamma((\nu_0 + k)/2)}{\Gamma(\nu_0/2) [(\nu_0 - 2)\pi]^{k/2} |V_t|^{1/2}} \times \left(1 + \frac{y_t V_t^{-1} y_t}{\nu_0 - 2}\right)^{-(\nu_0 + k)/2}, \quad (1)$$

where \mathcal{F}_t denotes the information set containing all returns and realized covariances up to and including time t . We assume that $\nu_0 > 2$, such that the covariance matrix exists.

Whereas a Student's t distribution for y_t is fairly standard in the literature, the distribution of the realized covariance matrix RC_t has received much less attention. Typically, one assumes a Wishart distribution for RC_t ; see for example Noureldin *et al.* (2012), Gouriéroux *et al.* (2009), and Golosnoy *et al.* (2012). As the empirical data used in Section 4 show, however, the Wishart distribution is strongly rejected by the data. In particular, for the diagonal elements of RC_t we strongly reject the scaled χ^2 (i.e. univariate Wishart) distribution, whereas we cannot reject a scaled F -distribution. Using this preliminary finding, we assume in this paper that the realized covariance matrices RC_t follow a conditional matrix- F distribution with ν_1 and ν_2 degrees of freedom. The corresponding conditional observation density is

$$p_{RC}(RC_t|V_t, \mathcal{F}_{t-1}; \nu_1, \nu_2) = K(\nu_1, \nu_2) \times \frac{\left|\frac{\nu_1}{\nu_2 - k - 1} V_t^{-1}\right|^{\nu_1/2} |RC_t|^{(\nu_1 - k - 1)/2}}{\left|I_k + \frac{\nu_1}{\nu_2 - k - 1} V_t^{-1} RC_t\right|^{(\nu_1 + \nu_2)/2}}, \quad (2)$$

with positive definite expectation $\mathbb{E}_t[RC_t|\mathcal{F}_{t-1}] = V_t$, degrees of freedom parameters $\nu_1, \nu_2 >$

²Alternatively, one could use realized kernels as in Barndorff-Nielsen *et al.* (2011). For high dimensions such as 15 or 30 stocks as used later in the paper, such methods suffer from increasing biases due to increasingly sparse refresh-time sampling.

$k - 1$, and

$$K(\nu_1, \nu_2) = \frac{\Gamma_k((\nu_1 + \nu_2)/2)}{\Gamma_k(\nu_1/2)\Gamma_k(\nu_2/2)}, \quad \Gamma_k(x) = \pi^{k(k-1)/4} \cdot \prod_{i=1}^k \Gamma(x + (1 - i)/2), \quad (3)$$

where $\Gamma_k(x)$ denotes the multivariate Gamma function; see Konno (1991), Tan (1969), and Gupta and Nagar (2000). The matrix- F distribution is the multivariate analogue of the univariate F distribution, which in turn is a ratio of two independent χ^2 distributions. Similarly, the matrix- F distribution is obtained by considering a Wishart times an inverted-Wishart distributed random matrix. When $\nu_2 \rightarrow \infty$, the matrix- F distribution degenerates to the Wishart distribution (the multivariate analogue of a χ^2 distribution) with ν_1 degrees of freedom.

The conditional observation densities for y_t and RC_t both depend on the common time varying covariance matrix V_t . To describe the dynamics of the unobserved matrix V_t , we use the generalized autoregressive score (GAS) framework of Creal *et al.* (2011, 2013); see also Harvey (2013), Creal *et al.* (2014), Lucas *et al.* (2014), and Hansen *et al.* (2014). The approach is observation driven in the classification of Cox (1981). An important advantage of this approach over parameter driven approaches is that the likelihood function is available in closed form, and therefore estimation and inference by means of maximum likelihood methods are straightforward.

The GAS recursion for V_t is given by

$$V_{t+1} = \Omega + \alpha S_t + \beta V_t, \quad (4)$$

where $S_t \in \mathbb{R}^{k \times k}$ is the scaled score as derived further below, α and β are scalars, and $\Omega \in \mathbb{R}^{k \times k}$ is a matrix of intercepts. A straightforward extension is to let α and β be diagonal matrices, or to follow the set-up of Golosnoy *et al.* (2012) and consider $V_{t+1} = CC' + AS_tA' + BV_tB'$ for triangular parameter matrices A , B , and C .

The recursion in (4) is reminiscent of the dynamic conditional correlation (DCC) recursion of Engle (2002). The main difference is that we use the scaled score S_t rather than the outer products $y_t y_t'$ of past returns. Unlike Creal *et al.* (2011), Lucas *et al.* (2014), and Hansen *et al.* (2014), our score is a matrix-valued rather than a vector-valued variable.

This substantially increases the numerical efficiency of our procedure in high dimensions. Whereas we only need to keep track of matrices of size $k \times k$, these other approaches need to track $\frac{1}{2}k(k+1) \times \frac{1}{2}k(k+1)$ sized matrices due to vectorization and scaling manipulations. Note that further lags of V_t and S_t can be added on the right-hand side of (4), as well as ‘asymmetry’ effects as in Cappiello *et al.* (2006).

We assume that conditional on V_t and \mathcal{F}_{t-1} , returns y_t and realized covariances RC_t are independent.³ The total log-likelihood at time t , denoted by \mathcal{L}_t , equals the sum of the log-likelihood contributions corresponding to the Student’s t and matrix- F densities. Likewise, the score ∇_t of the conditional observation density of (y_t, RC_t) with respect to V_t is equal to the sum of the scores of (1) and (2), i.e.,

$$\mathcal{L}_t = \log p_y(y_t|V_t, \mathcal{F}_{t-1}; \nu_0) + \log p_{RC}(RC_t|V_t, \mathcal{F}_{t-1}; \nu_1, \nu_2), \quad (5)$$

$$\nabla_t = \partial \mathcal{L}_t / \partial V_t = \nabla_{y,t} + \nabla_{RC,t}, \quad (6)$$

$$\nabla_{y,t} = \partial \log p_y(y_t|V_t, \mathcal{F}_{t-1}; \nu_0) / \partial V_t,$$

$$\nabla_{RC,t} = \partial \log p_{RC}(RC_t|V_t, \mathcal{F}_{t-1}; \nu_1, \nu_2) / \partial V_t.$$

This leads to the following result.

Proposition 1. *For the Student’s t density (1) and the matrix- F distribution (2), the corresponding $k \times k$ score matrices $\nabla_{y,t}$ and $\nabla_{RC,t}$ are*

$$\nabla_{y,t} = \frac{1}{2} V_t^{-1} [w_t y_t y_t' - V_t] V_t^{-1} \quad (7)$$

$$\nabla_{RC,t} = \frac{\nu_1}{2} V_t^{-1} \left[\frac{\nu_1 + \nu_2}{\nu_2 - k - 1} RC_t \left(I_k + \frac{\nu_1}{\nu_2 - k - 1} V_t^{-1} RC_t \right)^{-1} - V_t \right] V_t^{-1} \quad (8)$$

where $w_t = (\nu_0 + k) / (\nu_0 - 2 + y_t' V_t^{-1} y_t)$, and where derivatives have been taken with respect to a general non-symmetric matrix V_t rather than a positive definite symmetric matrix V_t .

Proofs of all propositions are provided in the appendix.

Creal *et al.* (2013) propose to scale the score ∇_t to account for the curvature in the log conditional observation density with respect to V_t . They do so using powers of the the

³Extensions to include conditional dependence between a *vector*-valued random variable y_t and a *matrix*-valued random variable RC_t currently appear non-trivial, and we leave such extensions to a future paper.

inverse conditional Fisher information matrix. In the context of our current model with vector-valued and matrix-valued random variables this would lead to a cumbersome and numerically inefficient procedure. We therefore propose a much more straightforward and numerically efficient way to scale the score expressions in equations (7) and (8) while still accounting for curvature of the raw score ∇_t as a function of V_t . In particular, we scale $\text{vec}(\nabla_t)$ by a scalar multiple of $(V_t \otimes V_t)$ and obtain

$$\text{vec}(S_t) \equiv \frac{2}{\nu_1+1} (V_t \otimes V_t) \text{vec}(\nabla_t) = \frac{2}{\nu_1+1} \text{vec}(V_t \nabla_t V_t) \quad \Rightarrow \quad (9)$$

$$S_t = \frac{w_t y_t y_t' - V_t}{\nu_1 + 1} + \frac{\nu_1}{\nu_1 + 1} \left[\frac{\nu_1 + \nu_2}{\nu_2 - k - 1} RC_t \left(I_k + \frac{\nu_1 V_t^{-1} RC_t}{\nu_2 - k - 1} \right)^{-1} - V_t \right], \quad (10)$$

where w_t is defined in Proposition 1, $\text{vec}(\cdot)$ stacks the columns of a matrix into a vector, and \otimes denotes the Kronecker product. Alternative forms of scaling may also be considered, but the computational advantages of (10) are substantial and typically outweigh the numerical complications incurred when considering more complex forms of scaling. This appears most clearly in Sections 3 and 4 when comparing the new approach with the approach of Hansen *et al.* (2014) based on $\frac{1}{2}k(k+1) \times \frac{1}{2}k(k+1)$ matrices compared to our current $k \times k$ recursion in (4). We label the model given by equations (1), (2), (4), and (10) the HEAVY GAS tF model.

Let us now discuss the basic intuition underlying the score equation (10). The first term in (10) relates to the multivariate Student's t distribution and has two important features. First, this score considers the deviations of the weighted outer product $w_t y_t y_t'$ from the local covariance matrix V_t . When $\nu_0 \rightarrow \infty$, i.e., when the Student's t distribution collapses to the normal distribution, the weights collapse to $w_t \equiv 1$ for all t and the dynamics of V_t resemble the covariance dynamics of a multivariate GARCH model, i.e., $y_t y_t' - V_t$. In that case also the scaling matrix $(V_t \otimes V_t)$ is directly proportional to the inverse conditional Fisher information matrix. Second, as discussed in Creal *et al.* (2011), the impact of 'large values' $y_t y_t'$ on V_t is tempered by w_t if the density for y is heavy-tailed, i.e., if $1/\nu_0 > 0$. Put differently, w_t decreases when $y_t' V_t^{-1} y_t$ explodes. This gives the covariance matrix dynamics as driven by the multivariate Student's t distribution an attractive robustness feature. The interpretation is that if y_t is drawn from a heavy-tailed distribution, large values of $y_t y_t'$

could arise as a result of the heavy-tailed nature of the distribution rather than as a result of a substantial change in the underlying covariance matrix. The score based approach automatically accounts for this.

The second term in (10) is new and due to the matrix- F distribution. The expression has a highly similar form and interpretation as the Student's t score discussed before. The main difference is that RC_t is a matrix-valued rather than a vector-valued random variable. Due to the fat-tailedness of the matrix- F distribution, 'large' values of RC_t as measured by $V_t^{-1}RC_t$ do not automatically lead to substantial changes in the covariance matrix V_t . Instead, the matrix 'weight' $(\mathbf{I}_k + \nu_1 V_t^{-1}RC_t / (\nu_2 - k - 1))^{-1}$ takes the same role as w_t in (7) and downweights the impact of a large $V_t^{-1}RC_t$ s on future values of V_t . When $\nu_2 \rightarrow \infty$, the matrix- F distribution collapses to the Wishart distribution with ν_1 degrees of freedom and the second term of the scaled score S_t in (10) collapses to $\nu_1 (\nu_1 + 1)^{-1} (RC_t - V_t)$, which is directly in line with the expressions in Hansen *et al.* (2014). The parameter ν_2 thus takes the same robustification role for the realized covariance measures RC_t as the parameter ν_0 takes for the returns y_t .

Looking at the two terms in (10) simultaneously, the value of ν_1 clearly trades off the relative contributions of the Student's t score and the matrix- F score when updating V_t . If ν_1 is large, the information in RC_t is deemed relatively precise compared to that in the daily return vector y_t . In that case, the score is dominated by the second term in the scaled score. The converse holds if ν_1 is low. In the limit $\nu_1 \rightarrow \infty$, RC_t measures V_t exactly, and the weight of the score part due to y_t drops out entirely. Each of the three parameters ν_0 , ν_1 , and ν_2 thus has a clear and intuitive interpretation in the construction of the score.

We end the model presentation by providing parameter restrictions to ensure positivity of the covariance matrices V_t . We also prove a result on the suitability of the derivative concept used in Proposition 1. Using (9), Proposition 2 provides a simple set of parameter restrictions ensuring that V_t is positive for each time t .

Proposition 2. *Consider the sequence of covariance matrices $\{V_t\}$ generated by equation (4). Assume that the realized measures RC_t are positive semi-definite for each time t . Given the scaled score steps as in equation (10) and given an initial positive definite matrix V_1 and positive semi-definite matrix Ω , then V_t is positive definite for each $t \geq 1$ if $\beta > \alpha > 0$.*

The parameter restrictions $\beta > \alpha > 0$ and Ω positive definite can easily be imposed during the estimation stage.

To formulate our final result in this section, we define the operator $\text{devec}(\cdot)$ as the inverse of $\text{vec}(\cdot)$, i.e., $\text{devec}(\text{vec}(V)) = V$. Similarly, we define the operators $\text{vech}(\cdot)$ and $\text{devech}(\cdot)$, with $\text{vech}(V)$ stacking the lower triangular elements of a symmetric matrix V into a vector, and $\text{devech}(\text{vech}(V)) = V$. Finally, we define the selection matrix \mathcal{S} with 1s and 0s such that $\text{vec}(V) = \mathcal{S} \text{vech}(V)$ for a symmetric matrix V . We now obtain the following result.

Proposition 3. *For symmetric $k \times k$ matrices V and ∇ ,*

$$\text{devec} \left((V^{-1} \otimes V^{-1})^{-1} \text{vec}(\nabla) \right) = \text{devech} \left((\mathcal{S}' (V^{-1} \otimes V^{-1}) \mathcal{S})^{-1} \mathcal{S}' \text{vec}(\nabla) \right). \quad (11)$$

The result in Proposition 3 is important given that in Proposition 1 we ignored the fact that V_t is symmetric when taking derivatives. We subsequently scaled the resulting vectorized derivative $\text{vec}(\nabla_t)$ by a scalar multiple of the inverse of $(V_t^{-1} \otimes V_t^{-1})$ to obtain our score step S_t in (10). Analogously, we scale the score with respect to $\text{vech}(V_t)$, i.e., $\partial \mathcal{L}_t / \partial \text{vech}(V_t) = \mathcal{S}' \text{vec}(\nabla_t)$, by the inverse of $\mathcal{S}' (V_t^{-1} \otimes V_t^{-1}) \mathcal{S}$. Proposition 3 now states that these two approaches yield precisely the same score steps S_t as both ∇_t and V_t are symmetric. It thus appears immaterial whether or not we account for the fact that V_t is symmetric when taking derivatives, as long as we use the appropriate form of scaling. Clearly, this does not necessarily hold for other forms of scaling, and therefore provides a further advantage to our current definition of S_t .⁴

2.2 Stationarity and ergodicity

A useful feature of our HEAVY GAS tF model is that under the assumption of correct specification the scores $\nabla_{y,t}$ and $\nabla_{RC,t}$ are martingale differences by design and therefore have conditional expectation 0. This follows directly from the fact that they are scores of a correctly specified density. To obtain stationarity of V_t , however, we need to study the

⁴One may also follow a more direct line of argument and look upon the steps in Proposition 1 as steepest-ascent steps on the simpler manifold of real, possibly asymmetric matrices rather than on the more complex manifold of positive definite symmetric matrices. Even though the derivatives in Proposition 1 do not *a priori* force the next covariance matrix V_{t+1} to be positive definite, the recursions make clear that all matrices V_t will *ex post* be positive definite almost surely for all times t under appropriate initialization and parameter restrictions; see Proposition 2.

probabilistic properties of the new model as generated by the non-linear recursion (4). We obtain the following result.

Proposition 4. *If $0 < \alpha < \beta < 1$, the process generated by the HEAVY GAS tF model is stationary, geometrically ergodic, and β -mixing.*

A key step in the proof of Proposition 4 is to rewrite the scaled score as

$$V_t^{-1/2} S_t (V_t')^{-1/2} = \frac{(\nu_0 + k)}{(\nu_1 + 1)(\nu_0 - 2)} \varepsilon_{y,t} \varepsilon'_{y,t} \left(1 + \frac{1}{\nu_0 - 2} \varepsilon'_{y,t} \varepsilon_{y,t} \right)^{-1} + \frac{\nu_1 (\nu_1 + \nu_2)}{(\nu_1 + 1)(\nu_2 - k - 1)} \varepsilon_{RK,t} \left(\mathbf{I}_k + \frac{\nu_1}{\nu_2 - k - 1} \varepsilon_{RK,t} \right)^{-1} - \mathbf{I}_k, \quad (12)$$

where $\varepsilon_{y,t}$ has Student's t distribution with mean zero, covariance matrix \mathbf{I}_k , and degrees of freedom ν_0 , and $\varepsilon_{RK,t}$ has a matrix- F distribution with expectation \mathbf{I}_k , and degrees of freedom parameters ν_1 and ν_2 . The right-hand side of (12) does not depend on V_t . Moreover, the terms on the right-hand side are transformations of (matrix) Beta distributed random variables and have finite expectations and variances if $2 < \nu_0 < \infty$, $k - 1 < \nu_1 < \infty$, and $k - 1 < \nu_2 < \infty$; see Tan (1969).

A further inspection of the proof of Proposition 4 and Theorem 2 in Boussama (2006) shows that we can easily generalize the result to models with dynamics of the type

$$V_{t+1} = \Omega + A S_t A' + B V_t B', \quad (13)$$

for $k \times k$ matrices A and B . Such models allow for possible volatility spillover effects. It is also clear from (12) that we can establish the existence of moments for V_t using the feature that the (matrix) Beta random variables are 'bounded' in the appropriate matrix sense. For $0 < \beta < 1$, we then directly obtain the unconditional first moment of V_t as $\mathbb{E}[V_t] = (1 - \beta)^{-1} \Omega$. A number of these features are discussed for the univariate case in Harvey (2013). Proposition 4 generalizes these results to the fully multivariate matrix context.

2.3 Estimation

We collect matrix-valued Ω and scalar-valued $\alpha, \beta, \nu_0, \nu_1, \nu_2$ into the static parameter vector θ and estimate θ by maximum likelihood. To do so, we maximize the log-likelihood $\mathcal{L}_T(\theta) = \sum_{t=1}^T \mathcal{L}_t$, where \mathcal{L}_t was defined in equation (5). The starting value V_1 can be either estimated or set equal to RC_1 . We further reduce the number of parameters following Hansen *et al.* (2014) by using a covariance targeting approach to estimate Ω . As $\Omega = (1 - \beta) \mathbb{E}[V_t]$ for $0 < \beta < 1$, we replace Ω during estimation by $(1 - \beta)$ times the sample mean of RC_t . This should be a consistent estimator for the expectation under a standard ergodicity assumption. Hence we are left only with 5 scalar-valued parameters: $\alpha, \beta, \nu_0, \nu_1$ and ν_2 . The resulting maximum likelihood estimation procedure is fast and numerically efficient. In our empirical section, we use it to estimate the parameters of dynamic systems up to 30 dimensions. Proceeding to even higher dimensional systems should be feasible as well, but is probably better addressed by studying covariance models with factor structures.

3 Simulation experiment

3.1 Monte Carlo analysis based on the correctly specified model

We now perform a Monte Carlo study to investigate the statistical properties of the maximum likelihood estimator for θ . We simulate time series of T daily returns and daily realized covariances of dimension k . We use $T = 500, 1000$, and $k = 5, 15$. We generate data using the HEAVY GAS tF model as the true data generating process (DGP) and set $\alpha = 0.8$, $\beta = 0.97$, $\nu_0 = 12$, $\nu_1 = 22$, and $\nu_2 = 35$. In addition, V_0 is a matrix with $V_{jj} = 4$ ($j = 1, \dots, k$) and $V_{ij} = 4\rho$ ($i \neq j$) with $\rho = 0.7$. The parameters resemble values found in the empirical application of Section 4. For each simulated series, we estimate θ by maximum likelihood.

Table 1 presents the results. All parameters are estimated near their true values. Standard deviations shrink as either the sample size T or the dimension k grows. Interestingly, there appears to be a small bias in A for larger dimensions $k = 15$, especially when the sample size is small. This could be due to the incidental parameter problem, as shown by Engle *et al.* (2008) for the DCC model.

Table 1: Parameter estimations of HEAVY GAS DGP

This table shows Monte Carlo averages and standard deviations (in parentheses) of parameter estimates from simulated HEAVY GAS processes. The table reports the mean and the standard deviation in parentheses based on 4000 replications.

Coef.	True	$k = 5$		$k = 15$	
		$T = 500$	$T = 1000$	$T = 500$	$T = 1000$
Maximum Likelihood					
A	0.80	0.797 (0.036)	0.798 (0.025)	0.787 (0.010)	0.792 (0.007)
B	0.97	0.966 (0.005)	0.968 (0.004)	0.967 (0.003)	0.969 (0.002)
ν_0	12.00	12.419 (2.233)	12.179 (1.460)	12.178 (1.147)	12.096 (0.805)
ν_1	22.00	22.063 (0.794)	22.037 (0.559)	21.998 (0.129)	21.997 (0.092)
ν_2	35.00	35.127 (2.049)	35.054 (1.435)	35.040 (0.374)	35.012 (0.267)

3.2 Monte Carlo analysis based on mis-specified models

One of the main aims of the new HEAVY GAS tF model is to obtain estimates of the unobserved V_t and to do so robustly in the presence of heavy-tailed distributions for the observations y_t and RC_t . Given $\hat{\theta}$, such estimates follow directly from the recursion (4) of the GAS model. To see how well the model does in tracking unknown dynamics of the covariance matrix V_t , we perform the following experiment. First, we consider a deterministic process for the daily volatilities and correlation of a bivariate return vector y_t . Over the t -th day, we simulate n intra-day returns $y_{i,t}$, $i = 1, \dots, n$. The returns are i.i.d. with covariance matrix V_t/n :

$$y_{i,t} \stackrel{i.i.d.}{\sim} N(0, V_t/n),$$

$$\sigma_t = 4 + \sin(2\pi t), \quad \rho_t = 0.5 \sin(2\pi t),$$

where σ_t^2 and $\rho_t \sigma_t^2$ are the variance and covariance at day $t = 1, \dots, T$. Using the intra-day returns, we construct the daily return y_t and the realized covariance matrix RC_t , computed as $\sum_{i=1}^n y_{i,t} y_{i,t}'$. We set $T = 1000$ and $n = 50$.

In a second experiment, we let the (co)variances vary in a stochastic, rather than a deterministic way. This DGP combines the fat-tailedness of returns and realized covariance matrices with stochastic volatility dynamics for the covariance matrix V_t . It does so in the

following way:

$$\begin{aligned}
y_t | \mathcal{F}_{t-1} &\sim t(\nu_0, V_t), & RC_t | \mathcal{F}_{t-1} &\sim F(\nu_1, \nu_2, V_t), \\
V_t &= \bar{V} + \gamma V_{t-1} + \eta_t, & \eta_t &\sim F(\nu_{\eta,1}, \nu_{\eta,2}, V_{0,\eta}), \\
V_0 &= 4 \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}, & &
\end{aligned} \tag{14}$$

with η_t a 2×2 matrix drawn from a matrix- F distribution with mean $V_{0,\eta} = \kappa(1 - \gamma)V_0$, and $\nu_{\eta,1}$ and $\nu_{\eta,2}$ degrees of freedom. We set $\nu_0 = 5$, $\nu_1 = 20$, $\nu_2 = 18$, $\gamma = 0.98$, $T = 1000$, $\nu_{\eta,1} = 8$, $\nu_{\eta,2} = 7$, $\kappa = 5$, and $\bar{V} = (1/2)(1 - \gamma)V_0$. All these values are chosen such that we obtain reasonable volatility and correlation patterns.

We compare our model with two alternatives. First we demonstrate the difference between the fat-tailed matrix- F distribution and the Wishart distribution in the context of the GAS framework, by considering the model of Hansen *et al.* (2014), which we label HJK. Second, we consider the CAW model of Golosnoy *et al.* (2012). This model assumes a Conditional Wishart distribution for RC_t and specifies its dynamics as follows:

$$V_t = \Omega + \alpha RC_t + \beta V_{t-1} \tag{15}$$

which is in fact similar as the observation equation of the Multivariate HEAVY model of Noureldin *et al.* (2012). As in the HEAVY GAS tF model, α and β are scalars and we estimate the matrix Ω by means of covariance targeting. After simulating 1500 paths from the DGP of (14), we report the root mean squared error (RMSE), defined as

$$RMSE = \left[\frac{1}{T} \sum_{t=1}^T \|V_t - \hat{V}_t\|^2 \right]^{1/2} = \left[\frac{1}{T} \sum_{i,j,t} (V_{ij,t} - \hat{V}_{ij,t})^2 \right]^{1/2}, \tag{16}$$

with \hat{V}_t the estimated covariance matrix from a particular model, and $\|\cdot\|$ denoting the (matrix) Frobenius norm.

Figure 1 presents results for one particular realization of the deterministic DGP. The blue lines represents the true values of the volatility, correlation, or covariance. The figure shows that the new model easily recovers the patterns for the volatility, correlation, and

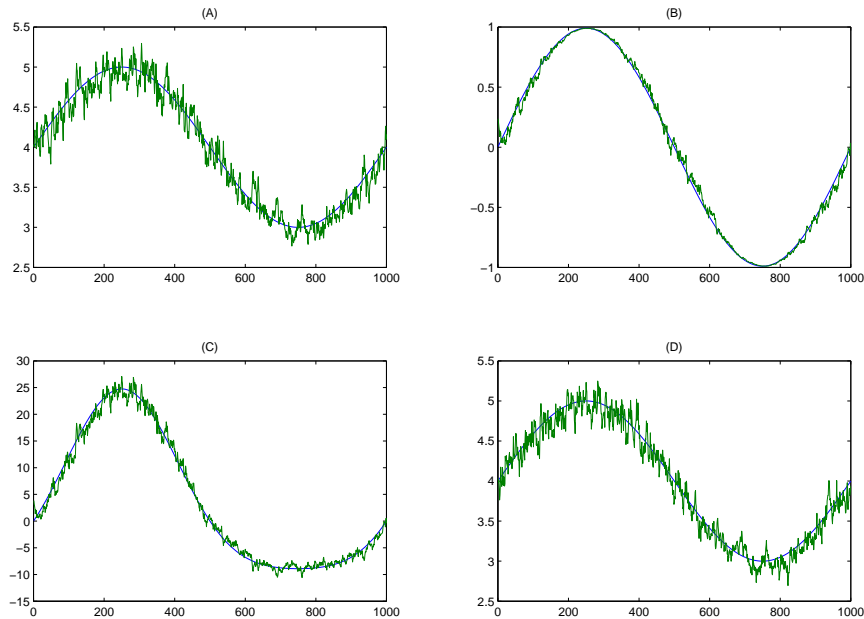


Figure 1: Model fit for a deterministic DGP

This figure shows a realization of the simulated (co)variance process of (14) (blue line) with $T = 1000$ and the fit from the HEAVY GAS tF model (green line). Panels A and D represent the volatilities, while panels B and C present the correlation and covariance, respectively

covariance.

Turning to the stochastic (co)variance case, Table 2 reports the means of the RMSE of the three models with respect to the true DGP. The standard deviations of the means are reported between parentheses. Note that all models are mis-specified in this case. The HEAVY GAS tF model produces the lowest RMSE. Compared to the HJK model, it does so by using the matrix- F distribution rather than the Wishart distribution when estimating the parameters. This also affects the score dynamics of the transition equation for V_t . Compared to the CAW model, both the HJK and the HEAVY GAS tF model do better. Apparently, using the information in the daily returns both for estimating the static model parameters and for filtering the covariance matrix is helpful for estimating the true covariance matrix dynamics. Overall, the results highlight the advantage of modeling the Realized Covariance matrix by a fat-tailed matrix- F distribution in combination with the generalized autoregressive score framework for the matrix dynamics. The impact of large incidental jumps are downweighted by the HEAVY GAS tF model, producing a lower RMSE than the Wishart based models that lack this property.

Table 2: Statistical fit on stochastic DGP

This table shows Monte Carlo averages and standard deviations (in parentheses) of the RMSE of three misspecified models with respect to the true bivariate covariance matrix, which is simulated from the SV process of (14). We compare the the HEAVY GAS tF model with the HJK model and the CAW model. The table reports the mean and the standard deviation of the mean in parentheses based on 1500 replications.

	HGAS tF	HJK	CAW
mean	11.42	12.66	13.49
st.dev.	(0.14)	(0.17)	(0.18)

4 Empirical application: U.S. equity returns

4.1 Data

We apply the HEAVY GAS tF model to daily realized (co)variances and daily returns of 30 randomly chosen U.S. equities from the S&P 500 index over the period January 2, 2001 until July 31, 2014, a total of 3415 observations. Table 3 lists the ticker symbols. For each stock, we observe consolidated trades (transaction prices) extracted from the Trade and Quote (TAQ) database with a time-stamp precision of one second. We first clean the high-frequency data following the guidelines of Brownlees and Gallo (2006) and Barndorff-Nielsen *et al.* (2009). Second, we follow Noureldin *et al.* (2012) and construct Realized Covariance matrices using 5-minute returns with subsampling.

In order to motivate the use of the matrix- F distribution, Table 3 lists p values for Kolmogorov-Smirnov (KS) tests. The tests take the sequence of realized variances for each stock and test whether its distribution is equal to a scaled χ^2 , i.e., the univariate version of the Wishart distribution. We also compute the tests for the null of a scaled F distribution. Table 3 indicates that in all cases the χ^2 or Wishart distribution is strongly rejected by the realized variance data. By contrast, the null hypothesis that the data come from an F distribution is rejected at the 5% significance level for only 5 out of the 30 stocks. Part of this is of course due to the fact that the unconditional distribution of $RC_{ii,t}$ is fatter tailed than a χ^2 due to the time variation in $V_{ii,t}$. However, our subsequent empirical results show that the F distribution also significantly improves upon the χ^2 distribution in a conditional distribution sense.

More insight in the rejection of the Wishart or χ^2 distribution is given in Figure 2. The

Table 3: Kolmogorov-Smirnov test on the distribution of realized variances

This table shows p-values associated with the Kolmogorov-Smirnov test on realized variances of 30 equities. The columns represent the Ticker symbol as well as p-values corresponding with the null hypothesis that RC_t is χ^2 or F distributed.

Ticker	H0: χ^2	H0: F	Ticker	χ^2	F	Ticker	χ^2	F
AA	0.00	0.05	MCD	0.00	0.01	BHI	0.00	0.82
AXP	0.00	0.00	PFE	0.00	0.86	BAC	0.00	0.00
BA	0.00	0.20	PG	0.00	0.22	C	0.00	0.00
CAT	0.00	0.44	WMT	0.00	0.25	DD	0.00	0.11
GE	0.00	0.05	XOM	0.00	0.84	DOV	0.00	0.93
HD	0.00	0.23	AIG	0.00	0.01	DUK	0.00	0.17
HON	0.00	0.19	AEP	0.00	0.48	F	0.00	0.06
IBM	0.00	0.06	ABT	0.00	0.86	JNJ	0.00	0.11
JPM	0.00	0.05	AEE	0.00	0.05	KEY	0.00	0.46
KO	0.00	0.15	BAX	0.00	0.87	LLY	0.00	0.97

left-hand panel shows the histogram of the $RC_{ii,t}$ series of Boeing (BA), as well as the best fitting χ^2 and F distributions. The left-hand panel is truncated at $RC_{ii,t} = 10$. The right panel shows the same three items, but for tail observations of $RC_{ii,t} > 10\%$. Combining the information in the two panels, the overall histogram is peaked at the left and suggests a fat right tail as values larger than 10% occur quite often. In addition, it is clear that the χ^2 distribution neither captures the peak at small values of $RC_{ii,t}$, nor the fat tail for large values of $RC_{ii,t}$. The F distribution on the other hand captures both of these features quite well. This example therefore strongly suggests that the matrix F distribution may lead to an empirically more congruent model than the familiar Wishart distribution when modeling the entire matrix RC_t .

4.2 In-sample performance

Using the full sample of 3415 trading days, we estimate the HEAVY GAS tF model. We compare the outcomes to the same two contemporary benchmarks as in the simulation section: the Wishart based CAW model of Golosnoy *et al.* (2012), see equation (15), and the HJK model of Hansen *et al.* (2014). The HJK model uses GAS dynamics for the the vech-torized form of the Cholesky decomposition of V_t and assumes a conditional normal distribution for the daily returns y_t and a Wishart distribution for the daily realized measures RC_t . The form of scaling adopted by Hansen *et al.* forces them to keep track of scaling

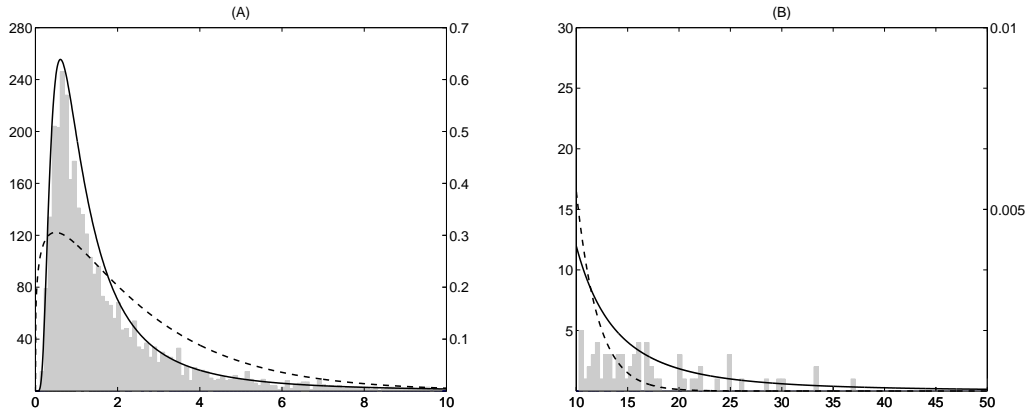


Figure 2: The fit of the probability distribution of the realized variance of BA
This figure shows a histogram of the realized variances (RV) of Boeing over the period 2001-2014. Panel A shows the histogram for $RV < 10$, while Panel B shows the remaining part of the histogram for values of RV larger than 10. The solid and dashed curves present the best fitting F and χ^2 distribution, respectively.

matrices of order $\frac{1}{2}k(k+1) \times \frac{1}{2}k(k+1)$, which makes the model hard to operationalize in higher dimensions. For $k = 15$, for example, this amounts to matrices of size 120×120 , while for $k = 30$, the sizes even become 480×480 . Because of the computational burden, we implement the HJK model only up to dimension $k = 15$.

Table 4 presents the parameter estimates and standard errors. Standard errors are based on the inverse negative hessian of the likelihood evaluated at the optimum. We show the results for two selections of $k = 5$ stocks, a selection of 15 stocks, and the full set of $k = 30$ equities. In addition, we present the log-likelihood values \mathcal{L}^* , the AIC and the BIC values corresponding to the RC_t observations. This is done to make the models comparable, as the CAW model does not have a model equation for the daily return observations.

The results in Table 4 suggest that allowing for fat tails in the distribution of the realized covariances improves the fit. The differences between the log-likelihood of the matrix- F based model and the Wishart based models are substantial and increase rapidly when the dimension becomes larger. Recall that the matrix- F distribution converges to the Wishart distribution if $\nu_2 \rightarrow \infty$. Also the differences in AIC and BIC values are also large and favor the HEAVY GAS tF model for all values of k considered.

Looking at the individual parameter estimates, we first note that the estimates of β are comparably high for the HEAVY GAS tF and the HJK model, and also similar to the persistence parameter $\beta + \alpha$ for the CAW model. This holds for all dimensions $k = 5, 15, 30$ considered. The α parameters cannot be compared directly between the different models.

Table 4: Parameter estimates, likelihoods and loss functions

This table reports maximum likelihood parameter estimates of the HEAVY GAS tF model, the HJK model of Hansen *et al.* (2014) and the CAW model of Golosnoy *et al.* (2012), applied to daily equity returns and/or daily realized covariances. Panels A.1 and A.2 list results for two randomly chosen sets containing five different assets. Panel B considers 15 assets and Panel C shows the results for the full set of 30 assets. Standard errors are provided in parenthesis. We report the total likelihood and the AIC and BIC of the CAW model and the likelihood associated with the Realized Covariance matrix (i.e. the matrix- F and the Wishart distributions) for the GAS models. Data are observed over the period January 2, 2001 until July 31, 2014 ($T = 3415$ observations).

	α	β	ν_0	ν_1	ν_2	\mathcal{L}^*	AIC	BIC
Panel A.1: BA/HD/JPM/PFE/PG								
GAS tF	0.87 (0.02)	0.99 (0.00)	10.20 (0.45)	69.46 (1.52)	34.90 (0.41)	-11,895	23,800	23,831
HJK	0.05 (0.00)	0.97 (0.00)		17.91 (0.10)		-21,747	43,500	43,519
CAW	0.34 (0.00)	0.65 (0.01)		19.74 (0.11)		-18,030	36,065	36,084
Panel A.2: CAT/HON/IBM/MCD/WMT								
GAS tF	0.83 (0.02)	0.99 (0.00)	9.11 (0.36)	72.65 (2.13)	34.03 (0.45)	-4,848	9,707	9,737
HJK	0.04 (0.00)	0.97 (0.00)		19.09 (0.11)		-12,927	25,861	25,879
CAW	0.31 (0.00)	0.67 (0.00)		20.26 (0.11)		-10,046	20,099	20,117
Panel B: AA/AXP/BA/CAT/GE/HD/HON/IBM/JPM/KO/MCD/PFE/PG/WMT/XOM								
GAS tF	0.67 (0.01)	0.99 (0.00)	12.21 (0.38)	140.19 (0.99)	63.98 (0.22)	149,152	-298,295	-298,264
HJK	0.03 (0.00)	0.98 (0.00)		37.47 (0.05)		88,986	-177,966	-177,948
CAW	0.23 (0.00)	0.75 (0.00)		39.32 (0.07)		104,666	-209,327	-209,308
Panel C: All equities (k=30)								
GAS tF	0.52 (0.00)	0.99 (0.00)	13.54 (0.38)	203.66 (0.40)	91.55 (0.14)	938,320	-1,876,631	-1,876,601
CAW	0.17 (0.00)	0.82 (0.00)		57.04 (0.05)		725,787	-1,451,568	-1,451,550

For example, the HJK model takes the vech of the Choleski decomposition of V_t as its time varying parameter, whereas the HEAVY GAS tF and CAW models take V_t itself as the time varying parameter. It is interesting to see that the parameter estimates in Panels A.1 and A.2 are highly similar, despite the fact that they use non-overlapping sets of stocks. The degree of persistence as well as the strength of the dependence of V_t on past values of y_t and RC_t thus seems a shared feature between stocks.

The three degrees of freedom parameters reveal that both the realized measures (ν_2) and the returns (ν_0) are fat-tailed. The degrees of freedom ν_2 may seem high at first sight, but one should realize that the matrix- F distribution requires $\nu_2 > k-1$. On top of that, already moderately large values of ν_2 cause a substantial change in the behavior of the matrix- F distribution compared to the Wishart. The values of ν_0 and ν_2 also moderate the impact of outliers and incidental large observations y_t and RC_t on future values of V_{t+1} . This is clearly seen by the estimated values of ν_1 between the HEAVY GAS tF model and the HJK model. The large values of ν_1 for the HEAVY GAS tF model signal that the model puts almost all attention on the realized kernels RC_t when determining the dynamics of V_t . The information in $y_t y_t'$ is hardly used, particularly in high dimensions ($k = 30$). By contrast, the HJK model still puts about 5% ($k = 5$) to 3% ($k = 15$) of the weight on the score of the distribution for y_t . We can attribute the difference to the fact that the robust filtering approach of the HEAVY GAS tF filter based on the matrix- F distribution provides a much better estimate of the time varying covariance matrix V_t .

Figure 3 plots a small selection of the fitted volatilities and correlations. We show the results for PG and PFE, according to the HEAVY GAS tF model (red line) and the HJK model (blue line). The upper-left and lower-right graphs show the estimated volatilities, while the upper-right and lower-left graphs present the estimated covariances and (implied) correlations, respectively.

The figure shows that the robust transition scheme based on the matrix- F GAS dynamics is successful in mitigating the impact of incidental large RC_t observations on the estimates of V_t . The HJK model, being based on thin-tailed densities, is much more sensitive to such observations. Important episodes where we see large differences are at the start of 2005 for Pfizer (PFE), or around the May 2010 flash crash for Procter & Gamble (PG). Note that in case of real stress periods (such as the financial crisis) the HEAVY GAS tF model

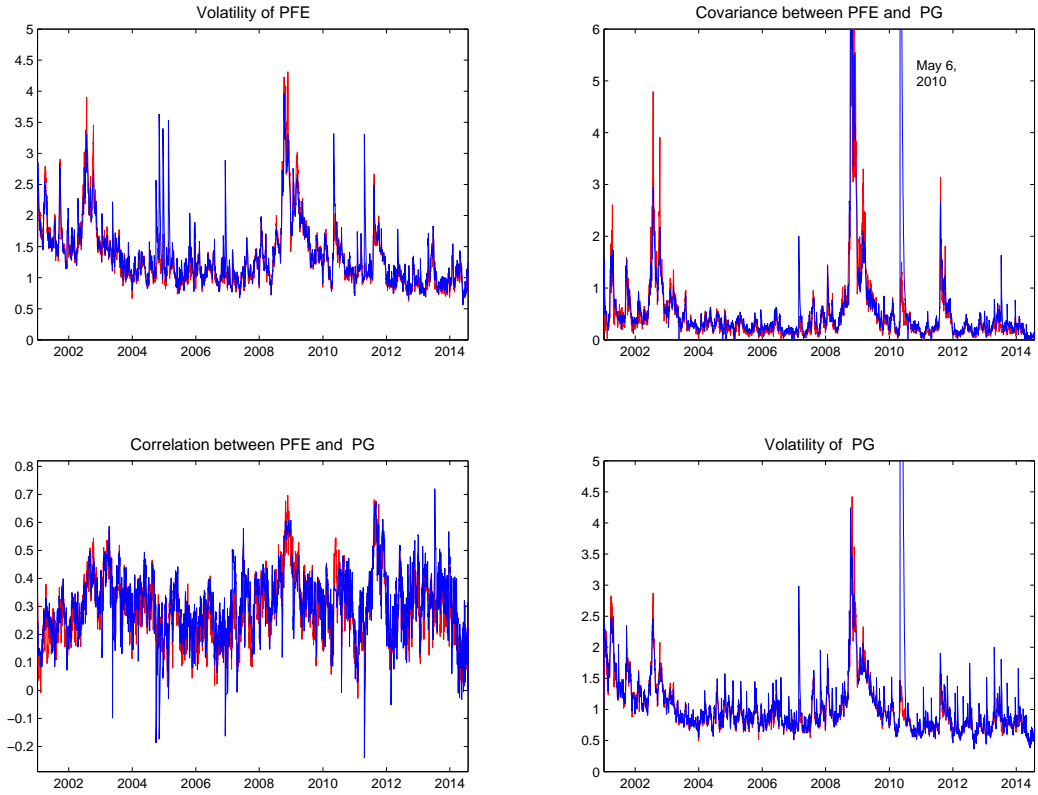


Figure 3: Estimated volatilities and correlations

This figure depicts estimated volatilities of PFE and PG at the main diagonal and their pairwise correlations and covariances at the off-diagonal, estimated by the HJK and HEAVY GAS tF model. The blue line corresponds with the HJK model, while the red line denotes the fit from the HEAVY GAS tF model. The estimation is based on the full sample, which runs from January 2, 2001 until July 31, 2014 (3415 observations).

produces larger volatilities than the HJK model. Interestingly, apart from the main striking differences for Pfizer and Procter & Gamble, we also see a range of other days where the HJK model produces a short-lived spike in the estimated V_t , whereas the fat-tailed HEAVY GAS tF model is much more stable around those times.

The patterns for the correlations and covariances reveal similar features. The correlation pair between PFE and PG clearly displays sudden incidental drops in correlations, for example around 2005, during the flash crash of May 2010, but also at the start of 2003 and the end of 2006. Incidental spikes in the covariances are visible for the HJK model in 2006, 2010 and 2013. Again, the robust HEAVY GAS tF model results in much more stable correlation and covariance patterns that are filtered from the data.

4.3 Out-of-sample performance

We assess the short-term forecasting performance of the models by considering 1-step ahead forecasts. Similar to the in-sample analysis of the previous subsection, we compare the HEAVY GAS tF model with the HJK model and the CAW model. We perform both a statistical and economic application. The former is based on 1-step ahead density forecasts of the realized covariance matrix. Recall that the HJK and CAW models assume a Wishart distribution for the realized covariance matrix, while the Heavy GAS tF model assumes a matrix- F distribution. In addition, we estimate V_t by a simple EWMA scheme

$$V_{t+1} = \beta V_t + (1 - \beta) RC_t,$$

with $\beta = 0.96$, and plug the resulting estimates into the matrix- F density as estimated by the HEAVY GAS tF model.

We use a moving-window approach in the forecasting exercise with an in-sample period of 1500 observations. This corresponds roughly to five calendar years. The out-of-sample period contains $N = 1914$ observations and contains the financial crisis, which therefore constitutes an important test for the robustness of the model. We re-estimate our model repeatedly after each 25 observations, which roughly corresponds to monthly updating of the parameters.

We use the log score (see Mitchell and Hall, 2005; Amisano and Giacomini, 2007) as a scoring rule to differentiate between the density forecasts of the models. Define the difference in log score between the two density forecasts M_1 and M_2 as

$$d_{ls,t} = S_{ls,t}(RC_t, M_1) - S_{ls,t}(RC_t, M_2), \quad (17)$$

for $t = R, R + 1, \dots, T - 1$ with R the length of the estimation window and $S_{ls,t}(RC_t, M_1)$ the log score of the density forecast corresponding to model M_1 at time t , which in turn is denoted as

$$S_{ls,t}(RC_t, M_1) = \log p_t(RC_t | V_t, \mathcal{F}_{t-1}, M_1), \quad (18)$$

where $p_t(\cdot)$ is the probability distribution function of the matrix- F or Wishart distribution respectively. The corresponding null hypothesis on equal predictive ability is given by $H_0 : \mathbb{E}[d_{ls}] = 0$ for all N out-of-sample forecasts. This null can be tested by means of a Diebold and Mariano (1995) (DM) statistic given by

$$t_{ls} = \frac{\bar{d}}{\sqrt{\hat{\sigma}^2/N}} \quad (19)$$

with \bar{d} the out-of-sample average of the log score differences and $\hat{\sigma}^2$ a HAC-consistent variance estimator of the true variance σ^2 of $d_{ls,t}$. Under the assumptions of the framework of Giacomini and White (2006) d_{ls} asymptotically follows a standard Normal distribution. A significantly positive value means that model M_1 has a superior forecast performance over model M_2 .

Rather than only presenting a statistical out-of-sample comparison of the different models, we also provide an economic comparison. Following Chiriac and Voev (2011) we consider global minimum variance portfolios (GMVP), motivated by the mean-variance optimization setting of Markowitz (1952). The model that provides the most accurate forecasts of the covariance matrix should give a lower portfolio variance than the portfolio variance of the competing models. Let us assume that the investor aims to minimize the 1-step ahead portfolio volatility over period $t + 1$ subject to a fully invested portfolio, given his best estimate of the covariance matrix at time t . The resulting GMVP weights $w_{t+1|t}$ are given by the solution of the quadratic problem:

$$\min w'_{t+1|t} V_{t+1|t} w_{t+1|t} \quad \text{s.t.} \quad w'_{t+1|t} \mathbf{1} = 1. \quad (20)$$

Similar as Chiriac and Voev (2011), we assess the predictive ability of the different models by comparing the *ex post* realizations of the conditional standard deviation, which are given by $\sigma_{p,t} = \sqrt{w'_{t+1|t} RC_{t+1} w_{t+1|t}}$. We again test whether the differences in portfolio standard deviation between the different models are significant using the DM test statistic as defined in (19).

Table 5 shows the average values of the log score and the ex-post portfolio standard deviations over the out-of-sample period for two sets of five assets, a set of 15 assets and the

Table 5: Out-of-sample log-scores and ex-post conditional standard deviations

This table shows the mean of log scores, defined in (18) and ex-post portfolio standard deviation, based on 1-step ahead predictions of the covariance matrix, according to the HEAVY GAS tF, HJK and the EMWA model for two pairs of five assets (Panel A), one pair of fifteen assets (Panel B) and for all equities (Panel C, $k = 30$). The highest (lowest) value of the predictive log-score (portfolio standard deviation) across the models are marked bold. In addition, we report HAC based test-statistics on the difference in predictive ability (DM_{DF}) and standard deviation (DM_{σ_p}) between the HEAVY GAS tF model and the other considered models. The superscripts ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The out-of-sample period goes from 2007 until July 2014 and contains 1914 observations.

	GAS tF	HJK	CAW	EMWA
Panel A: BA/HD/JPM/PFE/PG				
S^{ls}	-2.22	-6.39	-4.44	-4.42
DM_{DF}		(7.05)***	(3.41)***	(10.11)***
σ_p	0.791	0.811	0.794	0.794
DM_{σ_p}		(-5.76)***	(-2.77)***	(-2.27)**
Panel B: k = 15				
S^{ls}	57.55	33.70	42.03	48.86
DM_{DF}		(11.37)***	(7.85)***	(9.47)***
σ_p	0.630	0.649	0.632	0.631
DM_{σ_p}		(-10.63)***	(-2.83)***	(-2.34)**
Panel C: All equities (k = 30)				
S^{ls}	313.50		238.97	291.24
DM_{DF}			(16.20)***	(9.75)***
σ_p	0.572		0.574	0.574
DM_{σ_p}			(-2.61)***	(-2.29)**

complete set of 30 assets. In addition, we provide corresponding t -statistics for the difference in the log predictive density scores between the HEAVY GAS tF model and the competing models. Likewise, DM_{σ_p} represents the t -statistic based on the portfolio standard deviations of the HEAVY GAS tF model and the HJK, CAW, or EWMA model.

The results reinforce our earlier analysis, but now in an out-of-sample setting. The test statistics show that the HEAVY GAS tF model significantly outperforms the HJK, CAW and the EMWA models with respect to one step ahead density forecasts. This result has two implications. First, considering a matrix- F distribution for the realized covariance matrix is more accurate in terms of density fit than a Wishart distribution. This is in line with our in-sample results. The results are large and significant. Second, the GAS dynamics based on the matrix- F distribution provide improved forecasting densities. This is for example seen when comparing the EWMA and GAS tF approaches. The improvement from EWMA to GAS tF is roughly similar to the improvement from CAW to GAS tF. The HJK model

appears to be the most problematic in terms of out-of-sample density forecasts. Note that this model not only uses the Wishart rather than the matrix- F distribution for RC_t , but also parameterizes the Choleski matrix of V_t rather than V_t itself. Both features result in a worse density forecast.

Apart from the statistical evidence, the HEAVY GAS tF model also outperforms its competitors in the economic application. For all dimensions, the DM_{σ_p} statistics are negative and statistically significant at the 5% (1%) level in case of the EWMA (CAW/HJK) model, indicating that the HEAVY GAS tF model produces the lowest ex-post portfolio standard deviation compared to the competing models. We conclude that the new model also does well in an out-of-sample context, both statistically and economically.

5 Conclusions

We introduced a new dynamic multivariate HEAVY model that combines return observations and (ex-post) observed realized covariance matrices to estimate the unobserved common underlying covariance matrices. The proposed model explicitly acknowledges that both realized covariance matrices and returns are typically fat-tailed. The proposed setup is particularly suitable for cases where no explicit robustification methods are applied while estimating realized measures. Using the GAS dynamics of Creal *et al.* (2011, 2013) based on a matrix- F distribution for the realized covariance matrices and a Student's t distribution for the returns, we derived an observation driven model for the unobserved covariances with robust propagation dynamics. We proved that stationarity and ergodicity of the model and positive definiteness of the filtered covariance matrices could be ensured under simple and intuitive parameter restrictions.

An important feature of our model is that it retains the matrix format for the transition dynamics of the covariance matrices, unlike score driven models proposed earlier. This makes the model computationally highly efficient. We showed that the model adequately captures both deterministic and stochastic volatility (SV) dynamics. Using U.S. equity data over 2001–2014, the model also improves both the in-sample and out-of-sample fit of the covariance matrices for high-dimensional data sets of up to 30 dimensions. These improvements are both statistically and economically significant and persist over the episodes including the

recent financial crisis. We conclude that the model provides a valuable tool when modeling combinations of fat-tailed matrix-valued and vector-valued random variables. Moreover, the matrix- F distribution used here can also prove useful beyond the scope of the current paper, such as in for instance a Bayesian context.

References

- Abadir, K.M. and J.R. Magnus (2005), *Matrix algebra*, Cambridge University Press.
- Amisano, G. and R. Giacomini (2007), Comparing density forecasts via weighted likelihood ratio tests, *Journal of Business & Economic Statistics* **25**(2), 177–190.
- Andres, P. (2014), Computation of maximum likelihood estimates for score driven models for positive valued observations, *Computational Statistics and Data Analysis* **76**, 34–43.
- Asai, M., M. McAleer and J. Yu (2006), Multivariate stochastic volatility: a review, *Econometric Reviews* **25**(2-3), 145–175.
- Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2009), Realized kernels in practice: trades and quotes, *Econometrics Journal* **12**, 1–32.
- Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2011), Multivariate realized kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, *Journal of Econometrics* **12**, 1–32.
- Bauer, G.H. and K. Vorkink (2011), Forecasting multivariate realized stock market volatility, *Journal of Econometrics* **160**, 93–101.
- Bauwens, L., S. Laurent and J.V.K. Rombouts (2006), Multivariate GARCH models: a survey, *Journal of Applied Econometrics* **21**(1), 79–109.
- Blasques, C., S.J. Koopman and A. Lucas (2015), Information theoretic optimality of observation driven time series models for continuous responses, *Biometrika* **102**(2), 325–343.
- Boussama, F. (2006), Ergodicité des chaînes de markov à valeurs dans une variété algébrique: application aux modèles GARCH multivariés, *Comptes Rendus Mathématique, Académie Science Paris, Serie I* **343**(4), 275–278.
- Brownlees, C.T. and G.M. Gallo (2006), Financial econometric analysis at ultra-high frequency: Data handling concerns, *Computational Statistics and Data Analysis* **51**, 2232–2245.

- Cappiello, L., R.F. Engle and K. Sheppard (2006), Asymmetric dynamics in the correlations of global equity and bond returns, *Journal of Financial econometrics* **4**(4), 537–572.
- Chiriac, R. and V. Voev (2011), Modelling and forecasting multivariate realized volatility, *Journal of Applied Econometrics* **26**, 922–947.
- Cox, D.R. (1981), Statistical analysis of time series: some recent developments, *Scandinavian Journal of Statistics* **8**, 93–115.
- Creal, D., B. Schwaab, S.J. Koopman and A. Lucas (2014), Observation driven mixed-measurement dynamic factor models with an application to credit risk, *Review of Economics and Statistics* **96**(5), 898–915.
- Creal, D., S.J. Koopman and A. Lucas (2011), A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations, *Journal of Business and Economic Statistics* **29**, 552–563.
- Creal, D., S.J. Koopman and A. Lucas (2013), Generalized autoregressive score models with applications, *Journal of Applied Econometrics* **28**, 777–795.
- Diebold, F.X. and R.S. Mariano (1995), Comparing predictive accuracy, *Journal of Business & economic statistics* **13**, 253–263.
- Engle, R (2002), Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* **20**(3), 339–350.
- Engle, R.F. and G.M. Gallo (2006), A multiple indicators model for volatility using intra-daily data, *Journal of Econometrics* **131**, 3–27.
- Engle, R.F., N. Shephard and K. Sheppard (2008), Fitting and testing vast dimensional time-varying covariance models, Working Paper.
- Giacomini, R. and H. White (2006), Tests of conditional predictive ability, *Econometrica* **74**(6), 1545–1578.
- Golosnoy, V., B. Gribisch and R. Liesenfeld (2012), The conditional autoregressive wishart model for multivariate stock market volatility, *Journal of Econometrics* **167**, 211–223.
- Gourieroux, C., J. Jasiak and R. Sufana (2009), The wishart autoregressive process of multivariate stochastic volatility, *Journal of Econometrics* **150**, 167–181.
- Gupta, A.K. and D.K. Nagar (2000), *Matrix variate distributions*, Chapman & Hall/CRC.
- Hansen, P.R., P. Janus and S.J. Koopman (2014), Modeling Daily Covariance: a joint framework for low and high-frequency based measures, Working Paper.
- 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**, 877–906.

- Harvey, A.C. (2013), *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, Cambridge University Press.
- Harvey, A.C. and A. Luati (2014), Filtering with heavy tails, *Journal of the American Statistical Association* **109**, 1112–1122.
- Huang, X. and G. Tauchen (2005), The relative contribution of mumps to total price variance, *Journal of Financial Econometrics* **3**, 456–499.
- Janus, P., S.J. Koopman and A. Lucas (2014), "Long memory dynamics for multivariate dependence under heavy tails, *Journal of Empirical Finance* **29**, 187–206.
- Konno, Y. (1991), A note on estimating eigenvalues of scale matrix of the multivariate F-distribution, *Annals of the Institute of Statistical Mathematics* **43**, 157–165.
- Lee, S. and P.A. Mykland (2008), Jumps in financial markets: a new nonparametric test and jump dynamics, *Review of Financial Studies* **21**, 2535–2563.
- Lucas, A., B. Schwaab and X. Zhang (2014), Conditional euro area sovereign default risk, *Journal of Business and Economic Statistics* **32**(2), 271–284.
- Markowitz, H. (1952), Portfolio selection, *Journal of Finance* **7**(1), 77–91.
- Mitchell, J. and S.G. Hall (2005), Evaluating, comparing and combining density forecasts using the KLIC with an application to the bank of england and niesr fancharts of inflation, *Oxford bulletin of economics and statistics* **67**(s1), 995–1033.
- Noureldin, D., N. Shephard and K. Sheppard (2012), Multivariate high-frequency-based volatility (HEAVY) models, *Journal of Applied Econometrics* **27**, 907–933.
- Oh, D.H. and A.J. Patton (2013), Time-varying systemic risk: evidence from a dynamic copula model of CDS spreads, Duke University Working Paper.
- Shephard, N. and K. Sheppard (2010), Realising the future: forecasting with high-frequency-based volatility (HEAVY) models, *Journal of Applied Econometrics* **25**, 197–231.
- Tan, W.Y. (1969), Note on the multivariate and the generalized multivariate beta distributions, *Journal of the American Statistical Association* **64**(325), 230–241.

Appendix A: Proofs

We use the following matrix calculus results for a general matrix X ,

$$\begin{aligned} dX^{-1} &= -X^{-1}(dX)X^{-1}, & d \log |X| &= \text{tr}(X^{-1} dX), \\ \text{tr}(A'B) &= \text{vec}(A)' \text{vec}(B), & \text{vec}(ABC) &= (C' \otimes A) \text{vec} B, \\ b \otimes a &= \text{vec}(a'b), \end{aligned}$$

with $a, b \in \mathbb{R}^{k \times 1}$, and $\text{tr}(\cdot)$ denoting the trace. See for instance Abadir and Magnus (2005) for these and other useful results.

Proof of Proposition 1: The general form of the score is given by (7) and (8). The relevant parts of the log-likelihood that depend on V_t are

$$\begin{aligned} \ell_{y,t} &= -\frac{1}{2} \log |V_t| - \frac{\nu_0 + k}{2} \log \left(1 + \frac{y_t' V_t^{-1} y_t}{\nu_0 - 2} \right) &= -\frac{1}{2} \log |V_t| - \frac{\nu_0 + k}{2} \log(\tilde{w}_t) \\ \ell_{RC,t} &= -\frac{\nu_1}{2} \log |V_t| - \frac{\nu_1 + \nu_2}{2} \log \left| \mathbf{I}_k + \frac{\nu_1}{\nu_2 - k - 1} V_t^{-1} RC_t \right| &= -\frac{\nu_1}{2} \log |V_t| - \frac{\nu_1 + \nu_2}{2} \log |\tilde{W}_t|, \end{aligned}$$

with $\tilde{w}_t = (1 + (\nu_0 - 2)^{-1} y_t' V_t^{-1} y_t)$ and $\tilde{W}_t = (\mathbf{I}_k + \nu_1 (\nu_2 - k - 1)^{-1} V_t^{-1} RC_t)$. Using the matrix calculus results above, we obtain

$$\begin{aligned} d \ell_{y,t} &= -\frac{1}{2} \text{tr}(V_t^{-1} dV_t) - \frac{\nu_0 + k}{2} \tilde{w}_t^{-1} d\tilde{w}_t \\ &= -\frac{1}{2} \text{tr}(V_t^{-1} dV_t) - \frac{\nu_0 + k}{2} \tilde{w}_t d \frac{y_t' V_t^{-1} y_t}{\nu_0 - 2} \\ &= -\frac{1}{2} (\text{vec } V_t^{-1})' d \text{vec } V_t + \frac{\nu_0 + k}{2} \tilde{w}_t \frac{1}{\nu_0 - 2} y_t' V_t^{-1} dV_t V_t^{-1} y_t \\ &= -\frac{1}{2} (\text{vec } V_t^{-1})' d \text{vec } V_t + \frac{1}{2} \left[\frac{\nu_0 + k}{\nu_0 - 2} \tilde{w}_t y_t' V_t^{-1} \otimes y_t' V_t^{-1} \right] d \text{vec } V_t \\ &= -\frac{1}{2} (\text{vec } V_t^{-1})' d \text{vec } V_t + \frac{1}{2} \left[\frac{\nu_0 + k}{\nu_0 - 2} \tilde{w}_t \text{vec}(V_t^{-1} y_t y_t' V_t^{-1})' \right] d \text{vec } V_t, \end{aligned}$$

such that

$$\frac{\partial \ell_{y,t}}{\partial \text{vec } V_t} = -\frac{1}{2} \text{vec } V_t^{-1} + \frac{1}{2} \left[\frac{\nu_0 + k}{\nu_0 - 2} \tilde{w}_t \text{vec}(V_t^{-1} y_t y_t' V_t^{-1})' \right].$$

Note that we have dealt with V_t in the above derivations as a general rather than a symmetric matrix, for reasons explained in the main text. Omitting the vec operator and rewriting yields the desired result.

For $d\ell_{RC,t}$ we have

$$\begin{aligned}
d\ell_{RC,t} &= -\frac{\nu_1}{2} \text{tr}(V_t^{-1} dV_t) - \frac{\nu_1 + \nu_2}{2} \text{tr}\left(\tilde{W}_t^{-1} d\tilde{W}_t\right) \\
&= -\frac{\nu_1}{2} (\text{vec } V_t^{-1})' d \text{vec } V_t + \frac{\nu_1 + \nu_2}{2} \text{tr}\left(\tilde{W}_t^{-1} \frac{\nu_1}{\nu_2 - k - 1} V_t^{-1} dV_t V_t^{-1} RC_t\right) \\
&= -\frac{\nu_1}{2} (\text{vec } V_t^{-1})' d \text{vec } V_t + \frac{\nu_1 + \nu_2}{2} \text{tr}\left(\frac{\nu_1}{\nu_2 - k - 1} V_t^{-1} RC_t \tilde{W}_t^{-1} V_t^{-1} dV_t\right) \\
&= -\frac{\nu_1}{2} (\text{vec } V_t^{-1})' d \text{vec } V_t + \frac{\nu_1 + \nu_2}{2} \text{vec}\left(\frac{\nu_1}{\nu_2 - k - 1} V_t^{-1} RC_t \tilde{W}_t^{-1} V_t^{-1}\right)' d \text{vec } V_t.
\end{aligned}$$

Consequently,

$$\frac{\partial \ell_{RC,t}}{\partial \text{vec } V_t} = -\frac{\nu_1}{2} \text{vec } V_t^{-1} + \frac{\nu_1 + \nu_2}{2} \text{vec}\left(\frac{\nu_1}{\nu_2 - k - 1} V_t^{-1} RC_t \tilde{W}_t^{-1} V_t^{-1}\right).$$

Again, removing the vec operator yields the desired result. \square

Proof of Proposition 2: We can rewrite (4) using (10) as

$$V_{t+1} = \Omega + (\beta - \alpha) V_t + \alpha \frac{w_t y_t y_t'}{\nu_1 + 1} + \alpha \frac{\nu_1 (\nu_1 + \nu_2)}{(\nu_1 + 1)(\nu_2 - k - 1)} RC_t \left(\mathbf{I}_k + \frac{\nu_1 V_t^{-1} RC_t}{\nu_2 - k - 1}\right)^{-1}. \quad (\text{A.1})$$

As $\beta > \alpha > 0$, the sum of the first two terms $\Omega + (\beta - \alpha) V_t$ is positive definite if V_t is positive definite. As $\alpha > 0$, the third term is positive semi-definite. Finally, given that RC_t is positive semi-definite for every t , also the last term is positive semi-definite. To see this, note that $\nu_2 > k + 1$ and let k_t^* denote the rank of RC_t , with $RC_t = U_t S_t U_t'$ the singular value decomposition of RC_t , with U_t and S_t a $k \times k_t^*$ and a $k_t^* \times k_t^*$ matrix, respectively. For $a \in \mathbb{R}^{k \times 1}$, $c = \nu_1 (\nu_2 - k - 1)^{-1} > 0$, and positive definite V_t , we have

$$\begin{aligned}
a' RC_t (\mathbf{I}_k + c V_t^{-1} RC_t)^{-1} a &= a' U_t S_t U_t' (V_t + c U_t S_t U_t')^{-1} V_t a = \\
&= a' U_t S_t U_t' \left(V_t^{-1} - V_t^{-1} U_t (c^{-1} S_t^{-1} + U_t' V_t^{-1} U_t)^{-1} U_t' V_t^{-1}\right) V_t a = \\
&= a' U_t \left(S_t - S_t (U_t' V_t^{-1} U_t) (c^{-1} S_t^{-1} + U_t' V_t^{-1} U_t)^{-1}\right) U_t' a = \\
&= a' U_t \left(S_t - S_t \left(c^{-1} (U_t' V_t^{-1} U_t)^{-1} + S_t\right)^{-1} S_t\right) U_t' a = \\
&= a' U_t (S_t^{-1} + c U_t' V_t^{-1} U_t)^{-1} U_t' a \geq 0,
\end{aligned}$$

because both S_t and $U_t' V_t^{-1} U_t$ are positive definite for every t . The proof then follows by induction from the assumption that V_1 is positive definite. \square

Proof of Proposition 3: Let e_i and v_i^* denote the i th column of the $k \times k$ unit matrix and of V^{-1} , respectively. Define $\mathcal{J} = V^{-1} \otimes V^{-1}$. Note that the columns of \mathcal{S} are of the form $(1 + \delta_{i=j})(e_i \otimes e_j + e_j \otimes e_i)$, with $\delta_{i=j}$ denoting the Kronecker delta. If \mathcal{S}_\perp denotes the null space of \mathcal{S} , i.e., $\mathcal{S}'_\perp \mathcal{S} = 0$, then we can take

the columns of \mathcal{S}_\perp equal to $(e_i \otimes e_j - e_j \otimes e_i)$ for $i \neq j$. Define $\mathcal{S} = (\mathcal{S}, \mathcal{S}_\perp)$, where \mathcal{S} is invertible. Using these definitions, (11) holds if

$$\begin{aligned}
0 &= \mathcal{J}^{-1} \text{vec}(\nabla) - \mathcal{S}(\mathcal{S}'\mathcal{J}\mathcal{S})^{-1}\mathcal{S}' \text{vec}(\nabla) \\
&= (\mathcal{J}^{-1} - \mathcal{S}(\mathcal{S}'\mathcal{J}\mathcal{S})^{-1}\mathcal{S}') \text{vec}(\nabla) \\
&= \mathcal{J}^{-1} (\text{I} - \mathcal{J}\mathcal{S}(\mathcal{S}'\mathcal{J}\mathcal{S})^{-1}\mathcal{S}') \mathcal{S} \text{vech}(\nabla) \\
&= \mathcal{J}^{-1}(\mathcal{S}')^{-1} (\mathcal{S}'\mathcal{S} - \mathcal{S}'\mathcal{J}\mathcal{S}(\mathcal{S}'\mathcal{J}\mathcal{S})^{-1}\mathcal{S}'\mathcal{S}) \text{vech}(\nabla) \\
&= \mathcal{J}^{-1}(\mathcal{S}')^{-1} \left(\begin{pmatrix} \mathcal{S}'\mathcal{S} \\ 0 \end{pmatrix} - \begin{pmatrix} \mathcal{S}'\mathcal{S} \\ (\mathcal{S}'_\perp\mathcal{J}\mathcal{S})(\mathcal{S}'\mathcal{J}\mathcal{S})^{-1}\mathcal{S}'\mathcal{S} \end{pmatrix} \right) \text{vech}(\nabla) \\
&= \mathcal{J}^{-1}(\mathcal{S}')^{-1} \left(\begin{pmatrix} 0 \\ -(\mathcal{S}'_\perp\mathcal{J}\mathcal{S})(\mathcal{S}'\mathcal{J}\mathcal{S})^{-1}\mathcal{S}'\mathcal{S} \end{pmatrix} \right) \text{vech}(\nabla),
\end{aligned}$$

which is true for arbitrary $\text{vech}(\nabla)$ if and only if $\mathcal{S}'_\perp\mathcal{J}\mathcal{S} = 0$. Given the form of \mathcal{J} , \mathcal{S} , and \mathcal{S}_\perp , we have for $i \neq j$ that

$$\begin{aligned}
\mathcal{S}'_\perp\mathcal{J}\mathcal{S} &= (e_i \otimes e_j - e_j \otimes e_i)' \mathcal{J} \frac{(e_k \otimes e_\ell + e_\ell \otimes e_k)}{1 + \delta_{k=\ell}} = (v_i^* \otimes v_j^* - v_j^* \otimes v_i^*)' \frac{(e_k \otimes e_\ell + e_\ell \otimes e_k)}{1 + \delta_{k=\ell}} \\
&= \frac{v_{ik}^* v_{j\ell}^* - v_{j\ell}^* v_{ik}^* + v_{i\ell}^* v_{jk}^* - v_{jk}^* v_{i\ell}^*}{1 + \delta_{k=\ell}} = 0.
\end{aligned}$$

□

Proof of Proposition 4: Note that the recursion in (4) can be written as

$$V_{t+1} = \Omega + \beta V_t + \alpha (V_t)^{1/2} V_t^{-1/2} S_t (V_t')^{-1/2} (V_t')^{1/2} = \Omega + (\beta - \alpha) V_t + \alpha (V_t)^{1/2} \eta_t (V_t')^{1/2}, \quad (\text{A.2})$$

with $\eta_t = V_t^{-1/2} S_t (V_t')^{-1/2} + \text{I}$. From (12), we note that η_t is i.i.d. with expectation I. The recursion in (A.2) can now be recognized as a semi-polynomial Markov chain as defined in Boussama (2006). The result then follows directly by an application of his Theorem 2, noting that his requirement $|(\beta - \alpha) + \alpha| < 1$ together with the positivity constraint $\beta > 0$ implies $0 < \beta < 1$. We remark that the fact that Boussama uses $\tilde{\eta}_t \tilde{\eta}_t'$ for i.i.d. vector-valued $\tilde{\eta}_t$ with mean zero and covariance matrix I rather than a matrix valued η_t with expectation I, is immaterial for his result to hold. The only important feature in the proof of Boussama (2006) for our current purposes is that (A.2) is quadratic in $V_t^{1/2}$, and that η_t is i.i.d. with mean I. □