Modelling Covariance Using High-Frequency Data

5344K

July 2019

Abstract

In recent times, high frequency financial data has become easily available. This has led to the development of conditional covariance models that provide more accurate forecasts than traditional multivariate GARCH and Stochastic Volatility models. A recent, leading, example is the HEAVY GAS tF model of Opschoor et al (2018), which jointly models daily returns and realized covariances. The main contribution of this paper is to extend this model in two main directions. The first is to extend the model to a two component specification. Meanwhile, the second extension introduces "leverage effects". The empirical performance of the new model is evaluated using a set of 20 US equities. The dataset is extracted from the NYSE Trade and Quote database, and consists of transactions data for 3273 trading days, recorded at the 1 second frequency. [1]

1 Introduction

The covariance matrix of daily asset returns is an important quantity in many areas of finance and economics. Applications include asset allocation, option pricing and risk management. Traditionally, models for the covariance matrix have taken one of two avenues. The first avenue relates to multivariate generalized autoregressive conditional heteroskedasticity (GARCH) models and extensive surveys can be found in Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Teräsvirta (2009). The other strand of literature relates to multivariate stochastic volatility (SV) models and a useful survey can be found in Asai, McAleer and Yu (2006). Both types of models exploit daily return series to extract the dynamic process for the covariance matrix.

However, the recent availability of high frequency data has led to the development of models that can provide more accurate forecasts of the conditional covariance matrix. A key input to these models are realized measures of the covariance matrix, see for example Barndorff-Nielsen et al (2011). Typically, the models proceed by specifying dynamics for both the daily return and realized covariance series. Examples include the multivariate high-frequency-based volatility (HEAVY) model of Noureldin, Shephard, and Sheppard (2012); the HEAVY GAS tF model of Opschoor et al (2018), which extends the model of Noureldin et al to incorporate fat tails; and the dynamic component model of Jin and Maheu (2013), which is set in an SV framework. Associated models that focus solely on modelling the realised covariance

[1] The programs for this paper were written in the Python programming language and are available upon request.

matrix include the Wishart autoregressive (WAR) model of Gourieroux, Jasiak, and Sufana (2009) and the Conditional Autoregressive Wishart (CAW) model of Golosnoy, Gribisch and Leisenfeld (2012).

Thus, there are several approaches to modelling the covariance of assets using high frequency data. However, studies that extend these models to richer and more empirically realistic specifications are relatively rare, especially when compared to the univariate case. In this paper, I extend the HEAVY GAS tF model of Opschoor et al (2018) in two directions. First, I extend the model to include two components. This specification has the appealing interpretation of separating the covariance process into a long term and a short term component. The former component captures the slowly-changing "trend" in the covariances, while the latter picks up transitory changes in the covariance process. Furthermore, the two component model can also approximate long-memory behaviour (Andersen et al, 2006). The second extension is to allow for asymmetric responses in the conditional covariances to negative returns. This is often termed the "leverage effect" after Black (1976).

Note, an important challenge in the modelling of multivariate volatility is to guarantee positive definiteness of the conditional covariance matrix. Given this constraint, we will see that the one-component HEAVY GAS tF model faces a trade-off between incorporating long memory and leverage while tackling the "curse of dimensionality". The main contribution of this paper is therefore to show that a two component model with leverage confined to the short-run component can help to alleviate these trade-offs. I will argue that the two-component model provides a convenient framework to simultaneously model long memory and leverage, while also providing a rich interpretation.

The dynamics of the unobserved covariance matrix in the HEAVY GAS tF model are driven by the generalised autoregressive score (GAS) framework of Creal, Koopman, and Lucas (2011, 2013) and of Harvey (2013). In this framework, time varying parameters are driven by the score of the conditional density function. Continuing in the same spirit, all models estimated in this paper are driven by GAS dynamics. A justification for this is the fact that score driven dynamics exhibit information theoretic optimality properties, regardless of how badly the model is mispecified (Blasques, Koopman and Lucas, 2015). In particular, only score-based parameter updates will always reduce the local Kullback-Liebler divergence between the true conditional density and the model implied conditional density.

To my knowledge, there are no other papers that build a model with either two components or leverage for multivariate volatility with GAS dynamics. However, Harvey and Sucarrat (2014) do build such a model in the univariate framework. Furthermore, Opschoor and Lucas (2019) model the long memory in realised covariance using a fractional integration approach, while Opschoor et al (2018) suggest a HAR (Corsi, 2009) type extension to the HEAVY GAS tF model to account for long memory.

This paper proceeds by presenting the econometric framework in Section 2. This includes details on the computation of the realized covariance series; on the HEAVY GAS tF model and subsequent extensions; and on the Realized Wishart GARCH model of Gorgi et al (2018), which we use as a score-driven benchmark. Section 3 contains an empirical application of the models discussed. The dataset comprises intraday transactions data for 20 US equities at the 1 second frequency. The sample period is 1 January 2002 to 31 December 2014, which gives a total of 3273 trading days. Both the in-sample and out-of-sample performance of the models is evaluated. Finally, Section 4 concludes.

2 Econometric Framework

2.1 Realised Covariance

Define the k-dimensional log price process $p^*(\tau)$, where $\tau \in \mathbb{R}_+$ denotes continuous time. Suppose that p^* is a semi-martingale which can be decomposed as

$$p^*(\tau) = \alpha^*(\tau) + \int_0^\tau \Theta(u) dw(u) \tag{1}$$

where α^* is a finite variation drift term, Θ is the instantaneous covolatility process and w is a vector standard Brownian motion. The quantity of interest here is the integrated covariance:

$$\Sigma(\tau) = \int_0^\tau \Theta(u)\Theta(u)' du \tag{2}$$

which measures the ex-post covariation of p^* over the interval $[0, \tau]$.

Now, normalise τ to 1 day and suppose that we observe $m+1$ equally spaced, intraday prices. The j-th intraday vector of returns on day t is given by

$$r_{j,t} = p^*_{(t-1)+\frac{j}{m}} - p^*_{(t-1)+\frac{j-1}{m}}, \qquad j = 1, \ldots, m \quad t = 1, 2, \ldots \tag{3}$$

Using these intraday returns, we can compute the *realised covariance* matrix

$$RC_t = \sum_{j=1}^m r_{j,t} r'_{j,t} \tag{4}$$

It can be shown that, in the absence of market microstructure noise[2], RC_t is a consistent estimator of the integrated covariance of $p^*(\tau)$ as $m \to \infty$ (Barndorff-Nielsen and Shephard, 2004).

In practice however, intraday prices are contaminated with microstructure noise. RC_t is a biased estimator in this case. One possible solution is to sample sparsely and use the subsampling scheme of Aït-Sahalia et al (2005). Alternatively, noise-robust estimators are also available, for example the realised kernel of Barndorff-Nielsen et al (2011).

The vector of daily returns, meanwhile, can be computed as $r_t = \sum_{j=1}^m r_{j,t}$.

For the rest of this section, assume that we have a data series of daily returns and realised covariances. We will develop models for the joint dynamics of these series. The next subsection presents the baseline HEAVY GAS tF model of Opschoor et al (2018). Then, various extensions are incorporated into the model. These include HAR dynamics (Corsi, 2009), two components, and leverage. The latter two extensions are new to the HEAVY GAS tF model. Conditions for the positive definiteness and stationarity of the unobserved covariance matrix are also discussed.

Finally, the Realised Wishart-GARCH (RWG) model of Gorgi et al (2018) is presented. This will serve as a benchmark for the HEAVY GAS tF models. The main difference between the RWG and HEAVY GAS models is that the latter allow for fat tails in the distributions of returns and realised covariances, whereas the former does not.

[2]Sources of market microstructure noise include, for example, price discreteness and bid-ask bounce (Madhavan, 2000)

2.2 HEAVY GAS tF model

As above, $r_t \in \mathbb{R}^k$ is the vector of (demeaned) daily log returns and $RC_t \in \mathbb{R}^{k\times k}$ is the realised convariance matrix of k assets on day t, for $t = 1, \dots, T$. Opschoor et al (2018) assume that r_t is fat-tailed and that its conditional observation density is the standardized Student's t distribution:

$$p_r(r_t|V_t, \mathcal{F}_{t-1}; \nu_0) = \frac{\Gamma((\nu_0 + k)/2))}{\Gamma(\nu_0/2)\big[(\nu_0 - 2)\pi\big]^{k/2}|V_t|^{1/2}} \left(1 + \frac{r_t' V_t^{-1} r_t}{\nu_0 - 2}\right)^{-(\nu_0+k)/2} \tag{5}$$

where $\nu_0 > 2$ is the degrees of freedom parameter and V_t is the (positive definite) covariance matrix, which can vary over time. $\mathcal{F}_t$ is the time t information set containing all past values of r_t and RC_t, including at time t.

Contrary to the majority of other literature, Opschoor et al (2018) also account for fat tails in the realised covariance matrix, RC_t, by assuming that it follows the conditional matrix-F distribution:

$$p_{RC}(RC_t|V_t, \mathcal{F}_{t-1}; \nu_1, \nu_2) = K(\nu_1, \nu_2) \frac{\left|\frac{\nu_1}{\nu_2 - k - 1} V_t^{-1}\right|^{\nu_1/2} |RC_t|^{(\nu_1 - k - 1/2)}}{\left|I_k + \frac{\nu_1}{\nu_2 - k - 1} V_t^{-1} RC_t\right|^{(\nu_1+\nu_2)/2}} \tag{6}$$

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

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

where $\Gamma_k(x)$ is the multivariate Gamma function.

It is worth noting that as $\nu_2 \to \infty$, the matrix-F distribution converges to the Wishart distribution with ν_1 degrees of freedom, which is the multivariate counterpart of the χ^2 distribution. Another useful result is that if RC_t is matrix-$F(\nu_1, \nu_2)$ distributed, then any diagonal element, say $RC_{ii,t}$, follows a scalar $F(\nu_1, \nu_2 - (k + 2))$ distribution. Using these facts, Opschoor et al show that even with $\nu_2 \approx 100$, the matrix-F distribution is different to the Wishart and provides a larger dispersion around the mean.

Thus, the conditional distributions of r_t and RC_t both depend on the common time-varying covariance matrix V_t. The goal of the HEAVY GAS tF model is to filter for the dynamics of V_t. To this end, the generalised autoregressive (GAS) framework of Creal, Koopman and Lucas (2011, 2013) and Harvey (2013) is used. Here, the GAS framework uses the score of the joint conditional density of r_t and RC_t to drive the dynamics of the true, time varying covariance matrix V_t.

The baseline recursion adopted by Opschoor et al (2018) is

$$V_{t+1} = \Omega + \alpha S_t + \beta V_t \tag{7}$$

where $S_t \in \mathbb{R}^{k\times k}$ is the scaled score matrix (see below), while the coefficients $\alpha, \beta \in \mathbb{R}$ and matrix of intercepts $\Omega \in \mathbb{R}^{k\times k}$ are parameters to be estimated. More general specifications in which the scalars α and β are replaced by $k \times k$ matrices are also possible. However, in order to keep the dimensionality of the parameter space down, these specifications are not pursued here.

Now, in Cox (1981)'s terminology, the GAS framework is observation driven, as opposed to parameter driven. The advantage of the former is that the likelihood function is available in closed form. Thus, estimation and inference are computationally convenient when compared to the Bayesian estimation procedures that are usually required in parameter driven approaches.

The model assumes that conditional on V_t and $\mathcal{F}_{t-1}$, returns r_t and realised covariances RC_t are independent. Given this assumption, the total log likelihood can be decomposed into the contributions from the Student's t and matrix-F densities. That is,

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} (\mathcal{L}_{r,t} + \mathcal{L}_{RC,t}) \tag{8}$$

with

$$\mathcal{L}_{r,t} = \ln \Gamma\Big(\frac{\nu_0 + k}{2}\Big) - \ln \Gamma\Big(\frac{\nu_0}{2}\Big) - \frac{k}{2} \ln{(\nu_0 - 2)\pi} - \frac{1}{2} \ln |V_t| - \frac{\nu_0 + k}{2} \ln\Big(1 + \frac{r_t' V_t^{-1} r_t}{\nu_0 - 2}\Big),$$
$$\mathcal{L}_{RC,t} = \ln K(\nu_1, \nu_2) + \frac{\nu_1}{2} \ln \Big|\frac{\nu_1}{\nu_2 - k - 1} V_t^{-1}\Big| + \frac{\nu_1 - k - 1}{2} \ln |RC_t| - \frac{\nu_1 + \nu_2}{2} \ln \Big|I_k + \frac{\nu_1}{\nu_2 - k - 1} V_t^{-1} RC_t\Big|$$

where θ is the vector of parameters to be estimated.

The matrix series S_t in recursion (7) is derived by taking the score, $\partial \mathcal{L}_t(\theta)/\partial V_t$, of (8) wrt a general non-symmetric matrix V_t, and scaling by a scalar multiple of $(V_t \otimes V_t)$, where $\otimes$ denotes the Kronecker product. This gives

$$S_t = \frac{w_t r_t r_t' - V_t}{\nu_1 + 1} + \frac{\nu_1}{\nu_1 + 1}\Big[\frac{\nu_1 + \nu_2}{\nu_2 - k - 1} RC_t \Big(I_k + \frac{\nu_1 V_t^{-1} RC_t}{\nu_2 - k - 1}\Big)^{-1} - V_t\Big] \tag{9}$$

where $w_t = (\nu_0 + k)/(\nu_0 - 2 + r_t' V_t^{-1} r_t)$. Note, since S_t is the (scaled) score of a density function, it follows that $\mathbb{E}_{t-1}[S_t] = 0$. That is, S_t is a martingale difference sequence.

The scaled score in (9) has an intuitive interpretation. As discussed in Creal et al (2011), the first term in (9) emanates from the Student's t distribution. This term drives the dynamics of the covariance matrix according to the deviations of the weighted outer product $w_t r_t r_t'$ from the local covariance matrix V_t. For ν_0 finite (i.e. if the density of the returns, r_t, is heavy tailed), large values of $r_t r_t'$ do not necessarily lead to a large update in the elements of V_t. That is, the weight w_t dampens the effect of extreme values because it is decreasing in $r_t' V_t^{-1} r_t$. When $\nu_0 \to \infty$, $w_t = 1$ for all t. In this case, the dynamics correspond to the Gaussian multivariate GARCH model.

The second expression in (9) relates to the matrix-F distribution and was new in Opschoor et al (2018). The intuition for this term is similar to that for the Student's t. That is, large values of RC_t could arise because of the heavy-tailed nature of the matrix-F distribution rather than as a result of a substantial change in the underlying covariance matrix. Consequently, a significant update to the values of V_t may not necessarily be in order. The matrix weight $(I_k + (\nu_1 V_t^{-1} RC_t)/(\nu_2 - k - 1))^{-1}$ now takes the role of w_t and dampens the effect of a large RC_t through its negative dependence on $V_t^{-1} RC_t$. As $\nu_2 \to \infty$, the matrix-F converges to the Wishart distribution, and the matrix weight reduces to I_k.

In sum, ν_0 and ν_2 play a "robustification role" in the model. In fact, as $\nu_0, \nu_2 \to \infty$, the score becomes

$$\frac{1}{\nu_1 + 1}\big[r_t r_t' + \nu_1 RC_t\big] - V_t \tag{10}$$

which should be compared to the Realised Wishart-GARCH model of Gorgi et al (2018) presented below.

As for the role of ν_1 in (9), it determines the relative weights attached to the scores from the Student's t and matrix-F distributions. For higher values of ν_1, the signal derived from RC_t is judged to be increasingly accurate relative to the signal derived from the outer product of daily returns, $r_t r_t'$. The converse is true for lower values of ν_1. In the limit as $\nu_1 \to \infty$, RC_t provides a perfect signal for V_t, and $r_t r_t'$ falls out of the score in (9).

Opschoor et al complete the model description by providing necessary and sufficient conditions for the positive definiteness and stationarity of the series V_t generated by the recursion in (7). Given positive semidefinite realised measures RC_t, the scaled score in (9), an initial positive definite matrix V_1 and positive semidefinite matrix Ω, these conditions can be summarised as

$$0 < \alpha < \beta < 1 \tag{11}$$

These constraints can be implemented during estimation.

Now, note that computational burden is an important consideration when estimating multivariate models such as the current one, especially as the number of assets, k, gets large. One way of reducing the number of parameters to be estimated is by adopting the covariance targeting parametrization, as introduced by Engle and Mezrich (1996) for GARCH models. Given the condition for stationarity, (11), the unconditional mean of V_t can be derived from (7) as $\mathbb{E}[V_t] = (1-\beta)^{-1}\Omega$. We can then replace Ω in (7) as follows

$$V_{t+1} = (1-\beta)\bar{V} + \alpha S_t + \beta V_t \tag{12}$$

where $\bar{V} \equiv \mathbb{E}[V_t]$. Estimation of such a parametrization can be done using a two step procedure, see Section 2.5.

The next subsection deals with extensions to the HEAVY GAS tF model, namely, HAR dynamics, two components, and leverage.

2.3 Extensions

2.3.1 Long memory

One possible avenue to extend the model is to consider long-memory-type persistence in V_t. In this vein, Opschoor et al (2018) extend (7) by adding HAR dynamics similar to Corsi (2009) and Jin and Maheu (2016):

$$V_{t+1} = \Omega + \alpha S_t + \beta_1 V_{l_1,t} + \beta_2 V_{l_2,t} + \beta_3 V_{l_3,t} \tag{13}$$

with $V_{l,t} = l^{-1}\sum_{i=1}^{l} V_{t-i}$. I follow Jin and Maheu (2016) and Opschoor et al (2018) and set $l_1 = 1, l_2 = 12$ and $l_3 = 60$. This is labelled the "HEAVY GAS HAR tF" model. Necessary and sufficient conditions for the stationarity and positive definiteness of V_t are:

$$\begin{aligned} &\beta_j > 0 \qquad \text{for } j = 1,2,3, \\ &l_1^{-1}\beta_1 + l_2^{-1}\beta_2 + l_3^{-1}\beta_3 > \alpha > 0, \\ &\beta_1 + \beta_2 + \beta_3 < 1 \end{aligned} \tag{14}$$

The covariance targeting parmetrization is implemented by replacing $\Omega = (1 - \beta_1 - \beta_2 - \beta_3)\bar{V}$ in (13).

Now, I suggest a new method to incorporate long memory into the HEAVY GAS tF model. In the spirit of Engle and Lee (1999) who proposed a two component GARCH model, and Harvey (2013) who proposed a univariate, score-driven, two component model, I suggest separating the dynamic equation (7) into a long run and a short run component. That the sum of a few individually short-memory components can approximate long memory behaviour is discussed, for example, in Andersen et al (2006). The dynamic equation for V_t is now

$$\begin{aligned} V_{t+1} &= \Omega + V_{1,t+1} + V_{2,t+1}, \\ V_{i,t+1} &= \alpha_i S_t + \beta_i V_{i,t} \, , \qquad \text{for } i = 1, 2 \end{aligned} \tag{15}$$

Since S_t is a martingale difference, the moments of V_t can easily be derived from the infinite MA representation:

$$V_{t+1} = \Omega + \sum_{k=1}^{\infty} \psi_{1,k} S_{t-k} + \sum_{k=1}^{\infty} \psi_{2,k} S_{t-k} = \sum_{k=1}^{\infty} \psi_k S_{t-k}$$

where $\psi_{i,k} = \alpha_i \beta_i^k$ and $\psi_k = \psi_{1,k} + \psi_{2,k}$, see Harvey (2013). Therefore, $\mathbb{E}[V_t] = \Omega$ and the covariance targeting specification simply substitutes $\Omega = \bar{V} \equiv \mathbb{E}[V_t]$. Thus, stationarity and positive definiteness of V_t requires that the components $V_{i,t+1}$, $i = 1, 2$, are stationary and positive definite. It follows directly from Opschoor et al's proof for the baseline recursion in (7) that necessary and sufficient conditions for this are:

$$0 < \alpha_i < \beta_i < 1, \qquad i = 1, 2$$

A feature of two component models such as (15) is that the long-run component (with $i = 1$, say) exhibits a high level of persistence and has β_1 close to one. Meanwhile, the short-run component, which captures transitory changes in the covariance matrix, tends to have a lower value for β_2 along with a higher value for α_2. In fact, the parameter restriction

$$\beta_2 < \beta_1$$

is usually imposed to guarantee identifiability (Harvey, 2013).

An alternative to both the HAR and two component specifications of the HEAVY GAS model would be the Fractionally Integrated GAS model of Opschoor and Lucas (2019). Their framework also adopts the matrix-F distribution for realised covariances, but it drops the daily returns observations. Their model incorporates long memory in the same spirit as the FI-GARCH of Baillie, Bollerslev and Mikkelsen (1996).

Therefore, we have multiple competing specifications for incorporating long memory into score-driven models for realised covariances with fat-tailed distributions. However, a theoretical advantage of the two-component set up is its rich interpretation as decomposing the covariance process V_t into a long term "trend" and a transitory component. The latter can pick up temporary spikes in volatility following a large shock, for example The Flash Crash of May 2010. Indeed, to quote Harvey (2013), who quotes

Alizadeh et al (2002): "the interpretability of the two component model stands in sharp contrast to that of long-memory fractionally integrated volatility, which often appears mysterious and non-intuitive".

Furthermore, as we will see below, the true advantage of the two component set-up is realised when we extend the HEAVY GAS model to an even richer specification, one that incorporates both leverage and long memory.

2.3.2 Leverage

That conditional variances and covariances respond asymmetrically to positive and negative returns is a well documented empirical regularity, see for example Capiello, Engle and Sheppard (2006). In particular, stock volatility tends to increase more following a negative shock than following a positive shock of the same size. Furthermore, the covariance between stocks tends to increase following joint bad news, i.e. when both stocks experience negative returns. These asymmetric effects are often referred to as leverage effects, following Black (1976).

I incorporate leverage into the HEAVY GAS model in a similar spirit to Capiello et al (2006), who do this for the DCC GARCH model of Engle (2002). In particular, I add a term S_t^*, in which the ij-th element of RC_t is multiplied by $\mathbb{1}(r_{it} < 0)\mathbb{1}(r_{jt} < 0)$, to the recursion in (7).[3]

The recursion is now

$$V_{t+1} = \Omega + \alpha S_t + \alpha^* S_t^* + \beta V_t \tag{16}$$

where $S_t^* \equiv RC_t \odot Q_t$, $[Q_t]_{ij} = \mathbb{1}(r_{it} < 0)\mathbb{1}(r_{jt} < 0)$, and $\odot$ denotes the Hadamard product. Note, Q_t is positive semi-definite since it can be written as $Q_t = q_t q_t'$, where $q_t = [\mathbb{1}(r_{1t} < 0), \ldots, \mathbb{1}(r_{kt} < 0)]' \in \mathbb{R}^k$. Then, given that RC_t is also positive semi-definite, the Schur product theorem[4] implies that S_t^* is positive semi-definite.

Thus, given positive semi-definite Ω, the conditions for the positive definiteness of V_t in (16) are the same as those for the baseline recursion (7) and the additional condition that

$$\alpha^* > 0$$

This condition is empirically motivated in Section 3.1.

The covariance targetting parametrization is now

$$V_{t+1} = (\bar{V} - \beta\bar{V} - \alpha^*\bar{S}^*) + \alpha S_t + \alpha^* S_t^* + \beta V_t \tag{17}$$

where $\bar{S}^* \equiv \mathbb{E}[S_t^*]$. Positive semi definiteness of V_t in (17) now requires that the intercept is also positive semi-definite. A necessary and sufficient condition for this is that

$$1 - \beta - \alpha^*\delta > 0 \tag{18}$$

where δ is the largest eigenvalue of the matrix $\bar{V}^{-1/2}\bar{S}^*\bar{V}^{-1/2}$. This constraint can be evaluated on the sample data.[5].

[3] $\mathbb{1}(\cdot)$ denotes the indicator function

[4] The Schur product theorem states that if $A, B \in \mathbb{R}^{n\times n}$ are positive semi-definite matrices, then the Hadamard product $A \odot B \in \mathbb{R}^{n\times n}$ is also positive semi-definite. See Theorem 12.32 in Magnus and Abadir (2005).

[5] the definition of the matrix square root used here is based on the spectral decomposition, whereby $A^{1/2} = U\Lambda^{1/2}U'$ for some generic matrix A, where Λ contains the eigenvalues of A. If A is positive semidefinite, then $A^{1/2}$ is unique

The logic for the constraint in (18) can be seen by rewriting the intercept as

$$\bar{V}^{1/2}\big((1-\beta)I_k - \alpha^*\bar{V}^{-1/2}\bar{S}^*\bar{V}^{-1/2}\big)\bar{V}^{1/2}$$

Then, it can be seen that condition (18) simply requires that the smallest eigenvalue of the middle term in brackets is positive. If this is true, then the intercept term is positive semi-definite as the rewritten intercept is simply a matrix quadratic form.

Note that a feature (and perhaps a weakness) of the current specification of leverage is that the asymmetric response for covariances occurs only when both stocks in question experience a negative return. It might be interesting to relax this assumption. A possibility could be an additive specification for Q_t such that $[Q_t]_{ij} = \big(\mathbb{1}(r_{it} < 0) + \mathbb{1}(r_{jt} < 0)\big)/2$. However, note that in this case, Q_t would not be positive semi-definite and a sufficient condition for the positive definiteness of V_t in (16) currently appears non-trivial, if not infeasible.

Another potential drawback to the specification of leverage in (17) is that condition (18) might become restrictive in sample. That is, if β is close to 1 (i.e. if the series V_t is highly persistent), then α^* must necessarily be close to zero, even for moderate values of δ. The problem becomes more acute as the dimension k increases, as the expected value of δ also increases. This problem would still be present if we tried to add leverage to the HAR dynamic equation (13).

Thus, the one component HEAVY GAS tF model faces a trade-off between alleviating the curse of dimensionality (via covariance targetting), modelling long-memory and modelling leverage.

The two component model can help to circumvent these issues. In particular, the leverage term can be included exclusively in the short-run component. This assumption is supported by the evidence in Engle and Lee (1999) and Harvey and Sucarrat (2014) amongst others. The dynamic equations for V_t becomes

$$\begin{aligned} V_{t+1} &= \Omega + V_{1,t+1} + V_{2,t+1} \\ V_{1,t+1} &= \alpha_1 S_t + \beta_1 V_{1,t} \\ V_{2,t+1} &= \alpha_2 S_t + \alpha^* S_t^* + \beta_2 V_{2,t} \end{aligned} \tag{19}$$

Given stationarity, an infinite MA representation can again be used to show that $\mathbb{E}[V_t] = \Omega + \frac{\alpha^*}{(1-\beta_2)}\mathbb{E}[S_t^*]$. Therefore, the covariance targeting parametrization for V_{t+1} becomes

$$V_{t+1} = \left(\bar{V} - \frac{\alpha^*}{1-\beta_2}\bar{S}^*\right) + V_{1,t+1} + V_{2,t+1} \tag{20}$$

Following the same argument as before, the condition for positive semi-definiteness of the intercept is

$$1 - \beta_2 - \alpha^*\delta > 0 \tag{21}$$

Combined with the identifiability condition that $\beta_2 < \beta_1$, it is clear that the restriction in (21) is less likely to bite in an empirical setting than the restriction in (18). Thus, the two component specification helps to alleviate the trade-offs faced when trying to model leverage in the one component HEAVY GAS model.

2.4 Realised Wishart-GARCH: a benchmark

I also estimate the Realised Wishart-GARCH (RWG) model of Gorgi et al (2018). This model is also score-driven and considers daily returns and realised covariances jointly. However, the authors assume thin-tailed distributions for the daily returns and realized covariances. Therefore, it serves as a useful benchmark for the models described above.

The distributional assumptions are:

$$r_t|\mathcal{F}_{t-1} \sim N_k(0, H_t) \tag{22}$$

$$RC_t|\mathcal{F}_{t-1} \sim W_k(V_t/\nu_1, \nu_1) \tag{23}$$

where H_t is the covariance matrix of the k-dimensional multivariate normal distribution, and V_t is the mean of the k-dimensional Wishart distribution with degrees of freedom $\nu_1 \geq k$.

Gorgi et al (2018) assume that H_t and V_t share the same dynamic process, i.e.

$$H_t = \Lambda V_t \Lambda'$$

where $\Lambda = (\lambda_{ij})$ is a $k \times k$ non-singular matrix. The assumption $\lambda_{11} > 0$ is imposed to ensure identifiability. To further reduce the dimensionality of the problem, Λ is assumed to be a diagonal matrix.

The dynamics of V_t are modelled in terms of a vector process f_t, such that $V_t = V(f_t)$. In particular, Gorgi et al choose $f_t = \text{vech}(V_t)$, where the vech$(\cdot)$ operator stacks the lower triangular portion (including the diagonal) of the symmetric matrix V_t. Thus, f_t is a $k^* \times 1$ dimensional vector, with $k^* = k(k+1)/2$. Note the contrast with the HEAVY GAS model, which retains a matrix format in its dynamic equation.

The dynamic equation for f_t, with covariance targeting, is

$$f_{t+1} = (1-\beta)\mathbb{E}[f_t] + \alpha s_t + \beta f_t \tag{24}$$

where s_t is the scaled score vector as below. Sufficient conditions for positive definiteness and stationarity are:

$$0 < \alpha < \beta < 1$$

The log likelihood function is:

$$\mathcal{L}(\theta) = \sum_{t=1}^{T} (\mathcal{L}_{r,t} + \mathcal{L}_{RC,t}) \tag{25}$$

with

$$\mathcal{L}_{r,t} = -\frac{k}{2}\log 2\pi - \frac{1}{2}\log|\Lambda V_t \Lambda'| - \frac{1}{2}\text{tr}((\Lambda V_t \Lambda')^{-1} r_t r_t') \tag{26}$$

$$\mathcal{L}_{RC,t} = -\frac{1}{2}\nu k \log\frac{\nu}{2} + \log\Gamma_k\left(\frac{\nu}{2}\right) - \frac{\nu-k-1}{2}\log|RC_t| + \frac{\nu}{2}\log|V_t| + \frac{\nu}{2}\text{tr}(V_t^{-1} RC_t) \tag{27}$$

The scaled score is:

$$s_t = \frac{1}{\nu+1}\big(\text{vech}(\Lambda^{-1} r_t r_t' (\Lambda')^{-1}) + \nu\text{vech}(RC_t)\big) - \text{vech}(V_t) \tag{28}$$

which is similar to the (vectorized) scaled score for the HEAVY GAS tF model with $\nu_0, \nu_2 \to \infty$, see equation (10). Therefore, in a sense, the HEAVY GAS tF model nests the RWG as a special case.

2.5 Estimation

During estimation, the covariance targeting parametrization is used in all models. In practice, the expectations $\mathbb{E}[V_t]$, $\mathbb{E}[S_t^*]$ and $\mathbb{E}[f_t]$ are infeasible. Thus, a two step estimation procedure is used whereby these expectations are first replaced by a consistent estimator of the unconditional mean. In particular, $\mathbb{E}[V_t]$ ($\mathbb{E}[f_t]$) is replaced by (the vech$(\cdot)$ of) the sample mean of RC_t and $\mathbb{E}[S_t^*]$ is replaced by the sample mean of S_t^*. Second, the remaining parameters are estimated by maximum likelihood.

In the HEAVY GAS tF model, the log-likelihood function (8) is maximised with respect to the parameter vector $\theta = (\alpha, \beta, \nu_0, \nu_1, \nu_2)$. The starting values of V_1 and the scaled score S_1 are set at RC_1 and 0 respectively. The constraint, $0 < \alpha < \beta < 1$, is also imposed.

For the remaining specifications of the HEAVY GAS model, the parameter vector is adjusted in the obvious manner. The relevant constraints for positive definiteness and stationarity are also imposed, as discussed above.

For the RWG, the matrix Λ is assumed to be diagonal, following Gorgi et al (2018). The parameter vector is therefore $\theta = (\alpha, \beta, \nu_1, \lambda_{11}, \ldots, \lambda_{kk})$. The log-likelihood in (25) is maximised with respect to θ, subject to the constraints $\lambda_{11} > 0$ and $0 < \alpha < \beta < 1$. Initialisation is as in the HEAVY GAS model.

3 Empirical Application

3.1 Data

The data used in this paper was extracted from the NYSE Trade and Quote (TAQ) database. The dataset contains consolidated trades (transaction prices) for a collection of 20 liquid stocks in the S&P 500 (see Table 1). The sample period is from 1 January 2002 to 31 December 2014, which gives a total of 3273 trading days. The observations have timestamps accurate to the 1-second frequency.

The TAQ data is raw. Therefore, it must be cleaned before being used for econometric analysis. Here, I follow a cleaning procedure based on Brownlees and Gallo (2006) and Barndorff-Nielsen et al (2009). In particular, the cleaning steps taken (in sequential order) were:

- Delete entries with a timestamp outside the interval 9:30 am - 4 pm
- Delete entries with a transaction price equal to zero
- Delete entries with corrected trades (Trades with a *Correction Indicator*, CORR $\neq 0$)
- Delete entries that were reported with a delay (Trades with a *Sale Condition*, COND $= Z$)
- If multiple entries have the same timestamp, use the median price
- Delete entries for which the price deviated by more than 10 mean absolute deviations from a rolling centred median (excluding the observation under considertation) of 50 observations (25 observations before and 25 after)

- Since the data is not equally spaced, the previous tick interpolation method of Dacorogna et al (2002) is used. That is, for any timestamp at which a price is not observed, the most recent price is taken

Subsequently, the daily return and realised covariance series were computed using the clean data.

The daily return series was computed by taking open-to-close returns over the interval 9:30 am - 4 pm (same interval is used for the realised covariances). The economic interpretation of this is that agents open their positions at the beginning of the trading day and unwind at the close. This is a reasonable assumption in many cases. However, an implication is that we ignore overnight risk, which may be important to some market-makers and liquidity providers.
The summary statistics for the daily return series are presented in the first four numeric columns of Table 1. The (annualised) mean daily return is less than 1 standard deviation away from zero for all stocks. Furthermore, all the series exhibit fat tails (excess kurtosis) while the majority of them exhibit skewness.

Ticker	Mean[a]	Std.[a]	Skewness	Kurtosis[b]	$H_0 : \chi^2$	$H_0 : F$
BA	3.97	27.08	0.07	4.32	0.00	0.37
BAC	-38.69	46.54	-4.49	105.91	0.00	0.15
DUK	-0.59	23.56	-0.29	16.73	0.00	0.12
GS	5.65	33.60	0.12	11.33	0.00	0.00
HSY	4.21	28.20	-18.10	744.26	0.00	0.03
IBM	8.30	21.94	-0.02	7.46	0.00	0.05
JNJ	2.76	17.25	-1.10	28.07	0.00	0.06
JPM	-0.53	37.14	0.40	12.54	0.00	0.15
KO	2.82	18.16	-0.29	8.54	0.00	0.12
MRK	4.81	24.67	0.08	7.43	0.00	0.94
NKE	8.05	31.73	-12.70	443.22	0.00	0.16
ORCL	12.95	28.98	0.18	3.79	0.00	0.32
PG	2.37	25.12	-24.97	1081.38	0.00	0.06
RL	14.86	34.24	0.20	4.54	0.00	0.21
TGT	1.76	29.59	0.12	6.73	0.00	0.31
UNH	4.53	41.98	-12.14	334.66	0.00	0.14
VZ	-4.99	23.16	0.16	7.72	0.00	0.04
WMT	5.31	19.13	0.35	5.50	0.00	0.63
XOM	12.91	22.41	-0.26	12.42	0.00	0.33
XRX	0.86	36.20	0.48	7.97	0.00	0.00

a: annualised assuming 252 market days per year; b: excess kurtosis

Table 1: Summary statistics of daily return series and p-values corresponding to Kolomogorov Smirnov tests for the distribution of $RC_{ii,t}$ (last two columns)

The realised covariance series was computed as in (4), i.e. $RC_t = \sum_{j=1}^{m} r_{j,t} r'_{j,t}$. In order to mitigate bias due to microstructure noise, returns $r_{j,t}$ were computed by sampling every 5 minutes, so $m = 78$. Furthermore, the subsampling scheme of Aït-Sahalia et al (2005) was used to increase efficiency. This involves computing several estimates of RC_t for each day and averaging. In particular, 300 estimates of the realised covariance matrix were computed for each day by moving the starting time for each estimate forward in one second increments. Thus, $300 \times 3273 = 981,900$ covariance matrices were computed in

total.
Table 2 displays the annualised average realised volatility (diagonal elements) and average realised correlations (off diagonal elements). The average realised correlations are all positive and range from 0.173 to 0.540. This is unsurprising given that all the stocks were taken from the S&P 500 index. Meanwhile, the mean realised volatility ranges between 17.40 - 38.32%.

Furthermore, as in Opschoor et al (2018), I conduct Kolmogorov-Smirnov (KS) tests to test between candidate distributions for the realised variances, which correspond to the diagonal elements, $RC_{ii,t}$, of RC_t. The last two columns of Table 1 report p-values corresponding to the null hypotheses that the realised variance follows a univariate Wishart (χ^2) and a univariate-F distribution respectively. While the null of the Wishart distribution is rejected for all stocks, the null hypothesis that the realised variances follow a F-distribution is rejected in only 4 out of 20 cases at the 5% significance level. Note, Opschoor et al point out that the time variation of $V_{ii,t}$ partly causes the unconditional distribution of $RC_{ii,t}$ to have fatter tails than the χ^2. However, as shown by both their empirical results and my empirical results below, the F-distribution also provides significant gains in the conditional distribution sense.

Additionally, following Capiello et al (2006), I conduct non-parametric tests for asymmetries in the conditional variances and covariances. To check whether the variance of returns is higher after a negative return than after a positive return, the appropriate null hypothesis is $\mathbb{E}[RC_{ii,t}|r_{it} < 0] = \mathbb{E}[RC_{ii,t}|r_{it} > 0]$. This can be implemented via the linear regression model

$$RC_{ii,t} = \delta + \gamma \mathbb{1}(r_{i,t-1} < 0) + \varepsilon_{it}$$

with the null $\gamma = 0$. Similarly, we can test whether covariances are, on average, different after joint negative returns than after joint positive returns using the model

$$RC_{ij,t} = \delta + \gamma \mathbb{1}(r_{i,t-1} < 0)\mathbb{1}(r_{j,t-1} < 0) + \phi \mathbb{1}(r_{i,t-1} > 0)\mathbb{1}(r_{j,t-1} > 0) + \varepsilon_{ij,t}$$

with the null $\gamma = \phi$, for $i \neq j$. Table 3 reports the results. The values reported are the p-values multiplied by $sign(\gamma)$ if $i = j$ and by $sign(\gamma - \phi)$ otherwise. The results show that for 18 out of 20 stocks, the average realised variance is higher after a negative return than after a positive return, with the difference significant at the 10% level in 16 of these cases. Meanwhile, the average covariance increases more after joint negative returns than joint positive returns in 179 out of 190 cases, with the difference significant at the 10% level in 95 of these cases. Thus, there is strong evidence of the leverage effect in our sample.

Finally, note that alternative estimators to the realised covariance, RC_t, used here were also available. One that is commonly used in the literature is the noise robust realized kernel estimator of Barndorff-Nielsen et al (2011). However, the "refresh time sampling" scheme used in this estimator means that sampling must be done according to the slowest trading asset. This may lead to a large loss in data in high dimensions. In fact, in Gorgi et al (2018)'s sample, the computation of a 15×15 realized kernel involved discarding roughly 80% of the data. Given that this paper uses covariance matrices up to a dimension of 20×20, the realised covariance estimator of (4) was deemed more appropriate.

	BA	BAC	DUK	GS	HSY	IBM	JNJ	JPM	KO	MRK	NKE	ORCL	PG	RL	TGT	UNH	VZ	WMT	XOM	XRX
BA	25.37	0.339	0.286	0.364	0.247	0.402	0.331	0.366	0.349	0.320	0.339	0.354	0.326	0.293	0.344	0.222	0.353	0.352	0.400	0.297
BAC		38.32	0.230	0.499	0.213	0.366	0.265	0.584	0.279	0.266	0.309	0.311	0.269	0.303	0.341	0.199	0.312	0.322	0.352	0.284
DUK			24.13	0.259	0.230	0.306	0.284	0.267	0.304	0.274	0.245	0.252	0.283	0.197	0.259	0.173	0.308	0.294	0.342	0.214
GS				31.55	0.233	0.402	0.292	0.540	0.307	0.294	0.329	0.343	0.291	0.313	0.362	0.212	0.344	0.358	0.399	0.295
HSY					20.33	0.266	0.255	0.232	0.326	0.242	0.231	0.225	0.280	0.203	0.235	0.173	0.253	0.254	0.258	0.192
IBM						20.40	0.366	0.409	0.383	0.349	0.351	0.444	0.364	0.298	0.371	0.230	0.400	0.402	0.437	0.323
JNJ							17.70	0.306	0.380	0.405	0.280	0.314	0.371	0.214	0.290	0.226	0.351	0.347	0.367	0.243
JPM								34.16	0.320	0.301	0.334	0.349	0.296	0.303	0.370	0.209	0.359	0.368	0.386	0.301
KO									17.93	0.348	0.298	0.325	0.394	0.225	0.307	0.221	0.365	0.375	0.383	0.249
MRK										23.86	0.278	0.299	0.332	0.221	0.289	0.229	0.342	0.326	0.355	0.242
NKE											25.56	0.306	0.284	0.377	0.356	0.196	0.309	0.348	0.335	0.272
ORCL												29.08	0.308	0.249	0.319	0.196	0.345	0.354	0.386	0.283
PG													18.68	0.220	0.296	0.207	0.341	0.351	0.369	0.238
RL														31.53	0.337	0.173	0.253	0.288	0.278	0.253
TGT															28.82	0.196	0.329	0.503	0.341	0.273
UNH																36.21	0.208	0.204	0.233	0.176
VZ																	23.37	0.373	0.395	0.282
WMT																		19.71	0.376	0.265
XOM																			21.82	0.298
XRX																				35.32

Table 2: Annualised realised variance (%) (diagonal) and realised correlation (off-diagonal)

	BA	BAC	DUK	GS	HSY	IBM	JNJ	JPM	KO	MRK	NKE	ORCL	PG	RL	TGT	UNH	VZ	WMT	XOM	XRX
BA	0.000	0.025	0.027	0.006	0.048	0.015	0.135	0.077	0.001	0.048	0.010	0.014	0.007	0.073	0.123	0.047	0.009	0.094	0.002	0.000
BAC		0.100	0.488	0.091	0.192	0.060	0.177	0.453	0.023	0.126	0.070	0.189	0.021	0.200	0.517	0.102	0.105	0.586	0.064	0.002
DUK			0.010	0.080	0.360	0.313	0.504	0.499	0.108	0.781	0.266	0.382	0.141	0.333	0.336	0.644	0.077	0.768	0.139	0.004
GS				0.006	0.029	0.006	0.026	0.322	0.004	0.086	0.044	0.038	0.002	0.114	0.144	0.064	0.013	0.073	0.005	0.000
HSY					-0.625	0.123	0.484	0.264	0.010	0.328	0.148	0.209	0.028	0.061	0.401	0.051	0.111	0.426	0.033	0.000
IBM						0.000	0.256	0.109	0.008	0.150	0.053	0.093	0.010	0.089	0.237	0.040	0.029	0.214	0.013	0.000
JNJ							0.070	0.390	0.054	0.677	0.336	0.618	0.057	0.246	0.547	0.420	0.204	-0.747	0.088	0.002
JPM								0.002	0.067	0.389	0.207	0.455	0.096	0.480	0.779	0.152	0.049	0.597	0.077	0.006
KO									0.001	0.088	0.020	0.030	0.002	0.037	0.077	0.007	0.004	0.085	0.001	0.000
MRK										0.004	0.243	0.237	0.089	0.405	0.377	0.484	0.088	0.821	0.099	0.004
NKE											0.987	0.150	0.064	0.183	0.285	0.187	0.054	0.377	0.039	0.000
ORCL												0.000	0.037	0.225	0.850	0.075	0.061	0.790	0.095	0.000
PG													0.250	0.039	0.098	0.035	0.005	0.198	0.006	0.000
RL														0.015	0.776	0.552	0.242	0.470	0.055	0.001
TGT															0.008	0.246	0.094	0.876	0.130	0.004
UNH																-0.778	0.041	0.572	0.020	0.001
VZ																	0.001	0.238	0.016	0.000
WMT																		0.009	0.239	0.015
XOM																			0.001	0.000
XRX																				0.002

Table 3: Tests for leverage. Entries are p-val $\times sign(\gamma)$ for diagonal elements, and p-val $\times sign(\gamma - \phi)$ for off diagonal

3.2 In-sample results

Now I present the estimation results for the RWG model of Gorgi et al (2018) and various specifications of the HEAVY GAS tF model. In particular, the specifications estimated for the latter include the baseline recursion in (7) (HGtF); the HAR specification in (13) (HGtF-HAR), the two-component specification (15) (HGtF-TC); the specification with leverage (16) (HGtF-L); and finally the specification with two-components and leverage (HGtF-TC-L) (17).

The full sample of 3273 trading days is used for estimation. Since the sample contains observations from the low volatility regime of the early 2000s, the Financial Crisis, and the increasingly electronic equity markets of recent times, it provides a valuable test for the robustness of our model.

Table 4 presents the estimation results for four mutually exclusive cross sections of size $k = 5$ and for the full set of equities with $k = 20$. Apart from the partition (BAC/GS/JPM/MRK/PG), which is intended to be heavy on financial stocks, the other partitions are chosen randomly. The table includes parameter estimates, standard errors, log likelihood (LL) and Bayesian information criterion (BIC). Standard errors are computed using the inverse of the Hessian of the log-likelihood.

Firstly, note that allowing for fat tails in the conditional distributions of daily returns and realized covariances improves model fit dramatically. That is, moving from the RWG to the HGtF leads to a gain in log likelihood between c.17,000 - 33,000 for $k = 5$. The reduction in the BIC is correspondingly large, given that the HGtF only requires the additional estimation of two more parameters relative to the RWG. The gain for $k = 20$ is even larger, suggesting that the superiority of the HGtF model over the RWG is increasing in the cross-sectional dimension, k. These findings are in line with those of Opschoor et al (2018). Some intuition can be drawn from Figure 1, which plots RC_t and the fitted values, $\hat{V}_t$, for two stocks, GS and MRK. While the fitted values from both models (right) are less noisy than the RC_t series (left), the HGtF series tend to be more robust to short term spikes/noise in volatility and covariance than the RWG. This fact stands out in the volatility series of MRK and in the correlation series in particular. However, a byproduct is that the HGtF model seems to under-predict the level of volatility during the Financial Crisis.

Furthermore, the results show that incorporating long-memory into the HGtF model leads to a significant improvement in the fit, albeit much less than allowing for fat-tails. Adding HAR dynamics to the recursion for V_t reduces the BIC by between 1300 - 2000 for $k = 5$ and by c.20,000 for $k = 20$. Meanwhile, the improvement due to the two component specification is roughly half as much. This suggests that HAR dynamics are more effective in approximating the long-memory behaviour of V_t than two components.

As for the question of leverage, first define the standardized residuals $\hat{\zeta}_t = \hat{V}_t^{-1/2} r_t$ and $\hat{\eta}_t = \hat{V}_t^{-1/2} RC_t \hat{V}_t^{-1/2}$ for the baseline HGtF model, where $\hat{V}_t$ are the fitted values. Now consider Figure 2 (inspired by Noureldin et al (2011)). Figure 2a is a scatter plot of the element $(\hat{\zeta}_t)_i$ against the diagonal element $(\hat{\eta}_t)_{ii}$, where i corresponds to the index for the GS stock. Note, the residuals have been mapped to the copula space by applying the probability integral transform based on the empirical distribution function. Similarly, Figure 2b is the equivalent plot for the MRK stock. In both Figures 2a and 2b, the scatter plot is particularly dense near the bottom right corner. This suggests that large negative

	α_1	α_2	α^*	β_1	β_2	β_3	ν_0	ν_1	ν_2	LL	BIC
Panel A: HSY/JNJ/RL/TGT/XOM											
RWG	0.343			0.942				12.81		-51,746	103,556
	(0.005)			(0.003)				(0.059)			
HGtF	0.989			0.989			9.38	84.56	27.75	-32,384	64,809
	(0.019)			(0.030)			(0.685)	(1.429)	(0.482)		
HGtF-HAR	0.906			0.900	0.056	0.037	9.60	71.64	30.62	-31,644	63,344
	(0.024)			(0.040)	(0.000)	(0.004)	(0.621)	(1.335)	(0.559)		
HGtF-TC	0.566	0.652		0.996	0.758		9.52	90.32	27.60	-32,042	64,141
	(0.005)	(0.068)		(0.016)	(0.023)		(0.620)	(1.398)	(0.442)		
HGtF-L	0.967		0.040	0.967			9.27	89.45	27.38	-32,296	64,640
	(0.039)		(0.000)	(0.033)			(0.677)	(1.690)	(0.517)		
HGtF-TC-L	0.521	0.649	0.192	0.998	0.712		9.49	93.35	27.21	-31,772	63,608
	(0.004)	(0.059)	(0.002)	(0.031)	(0.026)		(0.684)	(1.443)	(0.569)		
Panel B: DUK/KO/NKE/ORCL/XRX											
RWG	0.242			0.976				14.98		-58,011	116,087
	(0.002)			(0.007)				(0.068)			
HGtF	0.988			0.988			9.51	88.95	32.21	-41,092	82,224
	(0.026)			(0.029)			(0.615)	(1.721)	(0.684)		
HGtF-HAR	0.899			0.893	0.063	0.038	9.89	76.31	35.96	-40,097	80,251
	(0.020)			(0.024)	(0.003)	(0.001)	(0.548)	(1.735)	(0.695)		
HGtF-TC	0.554	0.767		0.997	0.767		9.70	105.86	31.19	-40,584	81,225
	(0.006)	(0.072)		(0.045)	(0.021)		(0.732)	(1.509)	(0.583)		
HGtF-L	0.970		0.032	0.970			9.38	94.38	31.68	-41,024	82,097
	(0.036)		(0.000)	(0.028)			(0.703)	(1.351)	(0.622)		
HGtF-TC-L	0.543	0.719	0.188	0.998	0.723		9.57	109.57	31.56	-40,277	80,619
	(0.009)	(0.066)	(0.003)	(0.037)	(0.021)		(0.711)	(1.644)	(0.701)		
Panel C: BA/IBM/UNH/VZ/WMT											
RWG	0.260			0.835				9.16		-65,271	130,607
	(0.000)			(0.002)				(0.055)			
HGtF	0.990			0.990			10.21	87.83	28.89	-32,509	65,059
	(0.022)			(0.021)			(0.583)	(1.301)	(0.671)		
HGtF-HAR	0.911			0.907	0.046	0.042	10.48	75.82	31.57	-31,817	63,690
	(0.028)			(0.031)	(0.003)	(0.000)	(0.626)	(1.364)	(0.665)		
HGtF-TC	0.501	0.736		0.998	0.826		10.37	98.21	28.34	-32,179	64,415
	(0.008)	(0.029)		(0.029)	(0.016)		(0.641)	(1.378)	(0.586)		
HGtF-L	0.990		0.000	0.990			10.20	87.89	28.88	-32,509	65,066
	(0.024)		(0.000)	(0.021)			(0.627)	(1.369)	(0.595)		
HGtF-TC-L	0.509	0.731	0.146	0.998	0.789		10.55	99.52	28.79	-32,056	64,176
	(0.006)	(0.061)	(0.004)	(0.028)	(0.030)		(0.639)	(1.309)	(0.599)		
Panel D: BAC/GS/JPM/MRK/PG											
RWG	0.682			0.880				11.05		-62,818	125,702
	(0.001)			(0.000)				(0.062)			
HGtF	0.994			0.994			8.15	75.58	31.07	-33,786	67,613
	(0.031)			(0.035)			(0.597)	(1.395)	(0.613)		
HGtF-HAR	0.910			0.906	0.047	0.045	8.42	65.92	34.53	-32,929	65,914
	(0.025)			(0.026)	(0.003)	(0.000)	(0.631)	(1.413)	(0.607)		
HGtF-TC	0.505	0.774		0.999	0.811		8.29	87.95	30.12	-33,260	66,577
	(0.007)	(0.058)		(0.021)	(0.017)		(0.605)	(1.267)	(0.607)		
HGtF-L	0.984		0.017	0.984			8.09	77.85	30.80	-33,753	67,554
	(0.018)		(0.000)	(0.019)			(0.613)	(1.493)	(0.626)		
HGtF-TC-L	0.501	0.759	0.169	0.999	0.803		8.34	91.14	31.03	-33,085	66,234
	(0.006)	(0.053)	(0.001)	(0.031)	(0.021)		(0.648)	(1.452)	(0.602)		
Panel E: BA/BAC/DUK/GS/HSY/IBM/JNJ/JPM/KO/MRK/NKE/ORCL/PG/RL/TGT/UNH/VZ/WMT/XOM/XRX											
RWG	0.199			0.943				31.16		-70,243	140,672
	(0.000)			(0.000)				(0.009)			
HGtF	0.792			0.994			12.13	220.68	64.07	223,868	-447,695
	(0.001)			(0.000)			(0.132)	(0.517)	(0.112)		
HGtF-HAR	0.888			0.881	0.072	0.043	12.62	207.94	67.72	233,769	-467,481
	(0.000)			(0.000)	(0.000)	(0.000)	(0.142)	(0.599)	(0.120)		
HGtF-TC	0.393	0.570		0.998	0.815		12.41	224.12	64.48	228,479	-456,901
	(0.000)	(0.000)		(0.000)	(0.000)		(0.135)	(0.529)	(0.118)		
HGtF-L	0.835		0.014	0.985			11.94	241.00	62.84	224,077	-448,104
	(0.000)		(0.000)	(0.002)			(0.122)	(0.624)	(0.116)		
HGtF-TC-L	0.396	0.568	0.194	0.999	0.797		12.34	242.53	63.91	230,881	-461,697
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		(0.144)	(0.592)	(0.121)		

Table 4: Insample results. T = 3273. Standard errors in parentheses

innovations to returns tend to be associated with large positive innovations to the realised variance. Thus, the basic HGtF model does not sufficiently capture leverage.

Now, consider the estimation results for the HGtF-L model in Table 4. Compared to the introduction of long memory into the HGtF model, the introduction of leverage leads to a relatively small reduction in BIC. For $k = 5$, the reduction ranges 59 - 168 in three out of four cases. However, in Panel C, the BIC actually increases by 7 points. For $k = 20$, the BIC decreases by 409 points.

Nevertheless, recall from the discussion in Section 3.2.2 that the positive definiteness condition (18) can become restrictive in sample. That is, if β is close to one (i.e. if V_t is highly persistent), then α^* must necessarily be close to zero, even for moderate values of δ. Clearly, this is an issue in the current sample. The smallest estimate for β_1 in the baseline HGtF specification is 0.989. Furthermore, in some cases, the eigenvalue δ was as large as 0.8 in sample. Therefore, it is unsurprising that the magnitude of the estimates of α^* tend to be very small in all of the HGtF-L specifications.

As discussed, the solution is to estimate a two component model with the leverage effect confined to the short term component - the HGtF-TC-L model in Table 4. Now, the value of α^* is an order of magnitude higher than in the one component model. Furthermore, relative to the two component model without leverage (HGtF-TC), introducing leverage leads to a drop in BIC between 239 - 606 for $k = 5$ and of 4796 for $k = 20$. This is much larger than the improvement when leverage is introduced into the one component model. Therefore, the two component model with leverage provides a practical way to circumvent the trade-off between modelling the long-memory of V_t and its asymmetric dynamics.

Another point worth noting is that the HGtF-HAR specification outperforms the HGtF-TC-L specification for all cross sections. This is not necessarily a criticism of the latter model however. The point of introducing the two component model with leverage was to provide a way to model leverage and long memory in V_t simultaneously. The long term component could be extended to include HAR dynamics, and it is expected that such a model would perform even better than the HGtF-HAR model here.

Finally, a discussion of the degrees of freedom parameters is also insightful. Firstly, the parameters ν_0 and ν_2 are sufficiently small so that both the returns and realised covariances exhibit fat tails. Furthermore, the large values for ν_1 in the HGtF models mean that the score, (9), almost entirely puts all the weight on RC_t (as opposed to $r_t r_t'$) when updating V_{t+1}.

3.3 Out-of-sample results

In this section, I evaluate the out of sample forecasting performance of a selection of the models considered above. In particular, the forecasting performance of the two component HEAVY GAS tF model with leverage (HGtF-TC-L) is benchmarked against the RWG, HGtF and the HGtF-HAR. Additionally, following Gorgi et al (2018) and Opschoor et al (2018), I also construct forecasts using an EWMA strategy:

$$V_{t+1} = \beta V_t + (1 - \beta) RC_t \tag{29}$$

with $\beta = 0.96$.

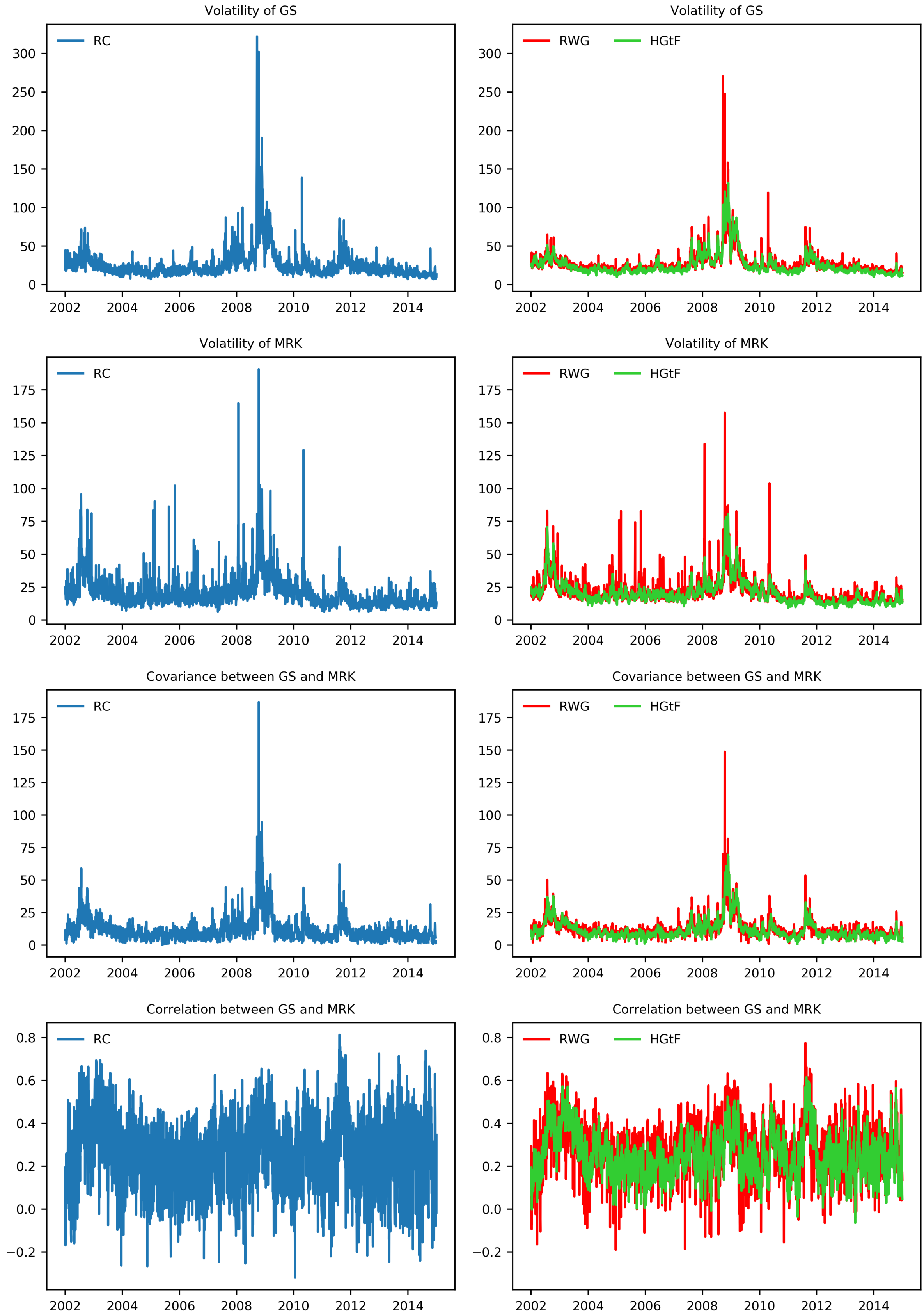
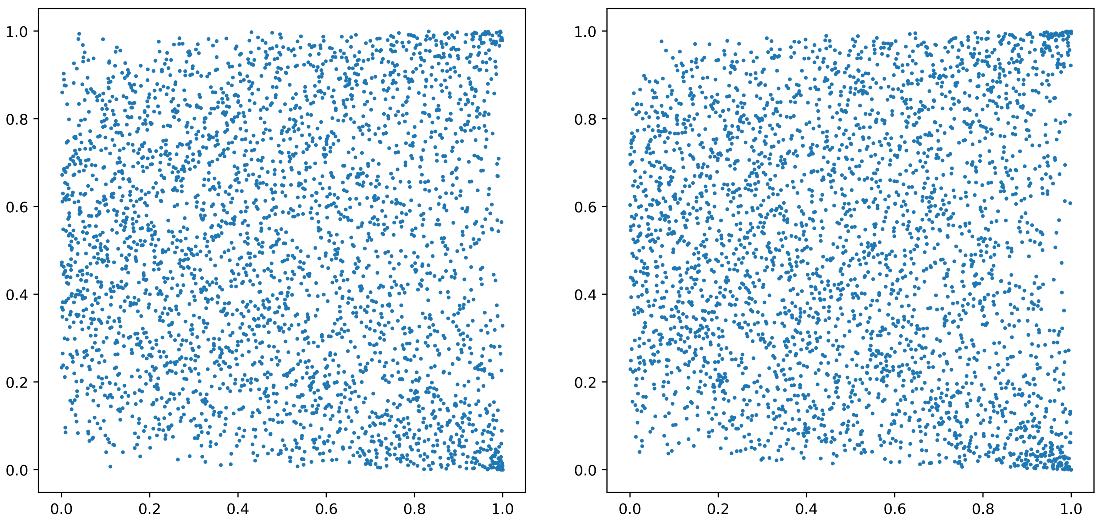


Figure 1: RC_t (left) and fitted values (right). The volatilities and covariance are annualised



(a) GS

(b) MRK

Figure 2: Residual analysis

Here, the analysis is confined to one-step ahead forecasts. Furthermore, a rolling estimation window is deployed, whereby only the 1500 most recent observations are used for estimation. In addition, the model parameters are updated every 252 observations, which is roughly equal to a calendar year. The hold-out sample therefore contains 1773 observations and includes the Financial Crisis. The estimation and updating strategy just mentioned is similar to that of Opschoor et al (2018).

Evaluation of the forecasts is conducted through both a statistical and portfolio optimisation perspective. In both cases, the Diebold-Mariano (1995) (DM) test is used to test whether the differences in forecasting performance of models are statistically significant. To be precise, define the loss function for model q at time t as $L_{q,t}(S_t, V_t)$, where S_t is an observed measure of the covariance matrix and V_t is the forecast. Further, define λ_t as the difference between the loss functions for two models. Under the null of equal forecasting performance ($\mathbb{E}[\lambda_t] = 0 \ \ \forall t$), the DM test statistic is

$$DM = \frac{\bar{\lambda}}{\sqrt{\hat{\sigma}^2/N}} \tag{30}$$

where $\bar{\lambda}$ is the out-of-sample sample mean of λ_t and $\hat{\sigma}^2$ is a HAC-consistent estimator of the variance of λ_t. Under the null, the DM statistic is asymptotically standard normal.

For the statistical evaluation, the loss function used is the matrix root mean squared error (RMSE):

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

As suggested by Gorgi et al (2018), the forecasting performances are evaluated in terms of forecasting the density of daily returns ($Z_t = r_t r_t'$) and in terms of forecasting the realized covariance matrix ($Z_t = RC_t$). This is motivated by the fact that our models jointly analyze both RC_t and r_t.

An economic evaluation of the model forecasts is also undertaken. As in Chiriac and Voev (2011), Jin and Maheu (2013) and Opschoor et al (2018), I focus on the problem of finding the global minimum variance portfolio (GMVP). The 1-step ahead GMVP is computed as the solution to the problem

$$\min_{w_{t+1|t}} w'_{t+1|t} V_{t+1|t} w_{t+1|t}$$
$$\text{s.t. } w'_{t+1|t}\iota = 1$$

where $\iota \in \mathbb{R}^k$ is a vector of ones, and $V_{t+1|t}$ is the one-step ahead forecast of the covariance matrix. The interpretation of the problem is that, given her information set at time t, the investor chooses portfolio weights, $w_{t+1|t}$ to minimise the portfolio volatility over period $t+1$. Engle and Colacito (2006) show that the portfolio weights selected according to the true conditional covariance matrix generate a lower bound on the volatility of portfolios computed using the GMVP. Thus, the more accurately a model forecasts the covariance matrix, the lower should be the subsequent portfolio variance.

Following Chiriac and Voev (2011), the forecasting performance of competing models is evaluated by comparing ex-post realizations of the conditional standard deviation, $\sigma_t^p = \sqrt{w'_{t+1|t} RC_{t+1} w_{t+1|t}}$. As before, the DM statistic, (30), is used to test whether the differences in σ_t^p across models are statistically significant.

The results for the out-of-sample forecasting analysis are summarised in Table 5. The first column in the top panel of Table 5 displays the mean RMSE for the HGtF-TC-L model with $Z_t = RC_t$. The mean RMSE for the other models is presented as a proportion of this value. A proportion larger than 1 means that the model in question has a larger out-of-sample loss than the HGtF-TC-L. The opposite is true for a proportion less than 1. The table also indicates the statistical significance of the differences in out-of-sample loss between the HGtF-TC-L model and the other models, as per the Diebold-Mariano test. The second panel of Table 5 presents the out-of-sample RMSE in a similar fashion, but with $Z_t = r_t r'_t$.

The bottom panel of Table 5 presents the mean values of the ex-post portfolio standard deviations. Note, they are not presented as proportions in this case. Again, the statistical significance of the difference in standard deviation between the HGtF-TC-L model and the other models is indicated.

The out-of-sample results provide a similar narrative to the in-sample results. Firstly, note that the HGtF-TC-L model comfortably outperforms the RWG and EWMA models in all comparisons and for all cross sections.

Next, it is interesting to note the differences in results when $Z_t = RC_t$ versus when $Z_t = r_t r'_t$. In the former case, the HGtF-TC-L model tends to outperform the baseline HGtF model, and tends to be outperformed by the HGtF-HAR specification. However, when $Z_t = r_t r'_t$, the out-of-sample forecasting performance of the different specifications of the HEAVY GAS model are indistinguishable. A possible explanation for this might be gleaned from our in-sample results. There, we saw that the HEAVY GAS model hardly attaches any weight to the information in $r_t r'_t$, as this provides a relatively imprecise estimate of the association between different stocks. Therefore, when $Z_t = r_t r'_t$, there is hardly any forecasting gain to be had when extending the HEAVY GAS tF model to richer specifications.

As for the economic evaluation, the results again suggest that the HGtF-TC-L model lies somewhere between the baseline HGtF specification and HAR specification in terms of performance.

RMSE, $Z_t = RC_t$	TC-L	RWG	HGtF	HAR	EWMA
BAC/GS/JPM/MRK/PG	**6.698**	1.125***	1.011*	0.994	1.215***
BA/IBM/UNH/VZ/WMT	**7.549**	1.126***	1.013*	0.991*	1.263***
DUK/KO/NKE/ORCL/XRX	**10.09**	1.159***	1.015*	0.998	1.237***
JNJ/RL/T/TGT/XOM	**8.923**	1.112***	1.021**	0.989*	1.229***
All 20 stock	**122.91**	1.162***	1.033***	0.994*	1.222***

RMSE, $Z_t = r_t r_t'$	TC-L	RWG	HGtF	HAR	EWMA
BAC/GS/JPM/MRK/PG	**22.17**	1.018***	1.002	0.999	1.032***
BA/IBM/UNH/VZ/WMT	**31.46**	1.021***	1.002	0.999	1.039***
DUK/KO/NKE/ORCL/XRX	**22.81**	1.015***	1.001	1.000	1.041***
JNJ/RL/T/TGT/XOM	**24.09**	1.020***	1.001	0.999	1.035***
All 20 stocks	**378.12**	1.027***	1.001	1.000	1.033***

σ^p	TC-L	RWG	HGtF	HAR	EWMA
BAC/GS/JPM/MRK/PG	**0.963**	0.971***	0.963	0.962***	0.973***
BA/IBM/UNH/VZ/WMT	**0.891**	0.897***	0.892***	0.891	0.901***
DUK/KO/NKE/ORCL/XRX	**0.856**	0.860***	0.857***	0.856	0.864***
JNJ/RL/T/TGT/XOM	**0.882**	0.890***	0.883***	0.881***	0.893***
All 20 stocks	**0.698**	0.707***	0.699***	0.697***	0.711***

***,**, and * indicate significance at the 1%, 5% and 10% level respectively (DM test)

Table 5: Out-of-sample results

4 Conclusion

To conclude, the main contribution of this paper has been to extend the HEAVY GAS tF model of Opschoor et al (2018) to a specification that can accommodate both long memory and leverage effects. The proposed model splits the dynamics of the covariance matrix into a long term and short term component and adds a leverage effect to the short term component only. In the empirical application, the proposed model outperformed the baseline HEAVY GAS tF specification, both in and out of sample. Although the two component model was outperformed by the HAR specification suggested by Opschoor et al (2018), this paper argued that this is not necessarily a criticism of the two component model. Indeed, the point of the two component specification is to enable the simultaneous modelling of leverage and long memory. Future work should look to combine the HAR and two component specifications.

Bibliography

Ait-Sahalia, Y., P. A. Mykland, and L. Zhang. 2005. A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. Journal of the American Statistical Association 100(472):1394–1411

Alizadeh, Sassan, Michael W. Brandt, and Francis X. Diebold. "Range-Based Estimation of Stochastic Volatility Models." Journal of Finance 57.3 (2002): 1047-091. Web.

Andersen, Bollerslev, Christoffersen, & Diebold. (2006). Chapter 15 Volatility and Correlation Forecasting. Handbook of Economic Forecasting, 1, 777-878.

Asai, M., McAleer, M., and Yu, J. (2006), "Multivariate Stochastic Volatility: A Review," Econometric Reviews, 25, 145–175

Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996) Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 74, 3-30.

Barndorff-Nielsen, O., Hansen, P., Lunde, A., and Shephard, N. (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. 2011a. Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading. Journal of Econometrics 162(2):149–169

Barndorff-Nielsen OE, Shephard N. 2004. Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. Econometrica 72: 885–925.

Bauwens, L., Laurent, S., and Rombouts, J. (2006), "Multivariate Garch Models: A Survey," Journal of Applied Econometrics, 21, 79–10

Black, F. (1976). ''Studies of Stock Prices Volatility Changes.'' Proceeding from the American Statistical Association, Business and Economics Statistics Section 177–181

Blasques, C., Koopman, S., and Lucas, A. (2015), "Information Theoretic Optimality of Observation Driven Time Series Models for Continuous Responses," Biometrika, 102, 325–343

Brownlees, C., and Gallo, G. (2006), "Financial Econometric Analysis at Ultra-High Frequency: Data Handling Concerns," Computational Statistics and Data Analysis, 51, 2232–2245

Cappiello, L., Engle, R., and Sheppard, K. (2006), "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns," Journal of Financial Econometrics, 4, 537–572

Chiriac, R., and Voev, V. (2011), "Modelling and Forecasting Multivariate Realized Volatility," Journal of Applied Econometrics, 26, 922–947

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

Cox, D. (1981), "Statistical Analysis of Time Series: Some Recent Developments," Scandinavian Journal of Statistics, 8, 93–115

Creal, D., Koopman, S., and Lucas, A. (2011), "A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations," Journal of Business and Economic Statistics, 29, 552–563

Creal, D., Koopman, S., and Lucas, A. (2013), "Generalized Autoregressive Score Models with Applications," Journal of Applied Econometrics, 28, 777–795

Diebold, F., and Mariano, R. (1995), "Comparing Predictive Accuracy," Journal of Business and Economic Statistics, 13, 253–263

Engle, R. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroscedasticity Models," Journal of Business and Economic Statistics, 20, 339–350

Robert Engle & Riccardo Colacito (2006) Testing and Valuing Dynamic Correlations for Asset Allocation, Journal of Business & Economic Statistics, 24:2, 238-253, DOI: 10.1198/073500106000000017

Engle, R., and Gallo, G. (2006), "A Multiple Indicators Model for Volatility using Intra-Daily Data," Journal of Econometrics, 131, 3–27

Engle, R.F. and Lee, G. (1999) A Permanent and Transitory Component Model of Stock Return Volatility. Cointegration, Causality and Forecasting: A Festschrift in Honor of Clive W.J. Granger. Oxford University Press, New York.

Engle, R. F., and J. Mezrich. 1996. GARCH for Groups. Risk 9:36–40

Golosnoy, V., Gribisch, B., and Liesenfeld, R. (2012), "The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility," Journal of Econometrics, 167, 211–223

Gorgi, P, P R Hansen, P Janus, S J Koopman, Realized Wishart-GARCH: A Score-driven Multi-Asset Volatility Model, Journal of Financial Econometrics, Volume 17, Issue 1, Winter 2019, Pages 1–32

Gourieroux, C., Jasiak, J., and Sufana, R. (2009), "The Wishart Autoregressive Process of Multivariate Stochastic Volatility," Journal of Econometrics, 150, 167–181

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

Harvey, A., Sucarrat, G. (2014). EGARCH models with fat tails, skewness and leverage. Computational Statistics and Data Analysis, p. 320-338.

Jin, X., and Maheu, J. (2013), "Modeling Realized Covariances and Returns," Journal of Financial Econometrics

Jin, X., and Maheu, J. (2016), "Bayesian Semiparametric Modeling of Realized Covariance Matrices," Journal of Econometrics, 192, 19–39

Madhavan, A. (2000). Market microstructure: A survey. Journal of Financial Markets, 3(3), 205-258

Abadir, K.M., Magnus, J.R., and Tilburg School of Economics Management. Matrix Algebra. 2005. Web.

Noureldin, D., Shephard, N., and Sheppard, K. (2012), "Multivariate High-Frequency-Based Volatility (Heavy) Models," Journal of Applied Econometrics, 27, 907–933.

Opschoor, A., Lucas, A., Fractional Integration and Fat Tails for Realized Covariance Kernels, Journal of Financial Econometrics, Volume 17, Issue 1, Winter 2019, Pages 66–90, https://doi.org/10.1093/jjfinec/nby029

Silvennoinen, A., and T. Tera¨svirta. 2009. “Multivariate GARCH Models”. In T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch (eds.), Handbook of Financial Time Series, 201–229. Berlin: Springer