Fractionally Integrated Multivariate Models for Fat-Tailed Realized Covariance Kernels and Returns

Andre Lucas[a] and Anne Opschoor[a*]

[a] *Vrije Universiteit Amsterdam and Tinbergen Institute*

This version: January 29, 2016

Abstract

We introduce a new fractionally integrated model for multivariate covariance matrix dynamics based on the long-memory behavior of daily realized covariance matrix kernels and daily return observations. We account for fat-tailedness in both types of data by assuming a matrix-F distribution for the realized kernels and a multivariate Student's t distribution for the returns. In addition, the score-driven propagation mechanism adopted for the covariance matrix dynamics endows our model with further robustness properties. Using intraday stock data over the period 2001-2012, we construct realized covariance kernels and show that the new fractionally integrated model outperforms recent alternatives such as the Multivariate HEAVY model and the Riskmetrics 2006 (long-memory) model both statistically and economically.

Keywords: multivariate volatility; fractional integration; realized covariance matrices; heavy tails; matrix-F distribution; score dynamics.

*Corresponding author, e-mail address: a.opschoor@vu.nl. Postal address: Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands. Phone number: +31(0)20-5982663.

1 Introduction

The recent financial crisis reopened the interest for adequate multivariate volatility models in various areas of risk- and portfolio management. The econometric literature on these models has mainly developped along two lines, namely that of multivariate GARCH models (for an overview, see Silvennoien and Teräsvirta, 2009) and stochastic volatility models (for an overview, see Asai *et al.*, 2006). More recently, the availability of intraday high-frequency data has led to a new class of volatility models. These models capture multivariate volatility dynamics by including realized (co)variance measures, which helps to measure and forecast volatility more precisely than do traditional squares and cross-products of returns; see for instance Andersen *et al.* (2001). Typically, these models include either the realized variance measures of Barndorff-Nielsen and Shephard (2002), or the realized kernel measures of Barndorff-Nielsen *et al.* (2008). As an example of the latter, we refer to the HEAVY model of Shephard and Sheppard (2010). Also Noureldin *et al.* (2012) develop the multivariate analogue of the HEAVY model that incorporates the realized covariance or realized kernel covariance into the model specification.

Typically, volatilities are strongly persistent. This has led to the further introduction of models that also account for this data feature. A salient example is the fractionally integrated GARCH (FI-GARCH) models of Baillie *et al.* (1996), which captures the strong persistence in volatilities by a model with long-memory features. Andersen *et al.* (2001) argue that realized volatility measures are also highly persistent and behave as slowly mean-reverting or fractionally integrated processes, which can be modeled by ARFIMA models; see also Koopman *et al.* (2005) and Janus *et al.* (2014a). As an alternative to the fractionally integrated lag structure, Corsi (2009) develops the HAR model, which relates realized volatility to a linear combination of lagged daily, weekly and monthly realized volatility to incorporate the long-memory effect of volatility.

Despite the abundance of available univariate volatility models that use daily returns and/or realized measures with long-memory features, we note two main shortcomings of these models that impede their application to the typical multivariate context. First, these models do not account for fat-tailed returns and outliers in both the realized measures and the returns. Though fat-tailed distributions are often used to describe the returns, thin-tailed distributions are typically used for the realized measures despite the fact that also data for the realized measures can be subject to outliers and influential observations. For example, the Flash Crash in 2010 led to a spike in the realized (co)variance of a large number of assets. Ignoring these data features both in the likelihood and in the volatility propagation dynamics may have a huge impact on the estimated dynamics for each of the recently proposed volatility models discussed above. Second, multivariate models that incorporate the long-memory feature of (realized) (co)variances face the challenge to simultaneously avoid the curse of dimensionality and solve the requirement of ensuring positive definite covariance matrices. Chiriac and Voev (2011) deal with these issues by proposing VARFIMA models for the cholesky decomposition of the realized covariance matrix, while Bauer and Vorkink (2011) consider the matrix log transformation of the realized covariance matrix. Both studies, however, model the vectorized (vech) matrix of interest, which may become computationally intensive when the dimension increases.

In this paper, we solve both of the above issues by introducing a new multivariate volatility model for realized (kernel) covariance matrices and daily return vectors. We allow for both the long-memory behavior and the fat-tailedness of (realized) covariances and returns by combining fractionally integrated processes with the generalized autoregressive score (GAS) dynamics of Creal *et al.* (2011, 2013). The only paper to our knowledge that combines long-memory and GAS is Janus *et al.* (2014a), but this paper is set entirely in a univariate context and does not incorporate realized measures. The generalized autoregressive score-driven framework uses the derivative of the log conditional probability density function to drive the dynamics of the time-varying parameters, which in our case is the covariance matrix. The framework has been applied to many settings, including volatility and location modeling (Harvey, 2013; Harvey and Luati, 2014), credit risk management (Creal *et al.*, 2014), systemic risk management (Oh and Patton, 2013; Lucas *et al.*, 2014). The availability of a closed-form expression for the likelihood function and the optimality of score-driven steps (see Blasques *et al.*, 2015) make the GAS framework a good starting point for combining long-memory, fat tails, robust time-varying parameter dynamics, and ease of estimation.

To account for fat tails, we assume a matrix-F distribution for the realized covariance matrix and a Student's t distribution for the daily returns. The use of the matrix-F distribution for realized volatility models was first propagated in Janus *et al.* (2014b) in a short-memory context. The combination of the matrix-F and vector-valued Student's t distribution allows for a tractable analytic expression for the score with respect to the unknown, dynamic covariance matrix. The score expressions automatically account for a reduced impact of outlying realized covariance matrices and/or return vectors in an intuitive way. This is important, as such influential observations can otherwise corrupt our estimates of the dynamics of the volatility matrix. To incorporate the long-memory feature into our model, we replace the usual short-memory lag polynomials in Creal *et al.* (2013) by their long-memory counterparts. Due to the matrix formulation of our volatility dynamics, this can be done in a parsimonious yet flexible way that allows for generalizations of the model in many directions of empirical interest. The parsimony of the approach is a major asset in the multivariate context, where the curse of dimensionality looms large.

We provide an empirical application of our multivariate Fractionally Integrated GAS model based on the matrix-F and Student's t distribution (FIGAS tF model from now on) on daily realized kernels and daily returns of 15 equities from the S&P 500 index. Our sample spans the period January 2001 to December 2012. Using a forecasting perspective for 1, 5, 10, and 22 days ahead, we compare both statistically and economically the performance of our new dynamic covariance matrix model to several benchmarks, such as the HEAVY model (Noureldin *et al.*, 2012), the GAS tF model (Janus *et al.*, 2014b) and the RM 2006 approach (Zumbach, 2006). Using a quasi-likelihood loss function, the FIGAS model outperforms the competing models, especially for long horizons. We assess the economic significance of our results by considering mean-variance efficient portfolios based on the forecasts. Again we find that the FIGAS tF model outperforms its competitors by producing significantly lower ex-post conditional portfolio standard deviations.

The rest of this paper is set up as follows. In Section 2, we introduce the new FIGAS tF model for realized covariance matrices and return vectors under fat-tails. In Section 3, we apply the model to a panel of daily realized kernels and equity returns. We conclude in Section 4.

2 Modeling Framework

2.1 The Multivariate FIGAS tF model

Consider a $(k \times 1)$ vector process y_t and a $(k \times k)$ matrix process RK_t, $t = 1, \ldots, T$, generated by

$$y_t = \mu + V_t^{1/2}\, z_t, \qquad\qquad z_t | \mathcal{F}_{t-1} \sim D_z(0, \mathrm{I}_k), \tag{1}$$

$$RK_t = V_t^{1/2}\, Z_t\, (V_t^{1/2})', \qquad\qquad Z_t | \mathcal{F}_{t-1} \sim D_Z(\mathrm{I}_k), \tag{2}$$

where $\mathcal{F}_{t-1}$ is the information set containing all information up to time $t-1$, μ denotes the conditional mean vector of the return vector y_t, V_t denotes the conditional covariance matrix, RK_t denotes the realized kernel covariance matrix measure, and z_t and Z_t denote a $(k \times 1)$ vector-valued and $(k \times k)$ matrix-valued innovation with possibly fat-tailed distribution $D_z(\cdot)(0, \mathrm{I}_k)$ and $D_Z(\mathrm{I}_k)$, respectively, such that $\mathbb{E}_t[z_t] = 0$ and $\mathbb{E}_t[Z_t] = \mathrm{Var}[z_t] = \mathrm{I}_k$. The matrix root $V_t^{1/2}$ is defined such that $V_t^{1/2}(V_t^{1/2})' = V_t$. The realized kernel RK_t is a consistent and robust estimator of V_t correcting for market-microstructure noise; for more details, see Barndorff-Nielsen *et al.* (2011). For simplicity and ease of notation, we set $\mu = 0$. Note, however, that we can easily allow for time-varying conditional means μ_t that incorporate for example autoregressive or moving average dynamics into the specification of y_t.

As shown by Robinson (1991), Baillie *et al.* (1996), and Andersen *et al.* (2001), the covariance matrix V_t typically follows a highly persistent stationary process. Following this line of literature, we introduce the Fractionally Integrated Genenerally Autoregressive Score (FIGAS) model

$$(1 - L)^d V_{t+1} = \Omega + B\,(1 - L)^d V_t + A\, s_t \tag{3}$$

where A and B are scalar parameters, Ω is a $(k \times k)$ parameter matrix, s_t is a $(k \times k)$ matrix-valued volatility innovation term, L is the lag operator $L\,V_t = V_{t-1}$, and $(1 - L)^d$ is the fractional difference operator defined by the binomial expansion

$$(1-L)^d = 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + \ldots, \tag{4}$$

for any real order of fractional integration $d > -1$. The innovation term s_t follows the generalized autoregressive score specification of Creal *et al.* (2011, 2013). One of the features of this approach is that s_t is a deterministic function of current and past data that depends on the specification of the conditional observation densities $D_z(\cdot)$ and $D_Z(\cdot)$ in (1) and (2), respectively. In particular, s_t is a combination of scaled derivatives of the log conditional observation densities $D_z(\cdot)$ and $D_Z(\cdot)$, such that it automatically accounts for the tail-behavior and other higher-order moments of both the returns y_t and the realized measures RK_t. The fact that s_t only depends on current and past data makes the model observation-driven in the classification of Cox (1981) and allows for an analytical expression of the likelihood function via a standard prediction error decomposition. This is highly convenient for both estimation and inference. We also note that the score based step results in local improvements in the Kullback-Leibler divergence between the model (1)–(2) and the unknown true data generating process, even in cases where the model is mis-specified; see Blasques *et al.* (2015).

Before proceeding to our matrix-F and Student's t specification of s_t, we comment on some relevant parameter restrictions for the FIGAS model. First, we note that Sowell (1992) shows that in the univariate GARCH case V_t is stationary if $|d| < 0.5$. Empirical values outside this range are therefore of particular concern. We check for this later in the empirical application. Next, we note that the dynamic process in (3) falls within the class of ARMIFA processes for the conditional mean as introduced by Granger and Joyeaux (1980) and Hosking (1981). Therefore, most statistical properties for ARFIMA processes also hold for the univariate FIGAS process, conditional on an assumption of correct specification. The latter ensures that s_t constitutes a martingale difference sequence; see Creal *et al.* (2013). For stationarity in this multivariate context, we therefore conjecture that a necessary restriction is $|B| < 1$, in line with the univariate analysis of Blasques *et al.* (2014). To ensure positive definiteness of the covariance matrices, we further need the restriction $B > A > 0$, which will follow from the particular distributional and scaling choices for s_t made further below.

An important feature of long-memory processes is the variety of different autocovariance functions that is possible; see Janus, Koopman and Lucas (2014a) for some illustrative examples. Although the theoretical properties of the multivariate FIGAS models have not been studied yet, the variety of shapes of autocovariance functions *a fortiori* holds for the multivariate case.

Given the basic structure of the multivariate FIGAS, we now make a particular choice for the conditional observation densities $D_z(\cdot)$ and $D_Z(\cdot)$. In particular, we assume that the return vector y_t follows a Student's t distribution,

$$p_y(y_t|V_t, \mathcal{F}_{t-1}; \nu_0) = \frac{\Gamma((\nu_0 + k)/2)}{\Gamma(\nu_0/2)[(\nu_0 - 2)\pi]^{k/2}|V_t|^{1/2}} \times \left(1 + \frac{y_t V_t^{-1} y_t}{\nu_0 - 2}\right)^{-(\nu_0+k)/2}, \tag{5}$$

with degrees of freedom parameter $\nu_0 > 2$ and V_t a positive definite covariance matrix at time t. Next, we assume that the realized kernel matrix RK_t has a matrix-F distribution, given by

$$p_{RK}(RK_t|V_t, \mathcal{F}_{t-1}; \nu_1, \nu_2) = K(\nu_1, \nu_2) \times \frac{\left|\frac{\nu_1}{\nu_2-k-1} V_t^{-1}\right|^{\frac{\nu_1}{2}} |RK_t|^{(\nu_1-k-1)/2}}{\left|\mathrm{I}_k + \frac{\nu_1}{\nu_2-k-1} V_t^{-1} RK_t\right|^{(\nu_1+\nu_2)/2}}, \tag{6}$$

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

$$K(\nu_1, \nu_2) = \frac{\Gamma_k((\nu_1 + \nu_2)/2)}{\Gamma_k(\nu_1/2)\Gamma_k(\nu_2/2)}, \tag{7}$$

and $\Gamma_k(x)$ is the multivariate Gamma function

$$\Gamma_k(x) = \pi^{k(k-1)/4} \cdot \prod_{i-1}^{k} \Gamma(x + (1 - i)/2); \tag{8}$$

see for example Konno (1991). Hence both observation densities depend on the common time varying covariance matrix V_t. We assume that conditional on V_t and $\mathcal{F}_{t-1}$, returns y_t and realized covariances RK_t are independent. Preliminary data analysis for bivariate cases reveal that conditional correlations, if any, are typically very small such that this is a reasonable assumption for the purpose at hand. The use of a matrix-F distribution for realized kernels was first proposed in Janus *et al.* (2014b) for a short-memory context and short-term forecasting purposes without benchmarking the performance of the model to the HEAVY model of Noureldin *et al.* (2012) or the RiskMetrics 2006 methodology of Zumbach (2006).

Given the two observation densities (5) and (6) and the conditional independence assumption, the time t predictive log likelihood function and its derivatives become

$$\begin{aligned} \mathcal{L}_t &= \log p_y(y_t|V_t, \mathcal{F}_{t-1}; \nu_0) + \log p_{RK}(RK_t|V_t, \mathcal{F}_{t-1}; \nu_1, \nu_2), && (9)\\ s_t &= V_t \left(\nabla_{y,t} + \nabla_{RK,t}\right) V_t / \nu_1 + 1. && (10)\\ \nabla_{y,t} &= \partial \log p_y(y_t|V_t, \mathcal{F}_{t-1}; \nu_0)/\partial V_t, \\ \nabla_{RK,t} &= \partial \log p_{RK}(RK_t|V_t, \mathcal{F}_{t-1}; \nu_1, \nu_2)/\partial V_t, \end{aligned}$$

where $\nabla_{y,t}$ and $\nabla_{RK,t}$ are given by

$$\begin{aligned} \nabla_{y,t} &= \tfrac{1}{2} V_t^{-1} \left[w_t \cdot y_t y_t' - V_t \right] V_t^{-1}, && (11)\\ w_t &= (\nu_0 + k) \cdot (\nu_0 - 2 + y_t' V_t^{-1} y_t)^{-1} \end{aligned}$$

$$\begin{aligned} \nabla_{RK,t} &= \tfrac{1}{2} \nu_1 V_t^{-1} \left[W_t \cdot RK_t - V_t \right] V_t^{-1}, && (12)\\ W_t &= \frac{\nu_1 + \nu_2}{\nu_2 - k - 1} \cdot \left(\mathrm{I}_k + \frac{\nu_1}{\nu_2 - k - 1} RK_t V_t^{-1} \right)^{-1}, \end{aligned}$$

and the score is scaled by the matrix $2(V_t \otimes V_t)/(\nu_1 + 1)$ to account of for the curvature of the log conditional density with respect to V_t; see Janus *et al.* (2014b) for further details on this part of the model. We refer to the complete model as the FIGAS tF model.

The specification of s_t in equations (10)–(12) has a number of interesting features for our fractionally integrated specification. First, given the model's assumptions, s_t forms a martingale difference. This follows directly from the fact that s_t is an $\mathcal{F}_{t-1}$-measurable transformation of the derivative of the model's log conditional density with respect to V_t. This brings the model close the specifications of Granger and Joyeaux (1980) and Hosking (1981) as formulated for the mean in that we have an infinite weighted sum of martingale differences, with the weights co-determinedby the fractional difference polynomial. Second, both the score for the return equation and for the realized measure equation hold familar terms of the form $w_t \; y_t y_t' - V_t$ and $W_t \; RK_t - V_t$, respectively. For the normal distribution, $w_t \equiv 1$, such that V_{t+1} directly reacts to the unweighted deviations of the squared returns $y_t y_t'$ from their expected values in V_t. This is similar to a standard multivariate GARCH model. For fat-tailed distributions, the weights w_t automatically downplay the importance of outlying values of y_t for the future evolution of V_t in accordance with the estimated fatness of the tails (ν_0) of y_t and the current estimate of the covariance matrix V_t. For the realized measure (RK_t) part of the score, we obtain a highly similar result. First consider the case of a Wishart distribution, which is obtained by setting $\nu_2 \to \infty$. In that case, $W_t \equiv \mathrm{I}_k$, and V_t directly reacts to the deviations of the realized measure RK_t from its expected value V_t. This is similar to a matrix-valued model for a time-varying mean. For fat-tailed matrix distributions ($\nu_2 < \infty$), the matrix weight W_t automatically downplays outliers in RK_t in accordance with the tail behavior (ν_2) of the distribution estimated for RK_t. The presence of both w_t and W_t thus gives the model a doubly robust feature for both types of measurements of V_t.

A final ingredient of the model is the parameter ν_1 in (10). This parameter determines the relative weights of $\nabla_{y,t}$ and $\nabla_{RK,t}$ in the evolution of V_t. If ν_1 decreases to its lower limit, the RK_t measurements become increasingly fat-tailed and, as a result, increasingly less reliable as a measurement for the current V_t. This results in a correspondingly lower weight of the realized measure's score in s_t. By contrast, if ν_1 increases without bounds, the realized measure becomes a precise measurement of the current value of V_t. Consequently, the realized measure's score in that case received the full weight in s_t, and the squared daily returns no longer contribute to the dynamics of V_t.

2.2 Estimation

For the estimation of the $(k \times k)$ intercept ω, we use a covariance targeting approach. If $0 < B < 1$ and the model is stationary, taking expectations of (3) and using the martingale difference property of s_t, we obtain

$$(1-L)^d \mathbb{E}_t[V_{t+1}] = \Omega + B \; (1-L)^d \mathbb{E}_t[V_t] \quad \Leftrightarrow \quad \Omega = (1-B) \cdot \mathbb{E}_t[V_{t+1}].$$

We therefore set $\hat{\Omega} = (1 - B)\bar{V}$ for $0 < B < 1$, with $\bar{V}$ the sample average of RK_t. This should be a consistent estimator for Ω if V_t is a stationary process with finite mean. We estimate the remaining static parameter vector $\theta = \{A, B, \nu_0, \nu_1, \nu_2, d\}$ of the FIGAS model by maximum likelihood. To do so, we maximize the log-likelihood $\mathcal{L}_{\mathrm{tF}}(\theta) = \sum_{t=1}^{T} \mathcal{L}_t$, where $\mathcal{L}_t$ is defined in equation (9). The starting value V_1 can be either estimated or set equal to RK_1.

The maximum likelihood estimation for the fractionally integrated model requires truncation of the infinite distributed lags of (4). We follow Baillie *et al.* (1996) and use a fixed truncation at lag 1000. As indicated by Bollerslev and Mikkelsen (1996) the effect of initial conditions for the starting process of the recursions has almost no effect on the parameter estimates, provided that the sample size is sufficiently large. We therefore follow their suggestion to put the pre-innovations to zero.

3 Empirical Application

In this section we apply the FIGAS model to an empirical data set of 15 US equities. All equities are part of the S&P 500 index. We first describe some of the stylized facts of the data. Next, we introduce our two competing models that serve as our benchmark. Finally, we test the in-sample and out-of-sample performance of our model and the competing benchmarks.

3.1 Data

The data consist of daily returns and daily realized covariances measures for 15 US equities. Table 1 provides a overview of the companies considered in our data set. The data spans the period January 2, 2001 until December 31, 2012 and contains $T = 3017$ trading days. We observe consolidated trades (transaction prices) extracted from the Trade and Quote (TAQ) database from 9:30 until 16:00 with a time-stamp precision of one second. We first clean the high-frequency data following the guidelines of Barndorff-Nielsen *et al.* (2009) and Brownlees and Gallo (2006). Next, we construct realized kernels using the refresh-time-sampling methods of Barndorff-Nielsen *et al.* (2011) with the same hyper-parameters as used by Hansen *et al.* (2014).

Figure 1 shows a snapshot of the data by plotting the realized variances (based on the

Table 1: S&P 500 constituents

This table lists 15 companies listed at the S&P 500 index during the period January 2, 2001 until December 31, 2012. Ts denotes the Ticker Symbol and PERMNO is the CRSP identifier.

Nr.	Ts	Permno	Name	Subsector
1	AA	24643	Alcoa Inc.	Materials
2	AXP	59176	American Express Company	Financials
3	BA	19561	The Boeing Company	Industrials
4	CAT	18542	Caterpillar Inc.	Industrials
5	GE	12060	General Electric Company	Industrials
6	HD	66181	The Home Depot	Consumer discretionary
7	HON	10145	Honeywell International	Industrials
8	IBM	12490	International Business Machines	IT
9	JPM	47896	JP Morgan	Financials
10	KO	11308	Coca-Cola	Consumer staples
11	MCD	43449	McDonald's	Consumer discretionary
12	PFE	21936	Pfizer	Health care
13	PG	18163	Procter & Gamble	Consumer staples
14	WMT	55976	Wal-Mart Stores Inc.	Consumer staples
15	XOM	11850	Exxon Mobil	Energy

kernel approach) of Alcoa Inc. (AA) and Caterpillar Inc. (CAT) on the diagonal panels, and the realized correlation and covariance in the off-diagonal panels. The figure shows that both the realized (co)variance(s) and the realized correlation contain a substantial number of spikes. The spikes do not only occur during the global financial crisis, but also during other periods such as the early 2000s. This motivates the use of our GAS framework based on the fat-tailed matrix-F and Student's t distributions, which automatically downweights the impact of such incidental observations on the volatility and covariance dynamics.

The autocorrelation functions in Figure 2 strongly suggest that the realized covariance matrix displays long-memory behavior. After lag 50, the autocorrelation is around 0.4 for the realized kernel volatilities of AA and CAT. Likewise, the autocorrelation of the realized covariance and correlation is equal to 0.25 and 0.3 at this long lag length. This provides an empirical motivation for incorporating long-memory features into the model on top of the fat-tailed, robust volatility dynamics discussed earlier.

3.2 Alternative forecasting models

To benchmark the performance of our FIGAS tF model, we benchmark it against three relevant alternative models: the multivariate HEAVY model of Noureldin *et al.* (2012),

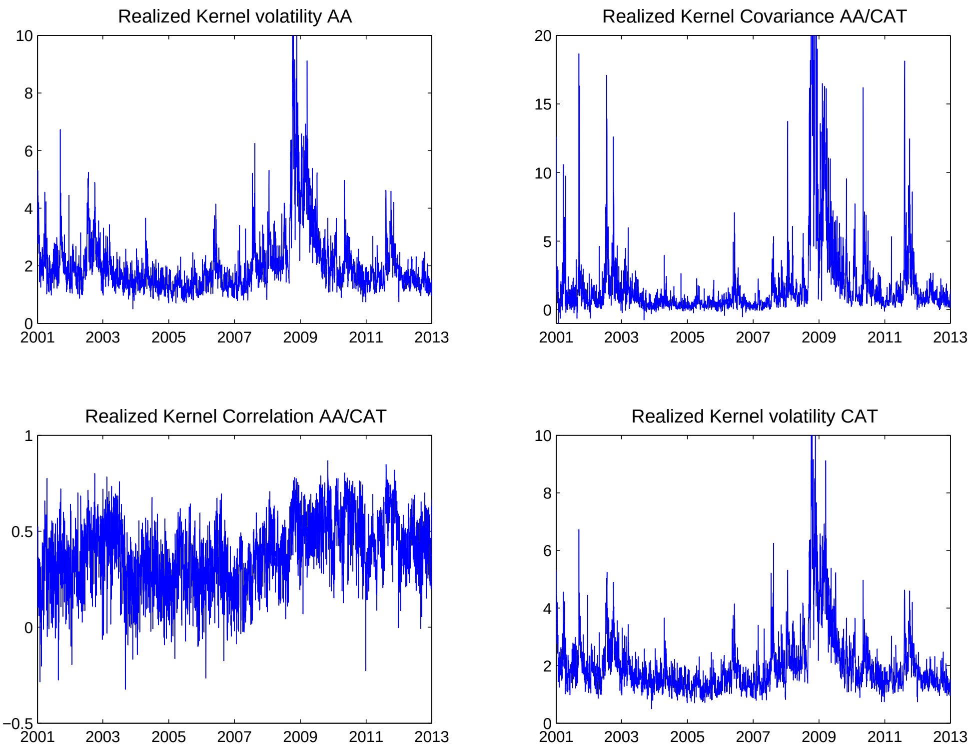

Figure 1: Realized Kernel estimates of AA/CAT

This figure shows daily realized kernel volatilities (square root of the variance) of Alcoa Inc. (AA) and Caterpillar Inc. (CAT) returns on the diagonal panels. The off-diagonal panels contain the realized kernel covariance (upper-right) and correlation (lower-left) between the two asset returns. The sample spans the period from January 2, 2001 until December 31, 2012 ($T = 3017$ days).

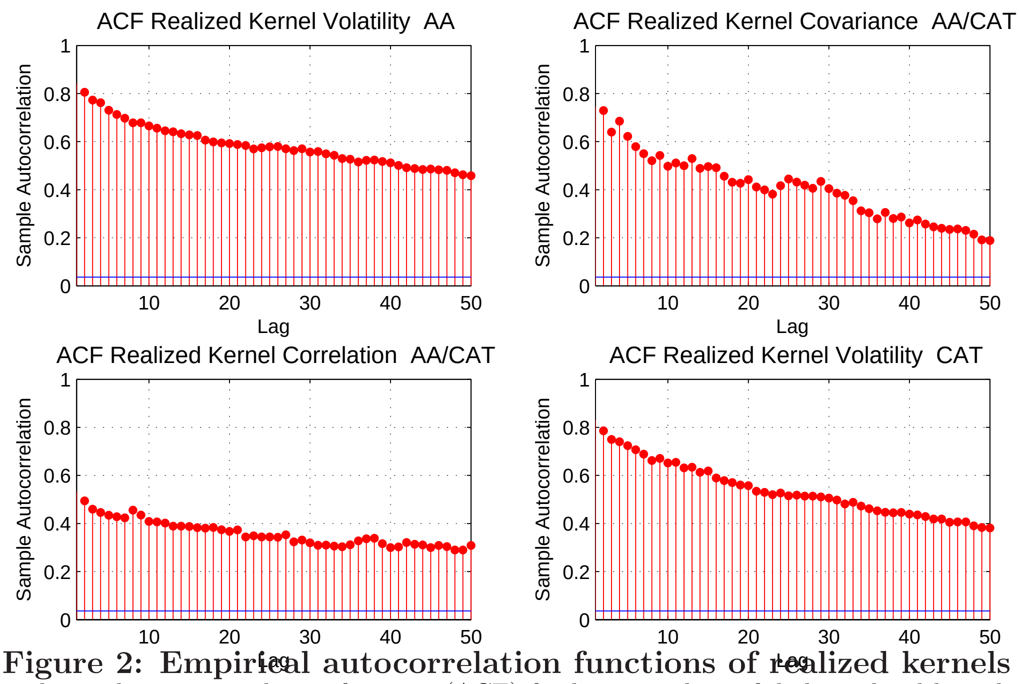

Figure 2: Empirical autocorrelation functions of realized kernels

This figure shows the autocorrelation function (ACF) for lag 1 until 50 of daily realized kernel volatilities (square root of variance) of Alcoa Inc. (AA) and Caterpillar Inc. (CAT) in the diagonal panels. The off-diagonal panels contain the ACF of the realized kernel covariance (upper-right) and correlation (lower-left) between the two asset returns. The sample is January 2, 2001 until December 31, 2012 ($T = 3017$ days).

the GAS tF model of Janus *et al.* (2014b) and a multivariate analogue of the Riskmetrics 2006 approach of Zumbach (2006). The multivariate HEAVY model incorporates realized measures into the volatility specification, by proposing a system of two multivariate GARCH equations for the quantities $V_t = \mathbb{E}_t[y_t y_t' | \mathcal{F}_{t-1}]$ and $M_t = \mathbb{E}_t[RK_t | \mathcal{F}_{t-1}]$. The innovations in both of these equations are the realized (co)variance measures as gathered in the matrix RK_t. The dynamics are given by

$$V_{t+1} = C_V C_V' + A_V RK_t + B_V V_t, \quad (13)$$

$$M_{t+1} = C_M C_M' + A_M RK_t + B_M M_t, \quad (14)$$

where A_V, A_M, B_V, and B_M are scalar parameters, and C_V and C_M are lower triangular matrices. The scalar parameters of both equations are estimated separately by Maximum Likelihood, assuming a Singular Wishart distribution for $y_t y_t'$ and a standardized Wishart distribution with k degrees of freedom for RK_t. The matrices C_V and C_M are typically estimated by covariance targeting, as discussed by Noureldin *et al.* (2012). We follow this approach when implementing the model in the remaining analysis.

The GAS tF model is our second benchmark. This model follows directly from the FIGAS tF model by setting d to zero in (3). This benchmark enables us to differentiate between short- and long-memory GAS models.

Our final benchmark is the RiskMetrics 2006 approach. This is a less well-known extension of the classical Exponentially Weighted Moving Average (EWMA) model also introduced by RiskMetrics. The 2006 model is a sum of EWMA models over increasing time horizons and in this sense is close in spirit to the HAR models of Corsi (2009). The goal of the RiskMetrics 2006 model is to capture the multi-scale trading structure of markets by averaging over intra-day, daily, weekly and monthly time horizons. In the univariate approach introduced by Zumbach (2006), a set of n_{max} historical conditional volatilities $\sigma_{n,t}$ are modeled by an exponentially weighted moving average of the squared returns y_t^2. This is done using the recursive equations

$$\sigma^2_{n,t+1} = \alpha_n \sigma^2_{n,t} + (1-\alpha_n) y_t^2, \tag{15}$$

$$\alpha_n = \exp(-1/\tau_n), \tag{16}$$

$$\tau_n = \tau_1 \, \rho^{n-1}, \tag{17}$$

for $n = 1, \ldots, n_{max}$ and the two tuning constants $\bar{\tau} > 0$ and $0 < \rho < 1$. The 'effective' volatility $\sigma_{eff,t}$ is then defined as the sum over the historical volatilities, with logarithmically decaying weights, i.e.,

$$\sigma^2_{eff,t} = \sum_{n=1}^{n_{max}} w_n \sigma^2_{n,t}, \tag{18}$$

$$w_n = \frac{1}{Q}\left(1 - \frac{\log(\tau_n)}{\log(\tau_0)}\right), \tag{19}$$

where Q is a normalizing constant such that the weights w_n sum to one. As in Zumbach (2006), we set the tuning parameters to $\tau_1 = 4 \Leftrightarrow \alpha_1 \approx 0.819$; $\tau_{n_{max}} = 512 \Leftrightarrow \alpha_{n_{max}} \approx 0.998$; $n_{max} = 15$; $\tau_0 = 1560$; and $\rho = \sqrt{2}$. To accommodate our multivariate analysis, we extend the original univariate approach of Zumbach (2006) to the multivariate setting by replacing the squared return y_t^2 by the outer product $y_t y_t'$ and and the effective variance $\sigma^2_{eff,t}$ by the covariance matrix V_t.

All benchmark models allow for easy h-step ahead prediction of V_t. In case of the HEAVY model, the second recursively equation delivers forecasts of RK_{t+h} for $h = 1, 2, \ldots$, which can subsequently be inserted into the first equation to obtain V_{t+h}. In the univariate RM 2006 approach, we use the recursive equations

$$\mathbb{E}_t[\sigma^2_{n,t+h}|\mathcal{F}_t] = \alpha_n \mathbb{E}_t[\sigma^2_{n,t+h-1}|\mathcal{F}_t] + (1-\alpha_n)\mathbb{E}_t[\sigma^2_{eff,t+h-1}|\mathcal{F}_t], \tag{20}$$

$$\mathbb{E}_t[\sigma^2_{eff,t+h}|\mathcal{F}_t] = \sum_{n=1}^{n_{max}} w_n \mathbb{E}_t[\sigma^2_{n,t+h}|\mathcal{F}_t]. \tag{21}$$

We can easily generalize these equations to the multivariate setting by replacing $\sigma^2_{n,t}$ by matrices, and $\sigma^2_{eff,t}$ by V_t. To deduct the h-step ahead forecast of V_t of the FIGAS model, we first rewrite (3) as follows:

$$V_{t+1} = \Omega + B\,V_t + A\,(1-L)^{-d} s_t. \tag{22}$$

Given the property that $\mathbb{E}_t[s_{t+h}|\mathcal{F}_t] = 0_k$ for any value of $h > 1$, V_{t+h} is obtained recursively while setting the future values of the score to the zero matrix. Similar results hold for the GAS tF model.

3.3 Model Evaluation Procedure

We follow Noureldin *et al.* (2012) and compare the in- and out-of-sample statistical fit of the models by computing the quasi-likelihood loss function:

$$QLIK_{t,h}(RK_{t+h}, V^a_{t+h}) = \log|V^a_{t+h}| + \mathrm{tr}((V^a_{t+h})^{-1} RK_{t+h}), \tag{23}$$

with V^a_{t+h} the covariance matrix estimate/forecast based on model a. Note that we use RK_{t+h} as proxy of the true covariance matrix. In-sample, h is set to zero and since V_t is known at time $t-1$, the criteria can also be interpreted as one-step ahead forecasting criteria. As indicated by Patton (2011), the QLIK loss-function implies a consistent ranking of volatility models since it is robust to noise in the proxy RK_t. To assess the in-sample performance, we set $h = 0$ and note that V_t only depends on the information in $\mathcal{F}_{t-1}$ for al three models considered. For the out-of-sample performance, we set $h > 0$.

We additionally test the predictive performance of the models using the framework of Giacomini and White (2006). We start by computing the difference in loss functions between two competing models a and b,

$$d_{t,h}(a,b) = QLIK_{t,h}(RK_{t+h}, V^a_{t+h}) - QLIK_{t,h}(RK_{t+h}, V^b_{t+h}), \tag{24}$$

for $t = R+1, \ldots T-h$, where the parameters are estimated based on a rolling window of $T_w = 1500$ observations. The difference d_t can be interpreted as a difference between two Kullback-Leibner (KL) divergences. Even if the underlying two models are both misspecified, the difference in their KL divergences still provides a valid assessment criterion. The corresponding null-hypothesis of equal predictive ability is given by $H_0 : \mathbb{E}[d_{t,h}(a,b)] = 0$ for all $T - h - R$ out-of-sample forecasts, which can be tested using the Diebold and Mariano (1995) (DM) test-statistic given by

$$DM_h(a,b) = \frac{\bar{d}_h}{\sqrt{\hat{s}_h^2/(T-h-R)}}, \tag{25}$$

with $\bar{d}_h$ the out-of-sample average of the loss differences, and $\hat{s}_h^2$ a HAC-consistent variance estimator $d_{t,h}(a,b)$. A significantly negative value of $DM_h(a,b)$ means that model a has a superior forecast performance over model b. The QLIK test can be used in-sample (interpreted as a 'one-step-ahead prediction') and out-of-sample. In the out-of-sample test, we choose $h = 1, 5, 10$ and 22. In addition, we consider the cumulative forecasts $V_{t:t+N} = \sum_{i=1}^{N} V_{t+i}$, where N equals 5 and 10 respectively.

A second statistical test we perform relates to the density forecasts of the realized covariance matrices. We use the log-score (see Mitchell and Hall, 2005; Amisano and Giacomini, 2007) as a scoring rule to differentiate between (1) the matrix-F (FIGAS tF model) and the Wishart (HEAVY model) distribution and (2) between the FIGAS tF and GAS tF model. We define the difference in log-scores between the density forecasts of models a and b as by replacing $QLIK_{t,h}$ by the log-score

$$S^{ls}(RK_t, a) = \log p(RK_t | V_t, \mathcal{F}_{t-1}, a), \tag{26}$$

both for models a and b, where $p(\cdot)$ denotes the probability distribution function of the matrix-F or Wishart distribution respectively.

Both evaluation criteria are statistical in nature. Motivated by the mean-variance optimization setting of Markowitz (1952), we therefore finally also assess the forecasting performance from a economic point of view. We do so by considering global minimum variance portfolios (GMVP); see for example Chiriac and Voev (2011); Engle and Kelly (2012), among others, who perform a similar analysis. The best forecasting model should provide portfolios with the lowest ex-post variance. Assuming that the investor's aim is to minimize the h-step portfolio volatility at time t subject to a fully invested portfolio, the resulting

GMVP weights $w_{t+h|t}$ are obtained by the solution of the quadratic programming problem

$$\min w'_{t+h|t} V_{t+s|t} w_{t+h|t}, \qquad \text{s.t. } w'_{t+h|t}\iota = 1. \tag{27}$$

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 $\sigma_{p,t} = \sqrt{w'_{t+h|t} RK_{t+h} w_{t+h|t}}$. We can again test the differences in portfolio standard deviations by means of the DM test statistic.

3.4 In-sample results

Table 2 shows parameter estimates and standard errors based on the inverse hessian of the likelihood evaluated at the optimum. We show the results for two selections of $k = 5$ stocks, as well as for the complete set of all 15 equities. In addition, we presents the total log-likelihood values corresponding with the (FI)GAS tF model and the HEAVY model as well as the average loss function for all competing models. Note that we present a different specifications of the HEAVY model: we found that that standard covariance targeting does not work appropriately for the HEAVY model. Hence HEAVY denotes here the results for a HEAVY model with

$$V_{t+1} = c\,\Omega_V + A_P RK_t + B_P V_t, \tag{28}$$

where c is a scalar static parameter that is estimated together with the other static parameters, and Ω_V is the unconditional covariance matrix, estimated by its sample analogue $\Omega_V = \sum_{t=1}^{T} y_t y_t'$. This slightly increased flexibility in the specification of the HEAVY model substantially increases the performance of the HEAVY model and makes it an even stricter benchmark for our FIGAS tF model.

The results in Table 2 show that the FIGAS tF model has the best fit to the data compared to the other models. Based on the QLIK loss function, the FIGAS tF has the lowest value, followed by the GAS tF, HEAVY and the RM 2006 model respectively. Comparing the (FI)GAS and HEAVY models, the values suggest that the largest gain is obtained by introducing the GAS framework, as the average QLIK drops by 0.10 (k=5) and 0.2 (all equities).

Table 2: Parameter estimates, likelihoods and loss function

This table reports maximum likelihood parameter estimates of the FIGAS tF, HEAVY and the GAS tF model, applied to daily equity returns and daily realized kernels of 5 and 15 assets. Asset identifiers are explained in Table 1. Standard errors are provided in parenthesis. The first two rows (A, B) correspond with the parameters of the equation for V_t of the (FI)GAS tF and the HEAVY model, while the third, fourth and fifth row (A_M, B_M, c) are related to the HEAVY equation (14) of RK_t with covariance targeting (CT). HEAVY represents the model defined in (28) with the additional scaling parameter c_P. For the RM 2006 methodology, we use the parameters $\tau_0 = 1500$,$\tau_1 = 4$, $\tau_{n_{max}} = 512$, $\rho = \sqrt{2}$, and $n_{max} = 15$. The table reports the log-likelihood as well as the the QLIK loss function, which is defined in (23). The likelihoods of the models are decomposed into their constituents, i.e., the part for the vector-valued y_t ($\mathcal{L}_\mathrm{t}$) and the matrix-valued RK_t ($\mathcal{L}_\mathrm{F}$) for the (FI)GAS tF model, and the part for the singular $y_t y_t'$ ($\mathcal{L}_\mathrm{SW}$) and the matrix-valued RK_t ($\mathcal{L}_\mathrm{W}$) for the HEAVY model. The sample is January 2, 2001, until December 31, 2012 (3017 observations).

	AA/BA/CAT/GE/KO				CAT/HON/IBM/MCD/WMT				All equities			
Coef.	FIGAS	HEAVY	GAS	RM	FIGAS	HEAVY	GAS	RM	FIGAS	HEAVY	GAS	RM
A	0.735	0.419	0.619		0.747	0.462	0.635		0.498	0.265	0.388	
	(0.014)	(0.035)	(0.012)		(0.014)	(0.032)	(0.012)		(0.004)	(0.011)	(0.004)	
B	0.999	0.597	0.986		0.998	0.554	0.985		0.999	0.743	0.991	
	(0.001)	(0.033)	(0.001)		(0.000)	(0.029)	(0.001)		(0.000)	(0.010)	(0.000)	
c		0.046				0.060				0.026		
		(0.006)				(0.007)				(0.002)		
A_M		0.286				0.286				0.196		
		(0.009)				(0.008)				(0.003)		
B_M		0.698				0.696				0.792		
		(0.010)				(0.009)				(0.003)		
ν_0	10.37		10.01		9.134		8.973		12.077		11.61	
	(0.504)		(0.469)		(0.395)		(0.497)		(0.403)		(0.362)	
ν_1	46.27		46.61		48.42		49.10		66.33		66.66	
	(0.925)		(0.911)		(0.982)		(0.896)		(0.373)		(0.143)	
ν_2	36.22		34.97		34.73		33.65		62.60		61.16	
	(0.577)		(0.521)		(0.519)		(0.518)		(0.332)		(0.088)	
d	-0.241				-0.222				-0.226			
	(0.006)				(0.007)				(0.003)			
$\mathcal{L}_t$	-26,436		-26,474		-25,063		-25,085		-72,211		-72,343	
$\mathcal{L}_{SW}$	-43,838				-40,695				-150,072			
$\mathcal{L}_F/\mathcal{L}_W$	-20,788	-45,750	-21,243		-12067	-37,114	-12,420		67,221	-42,958	64,774	
QLIK	7.694	7.806	7.712	51.43	6.761	6.873	6.774	93.15	19.06	19.25	19.13	602.9

Hence allowing for fat-tailedness in both the return observations and realized covariance kernels improves the fit substantially. In addition, allowing for long-memory does play a role, albeit less strong. This is not surprising, as the QLIK is equivalent with a one-step ahead forecast. As we will show in the out-of-sample analysis, the fractionally-integrated part of the model becomes more important for larger horizons.

The likelihoods of the models are decomposed into their constituents. For the (FI)GAS tF model, we distinguish the part of the likelihood attributable to the Student's t vector-valued observations y_t ($\mathcal{L}_{\text{t}}$) and to the matrix-F distributed observations RK_t ($\mathcal{L}_{\text{F}}$). Similarly, for the HEAVY model we distinguish the likelihood part $\mathcal{L}_{\text{SW}}$ attributable to the singular Wishart observations $y_t y_t'$ and the part $\mathcal{L}_{\text{W}}$ attributable to RK_t. It is clear that the (FI)GAS tF model performs much better for both parts of the model than the HEAVY model, i.e., for both y_t and RK_t. For the return observations y_t this is well-known. The results underline again, however, that it is also important to account for the fat-tailedness of the realized realized covariance kernels. The (FI)GAS tF model deals with this by adopting the matrix-F distribution for RK_t. Finally, the log-likelihood of the matrix-F distribution increases by 400 or even 2,500 points (all equities) when allowing for long-memory effects in the GAS framework. This illustrates that including one extra d parameter has a considerable effect on the statistical fit of the model.

Looking at the individual parameter estimates, we first note that the estimates of A cannot be compared between the FIGAS tF and the HEAVY model. This is due to the fact that the former shows the effect on the scaled score, that exists of both a matrix-weighted RK_t and a scalar-weighted $y_t y_t'$, whereas the latter focuses solely on the realized covariance matrix. Also the distributional assumptions differs between the two models. The FIGAS tF model shows a high degree of persistence in V_t via the estimate of B, which is close to 1. Note that the sum of A_M and B_M (the parameters corresponding with the second HEAVY equation) is also very close to 1, confirming the high persistence of RK_t. The degrees of freedom parameter ν_2 is estimated at around 35 and 65 for 5 and 15 dimensions, respectively. Despite that the value of $\hat{\nu_2}$ may appear high, we find that these values of $\hat{\nu_2}$ already result in a substantial moderation of the effect of incidentally large observations RK_t in (12) through the matrix weighting scheme.

The negative estimated long-memory coefficient d is highly significant, indicating the presence of long-memory effects in the volatility. As is more often the case when estimating long-memory models, the likelihood has two local optima: one with a value of B close to 1 and $d < 0$, and a second optimum with $0 < d < 0.5$ and B much smaller than 1. When looking at the implied correlation functions corresponding to these two local optima we find that the implied correlation functions are quite close. We therefore choose the optimum with the highest likelihood value.

Figure 3 plots some of the fitted volatilities and correlations. We show the results for AA and BA for the FIGAS tF model (blue line) and the HEAVY model of equation (28) (red line). The figure shows remarkable differences between the two models for both the volatility and the covariances/correlations. Focusing first on the volatilities and covariances, the robust transition scheme based on the Student's t and matrix-F GAS dynamics produces considerably less spikes in the paths. Put differently, the GAS framework is able to mitigate the impact of temporary RK_t and $y_t y_t'$ observations on the estimates of V_t. The HEAVY model, based on the thin-tailed (singular) Wishart distribution, produces much more spikes. Notable differences are apparent for both considered companies during the period 2001-2003, 2007-2008 and the period 2010–2011. The patterns for the correlations reveal similar remarkable differences. Again, the number of spikes in the correlation patterns for the HEAVY is much higher than for the FIGAS model.

3.5 Out-of-sample results

We also assess both the short-term and long-term forecasting performance of the FIGAS tF model by considering s-step ahead forecasts, with $s = 1, 5, 10$, and 22. In addition, we consider aggregated covariance forecasts for the next (two) trading weeks, i.e. $V_{t:t+s} = V_{t+1} + V_{t+2} + \ldots + V_{t+s}$ with $s = 5, 10$. Similar to the in-sample analysis of the previous subsection, we compare the FIGAS tF model with the HEAVY model, the GAS tF model and the RM2006 approach. We now test on predictive ability based on the loss-differences of the QLIK loss function (23) using the test-statistic defined in (25). We use a moving window of 1500 observations and re-estimate the parameters after 25 observations, which roughly equals one month. The first in-sample period corresponds to the period January, 2001 until December 2006. Hence the out-of-sample period starts before the heat of the

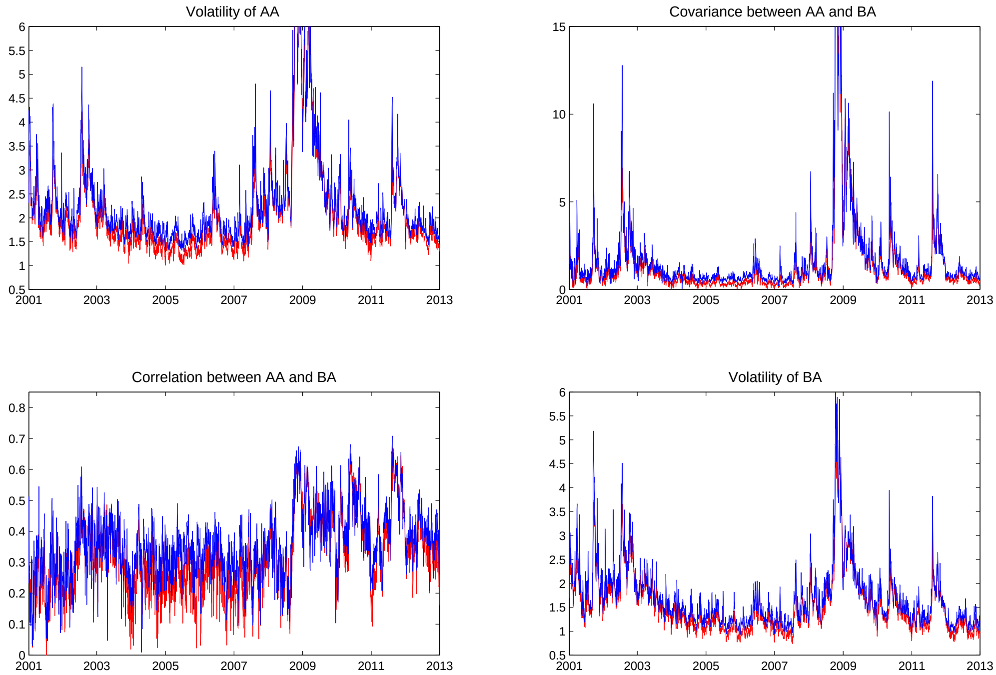

Figure 3: Estimated volatilities and correlations

This figure plots the estimated volatilities of AA and BA (see Table 1) in the upper-left and lower-right panels, and the pairwise covariances and correlations in the upper-right and lower-left panels respectively. Time varying parameter paths are estimated using the FIGAS tF model (red line) and HEAVY model (blue line). The estimates are based on the full sample, January 2, 2001 until December 31, 2012 (3017 observations).

financial crisis (October 2008). This current forecasting experiment therefore constitutes a major robustness test for all the models considered.

Table 3 shows the test-statistic corresponding with the QLIK loss-function over the out-of-sample period for our two combinations of 5 assets as well as for all 15 assets. We present the results for different forecast horizons. Negative t-test statistics (in parentheses) indicate that the FIGAS tF model performs better. The table indicates that the FIGAS model outperforms the competing models, in particular for long horizons. For example, the t-statistics -3.8, -5.2 and -3.6 associated with 22-step ahead predictions confirm that the HEAVY model is significantly outperformed. Similar results hold with respect to the GAS model. Hence, in contrast to the in-sample results discussed in the previous section, we clearly see the importance of long-memory in addition to the GAS framework for long-term predictions. Note that the FIGAS model also outperforms the GAS model in panel A for 1- and 10-step ahead forecasts (t-stats of -2.7 and -2.3) and in panel C for 1-step ahead forecasts. The same result holds with respect to the HEAVY model, as panel B indicates that this model is also significantly outperformed for short horizons. Finally, we clearly

Table 3: Test-statistics on predictive ability

This table shows test statistics on superior predictive ability between the FIGAS tF model and the HEAVY, GAS tF or RM 2006 model respectively, based on the QLIK loss function defined in (23). The test is based based on 1, 5, 10 and 22-step ahead predictions of the covariance matrix, applied to 5 and 15 (all) equities. In addition, we show results of the aggregated forecast of 5 and 10 consecutive days (1:5 and 1:10). We report the average QLIK loss for each model with the associated DM-type of test statistic in parentheses. A negative test statistic indicates superior predictive ability of the FIGAS tF model. We use a moving window of 1500 observations. The prediction period runs from December, 2006 until December, 2012 and contains 1495 observations. The number of observations corresponding with the aggregated forecasts are equal to 300 (1:5) and 150 (1:10) respectively.

	1	5	10	22	1:5	1:10
Panel A: AA/BA/CAT/GE/KO						
FIGAS tF	8.04	8.41	8.67	9.11	16.30	19.94
HEAVY	8.07	8.44	8.73	9.31	16.58	20.21
	(-1.4)	(-0.9)	(-1.5)	(-3.8)	(-5.9)	(-3.9)
GAS tF	8.05	8.43	8.75	9.32	17.06	20.86
	(-2.7)	(-1.1)	(-2.3)	(-3.6)	(-6.5)	(-4.5)
RM 2006	8.84	9.30	10.14	12.67	16.40	20.04
	(-13.7)	(-8.0)	(-7.1)	(-7.7)	(-2.5)	(-1.9)
Panel B: CAT/HON/IBM/MCD/WMT						
FIGAS tF	6.12	6.44	6.67	7.08	14.36	17.97
HEAVY	6.20	6.57	6.84	7.37	14.49	18.12
	(-4.0)	(-3.8)	(-3.6)	(-5.2)	(-4.4)	(-3.5)
GAS tF	6.12	6.44	6.70	7.23	14.36	17.97
	(-0.1)	(0.6)	(-1.4)	(-4.1)	(0.6)	(-0.2)
RM 2006	7.01	7.44	8.28	10.66	15.23	18.94
	(-11.7)	(-6.5)	(-5.6)	(-6.8)	(-5.3)	(-3.8)
Panel C: all equities						
FIGAS tF	19.07	20.04	20.75	21.84	43.78	54.65
HEAVY	19.12	20.11	20.93	22.33	43.87	54.84
	(-0.8)	(-0.7)	(-1.3)	(-3.6)	(-0.9)	(-1.5)
GAS tF	19.12	20.06	20.89	22.23	43.79	54.72
	(-2.3)	(-0.3)	(-1.5)	(-3.2)	(-0.2)	(-0.9)
RM 2006	24.58	26.62	29.74	38.20	49.61	61.27
	(-10.1)	(-8.6)	(-8.0)	(-8.4)	(-5.8)	(-4.5)

Table 4: Density forecasts of the realized covariance

This table shows results on predicted density forecasts of the realized covariance matrix of 5 or 15 assets, using the log-score. The test is based on 1, 5, 10 and 22-step ahead predictions of the covariance matrix, according to the FIGAS tF, the GAS tF and the HEAVY model, applied to five and 15 (all) equities. The associated distributions are the matrix F and Wishart distribution respectively. We show the mean difference of the log-score, as well as the test-statistic in parentheses, based on HAC-standard errors. A positive difference in the log-score indicates the density forecast of the FIGAS model is better than the density forecast of the HEAVY or GAS model. We use a moving window of 1500 observations. The prediction period runs from December, 2006 until December, 2012 (1495 observations).

	1	5	10	22
Panel A: AA/BA/CAT/GE/KO				
FIGAS vs HEAVY	8.77	7.31	6.67	5.90
	(50.1)	(26.9)	(21.4)	(15.9)
FIGAS vs GAS	0.16	0.24	0.42	0.98
	(4.0)	(2.4)	(2.9)	(5.0)
Panel B: CAT/HON/IBM/MCD/WMT				
FIGAS vs HEAVY	9.39	8.08	7.23	6.16
	(58.6)	(31.2)	(22.8)	(15.2)
FIGAS vs GAS	0.09	0.17	0.53	1.53
	(2.8)	(1.9)	(4.0)	(7.9)
Panel C: all equities				
FIGAS vs HEAVY	43.55	40.93	40.92	42.08
	(49.1)	(36.3)	(24.3)	(13.5)
FIGAS vs GAS	0.78	0.80	1.52	3.00
	(4.1)	(2.0)	(2.7)	(4.3)

outperform the RM 2006 model for all considered horizons. To summarize both results, taking into account fat-tailedness of returns and realized covariance kernels at the one end, plus allowing for long-memory effects at the other hand provides superior performance of the FIGAS tF model.

Table 4 provides further insight in the quality of the predictive densities of the realized covariance matrices using the (FI)GAS (matrix-F distribution) models and the HEAVY (Wishart distribution) model, respectively. The table shows the average difference of the log-score of the HEAVY or GAS model relative to the FIGAS model, with HAC-based test-statistics in parentheses. Positive numbers indicate superior density forecasts of the FIGAS model. Without any doubt, the density forecasts corresponding to the FIGAS tF model are superior compared to Wishart densities of the HEAVY model. Hence accounting for fat-tails in the realized covariance kernels also delivers better out-of-sample density forecasts.

Even more interesting is the result that the density forecasts of the FIGAS model also

Table 5: Ex-post portfolio standard-deviations

This table shows results on a global minimum variance portfolio, based on 1, 5, 10 and 22-step ahead predictions of the covariance matrix, according to the FIGAS tF, HEAVY, GAS and the RiskMetrics 2006 approach, applied to 5 and 15 equities. For each model, the table shows the ex-post mean of the daily portfolio volatility, whereas below the HEAVY, GAS and RM 2006 model. The number between parentheses shows the test-statistic on equal portfolio volatility between the FIGAS tF model and the HEAVY, GAS or RM 2006 model. We use a moving window of 1500 observations. The prediction period runs from December, 2006 until December, 2012 (1495 observations).

	1	5	10	22	1:5	1:10
Panel A: AA/BA/CAT/GE/KO						
FIGAS tF	0.925	0.933	0.938	0.946	2.118	3.029
HEAVY	0.927	0.936	0.942	0.951	2.122	3.040
	(-4.6)	(-3.9)	(-4.1)	(-4.7)	(-2.9)	(-3.6)
GAS tF	0.926	0.936	0.942	0.951	2.122	3.042
	(-4.9)	(-6.6)	(-5.3)	(-5.9)	(-4.4)	(-4.3)
RM 2006	1.013	1.010	1.008	1.005	2.303	3.280
	(-14.2)	(-13.2)	(-12.2)	(-12.0)	(-7.9)	(-5.8)
Panel B: CAT/HON/IBM/MCD/WMT						
FIGAS tF	0.848	0.856	0.859	0.866	1.941	2.771
HEAVY	0.850	0.858	0.862	0.868	1.949	2.781
	(-2.5)	(-3.6)	(-2.7)	(-2.1)	(-2.0)	(-2.9)
GAS tF	0.849	0.858	0.861	0.866	1.944	2.775
	(-3.0)	(-4.5)	(-3.4)	(-1.1)	(-4.8)	(-2.6)
RM 2006	0.907	0.900	0.895	0.891	2.055	2.921
	(-13.0)	(-12.8)	(-12.9)	(-9.7)	(-7.8)	(-6.0)
Panel C: all equities						
FIGAS tF	0.688	0.700	0.707	0.718	1.586	2.278
HEAVY	0.690	0.703	0.711	0.723	1.592	2.289
	(-2.6)	(-4.1)	(-4.6)	(-4.5)	(-3.3)	(-4.6)
GAS tF	0.689	0.703	0.711	0.723	1.590	2.286
	(-4.6)	(-5.4)	(-5.5)	(-6.4)	(-3.9)	(-4.0)
RM 2006	0.830	0.817	0.805	0.796	1.872	2.646
	(-15.4)	(-14.5)	(-13.6)	(-12.6)	(-8.2)	(-6.2)

outperform the density forecasts of the GAS model. Recall that the difference between both density forecasts only comes from the model specification of the covariance matrix, as both models assume a matrix-F distribution with roughly the same parameter estimates for ν_1 and ν_2 (see Table 2). Hence incorporating long-memory into the model specification delivers better density forecasts.

Finally, we turn to the economic significance of the covariance matrix forecasts. Table 5 shows the mean of the ex-post conditional portfolio standard deviation, computed by implementing the ex-ante minimum variance portfolio weights as obtained by optimizing

(27) each period. Panels A, B and C display the average out-of-sample portfolio standard deviation and the associated DM test statistics vis-à-vis the FIGAS tF model (in parentheses). For all pairs of assets and all forecasting horizons considered, the FIGAS tF model produces the lowest ex-post portfolio standard deviation. The reductions in standard deviations are statistically significant, compared to the HEAVY model, the GAS tF model and the RM 2006 approach. Panel B shows one exception: the ex-post portfolio standard deviation corresponding to the 22-step ahead forecasts of the FIGAS tF and GAS tF model are not statistically different. We conclude that the forecasting performance of the FIGAS tF model is superior at both short and long horizons when compared to the HEAVY model, the GAS tF model and the RiskMetrics 2006 approach.

4 Conclusions

We introduced a new multivariate fractionally integrated model with score-driven volatility dynamics (FIGAS tF) that combines observed realized covariance matrices and vector-valued return observations to estimate the dynamics of unobserved common covariance matrices. The proposed model explicitly acknowledges that (co)variances display long-memory behavior. In addition, the model takes into account that both realized covariance matrices and financial return data are typically fat-tailed. The score-driven matrix-valued dynamics automatically correct for influential observations in either type of data by automatically assigning weights to the new observations in line with the fatness of the tails of the estimated conditional distributions. The proposed setup is particularly suitable for the cases where no explicit robustification methods are applied while estimating realized (co)variance measures. An important feature of our model is that it retains the matrix format for the transition dynamics of the covariance matrices, which makes the model computationally efficient and parsimonious. We showed that the model improves both the in-sample and out-of-sample fit of the covariance matrices for equity returns from the S&P500 during 2001-2012 when compared to the recently proposed HEAVY model of Noureldin *et al.* (2012), the GAS tF model of Janus *et al.* (2014b) and the RM 2006 model of Zumbach (2006).

Acknowledgements

We appreciate the comments of participants at the 2015 NBER-NSF Time Series Conference (Vienna, September 2015), the 2nd International Workshop on Financial Markets and Non-linear Dynamics (Paris, June 2015) and seminar participants at the Econometrics Brown Bag Seminar Series at Vrije Universiteit Amsterdam. Lucas and Opschoor thank the Dutch National Science Foundation (NWO, grant VICI453-09-005) for financial support.

References

Amisano, Gianni and Raffaella Giacomini (2007), Comparing density forecasts via weighted likelihood ratio tests, *Journal of Business & Economic Statistics* **25**(2), 177–190.

Andersen, T., T. Bollerslev, F.X. Diebold and P. Labys (2001), The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* **96**, 42–55.

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

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996), Forecasting multivariate realized stock market volatility, *Journal of Econometrics* **74**, 3–50.

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

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

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.

Blasques, C., S.J. Koopman and A. Lucas (2014), Maximum Likelihood Estimation for Generalized Autoregressive Score Models, Tinbergen Institute Discussion Paper, TI 14-029/III.

Blasques, C., S.J. Koopman and A. Lucas (2015), Information Theoretic Optimality of Observation Driven Time Series Models for Continuous Responses, **102**(2), 325–343.

Bollerslev, T. and H.O. Mikkelsen (1996), Modeling and pricing long memory in stock market volatility, *Journal of Econometrics* **73**, 151–184.

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.

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

Conrad, C. and B.R. Haag (2006), Inequality constraints in the fractionally integrated GARCH model, *Journal of Financial Econometrics* **4**, 413–449.

Corsi, F. (2009), A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* **7**, 174–196.

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

Creal, D., S.J. Koopman and A. Lucas (2011), A dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations, *Journal of Business 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.

Creal, D., S.J. Koopman, A. Lucas and B. Schwaab (2014), Observation Driven Mixed-Measurement Dynamic Factor Models with an Application to Credit Risk, *Review of Economics and Statistics (forthcoming)* .

Diebold, Francis X and Robert S Mariano (1995), Comparing predictive accuracy, *Journal of Business & economic statistics* **13**, 253–263.

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

Giacomini, Raffaella and Halbert White (2006), Tests of conditional predictive ability, *Econometrica* **74**(6), 1545–1578.

Granger, C.W.J. and R. Joyeaux (1980), An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis* **1**, 15–29.

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.

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 (forthcoming)* .

Hosking, J.R. (1981), Fractional differencing, *Biometrika* **68**, 165–176.

Janus, P., A. Lucas, A. Opschoor and D. van Dijk (2014b), New HEAVY models fat-tailed realized covariances and returns, *Tinbergen Institute Discussion Paper* **14-073/IV**.

Janus, P., S.J. Koopman and A. Lucas (2014a), 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.

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

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, Harry (1952), Portfolio selection, *Journal of Finance* **7**(1), 77–91.

Mitchell, James and Stephen G Hall (2005), Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR Fan Charts 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.

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

Robinson, P.M. (1991), Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression, *Journal of Econometrics* **47**(1), 67–84.

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

Silvennoien, A. and T. Teräsvirta (2009), Multivariate GARCH models, in T.G. Andersen, A. Davis, J.P. Kreib and T. Mikosch (eds.), *Handbook of Financial Time Series*, Springer-Verlag, pp. 201–229.

Sowell, F. (1992), Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics* **53**, 165–188.

Zumbach, G. (2006), The RiskMetrics 2006 methodology, Working Paper.