TI 2016-069/IV
Tinbergen Institute Discussion Paper

Fractional Integration and Fat Tails for Realized Covariance Kernels and Returns

André Lucas
Anne Opschoor

Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, the Netherlands.

Fractional Integration and Fat Tails for Realized Covariance Kernels and Returns

André Lucas[a] and Anne Opschoor[a*]

[a] *Vrije Universiteit Amsterdam and Tinbergen Institute*

This version: September 1, 2016

Abstract

We introduce a new fractionally integrated model for covariance matrix dynamics based on the long-memory behavior of daily realized covariance matrix kernels and daily return observations. We account for fat tails in both types of data by appropriate distributional assumptions. The covariance matrix dynamics are formulated as a numerically efficient matrix recursion that ensures positive definiteness under simple parameter constraints. Using intraday stock data over the period 2001-2012, we construct realized covariance kernels and show that the new fractionally integrated model statistically and economically outperforms recent alternatives such as the Multivariate HEAVY model and the 2006 "long-memory" version of the Riskmetrics model.

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 developed 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 help 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.

Volatilities are often found to be 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. Conrad and Haag (2006) ensure positivity of variances of the FI-GARCH class of models, which is particularly relevant for our multivariate setting to ensure positive definiteness of covariance matrices. 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.* (2014). 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 volatilities 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 either the realized measures, the returns, or both. Though fat-tailed distributions are often used to describe 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 modeling 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.* (2014), 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), and systemic risk management (Oh and Patton, 2016; 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 Opschoor *et al.* (2014) 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. In addition, we can directly apply the theoretical results of Conrad and Haag (2006) and obtain simple parameter restrictions to establish positive definiteness of the estimated covariance matrices over the entire sample period.

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 for 15 equities from the S&P 500 index. Our sample spans the period January 2001 to December 2012. Using a forecasting horizon of 1, 5, 10, and 22 days ahead, we compare both statistically and economically the performance of our new dynamic covariance matrix model to several strong benchmarks, such as the HEAVY model (Noureldin *et al.*, 2012), the GAS tF model (Opschoor *et al.*, 2014) and the long-memory version of the RiskMetrics (RM) model introduced by (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 statistically significantly lower ex-post conditional portfolio standard deviations, particularly at longer horizons.

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

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

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

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.

To drive the dynamics of V_t, we follow the score-driven approach of Creal *et al.* (2011, 2013). This approach adjust the time varying parameter V_t in the direction of steepest ascent of the local log likelihood function. The approach is computationally straightforward because of its explicit form for the likelihood function, and is known to possess information theoretic optimality properties; see Blasques *et al.* (2015). Let $p(y_t, RK_t \mid V_t)$ denote the conditional predictive density for (y_t, RK_t). Then the (short-memory) score dynamics for

V_t are given by

$$V_{t+1} = \Omega + B\, V_t + A\, s_t, \qquad s_t = S_t \cdot \Big(\partial \log p(y_t, RK_t \mid V_t)/\partial V_t\Big) \cdot S_t', \tag{3}$$

where Ω is a $k \times k$ matrix of parameters, and for the sake of simplicity A and B are scalar parameters. Further, s_t is the scaled score, and S_t is a scaling matrix to correct for the curvature of the log predictive density at time t. We come back to the precise form of the conditional observation density $p(y_t, RK_t \mid V_t)$ and the scaled score s_t in detail further below.

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. To capture this, we use fractionally integrated dynamics for V_t rather than the short-memory dynamics as in (3). To cast the score-driven model into its fractionally integrated form, we follow the approach of Baillie *et al.* (1996) and rewrite (3) as

$$\big(1 - B\, L\big)\, s^\star_{t+1} = \Omega + \big(1 - (B - A)\, L\big) s_{t+1}, \tag{4}$$

where L denotes the lag operator $L\, s_t = s_{t-1}$; $s^\star_t = s_t + V_t$; and where the scaled scores s_t follow a martingale difference sequence. We replace the left-hand side lag polynomial $(1 - B\, L)$ by the fractionally integrated lag polynomial $(1 - L)^d\, (1 - \phi\, L)$ for a scalar ϕ, and reparameterize the right-hand side polynomial $(1 - (B - A)\, L)$ as $(1 - \tilde{B}\, L)$, thus obtaining

$$\big(1 - L\big)^d\, \big(1 - \phi\, L\big)\, s^\star_{t+1} = \Omega + \big(1 - \tilde{B}\, L\big) s_{t+1}, \tag{5}$$

where $(1 - L)^d$ is the fractional difference operator defined by the binomial expansion

$$(1 - L)^d = 1 - dL + \frac{d(d-1)}{2!} L^2 - \frac{d(d-1)(d-2)}{3!} L^3 + \ldots, \tag{6}$$

for any real order of fractional integration parameter $|d| < 1$. As an example, consider the univariate case with the assumption of conditional Gaussian returns without realized measures. We can then show that $s_t = y_t^2 - V_t$ and $s^\star_t = y_t^2$ and we obtain the precise formulation of the FIGARCH model of Baillie *et al.* (1996). Using the definition $s^\star_t = s_t + V_t$, we can rewrite (5) as the FIGAS$(1, d, 1)$ model,

$$\begin{aligned} V_{t+1} &= \frac{\Omega}{1-\tilde{B}} + \left(1 - \frac{\left(1-L\right)^{d}\left(1-\phi\, L\right)}{1-\tilde{B}\, L}\right) s^{\star}_{t+1} \\ &= \tilde{\Omega} + \Psi(L)\, s^{\star}_{t+1}, \end{aligned} \tag{7}$$

with $\tilde{\Omega} = \Omega/(1-\tilde{B})$, and

$$\Psi(L) = 1 - \frac{(1-L)^d(1-\phi L)}{(1-\tilde{B}\, L)} = \sum_{i=1}^{\infty} \psi_i L^i. \tag{8}$$

Thus, the current conditional covariance matrix V_t is an infinite weighted sum of current and past $s^{\star}_t$s, where the weight assigned to each lag declines hyperbolically according to $\Psi(L)$. The FIGAS specification spans a variety of different autocovariance functions; see Janus *et al.* (2014) for some illustrative univariate examples. This flexibility to describe a range of autocovariance functions transfers *a fortiori* to the multivariate case.

An important feature of our current formulation of the FIGAS specification is that the sequence of covariance matrices V_t is automatically positive definite for all times t if (i) $\tilde{\Omega}$ is positive definite; (ii) the coefficients ψ_i are non-negative; and (iii) the terms $s^{\star}_t$ are non-negative definite. We show later that (iii) is automatically satisfied for the generalized autoregressive score specification of Creal *et al.* (2011, 2013) and the distributional choices made in this paper. Condition (i) is easily enforced through the model's parameterization. Condition (ii) is the most intricate. Given the analogy of our current model formulation to Baillie *et al.* (1996), however, we can directly draw upon for instance Corollary 1 of Conrad and Haag (2006, page 427) to check whether positivity of the ψ_is is ensured. This contrasts sharply with the FIGAS set-up of Janus *et al.* (2014), where positivity of the covariance matrix set-up cannot be easily ensured. In addition, Baillie *et al.* (1996) show that strict stationarity and ergodicity are obtained for $0 \leq d \leq 1$.

We now turn to the specific choice for the conditional observation densities $D_z(\cdot)$ and $D_Z(\cdot)$ in (1) and the completion of the FIGAS model specification under fat tails. To account for the possible fat-tailedness of the returns, we assume that y_t follows a (conditional)

Student's t distribution,

$$p_y(y_t|V_t,\mathcal{F}_{t-1};\nu_0)=\frac{\Gamma((\nu_0+k)/2)}{\Gamma(\nu_0/2)[(\nu_0-2)\pi]^{k/2}|V_t|^{1/2}}\times\left(1+\frac{y_tV_t^{-1}y_t}{\nu_0-2}\right)^{-(\nu_0+k)/2}, \tag{9}$$

with degrees of freedom parameter $\nu_0>2$ and V_t a positive definite covariance matrix at time t. Similarly, to account for possible fat tails of the realized kernel distribution, we assume that RK_t has a matrix-F distribution as given by

$$p_{RK}(RK_t|V_t,\mathcal{F}_{t-1};\nu_1,\nu_2)=K(\nu_1,\nu_2)\times\frac{\left|\frac{\nu_1}{\nu_2-k-1}V_t^{-1}\right|^{\frac{\nu_1}{2}}|RK_t|^{(\nu_1-k-1)/2}}{\left|\mathrm{I}_k+\frac{\nu_1}{\nu_2-k-1}V_t^{-1}RK_t\right|^{(\nu_1+\nu_2)/2}}, \tag{10}$$

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

$$K(\nu_1,\nu_2)=\frac{\Gamma_k((\nu_1+\nu_2)/2)}{\Gamma_k(\nu_1/2)\Gamma_k(\nu_2/2)}, \tag{11}$$

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

$$\Gamma_k(x)=\pi^{k(k-1)/4}\cdot\prod_{i-1}^{k}\Gamma(x+(1-i)/2); \tag{12}$$

see for example Konno (1991). 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 such that the joint conditional density $p(y_t,RK_t\mid V_t)$ is just the product of the conditional marginal densities (9) and (10). 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 Opschoor *et al.* (2014) for a short-memory context and short-term forecasting purposes. Here, we extend it to the long-memory context and benchmark its long-term forecasting performance to that of the HEAVY model of Noureldin *et al.* (2012) or the RiskMetrics 2006 methodology of Zumbach (2006). We do so in a way that ensures positivity of the covariance matrices despite the complications of the fractionally integrated lag polynomial.

Given the two observation densities (9) and (10) and the conditional independence assumption, the time t predictive log likelihood function and its derivatives become

$$
\begin{aligned}
\mathcal{L}_t &= \log p_y(y_t|V_t, \mathcal{F}_{t-1}; \nu_0) + \log p_{RK}(RK_t|V_t, \mathcal{F}_{t-1}; \nu_1, \nu_2), && (13)\\
s_t &= V_t\ (\nabla_{y,t} + \nabla_{RK,t})\ V_t/\,(\nu_1 + 1). && (14)\\
\nabla_{y,t} &= \partial \log p_y(y_t|V_t, \mathcal{F}_{t-1}; \nu_0)/\partial V_t,\\
\nabla_{RK,t} &= \partial \log p_{RK}(RK_t|V_t, \mathcal{F}_{t-1}; \nu_1, \nu_2)/\partial V_t,
\end{aligned}
$$

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

$$
\begin{aligned}
\nabla_{y,t} &= \tfrac{1}{2} V_t^{-1} \left[w_t \cdot y_t y_t' - V_t\right] V_t^{-1}, && (15)\\
w_t &= (\nu_0 + k) \cdot (\nu_0 - 2 + y_t' V_t^{-1} y_t)^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\nabla_{RK,t} &= \tfrac{1}{2} \nu_1 V_t^{-1} \left[W_t \cdot RK_t - V_t\right] V_t^{-1}, && (16)\\
W_t &= \frac{\nu_1 + \nu_2}{\nu_2 - k - 1} \cdot \left(\mathrm{I}_k + \frac{\nu_1}{\nu_2 - k - 1} RK_t V_t^{-1}\right)^{-1},
\end{aligned}
$$

such that after scaling the score by the matrix $2(V_t \otimes V_t)/(\nu_1 + 1)$ to account for the curvature of the log conditional density with respect to V_t (see Opschoor *et al.* (2014) for further details on this part of the model), the scaled score s_t reads

$$
\begin{aligned}
s_t &= \tfrac{1}{\nu_1+1} \left[w_t \cdot y_t y_t' - V_t\right] + \tfrac{\nu_1}{\nu_1+1} \left[W_t \cdot RK_t - V_t\right]\\
&= \tfrac{1}{\nu_1+1} \left[w_t \cdot y_t y_t'\right] + \tfrac{\nu_1}{\nu_1+1} \left[W_t \cdot RK_t\right] - V_t. && (17)\\
&= s_t^\star - V_t && (18)
\end{aligned}
$$

with w_t and W_t defined in (15) and (16) respectively. Due to the nature of V_t as a scale matrix, the shifted scaled score $s_t^\star = s_t + V_t$ is positive definite by construction given the definition of w_t and W_t and a positive definite matrix V_t. Recall that $s_t^\star$ is the key term in the infinite sum representation of the FIGAS model in (7). It can be interpreted as a multivariate and score-driven analogue of the univariate squared return ϵ_t^2 of the FIGARCH model. We refer to the complete model as the FIGAS tF model.

The specification of s_t in equations (14)–(18) 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-determined by the fractional difference polynomial. Second, both the score for the return equation and for the realized measure equation hold familiar 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 (14). 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

We estimate the parameters of the FIGAS tF model by maximum likelihood. In order to circumvent the number of estimated parameters corresponding to the $\tilde{\Omega}$ matrix, we propose two specifications: $\tilde{\Omega} = cI_k$ or $\tilde{\Omega} = c(1/T)\sum_{t=1}^{T} RK_t$ with c a scalar. The first specification restricts the matrix to a diagonal matrix with entries equal to c, while the second specification is related to the covariance targeting approach as Ω depends on the sample average of RK_t. We estimate the remaining static parameter vector $\theta = \{c, \phi, \tilde{B}, \nu_0, \nu_1, \nu_2, d\}$ of the FIGAS model by maximum likelihood. To do so, we maximize the log-likelihood $\mathcal{L}_{\text{tF}}(\theta) = \sum_{t=1}^{T} \mathcal{L}_t$, where $\mathcal{L}_t$ is defined in equation (13). This standard prediction error decomposition of the likelihood function is made possible due to the observation-driven nature of the FIGAS model in the classification of Cox (1981). 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 (6). 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 competing benchmark models. 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 an 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.

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

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 kernel approach) of Alcoa Inc. (AA) and Caterpillar Inc. (CAT) in 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

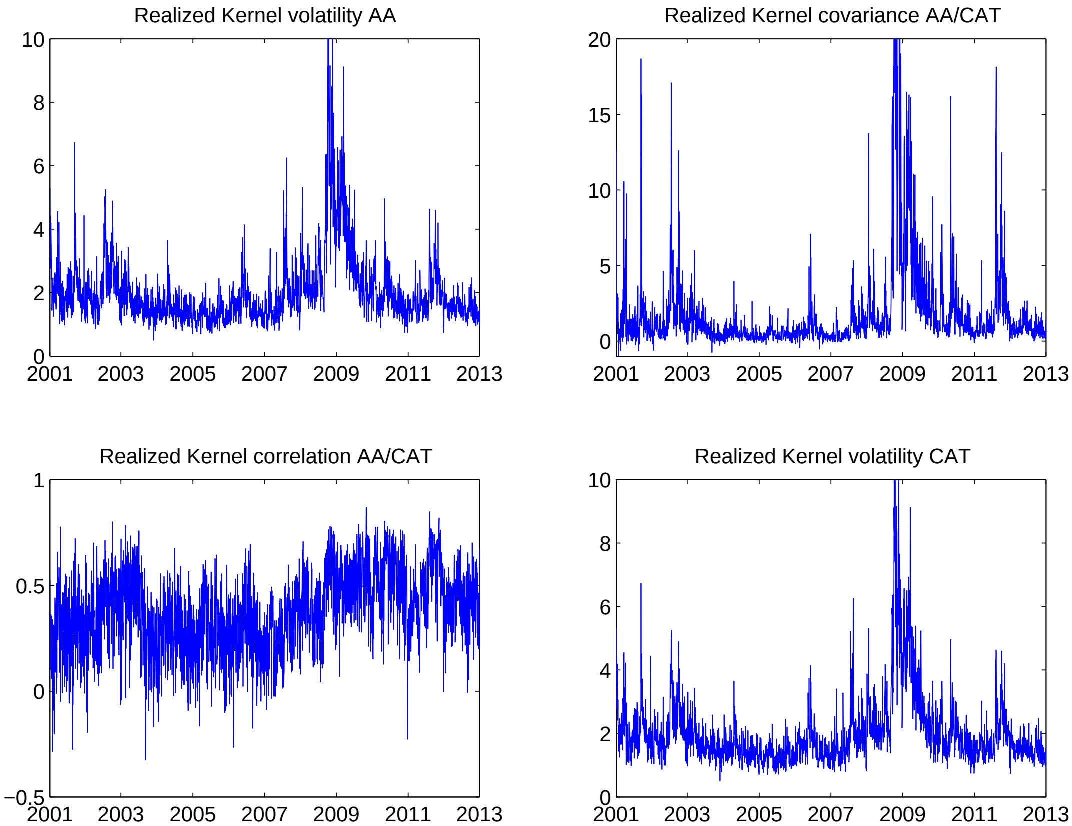

Figure 1: Realized Kernel estimates of AA/CAT
This figure shows daily realized kernel volatilities (square root of the variance) of Alcoa Inc. (AA) and Caterpillar Inc. (CAT) returns on the diagonal panels. The off-diagonal panels contain the realized kernel covariance (upper-right) and correlation (lower-left) between the two asset returns. The sample spans the period from January 2, 2001 until December 31, 2012 ($T = 3017$ days).

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 use three relevant alternative models: the multivariate HEAVY model of Noureldin *et al.* (2012), the short-memory GAS tF model of Opschoor *et al.* (2014) and a multivariate analogue of the "quasi long-memory" 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

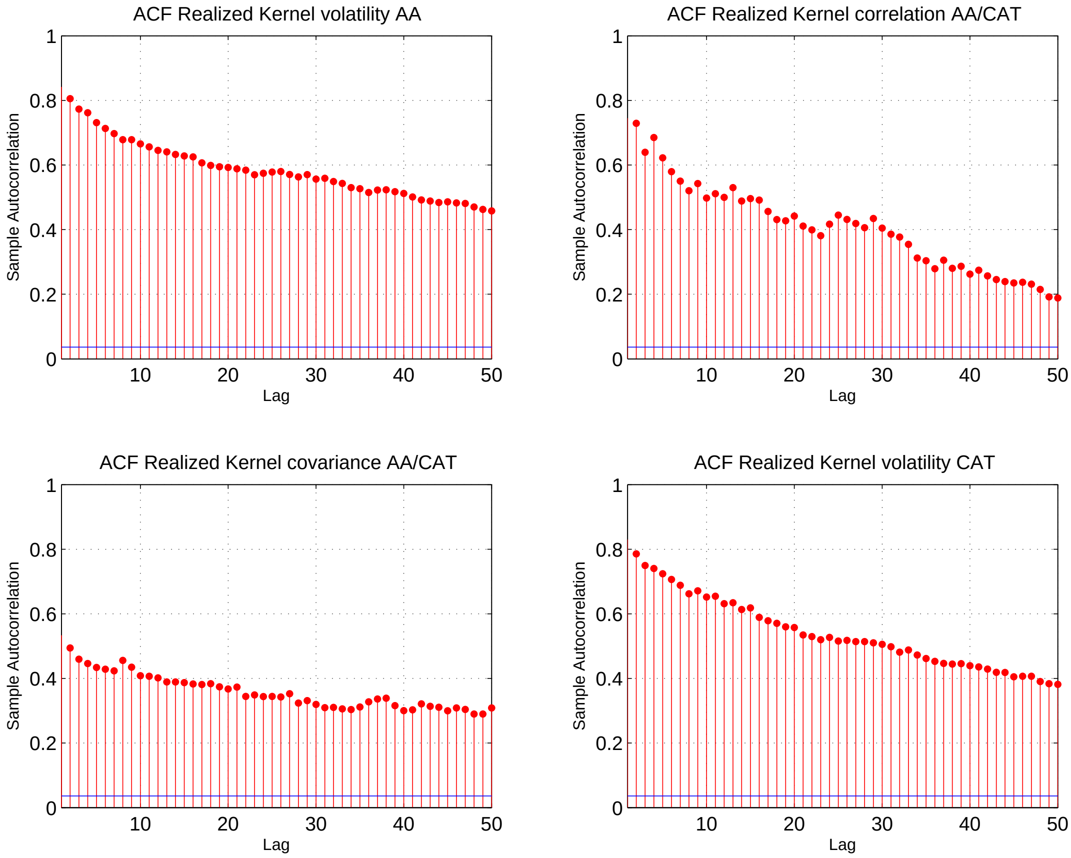

Figure 2: Empirical autocorrelation functions of realized kernels
This figure shows the autocorrelation function (ACF) for lag 1 until 50 of daily realized kernel volatilities (square root of variance) of Alcoa Inc. (AA) and Caterpillar Inc. (CAT) in the diagonal panels. The off-diagonal panels contain 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).

innovations in both of these equations are the realized (co)variance measures as gathered in the matrix RK_t. The dynamics are given by

$$V_{t+1} = C_V C_V' + A_V RK_t + B_V V_t, \quad (19)$$

$$M_{t+1} = C_M C_M' + A_M RK_t + B_M M_t, \quad (20)$$

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.

Our second benchmark is the GAS tF model. This model follows directly from the FIGAS tF model by setting d to zero in (7) and is given by

$$V_{t+1} = \Omega + A\, s_t + B\, V_t \tag{21}$$

with s_t defined as in (14). This benchmark enables us to differentiate between short-memory 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 introduced earlier 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

$$\sigma^2_{n,t+1} = \alpha_n \sigma^2_{n,t} + (1-\alpha_n) y_t^2, \tag{22}$$

$$\alpha_n = \exp(-1/\tau_n), \tag{23}$$

$$\tau_n = \tau_1\, \rho^{n-1}, \tag{24}$$

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.,

$$\sigma^2_{eff,t} = \sum_{n=1}^{n_{max}} w_n \sigma^2_{n,t}, \tag{25}$$

$$w_n = \frac{1}{Q}\left(1 - \frac{\log(\tau_n)}{\log(\tau_0)}\right), \tag{26}$$

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 transition 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

$$\mathbb{E}_t[\sigma^2_{n,t+h}|\mathcal{F}_t] = \alpha_n \mathbb{E}_t[\sigma^2_{n,t+h-1}|\mathcal{F}_t] + (1-\alpha_n)\mathbb{E}_t[\sigma^2_{eff,t+h-1}|\mathcal{F}_t], \tag{27}$$

$$\mathbb{E}_t[\sigma^2_{eff,t+h}|\mathcal{F}_t] = \sum_{n=1}^{n_{max}} w_n \mathbb{E}_t[\sigma^2_{n,t+h}|\mathcal{F}_t]. \tag{28}$$

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. The h-step ahead forecast of V_t of the FIGAS model follows directly from (7): V_{t+h} depends on $s^*_{t+h-1}, s^*_{t+h-2}, \ldots s^*_t, s^*_{t-1}, \ldots$, with $s^*_t = s_t + V_t$ by definition. Given the property that $\mathbb{E}_t[s_{t+h}|\mathcal{F}_t] = 0_k$ for any value of $h \geq 1$, V_{t+h} is obtained recursively by setting the values of future score matrices to the zero. 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:

$$QLIK_{t,h}(RK_{t+h}, V^a_{t+h|t}) = \log|V^a_{t+h|t}| + \mathrm{tr}((V^a_{t+h|t})^{-1} RK_{t+h}), \tag{29}$$

with $V^a_{t+h|t}$ the covariance matrix estimate/forecast for time $t+h$ given all information up to time t 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,

$$d_{t,h}(a,b) = QLIK_{t,h}(RK_{t+h}, V^a_{t+h|t}) - QLIK_{t,h}(RK_{t+h}, V^b_{t+h|t}), \tag{30}$$

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

$$DM_h(a,b) = \frac{\bar{d}_h}{\sqrt{\hat{s}^2_h/(T-h-R)}}, \tag{31}$$

with $\bar{d}_h$ the out-of-sample average of the loss differences, and $\hat{s}^2_h$ a HAC-consistent variance estimator of $d_{t,h}(a,b)$. 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 $V_{t:t+N|t} = \sum_{i=1}^{N} V_{t+i|t}$, where N equals 5 and 10 respectively.

As the above evaluation criteria are statistical in nature, we finally also assess the forecasting performance from a economic point of view. Motivated by the mean-variance optimization setting of Markowitz (1952), 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

$$\min w'_{t+h|t} V_{t+h|t} w_{t+h|t}, \qquad \text{s.t. } w'_{t+h|t}\iota = 1. \tag{32}$$

with ι a $k \times 1$ vector of ones. Similar as Chiriac and Voev (2011), we assess the predictive ability of the different models by comparing the results to the *ex-post* realizations or 'oracle forecasts' 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 again test for significantly different 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 the standard covariance targeting approach does not work well for the HEAVY model. Hence we improve the performance of the HEAVY benchmark model by estimating the HEAVY model with

$$V_{t+1} = c_V\, \Omega_V + A_V RK_t + B_V V_t, \tag{33}$$

where c_V 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. Note that we estimated a FIGAS$(1, d, 0)$ model as the ϕ coefficient turned out to be insignificant. The results therefore only report the parameters $\tilde{B}$ and d for this model. Based on the QLIK loss function, the FIGAS tF has the lowest value,

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 three rows (A, B, c) correspond with the parameters of the equation for V_t of the (FI)GAS tF and the HEAVY model, defined in (33) with the additional scaling parameter c. The fourth and fifth row (A_M, B_M) are related to the HEAVY equation (20) of RK_t with covariance targeting (CT). 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 (29). 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
d	0.660				0.681				0.645			
	(0.009)				(0.009)				(0.004)			
A_V/A		0.419	0.619			0.462	0.635			0.265	0.388	
		(0.035)	(0.012)			(0.032)	(0.012)			(0.011)	(0.004)	
$\tilde{B}/B_V/B$	-0.064	0.597	0.986		-0.053	0.554	0.985		0.164	0.743	0.991	
	(0.018)	(0.033)	(0.001)		(0.016)	(0.029)	(0.001)		(0.006)	(0.010)	(0.000)	
c/c_V	0.004	0.046			0.004	0.060			0.005	0.026		
	(0.001)	(0.006)			(0.001)	(0.007)			(0.001)	(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		9.141		8.973		12.10		11.61	
	(0.501)		(0.469)		(0.394)		(0.497)		(0.404)		(0.377)	
ν_1	46.28		46.61		48.36		49.10		67.00		66.66	
	(0.938)		(0.911)		(0.968)		(0.896)		(0.377)		(0.375)	
ν_2	36.21		34.97		34.75		33.65		61.95		61.16	
	(0.585)		(0.521)		(0.512)		(0.518)		(0.322)		(0.315)	
$\mathcal{L}_t$	-26,431		-26,474		-25,057		-25,085		-72,210		-72,343	
$\mathcal{L}_{SW}$		-43,838				-40,695				-150,072		
$\mathcal{L}_F/\mathcal{L}_W$	-20,795	-45,750	-21,243		-12,082	-37,114	-12,420		67,131	-42,958	64,774	
QLIK	7.692	7.806	7.712	51.43	6.758	6.873	6.774	93.15	19.04	19.25	19.13	602.9

followed by the GAS tF, HEAVY and the RM 2006 model respectively. Comparing the (FI)GAS and HEAVY models, the values suggest for $k = 5$ that the largest gain is obtained by introducing the GAS framework, as the average QLIK drops by 0.10. Hence allowing for fat-tailedness in both the return observations and realized covariance kernels improves the fit substantially. For all equities the combination of long-memory and fat-tailedness that plays a role, though the importance of the long-memory part is less strong for the $h = 1$ forecast horizon. This is not surprising, as the long-memory feature should be particularly relevant for longer forecast horizons. We turn to this in our out-of-sample analysis later on.

We also decompose the likelihoods of the models 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 ($\mathcal{L}_{\mathrm{t}}$) and to the matrix-F distributed observations RK_t ($\mathcal{L}_{\mathrm{F}}$). Similarly, for the HEAVY model we distinguish the likelihood part $\mathcal{L}_{\mathrm{SW}}$ attributable to the singular Wishart observations $y_t y_t'$ and the part $\mathcal{L}_{\mathrm{W}}$ 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 ($k = 5$) or even 2,500 points (all equities) when allowing for long-memory effects in the GAS framework. This illustrates that including one extra parameter (d) has a considerable effect on the statistical fit of the model.

Looking at the individual parameter estimates, we first note the positive and strongly significant long-memory coefficient d, indicating the presence of long-memory effects in volatility. The value of d is highly robust across the two sets of $k = 5$ equities considered, and also hardly changes if we consider all 15 equities. The value of $\tilde{B}$ changes from negative for $k = 5$ for both sets of equities, to positive for the case of all equities. Based on Corollary 3 of Conrad and Haag (2006, page 427), in all three cases we satisfy the constraint for positive definiteness of the resulting covariance matrices V_t. In particular, the Corollary states that if $-1 \leq \tilde{B} \leq 0$, V_t is always positive if $(d - \sqrt{2(2-d)}/2 \leq \tilde{B}$. If $0 \leq \tilde{B} \leq 1$, the variance is positive if $d - \tilde{B} \geq 0$. Both restrictions hold in our empirical application and imply the positive definiteness of the covariance matrices. Though the value of B is negative

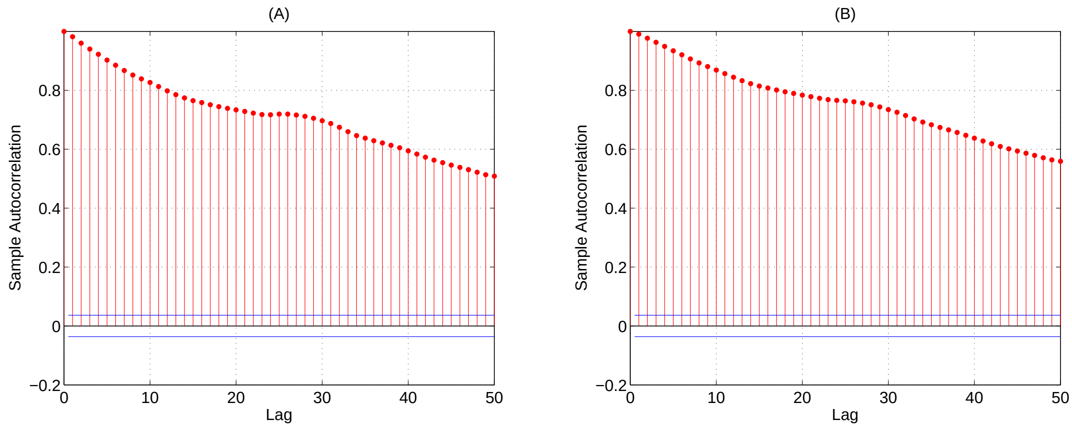

Figure 3: Implied correllograms for V_t
This figure plots the implied correllograms of the conditional variance of Alcoa Inc. (AA) corresponding with the FIGAS tF model based on parameter estimates in Table 2. The left (right) panel shows the correllogram implied by the estimated model on AA/BA/CAT/GE/KO (all equities).

for $k = 5$ and positive for $k = 15$, both values imply a highly similar set of autocovariance functions; see Figure 3. If anything, the increase of the dimension leads to a slightly stronger long-memory feature.

The high degree of persistence in the FIGAS tF model is mirrored by the other models. For example, the estimate of B for the short-memory GAS is very close to 1, indicating a strong persistence. Similarly, the sum of A_M and B_M (the parameters corresponding with the second HEAVY equation) is also very close to 1. 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, such values already result in a substantial moderation of the effect of incidentally large observations RK_t in (16) through the matrix weighting scheme and substantial fat-tailedness of the distribution of RK_t compared to the Wishart distribution.

Figure 4 plots some of the fitted volatilities and correlations. We show the results for Alcoa (AA) and Boeing (BA) for the FIGAS tF model (blue line) and the HEAVY model of equation (33) (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. 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 many more spikes. Notable differences are apparent for both companies during the periods 2001-

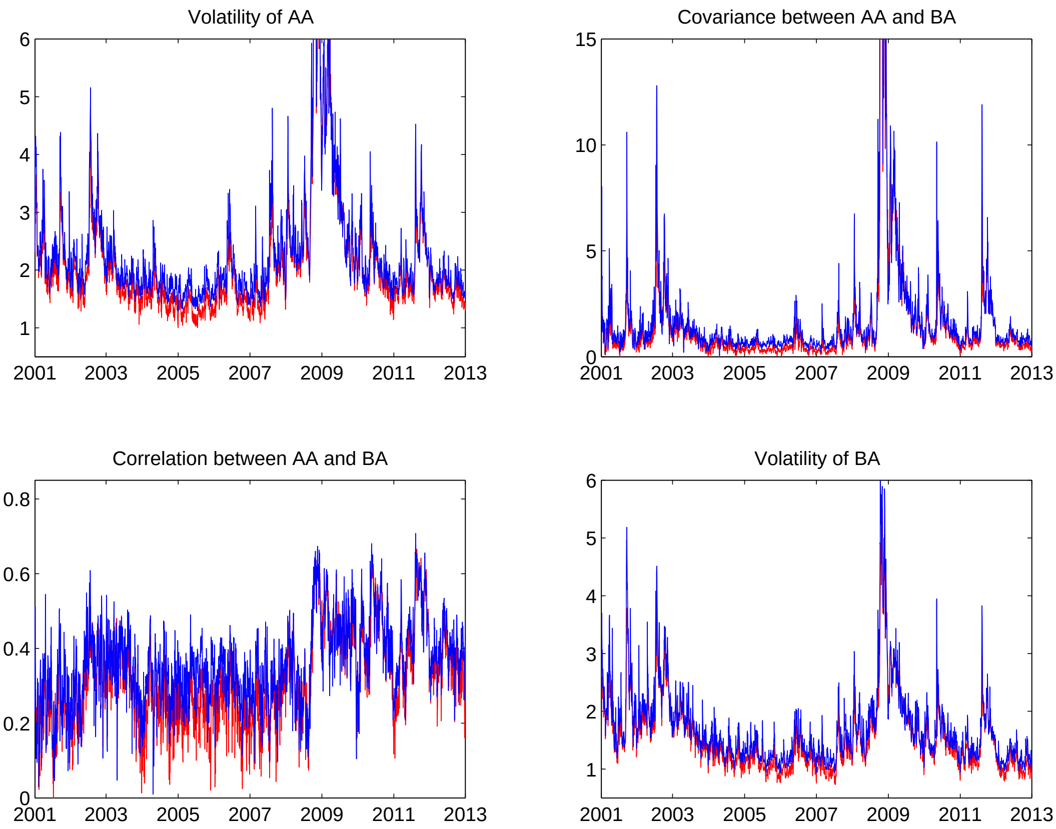

Figure 4: Estimated volatilities and correlations
This figure plots the estimated volatilities of AA and BA (see Table 1) in the upper-left and lower-right panels, and the pairwise covariances and correlations in the upper-right and lower-left panels respectively. Time varying parameter paths are estimated 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).

2003, 2007-2008, and 2010–2011. The patterns for the correlations reveal similar remarkable differences. Again, the number of spikes in the correlation patterns for the HEAVY model is much higher than for the FIGAS model.

3.5 Out-of-sample results

In our out-of-sample analysis, we assess both the short-term and long-term forecasting performance of the FIGAS tF model. We consider h-step ahead forecasts, with $h = 1, 5, 10$, and 22. In addition, we consider aggregated covariance forecasts for the next one or two trading weeks, i.e. $V_{t:t+h} = V_{t+1} + V_{t+2} + \ldots + V_{t+h}$ with $h = 5, 10$. 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.

Table 3: Test-statistics on predictive ability (QLIK criterion)

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 (29). 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	8.42	8.68	9.14	16.30	19.94
	(—)	(—)	(—)	(—)	(—)	(—)
HEAVY	8.07	8.44	8.73	9.31	16.33	19.99
	(-1.5)	(-0.8)	(-1.1)	(-3.1)	(-1.0)	(-1.1)
GAS tF	8.05	8.43	8.75	9.32	16.31	19.97
	(-2.7)	(-0.8)	(-1.7)	(-2.7)	(-0.6)	(-1.1)
RM 2006	8.84	9.30	10.14	12.67	17.06	20.86
	(-13.7)	(-8.0)	(-7.0)	(-7.6)	(-6.5)	(-4.4)
Panel B: CAT/HON/IBM/MCD/WMT						
FIGAS tF	6.13	6.46	6.71	7.12	14.37	17.98
	(—)	(—)	(—)	(—)	(—)	(—)
HEAVY	6.20	6.57	6.84	7.37	14.49	18.12
	(-3.9)	(-3.3)	(-2.7)	(-4.1)	(-4.0)	(-3.0)
GAS tF	6.12	6.44	6.70	7.23	14.36	17.97
	(0.5)	(1.4)	(0.0)	(-2.2)	(1.2)	(0.5)
RM 2006	7.01	7.44	8.28	10.66	15.23	18.94
	(-11.6)	(-6.4)	(-5.6)	(-6.7)	(-5.3)	(-3.7)
Panel C: all equities						
FIGAS tF	19.06	20.05	20.81	21.97	43.77	54.65
	(—)	(—)	(—)	(—)	(—)	(—)
HEAVY	19.12	20.11	20.93	22.33	43.87	54.84
	(-1.0)	(-0.6)	(-0.8)	(-2.4)	(-1.1)	(-1.5)
GAS tF	19.12	20.06	20.89	22.23	43.79	54.72
	(-2.9)	(-0.2)	(-0.8)	(-1.7)	(-0.4)	(-0.7)
RM 2006	24.58	26.62	29.74	38.20	49.61	61.27
	(-10.1)	(-8.6)	(-8.0)	(-8.4)	(-5.7)	(-4.5)

We test the predictive ability of the different models based on the loss-differences of the QLIK loss function (29) using the test-statistic defined in (31). We use a moving window of 1500 observations and re-estimate the parameters after each 25 observations ($\approx$ one month). The first in-sample period corresponds to the period January 2001 until December 2006, which is well before the financial crisis of October 2008. This current forecasting experiment therefore constitutes a major robustness test for all the models considered.

Table 3 contains the results. Negative t-test statistics (in parentheses) indicate that the FIGAS tF model performs better. The negative values in the table clearly show that the FIGAS model outperforms all competing models. This holds particularly for long horizons. For example, the t-statistics -3.1, -4.1, and -2.4 confirm that the FIGAS significantly outperforms the HEAVY model at the 22-step-ahead horizon. Similar results hold with respect to the GAS model and the RM2006 approach. Hence, in contrast to the in-sample results for $h = 1$ 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-step and 10-step ahead forecasts (t-stats of -2.7 and -1.7) and in panel C for 1-step ahead forecasts. The same result holds with respect to the HEAVY model, where panel B indicates that the FIGAS model performs better for short horizons. Compared to the RM 2006 model, FIGAS performs better for all horizons and equity combinations considered. To summarize, taking fat-tailedness of returns and realized covariance kernels into account while simultaneously allowing for long-memory effects provides superior performance to the FIGAS tF model.

Finally, we turn to the economic significance of the covariance matrix forecasts. Table 4 shows the mean of the ex-post conditional portfolio standard deviation, computed by implementing the period-by-period ex-ante minimum variance portfolio weights obtained from equation (32). 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. There appears only one exception (panel B): the ex-post portfolio standard deviation corresponding to the 22-step ahead forecasts of the FIGAS tF and GAS tF model are not statistically significantly 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.

Table 4: Ex-post minimum variance 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). 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	0.925	0.933	0.938	0.946	2.117	3.028
	(—)	(—)	(—)	(—)	(—)	(—)
HEAVY	0.927	0.936	0.942	0.951	2.122	3.040
	(-4.7)	(-4.1)	(-4.2)	(-4.8)	(-3.1)	(-3.7)
GAS tF	0.926	0.936	0.942	0.951	2.122	3.042
	(-5.1)	(-6.5)	(-5.0)	(-5.6)	(-4.6)	(-4.4)
RM 2006	1.013	1.010	1.008	1.005	2.303	3.280
	(-14.2)	(-13.3)	(-12.3)	(-12.3)	(-7.9)	(-5.8)
Panel B: CAT/HON/IBM/MCD/WMT						
FIGAS tF	0.848	0.856	0.859	0.865	1.941	2.771
	(—)	(—)	(—)	(—)	(—)	(—)
HEAVY	0.850	0.858	0.862	0.868	1.949	2.781
	(-2.4)	(-3.4)	(-2.6)	(-2.2)	(-2.0)	(-2.8)
GAS tF	0.849	0.858	0.861	0.866	1.944	2.775
	(-2.6)	(-4.0)	(-3.0)	(-1.3)	(-4.2)	(-2.4)
RM 2006	0.907	0.900	0.895	0.891	2.055	2.921
	(-13.0)	(-12.9)	(-13.1)	(-10.0)	(-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)	(-4.2)	(-4.7)	(-4.5)	(-3.2)	(-4.7)
GAS tF	0.689	0.703	0.711	0.723	1.590	2.286
	(-4.7)	(-5.3)	(-5.3)	(-5.8)	(-4.0)	(-3.9)
RM 2006	0.830	0.817	0.805	0.796	1.872	2.646
	(-15.4)	(-14.5)	(-13.6)	(-12.7)	(-8.2)	(-6.2)

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. For S&P500 equity returns over the period 2001–2012 we showed that both in-sample and out-of-sample and both statistically and economically the new model outperformed recent competitors such as the HEAVY model of Noureldin *et al.* (2012), the GAS tF model of Opschoor *et al.* (2014), and the Zumbach (2006) long-memory version of the RiskMetrics model.

Acknowledgements

We appreciate the comments of participants at the 69th European meeting of the Econometric Society (Geneva, August 2016), the Financial Econometrics and Empirical Asset Pricing Conference (Lancaster, July 2016), the 3rd annual meeting of the IAAE (Milan, June 2016), the 2015 NBER-NSF Time Series Conference (Vienna, September 2015), the 2nd International Workshop on Financial Markets and Nonlinear 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

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 (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, F.X. and R.S. Mariano (1995), Comparing predictive accuracy, *Journal of Business and Economic statistics* **13**, 253–263.

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

Giacomini, R. and H. 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., S.J. Koopman and A. Lucas (2014), Long memory dynamics for multivariate dependence under heavy tails, *Journal of Empirical Finance* **29**, 187–206.

Konno, Y. (1991), A Note on Estimating Eigenvalues of Scale Matrix of the Multivariate F-distribution, *Annals of the Institute of Statistical Mathematics* **43**, 157–165.

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

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 (2016), Time-Varying Systemic Risk: Evidence from a Dynamic Copula Model of CDS Spreads, *Journal of Business and Economic Statistics (forthcoming)* .

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

Patton, A.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.

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