Long Memory Dynamics for Multivariate Dependence under Heavy Tails

Pawel Janus[1]
Siem Jan Koopman[1]
André Lucas[1,2]

[1] *Amsterdam School of Economics, VU University Amsterdam,
and Tinbergen Institute;*
[2] *Duisenberg school of finance.*

Long memory dynamics for multivariate dependence under heavy tails

Paweł Janus [a,b] , *Siem Jan Koopman* [a,b] and *André Lucas* [b,c,d]*

(a) Department of Econometrics, VU University Amsterdam

(b) Tinbergen Institute

(c) Department of Finance, VU University Amsterdam

(d) Duisenberg school of finance

December 9, 2011

Abstract

We develop a new simultaneous time series model for volatility and dependence with long memory (fractionally integrated) dynamics and heavy-tailed densities. Our new multivariate model accounts for typical empirical features in financial time series while being robust to outliers or jumps in the data. In the empirical study for four Dow Jones equities, we find that the degree of memory in the volatilities of the equity return series is similar, while the degree of memory in correlations between the series varies significantly. The forecasts from our model are compared with high-frequency realised volatility and dependence measures. The forecast accuracy is overall higher compared to those from some well-known competing benchmark models.

JEL classification: C10; C22; C32; C51.
Some keywords: fractional integration; correlation; Student's t copula; time-varying dependence; multivariate volatility.

1 Introduction

We propose a new dynamic modeling framework for the simultaneous analysis of volatility and dependence in time series of financial returns. In many studies multivariate extensions of the generalized autoregressive conditional heteroskedasticity (GARCH) model are considered for this purpose such as reviewed by, for example, Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Teräsvirta (2009). When volatility possibly has long memory features, the class of fractionally integrated processes is often considered; see, for example, Robinson (1991), Baillie, Bollerslev, and Mikkelsen (1996), Bollerslev and Mikkelsen (1996), and Tse (1998) for the univariate context, or Teyssière (1997), Brunetti and Gilbert (2000), Pafka and Mátyás (2001), and Conrad, Karanasos, and Zeng (2011) for the multivariate setting. Teyssière (1997) and Brunetti and Gilbert (2000) consider bivariate models where the variance of each series is fractionally integrated, whereas the covariances are time-varying and fractionally integrated only due to their dependence on the volatilities. Pafka and Mátyás (2001) extend Teyssière (1997) by considering a trivariate model with the same fractional integration coefficient for each of the elements of the covariance matrix. Also Conrad, Karanasos, and Zeng (2011) estimate bivariate and trivariate fractionally integrated volatility models under the assumption of constant conditional correlations. Most of the results reported in these studies have indicated that the order of (fractional) integration is similar across the volatilities of different series. There is no evidence that similar results hold for the covariances and/or correlations.

Our contribution in this paper is to develop a new multivariate fractionally integrated model for volatilities and dependence by extending the model of Creal, Koopman, and Lucas (2011) to allow for long memory dynamics in both volatilities and correlations. Our new

model overcomes at least three drawbacks of the models proposed in the literature. First, the time-varying conditional covariance matrix implied by available fractionally integrated models for time-varying volatility and dependence cannot be subjected to analytical restrictions that ensure positive definiteness of the resulting covariance matrix; see Teyssière (1997). Each period, positive definiteness has to be validated and, if needed, restored. The latter is typically done in an *ad hoc* manner. To circumvent problems with positive definiteness of the covariance matrix, most authors impose constant correlations. The hypothesis of constant correlations is typically rejected for financial time series; see Engle (2002). We therefore allow the correlations to be time-varying with possibly long memory dynamics.

Second, most available models are directed towards modeling volatilities or variances. This requires constraints on the parameter space to ensure that the estimated variances are positive at all times. Such restrictions on the parameter space may lead to difficulties at the estimation stage. We circumvent this problem by modeling log-volatilities rather than volatilities themselves. This is close in spirit to the exponential GARCH specification (EGARCH) of Nelson (1992) and it leads to positive variance estimates by construction.

Third, most models integrate the dynamic behaviour of marginal distributions with that of the dependence structure. For example, Conrad et al. (2011) stress the need for heavy-tailed densities, but they use the same tail shape for each of the marginals. Bollerslev (1987) was amongst the first to show that the tail properties can be very different for different financial time series. To allow for different tail shapes, we model the marginals and the dependence structure separately using a dynamic copula approach as in Joe and Xu (1996). The copula framework enables us to have different long memory properties for the volatilities vis-à-vis the dependence measures.

Our final model consists of time-varying conditional marginal Student's t densities with time-varying fractionally integrated volatilities and possibly different degrees of freedom. In addition, the dependence between the marginals is captured by a time-varying Student's t copula function with long memory dynamics for the correlation parameter. Time-varying copulas have been introduced by Patton (2002, 2006) and applied and extended by Dias and Embrechts (2004, 2010), Jondeau and Rockinger (2006), Bartram, Taylor, and Wang (2007) and Ausin and Lopes (2010). In most studies, the dynamics of the copula are estimated as highly persistent processes; see, for example, Jondeau and Rockinger (2006, Table 6). Hafner and Manner (2010) even report that some dynamic copulas exhibit behavior close to a random walk. Our model, however, is the first to introduce fractional integration in the copula setting and to implement this in estimation and forecasting.

The key challenge is to define a fractionally integrated process for the dynamic copula parameter from the Student's t copula. In particular, we need to construct functions of past data that drive the dynamic correlation process in some appropriate way. The functions used in Engle (2002) or Patton (2006) do not account for the particular shape of the observation density. We find a solution in the generalized autoregressive score (GAS) mechanism of Creal, Koopman, and Lucas (2011). They argue that an appropriate function can be based on the scaled score of the conditional density. We adopt their approach in our current setting with Student's t marginals and Student's t copula functions. The resulting function has (conditional) zero mean by construction and facilitates the use of the infinite moving average expansion in the fractionally integrated setting. An additional advantage of this approach is that the resulting model automatically downweights possible outliers and thus limits their impact on volatility and correlation dynamics. The robustness property is crucial

for long memory models where the impact of innovations on future values dissipates at a slow hyperbolic rate.

We conduct both a Monte Carlo study and an empirical study to illustrate the merits of the new model. In the Monte Carlo study, we show that our new fractionally integrated model outperforms short memory alternatives in terms of accuracy when the copula dynamics exhibit a slow rate of decay. In the empirical study, we analyze high-frequency tick-by-tick data from the Trade and Quote (TAQ) database over a long period from January 1993 to May 2010. We construct realised measures as benchmarks for evaluating the forecast quality of different dynamic copula models at a daily frequency. In accordance with earlier studies we find evidence of long memory in the volatility of these series. Long memory features also appear important for the dynamic Student's t copula. However, whereas the long memory properties are comparable across the volatilities of the different series, the degree of long memory for the correlations differs substantially between different pairs of equities. Also, the long memory properties of the correlations and volatilities are substantially different. These findings underline the need for a flexible modeling framework that accounts for these empirical features.

The remainder of the paper is organized as follows. In Section 2 we introduce our long memory model for volatility and dependence. In Section 3 we conduct a Monte Carlo study to examine the performance of competing models for time-varying dependence under a variety of correlation dynamics. In Section 4 we present the results of our empirical study. Section 5 concludes.

2 Conditional volatility and dependence

2.1 Modeling returns and volatility

Let y_t^* denote a daily log return series for time $t = 1, \ldots, n$. The unconditional mean $\mathbb{E}[y_t^*]$ can be consistently estimated by the sample average $\bar{y}^*$. We define the demeaned series $y_t = y_t^* - \bar{y}^*$ for $t = 1, \ldots, n$, as our main quantity of interest. We assume that y_t can be conditionally modeled as a Student's t distributed white noise series. The density of y_t is given by

$$p(y_t | \sigma_t^2; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu-2)\pi\sigma_t^2}} \left[1 + \frac{y_t^2}{(\nu-2)\sigma_t^2} \right]^{-(\nu+1)/2}, \tag{2.1}$$

with time-varying variance $\sigma_t^2 > 0$, degrees of freedom parameter $\nu > 2$, $\Gamma(\cdot)$ denoting the gamma function. The Gaussian distribution is obtained as a special case for $\nu^{-1} \to 0$.

To enforce the volatility process to be positive, we model log volatility $h_t = \log \sigma_t^2$ rather than σ_t^2 itself. The evolution of h_t over time is typically stationary and we can, for example, consider a first-order autoregressive process

$$h_{t+1} = \delta^{(h)} + \beta^{(h)} h_t + \alpha^{(h)} \eta_t, \qquad t = 1, \ldots, n, \tag{2.2}$$

where $\delta^{(h)}$ is an unknown constant, $\beta^{(h)}$ is the autoregressive coefficient, $\alpha^{(h)}$ is a scaling constant, and η_t is an innovation term with mean zero. The superscript (h) indicates that coefficients correspond to the time-varying volatility process h_t. For a stationary process h_t, the unconditional mean is then given by $\mu^{(h)} = \delta^{(h)}/(1 - \beta^{(h)})$. Equation (2.2) can be rewritten as

$$h_{t+1} - \mu^{(h)} = \beta^{(h)}(h_t - \mu^{(h)}) + \alpha^{(h)} \eta_t, \qquad t = 1, \ldots, n. \tag{2.3}$$

6

The nature of the innovation term determines the class of the model. When we take $\eta_t = y_t^2 / \sigma_t^2 - 1$, the model is directly related to the GARCH model of Engle (1982) and Bollerslev (1986) and has several features in common with the exponential GARCH model of Nelson (1992). We base our choice for η_t on Creal et al. (2011) for which the details are given below.

To account for the high-persistence of h_t typically found in empirical work, stationary long memory processes have been proposed for modeling volatility, in particular fractionally integrated processes; see Robinson (1991), Baillie, Bollerslev, and Mikkelsen (1996), Bollerslev and Mikkelsen (1996) and Tse (1998). The fractionally integrated model can be written as

$$(1 - L)^{d^{(h)}} (h_{t+1} - \mu^{(h)}) = \beta^{(h)} (1 - L)^{d^{(h)}} (h_t - \mu^{(h)}) + \alpha^{(h)} \eta_t, \qquad t = 1, \dots, n, \qquad (2.4)$$

where L is the lag operator (with $L h_t = h_{t-1}$) and where the fractional difference operator $(1 - L)^{d^{(h)}}$ is defined by the binomial expansion

$$(1 - L)^{d^{(h)}} = 1 - d^{(h)} L + \frac{d^{(h)}(d^{(h)} - 1)}{2!} L^2 - \frac{d^{(h)}(d^{(h)} - 1)(d^{(h)} - 2)}{3!} L^3 + \dots, \qquad (2.5)$$

for any real order of fractional integration $d^{(h)} > -1$. To enforce stationarity, one can impose $d^{(h)} < 1/2$. For most financial time series positive values for $d^{(h)}$ are found. The time series process (2.4) for h_t is a special case of the autoregressive fractionally integrated moving average (ARFIMA) model as introduced by Granger and Joyeaux (1980) and Hosking (1981). The statistical properties for the ARFIMA process are also applicable to (2.4). For example, if $|\beta^{(h)}| < 1$, $d^{(h)} < 1/2$, and η_t is a martingale difference sequence with finite and time-invariant variance, a covariance stationary and invertible solution exists for (2.4); see

Palma (2007, Section 3.2).

The ARFIMA process substantially extends the range of autocovariance functions implied by short memory processes such as (2.2). The memory properties of h_t depend on both $\beta^{(h)}$ and $d^{(h)}$. For $d^{(h)} < 0$ the process has intermediate memory and all autocovariances (except for lag 0) are negative and decay hyperbolically to zero. For $d^{(h)} = 0$, the autocovariance function decays exponentially and the process is said to have short memory. For $0 < d^{(h)} < 1/2$, the process is persistent and has long memory with an autocovariance function that decays at a slow hyperbolic rate. The innovation η_t then has a long-lasting effect on future values of h_t. Finally, for $1/2 \leq d^{(h)} < 1$ the process is not covariance stationary, but still mean-reverting; see Baillie (1996, p.22). For $d^{(h)} = 1$ the log variance follows a unit root process.

The specification in (2.4) can be further generalized by including a leverage effect. An indicator function for h_t with a different response to positive and negative values of y_t can be considered as in Glosten, Jagannathan, and Runkle (1993). We adopt this generalization in our empirical application of Section 4.

The volatility innovation η_t in (2.2) or (2.4) is set equal to the scaled score function of the predictive logdensity of observation y_t with respect to h_t; see Creal, Koopman, and Lucas (2011). In this framework, the log variance h_t becomes a function of $y_{t-1}, y_{t-2}, \ldots$. The score function is scaled by the inverse of its asymptotic variance which under standard regularity conditions equals the expected information matrix. In the limiting case of a normal density function for y_t in (2.1), we recover a standard GARCH type model. Creal et al. (2011) show that a range of other conditional dynamic models used in financial econometrics can be obtained from the current set-up when considering other densities and parameterizations.

The martingale difference property of the score makes it a convenient choice for η_t in (2.2) or (2.4). A similar model was developed by Harvey and Chakravarty (2008) and is referred to as the Beta-t-EGARCH model. Here we extend the original GAS model with a fractionally integrated volatility (and later also correlation) process.

For the Student's t density in (2.1) with $\sigma_t^2 = \exp(h_t)$, we can express the logdensity as a function of h_t,

$$\log p(y_t|h_t;\nu) = a(\nu) - \frac{1}{2}h_t - \frac{\nu+1}{2}\log[1 + (\nu-2)^{-1}\exp(-h_t)y_t^2], \qquad (2.6)$$

where

$$a(\nu) = \log\big(\Gamma((\nu+1)/2)\big) - \log\big(\Gamma(\nu/2)\big) - \frac{1}{2}\log(\nu-2) - \frac{1}{2}\log\pi,$$

is a function with ν as its only argument. We differentiate (2.6) with respect to h_t and obtain

$$\nabla_t^{(h)} = \frac{\partial \log p(y_t|h_t;\nu)}{\partial h_t} = \frac{1}{2}\left[\omega_t \sigma_t^{-2} y_t^2 - 1\right], \qquad (2.7)$$

where the weight ω_t is given by

$$\omega_t = \frac{\nu+1}{(\nu-2) + \sigma_t^{-2}y_t^2} \geq 0. \qquad (2.8)$$

For given values of y_t and σ_t^2, the weight ω_t converges to unity for the limiting Gaussian case of $\nu^{-1} \to 0$. For given values of ν and σ_t^2, the weight ω_t is decreasing in $\sigma_t^{-2}y_t^2$. Hence relatively large observations (in absolute terms) are downweighted and have less impact on future volatilities. This is a consequence of the heavy tails of the Student's t distribution.

9

The Fisher information is given by

$$\mathcal{I}_t^{(h)} = \mathbb{E}_{t-1}\left(\nabla_t^{(h)\,2}\right) = \frac{1}{2}\frac{\nu}{\nu+3}, \tag{2.9}$$

where $\mathbb{E}_{t-1}$ is expectation with respect to density $p(y_t|h_t;\nu)$. We notice that h_t is a function of $y_{t-1}, y_{t-2}, \ldots$ in our framework. For the limiting Gaussian case, we have $\mathcal{I}_t = \frac{1}{2}$. We finally define the innovation η_t for the log volatility process h_t in (2.2) or (2.4) as

$$\eta_t = \left[\mathcal{I}_t^{(h)}\right]^{-\frac{1}{2}} \nabla_t^{(h)}. \tag{2.10}$$

The initial value h_1 can be set to the unconditional mean of the actual process h_t. Here we scale the score by the square root of the inverse of the Fisher information. Creal et al. (2011) consider different scaling variables for ∇_t. Since the information matrix is constant here, the scaling is less relevant.

An attractive feature of the current model is the explicit link between the density of y_t and the dynamics of the log volatility process h_t. The expression of the weight ω_t in (2.7) is a result of the Student's t distribution for y_t. The robustness property due to ω_t distinguishes our modeling framework from existing fractionally integrated models for volatility, in particular from the fractionally integrated GARCH and EGARCH models of Baillie, Bollerslev, and Mikkelsen (1996) and Bollerslev and Mikkelsen (1996). A volatility model that is robust to large positive or negative log returns is particularly important when log volatility is modeled as a long memory process: large values of η_t then have a long-lasting effect on many future values of h_t due to the slow hyperbolic decay rate of the autocorrelation function. Aberrant values of y_t may then corrupt our volatility estimates

10

for a substantial number of periods. Robustifying the impact of influential values of y_t may thus be particularly useful to improve the fit of long-memory models. We provide empirical illustrations in Section 4.

2.2   Modeling dependence

Our treatment of the cross-sectional dependence between series is based on copula functions and Sklar's theorem. We assume that for a k-dimensional time series the joint distribution function can be decomposed into (i) k marginal distributions and (ii) a copula function for the dependence structure. We consider the bivariate Student's t copula that enable us to capture the dependence between extreme events; see the discussion in Demarta and McNeil (2005). Since the modeling stages of the marginals and the dependence structure are separated, the degrees of freedom parameters for the marginal Student's t distributions can be different. In addition, these degrees of freedom parameters can be different from the degrees of freedom parameter of the Student's t copula that ties the marginals together. Hence the clustering of large absolute returns and of dependence in tail events are modeled separately.

Let $z_{it} = \sigma_{it}^{-1} y_{it}$ denote the ith standardized observation at time t, with $i = 1, \ldots, k$, and $t = 1, \ldots, n$. The probability integral transform of z_{it} is given by

$$u_{it} = \int_{-\infty}^{z_{it}} p_{it}(x) \, \mathrm{d}x,$$

where $p_{it}(\cdot)$ is the marginal density of z_{it}. It follows that u_{it} is uniformly distributed, $u_{it} \sim U(0, 1)$. The copula function is then defined as a distribution function on a k-dimensional hypercube with uniform marginals. We denote the copula as $C(u_{1t}, \ldots, u_{kt})$

11

and the copula density as $c(u_{1t}, \ldots, u_{kt})$. The bivariate Student's t copula density with time-varying correlation parameter ρ_t and degrees of freedom parameter κ is given by

$$c(u_{1t}, u_{2t}|\rho_t; \kappa) = \gamma(\kappa)(1 - \rho_t^2)^{-1/2} \frac{\left[1 + \kappa^{-1}(1 - \rho_t^2)^{-1}\left(x_{1t}^2 + x_{2t}^2 - 2\rho_t x_{1t} x_{2t}\right)\right]^{-(\kappa+2)/2}}{\left[(1 + x_{1t}^2/\kappa)(1 + x_{2t}^2/\kappa)\right]^{-(\kappa+1)/2}}, \quad (2.11)$$

where

$$\gamma(\kappa) = \Gamma\left(\frac{\kappa+2}{2}\right)\Gamma\left(\frac{\kappa}{2}\right)\Gamma^{-2}\left(\frac{\kappa+1}{2}\right), \qquad x_{it} = T_\kappa^{-1}(u_{it}), \qquad i = 1, 2,$$

with correlation coefficient $\rho_t \in (-1, 1)$, and $T_\kappa^{-1}(\cdot)$ the inverse cumulative distribution function of the univariate Student's t distribution with $\kappa > 0$ degrees of freedom. The Gaussian copula density is obtained as the limiting case $\kappa \to \infty$. The resulting joint distribution of the bivariate data is only elliptical if $\nu_1 = \nu_2 = \kappa$ such that it corresponds to the bivariate Student's t distribution.

The correlation coefficient ρ_t and degrees of freedom parameter κ together determine the dependence between the two components of y_t. We treat κ as an unknown fixed coefficient but we allow the correlation parameter ρ_t to be time-varying. To enforce that the correlation always lies in the interval $(-1, 1)$, we define the variable g_t as

$$g_t = \log\frac{1 + \rho_t}{1 - \rho_t} \quad \Leftrightarrow \quad \rho_t = \frac{\exp(g_t) + 1}{\exp(g_t) - 1}. \tag{2.12}$$

We let g_t evolve over time by the process

$$(g_{t+1} - \mu^{(g)}) = \beta^{(g)}(g_t - \mu^{(g)}) + \alpha^{(g)}\varepsilon_t, \qquad t = 1, \ldots, n, \tag{2.13}$$

where $\mu^{(g)}$, $\beta^{(g)}$, and $\alpha^{(g)}$ are unknown coefficients that need to be estimated. The superscript

(g) indicates that coefficients correspond to the time-varying dependence parameter g_t. The

innovation term ε_t has mean zero. Similar to the log-volatility process h_t, the dependence

parameter g_t can also be modeled as a fractionally integrated process

$$(1 - L)^{d^{(g)}}(g_{t+1} - \mu^{(g)}) = \beta^{(g)}(1 - L)^{d^{(g)}}(g_t - \mu^{(g)}) + \alpha^{(g)}\varepsilon_t, \qquad t = 1, \ldots, n, \qquad (2.14)$$

for any real order of fractional integration $d^{(g)} > -1$ with stationarity condition $d^{(g)} < 1/2$.

The discussions related to the processes (2.2) and (2.4) and its coefficients also apply to

the dynamic processes for dependence. The dependence innovation ε_t in (2.13) or (2.14) is

defined by the scaled score of the observation density with respect to g_t. For the case of our

Student's t copula density, we obtain

$$
\begin{aligned}
\nabla_t^{(g)} &= \frac{\partial \log c(u_{1t}, u_{2t}|g_t; \kappa)}{\partial g_t} \\
&= \frac{\dot{\rho}_t}{(1 - \rho_t^2)^2}\left[(1 + \rho_t^2)(\pi_t x_{1t} x_{2t} - \rho_t) - \rho_t(\pi_t x_{1t}^2 + \pi_t x_{2t}^2 - 2)\right], \qquad (2.15)
\end{aligned}
$$

with $\dot{\rho}_t = \partial\rho_t/\partial g_t$,

$$
\begin{aligned}
\pi_t &= (\kappa + 2)/(\kappa + m_t) \geq 0, \\
m_t &= \frac{1}{1 - \rho_t^2}(x_{1t}^2 + x_{2t}^2 - 2\rho_t x_{1t} x_{2t}) \geq 0,
\end{aligned}
$$

where π_t is a weight and m_t is the squared Mahalanobis distance of the vector $x_t = (x_{1t}, x_{2t})'$

to zero with respect to the correlation matrix

$$R_t = \begin{bmatrix} 1 & \rho_t \\ \rho_t & 1 \end{bmatrix}.$$

The weight π_t can be interpreted in a similar way as the weight ω_t for the log variance equation (2.7). For a finite κ, extreme observations x_{1t} and/or x_{2t} leading to a large distance m_t will, as the result of downweighting via π_t, have less impact on the correlation dynamics because they are partly attributed to the heavy-tailed nature of the copula rather than to local increases in the correlation. The degree of robustness increases with a decreasing κ. In the limiting case of $\kappa \to \infty$, we recover the Gaussian copula density with $\pi_t = 1$.

We obtain the expected information as

$$\mathcal{I}_t^{(g)} = \frac{\dot{\rho}_t^2}{(1-\rho_t^2)^2}\left(1 + \rho_t^2 - \frac{2\rho_t^2}{\kappa+2}\right)\frac{\kappa+2}{\kappa+4}, \tag{2.16}$$

see for example Lange, Little, and Taylor (1989) for a detailed derivation. The innovation ε_t in (2.13) or (2.14) is then given by

$$\varepsilon_t = \left[\mathcal{I}_t^{(g)}\right]^{-1/2}\nabla_t^{(g)}. \tag{2.17}$$

Other scaling terms for $\nabla_t^{(g)}$ are also possible as in Creal et al. (2011). From an (unreported) preliminary Monte Carlo study, however, we concluded that scaling as in (2.17) is slightly preferred, as it produces numerically the most stable results. Also, scaling as in (2.17) implies that ε_t has zero mean and constant unit variance.

2.3 Estimation

Parameter estimation is carried out by the method of maximum likelihood. All static model parameters, including the order of fractional integration, are collected in a parameter vector ψ. We maximize the loglikelihood with respect to ψ. The standard errors are computed by means of the square root of the inverse Hessian of the log-likelihood evaluated at the maximum.

Since our modeling framework is based on the copula function, we can estimate the parameters in two steps. First, we consider the marginal models for all series separately and rely on the univariate volatility specification as discussed in Section 2.1. Here we include a subscript i to denote the parameters and marginal densities of the ith series. For example, the log volatility h_t for the ith series is denoted by h_{it}. The ith marginal loglikelihood function is based on the log density (2.6) and is given by

$$
\begin{aligned}
\mathcal{L}_i^{(h)} &= \sum_{t=1}^{n} \log p(y_{it}|h_{it}; \nu_i) \\
&= n \cdot a(\nu_i) - \frac{1}{2} \sum_{t=1}^{n} h_{it} - \frac{\nu_i + 1}{2} \sum_{t=1}^{n} \log\left[1 + (\nu_i - 2)^{-1} \exp(-h_{it}) y_{it}^2\right], \quad (2.18)
\end{aligned}
$$

where the function $a(\nu)$ is given below equation (2.6). For each i, the parameter estimates related to h_{it} are obtained by maximizing (2.18).

Second, the dependence parameters are estimated as follows. For each series i, we compute $\hat{h}_{it}$ using the appropriate updating equations for h_{it} given by (2.2) or (2.4) with parameters replaced by their corresponding estimates. We compute the probability integral

transform for each observation by

$$\hat{u}_{it} = T_{\hat{\nu}_i}\left[\exp(-\frac{1}{2}\hat{h}_{it})\,y_{it}\right],$$

where $T_\nu(\cdot)$ is the cumulative distribution function of the univariate Student's t distribution with ν degrees of freedom. The likelihood for the (bivariate) copula is given by

$$
\begin{aligned}
\mathcal{L}^{(g)} &= \sum_{t=1}^{n} \log c(\hat{u}_{1t}, \hat{u}_{2t}|\rho_t; \kappa) \\
&= n \cdot \gamma(\kappa) + \frac{\kappa+1}{2}\sum_{t=1}^{n}\sum_{i=1}^{2}\log\left(1 + x_{it}^2/\kappa\right) \\
&\quad - \frac{1}{2}\sum_{t=1}^{n}\log(1-\rho_t^2) - \frac{\kappa+2}{2}\sum_{t=1}^{n}\log\left(1 + \frac{\left(x_{1t}^2 + x_{2t}^2 - 2\rho_t x_{1t} x_{2t}\right)}{\kappa(1-\rho_t^2)}\right), \quad (2.19)
\end{aligned}
$$

where $\gamma(\kappa)$ is defined below equation (2.11),

$$x_{it} = T_\kappa^{-1}(\hat{u}_{it}),$$

and ρ_t is a function of g_t as given by (2.12).

For the evaluation of h_t and g_t, it is convenient to express the model in its infinite moving average form. For example, the updating equation (2.4) for h_t can also be given by

$$h_{t+1} = \mu^{(h)} + (1 - \beta^{(h)}L)^{-1}(1-L)^{-d^{(h)}}\alpha^{(h)}\eta_t. \qquad (2.20)$$

Maximum likelihood estimation in fractionally integrated models requires the truncation of the infinite distributed lags polynomial $(1-L)^{-d^{(h)}}$. We follow Baillie, Bollerslev, and Mikkelsen (1996) and Bollerslev and Mikkelsen (1996) and use a fixed truncation lag of

16

1000. For the initialization of the volatility and correlation recursions, we follow Bollerslev and Mikkelsen (1996) who argue that initialization has a negligible effect on parameter estimation when the sample size is sufficiently large. In particular, we set the innovation term of the volatility and correlation recursions to zero for all pre-sample values. Similarly, we set h_1 and g_1 equal to their respective unconditional means.

3 Monte Carlo study

In this section we report the results of a Monte Carlo study to compare the performance of competing models. We concentrate on the newest feature of the model proposed in Section 2, namely the fractionally integrated correlation dynamics under fat-tailed distributions.

3.1 Monte-Carlo design

The design of the current experiment is similar to the studies in Engle (2002) and Hafner and Manner (2010). We simulate series of $n = 5,000$ observations from the bivariate Student's t copula with $\kappa \in \{2, 5, 15\}$ degrees of freedom. The correlation dynamics are generated by updating ρ_t directly or by updating g_t and using the transformation (2.12) to obtain ρ_t. The

following eight updating equations for the correlation process are considered:

1. Constant: $\rho_t = 0.9$;

2. Sine: $\rho_t = 0.5 + 0.4\mathrm{cos}(2\pi t/1000)$;

3. Fast Sine: $\rho_t = 0.5 + 0.4\mathrm{cos}(2\pi t/100)$;

4. Step: $\rho_t = 0.9 - 0.5(t > 2500)$;

5. Ramp: $\rho_t = \mathrm{mod}(t/1000)$;

6. ARMA: $(1 - 0.99L)g_t = 0.01 + 0.05\xi_t$;

7. ARFIMA: $(1 - 0.90L)(1 - L)^{0.45}(g_t - 1) = 0.05\xi_t$;

8. Random Walk: $(1 - L)g_t = 0.025\xi_t$;

where L is the lag operator and ξ_t is Gaussian white noise with mean zero and unity variance. Graphs of the deterministic processes 1 to 5 can be found in Engle (2002). We simulate $M = 1,000$ (bivariate) series (of length $n = 5,000$) for each correlation process and then estimate the parameters for five dynamic model specifications:

(a) GAS : the autoregressive process (2.13) with ε_t as the scaled score;

(b) IGAS : random walk process (2.14) with $\beta^{(g)} = 0$ and $d^{(g)} = 1$;

(c) FIGAS : long memory process (2.14);

(d) Fisher : benchmark model of Dias and Embrechts (2010);

(e) FIBase : benchmark model of Patton (2002, 2006).

The benchmark model (d) is given by

$$g_{t+1} = \delta^{*(g)} + \beta^{*(g)}g_t + \alpha^{*(g)}\,\mathrm{sgn}(x_{1t}\,x_{2t})|x_{1t}\,x_{2t}|^{1/2}, \tag{3.1}$$

18

with coefficients $\delta^{*(g)}$, $\beta^{*(g)}$ and $\alpha^{*(g)}$, where sgn($\cdot$) is the sign function and where x_{it} is defined below (2.11) for $i = 1, 2$. The use of the square root of the product $|x_{1t}\, x_{2t}|$ dampens the impact of extreme observations on the correlation dynamics. Dias and Embrechts (2010) refer to (3.1) as the Fisher(1, 1) dynamic model and they report that it outperforms a set of competing models.

The benchmark model (e) is a long memory extension of Patton (2002, 2006) for the Student's t copula,

$$(1-L)^{d^{\dagger(g)}}(g_{t+1}-\mu^{\dagger(g)}) = \beta^{\dagger(g)}(1-L)^{d^{\dagger(g)}}(g_t-\mu^{\dagger(g)})+\alpha^{\dagger(g)}H^{-1}\sum_{j=1}^{H}(x_{1t-j+1}x_{2t-j+1}-\rho_t), \quad (3.2)$$

with coefficients $d^{\dagger(g)}$, $\mu^{\dagger(g)}$, $\beta^{\dagger(g)}$ and $\alpha^{\dagger(g)}$, where H is a smoothing parameter and where ρ_t is a function of g_t as in (2.12). We set $H = 10$. The updating equation weights all observations equally and does not exploit the robust features of the Student's t density. We refer to equation (3.2) as the FIBase(1, 1) model.

The coefficients for each model are estimated by the method of maximum likelihood as discussed in Section 2.3. In our study we focus on the accuracy of models which is measured by the mean absolute error (MAE) and the mean squared error (MSE) as given by

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|\rho_t - \hat{\rho}_t|, \qquad \text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(\rho_t - \hat{\rho}_t)^2,$$

where $\hat{\rho}_t$ denotes the correlation estimate for each of the five statistical models considered. Both statistics are reported as averages over the $M = 1,000$ simulations. These statistics measure effectively the accuracy of a model in its ability to fit the underlying correlation pattern ρ_t. None of the five statistical models is embedded in any of the eight correlation

19

data generating processes and, therefore, none of the statistical models is *a priori* put at an advantage.

3.2   Results

Table 1 presents the main findings. For each combination of the model (row-wise), the correlation process (column-wise) and the degrees of freedom parameter κ (panels), the MAE and MSE statistics are reported relative to their smallest value across the five statistical models considered. Hence the best performing model is normalized to one and printed in bold. For example, for $\kappa = 2$ and the Step correlation process, the GAS model performs best in terms of MAE as well as MSE.

For $\kappa = 5$, the Student's t distribution has moderately heavy tails. The FIGAS specification produces the best results in half of the cases whereas the GAS model performs best in the other half of the cases. Particularly, if the correlation process has long memory or is smoothly varying (processes Sine, Fast Sine and ARFIMA), the performance of the FIGAS model is excellent. The GAS model produces better results for correlation processes Step, Ramp, ARMA and Random Walk. Overall, the GAS and FIGAS models display good and stable results. The Fisher and FIBase benchmark models perform considerably worse. The deteriorations in terms of MAE and MSE in relation to the GAS and FIGAS models can be as large as 90% for MAE and 150% for MSE.

These findings are amplified if we move to $\kappa = 2$ where the tails of the Student's t are more extreme. The GAS, IGAS and FIGAS models display stable results; their performances and rankings are comparable to the $\kappa = 5$ case. However, the performance of the benchmark models has deteriorated substantially with efficiency losses of more than 600% in

Table 1: *Simulation results*

	Constant	Sine	Fast Sine	Step	Ramp	ARMA	ARFIMA	Rand Walk
				$\kappa = 2$				
MAE								
GAS	1.119	1.044	1.008	**1.000**	1.001	**1.000**	1.004	**1.000**
IGAS	**1.000**	1.058	1.117	1.098	**1.000**	1.064	1.049	1.032
FIGAS	1.080	**1.000**	**1.000**	1.143	1.014	1.006	**1.000**	1.015
Fisher	1.355	1.741	1.326	2.258	1.322	1.109	1.112	1.240
FIBase	1.944	2.842	1.949	3.351	2.004	1.360	1.500	2.437
MSE								
GAS	**1.000**	1.078	1.010	**1.000**	**1.000**	**1.000**	1.009	**1.000**
IGAS	1.100	1.115	1.317	1.090	1.076	1.119	1.096	1.110
FIGAS	1.108	**1.000**	**1.000**	1.118	1.030	1.011	**1.000**	1.079
Fisher	1.849	3.299	1.675	4.563	1.710	1.249	1.217	1.586
FIBase	4.489	6.541	3.553	7.460	3.549	3.613	2.456	5.037
				$\kappa = 5$				
MAE								
GAS	1.130	1.033	1.008	**1.000**	**1.000**	**1.000**	1.004	**1.000**
IGAS	1.311	1.073	1.137	1.113	1.006	1.079	1.053	1.038
FIGAS	**1.000**	**1.000**	**1.000**	1.125	1.008	1.011	**1.000**	1.032
Fisher	1.289	1.442	1.221	1.735	1.157	1.083	1.092	1.209
FIBase	1.723	1.576	1.429	1.904	1.239	1.141	1.144	1.264
MSE								
GAS	1.174	1.049	1.006	**1.000**	**1.000**	**1.000**	1.006	**1.000**
IGAS	1.265	1.144	1.367	1.111	1.078	1.166	1.112	1.136
FIGAS	**1.000**	**1.000**	**1.000**	1.102	1.026	1.022	**1.000**	1.127
Fisher	1.427	1.897	1.355	2.367	1.292	1.164	1.171	1.493
FIBase	2.097	2.149	1.708	2.546	1.418	1.287	1.279	1.586
				$\kappa = 15$				
MAE								
GAS	1.341	1.031	1.009	**1.000**	**1.000**	**1.000**	1.005	**1.000**
IGAS	1.119	1.075	1.134	1.121	1.000	1.084	1.058	1.045
FIGAS	**1.000**	**1.000**	**1.000**	1.127	1.012	1.005	**1.000**	1.033
Fisher	1.047	1.433	1.231	1.749	1.148	1.086	1.094	1.215
FIBase	1.309	1.514	1.215	1.860	1.243	1.123	1.135	1.279
MSE								
GAS	1.360	1.043	1.005	**1.000**	**1.000**	**1.000**	1.009	**1.000**
IGAS	1.150	1.147	1.371	1.127	1.062	1.182	1.128	1.152
FIGAS	**1.000**	**1.000**	**1.000**	1.115	1.036	1.012	**1.000**	1.133
Fisher	1.089	1.874	1.341	2.457	1.344	1.169	1.178	1.529
FIBase	1.377	1.997	1.378	2.464	1.442	1.249	1.259	1.619

specific cases. The case of $\kappa = 15$ represents tail properties close to those for the Gaussian distribution. The performances of the Fisher and FIBase models become relatively better for $\kappa = 15$. The GAS and FIGAS specifications still produce more accurate results when compared to the benchmark models.

We conclude that the GAS and FIGAS specifications present an attractive robustness

feature. Heavy tails for the observation distribution appear to distort the dynamic correlations in a minor way compared to the other models. The models produce good fits to the data under a variety of data generating processes, and in particular produce good results under thin-tailed as well as fat-tailed data distributions.

4   Long memory in equity return dependence measures

In the empirical study we illustrate our modeling framework for an equity data set. Earlier contributions have primarily concentrated on the long memory properties of volatility and have found that these properties are comparable across equities. We consider the long memory properties of both volatility and dependence. Our main conclusion is that the memory properties of dependence measures such as correlations are more diverse across different pairs of stocks.

4.1   Data and proxies for the latent processes

We analyze four different stocks from the Dow Jones Industrial Average index: American Express (NYSE ticker symbol is AXP), General Electric (GE), Coca-Cola (KO), and Procter & Gamble (PG). The stocks represent different industries.

In contrast to the simulation experiment in Section 3, we clearly do not observe the volatility σ_t^2 and correlation ρ_t processes. This creates a complication in assessing the performance of our different statistical models. To solve this problem, we compare the model implied estimates with non-parametric estimates of volatility and correlation. In particular, we follow Andersen and Bollerslev (1998) and use realised (co)variance measures based on high-frequency intra-day data as proxies for their true, unobserved counterparts.

22

The intra-day tick data we use consists of consolidated trades extracted from the Trade and Quote (TAQ) database downloaded through the Wharton Research Data Services system. The data is taken from the NYSE core trading session between 9:30am and 4:00pm, Eastern Time. The sample period runs from January 4, 1993, to May 28, 2010. This yields a total of $n = 4,385$ trading days for all four equities. We clean the raw trade data by applying the methodology discussed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). All data is put to calendar time sampling using the previous tick method and aggregating records with the same time-stamp (in seconds) into one observation using the median price; see Hansen and Lunde (2006).

The statistical models from Section 2 are used to analyze daily return data which are defined as the difference between the logarithm of the close and the open price. The realised (co)variance measures are computed as logarithmic price increments sampled at 5 minute intervals. At this sampling frequency the impact of microstructure noise is benign and can be ignored for all practical purposes. The overnight returns are skipped in the computations.

To account for the possible presence of jumps, we use the threshold realised volatility estimator proposed by Mancini (2009) and defined as

$$TRV_N^k = \sum_{i=1}^{N} (\Delta X_{t_i}^k)^2 \mathbf{1}_{\{(\Delta X_{t_i}^k)^2 \leq r_N\}}, \qquad \Delta X_{t_i}^k = X_{t_i}^k - X_{t_{i-1}}^k, \tag{4.1}$$

where $\mathbf{1}_A$ is the indicator function for event A, r_N is a threshold value and $X_{t_i}^k$ is the log price of asset k at time t_i, for $i = 0, \ldots, N$ with $N = 78$ as the number of intraday returns sampled at the 5 minute frequency. The threshold value is set as $r_N = N^{-0.99}$; see Mancini (2009). The threshold estimator (4.1) is based on the statistical properties of the Brownian motion

23

whose variation is proportional to the time step N^{-1}. A price increment whose variation is larger than the regular Brownian variation might contain a jump component. The use of the threshold function r_N mitigates the influence of jumps on estimated volatility.

The threshold realised correlation between assets k and l is given by

$$TRCorr_N^{k,l} = \frac{\sum_{i=1}^N \Delta X_{t_i}^k \mathbf{1}_{\{(\Delta X_{t_i}^k)^2 \leq r_N\}} \Delta X_{t_i}^l \mathbf{1}_{\{(\Delta X_{t_i}^l)^2 \leq r_N\}}}{\sqrt{\sum_{i=1}^N (\Delta X_{t_i}^k)^2 \mathbf{1}_{\{(\Delta X_{t_i}^k)^2 \leq r_N\}} \sum_{i=1}^N (\Delta X_{t_i}^l)^2 \mathbf{1}_{\{(\Delta X_{t_i}^l)^2 \leq r_N\}}}}; \qquad (4.2)$$

see Mancini and Gobbi (2010). It serves as our benchmark for the model implied estimates of correlation. As in (4.1), the threshold estimator in (4.2) measures the correlation attributable to the continuous parts of the price processes. Classical realised volatility (RV) and realised correlation ($RCorr$) measures can be retrieved in the usual way as described by Andersen, Bollerslev, Diebold, and Labys (2001) and Barndorff-Nielsen and Shephard (2002). We compute the realised measures for each day in our sample.

4.2 Conditional volatility of equity returns

In Table 2 we report the maximum likelihood estimates of parameters for four different models. The equity return series y_t is modeled by

$$y_t = \sigma_t z_t, \qquad z_t \sim \text{Student's } t(\nu), \qquad t = 1, \ldots, n,$$

such that the density for y_t is given by (2.1). The unconditional mean of equity return is denoted as θ. We consider the following four updating processes for transformations of σ_t :

1. fractionally integrated GARCH (FIGARCH):

$$(1 - \beta^{(h)}L)\sigma_{t+1}^2 = \mu^{(h)}(1 - \beta^h) + \left[1 - \beta^{(h)}L - (1 - \alpha^{(h)}L)(1 - L)^{d^{(h)}}\right]y_{t+1}^2.$$

2. fractionally integrated EGARCH (FIEGARCH):

$$(1 - \beta^{(h)}L)(h_{t+1} - \mu^{(h)}) = (1 - L)^{-d^{(h)}}f(z_t), \qquad h_t = \log \sigma_t^2,$$

where

$$f(z_t) = \gamma^{(h)}z_t + \alpha^{(h)}\left(|z_t| - \mathbb{E}|z_t|\right), \qquad \mathbb{E}|z_t| = \frac{\sqrt{\nu - 2}\Gamma((\nu + 1)/2)}{(\nu - 1)\sqrt{\pi}\Gamma(\nu/2)}.$$

3. FIGAS : equations (2.4), (2.7), (2.9) and (2.10) with $h_t = \log \sigma_t^2$.

4. FIGAS with leverage effect (FIGAS+L) : as FIGAS with $\alpha^{(h)}\eta_t$ replaced by

$$\left(\alpha^{(h)} + \gamma^{(h)}\mathbf{1}_{y_t < 0}\right)\eta_t,$$

where $\mathbf{1}_A$ is the indicator function for event A.

We have also considered short memory processes for transformations of σ_t which are obtained from the above models by setting the fractional order of integration to zero; these results are available in the online appendix accompanying this paper.

The estimates of the fractional order of integration $d^{(h)}$ for the threshold realised volatility are computed using the log-periodogram regression method GPH of Geweke and Porter-Hudak (1983); for the details, see Maynard and Phillips (2001). The resulting estimates are all between 0 and 1, and they are similar across equities. It provides clear evidence of

Table 2: *Maximum likelihood estimates for four conditional volatility models*

Equity	$d^{(h)}_{TRV}$	Model	θ	$\alpha^{(h)}$	$\gamma^{(h)}$	$\beta^{(h)}$	ν	$d^{(h)}$	AIC	MAE_{TRV}
AXP	0.543	FIGARCH	0.046 (0.020)	0.250 (0.034)	-	0.748 (0.043)	10.338 (1.224)	0.595 (0.056)	16996.78	1.021
		FIEGARCH	0.049 (0.017)	0.187 (0.030)	-0.067 (0.015)	0.333 (0.133)	9.411 (1.169)	0.692 (0.031)	16945.88	0.955
		FIGAS	0.042 (0.020)	0.097 (0.017)	-	0.390 (0.141)	9.536 (1.213)	0.699 (0.034)	16965.03	0.978
		FIGAS+L	0.047 (0.018)	0.056 (0.015)	0.088 (0.024)	0.297 (0.159)	9.640 (1.237)	0.703 (0.032)	16941.17	0.945
GE	0.507	FIGARCH	0.006 (0.017)	0.313 (0.037)	-	0.743 (0.039)	11.419 (1.382)	0.544 (0.048)	15158.66	1.048
		FIEGARCH	0.006 (0.014)	0.216 (0.033)	-0.074 (0.016)	0.121 (0.157)	11.930 (1.794)	0.708 (0.037)	15077.14	0.978
		FIGAS	0.003 (0.017)	0.111 (0.017)	-	0.074 (0.168)	11.751 (1.722)	0.743 (0.039)	15103.16	0.969
		FIGAS+L	0.007 (0.016)	0.075 (0.016)	0.068 (0.021)	0.144 (0.157)	12.373 (1.883)	0.713 (0.040)	15093.18	0.956
KO	0.527	FIGARCH	0.046 (0.015)	0.304 (0.037)	-	0.785 (0.038)	8.219 (0.753)	0.509 (0.047)	13835.42	0.997
		FIEGARCH	0.040 (0.014)	0.151 (0.033)	-0.038 (0.013)	0.343 (0.177)	7.651 (0.786)	0.696 (0.046)	13817.07	0.975
		FIGAS	0.043 (0.015)	0.073 (0.019)	-	0.428 (0.199)	7.858 (0.834)	0.705 (0.052)	13829.63	0.982
		FIGAS+L	0.036 (0.013)	0.036 (0.013)	0.086 (0.024)	0.324 (0.179)	8.275 (0.922)	0.700 (0.043)	13805.18	0.957
PG	0.510	FIGARCH	0.084 (0.014)	0.400 (0.031)	-	0.712 (0.032)	8.368 (0.738)	0.439 (0.029)	13856.10	1.052
		FIEGARCH	0.090 (0.014)	0.192 (0.034)	-0.047 (0.015)	0.406 (0.148)	7.923 (0.823)	0.612 (0.044)	13781.02	0.948
		FIGAS	0.096 (0.015)	0.105 (0.021)	-	0.371 (0.175)	8.135 (0.870)	0.637 (0.047)	13790.22	0.952
		FIGAS+L	0.089 (0.015)	0.079 (0.017)	0.065 (0.023)	0.346 (0.159)	8.225 (0.895)	0.632 (0.047)	13780.84	0.932

The first column gives the NYSE ticker symbol of an equity. The second column is for $d^{(h)}_{TRV}$ and denotes the estimate of the fractional difference parameter for the threshold realised volatility (TRV) series as defined in (4.1). The $d^{(h)}_{TRV}$ is estimated based on the log-periodogram regression method of Geweke and Porter-Hudak (1983) and implemented as in Maynard and Phillips (2001). The third column indicates the process for σ_t as detailed in the text. The penultimate column is for the Akaike Information Criterion (AIC). The final column is for MAE_{TRV} and denotes mean absolute error between the model-based volatility process and the TRV series. We do not report the estimates of $\mu^{(h)}$ because it is not identified when $d^{(h)} \geq 1/2$. We report the estimates with their associated standard errors in parentheses below.

long range dependence in volatility; see Maasoumi and McAleer (2008) for more discussion. The persistence of the volatility processes can also be inferred from the estimates of the autoregressive coefficient that are all larger than 0.99 for the short memory formulations. The estimates of $d^{(h)}$ for the different models and different series are between 0.43 and 0.75; they are statistically different from zero and from unity based on conventional significance levels. This finding implies that innovations have a long-term effect on future volatility levels. The AIC criterion suggests that in most cases the volatility is best described by the non-stationary but mean-reverting fractionally integrated process. Confirmation of this finding is given by the reported MAE values in the last column of Table 2. It shows that the FIGAS+L model is most precise for all four equities. This finding is also consistent with earlier studies in which a similar degree of long memory in volatility is found for these equities. In addition, we find convincing evidence for a leverage effect; the estimates of γ are statistically significant for all equities. The degrees of freedom parameter ν is estimated in the range between 7.5 and 12.5; the tail shapes of the equity series are quite similar.

In Figure 1 we present the model-based volatility estimate of σ_t together with $TRV_t^{1/2}$. The realised measure is more jagged than the model-based volatilities but similar patterns can be detected. The time series are subject to volatile and tranquil periods with the average annualized volatility (right-hand scale) of around 25%. The other panels in Figure 1 present scatter plots of the four volatility estimates with long memory properties. A clear difference is apparent between FIGAS+L and FIGARCH volatility estimates; the estimates from the FIGAS+L, FIGAS and FIEGARCH models are similar. Noteworthy differences remain, however. To highlight these differences, we zoom in on some particular event days.
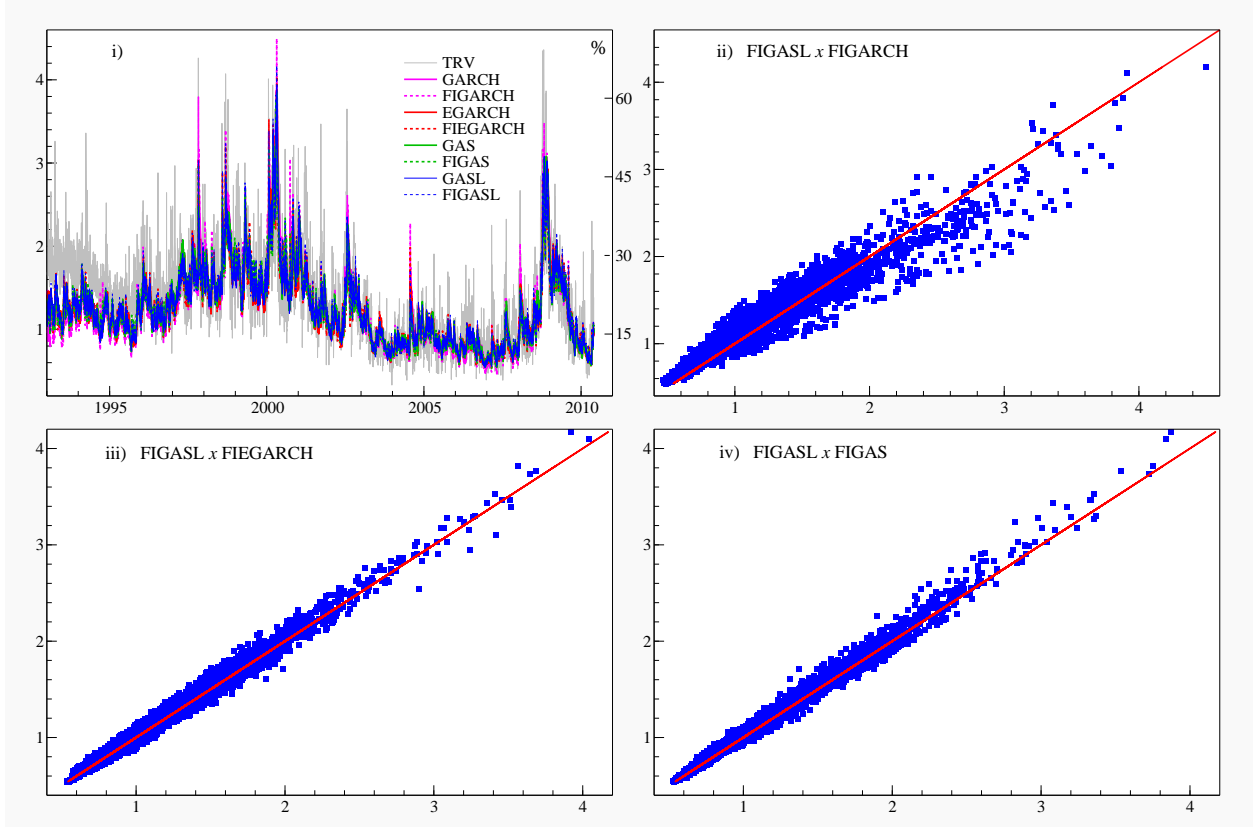
27

Figure 1: *Conditional volatility estimates of Procter & Gamble equity returns*

The upper-left panel presents a time series plot for Procter & Gamble (PG) equity returns of model implied volatilities plotted along with the threshold realised volatility estimator. The other panels contain FIGAS+L volatility estimates plotted against FIGARCH (upper-right), FIGAS (lower-left), and FIEGARCH (lower-right) volatility estimates.
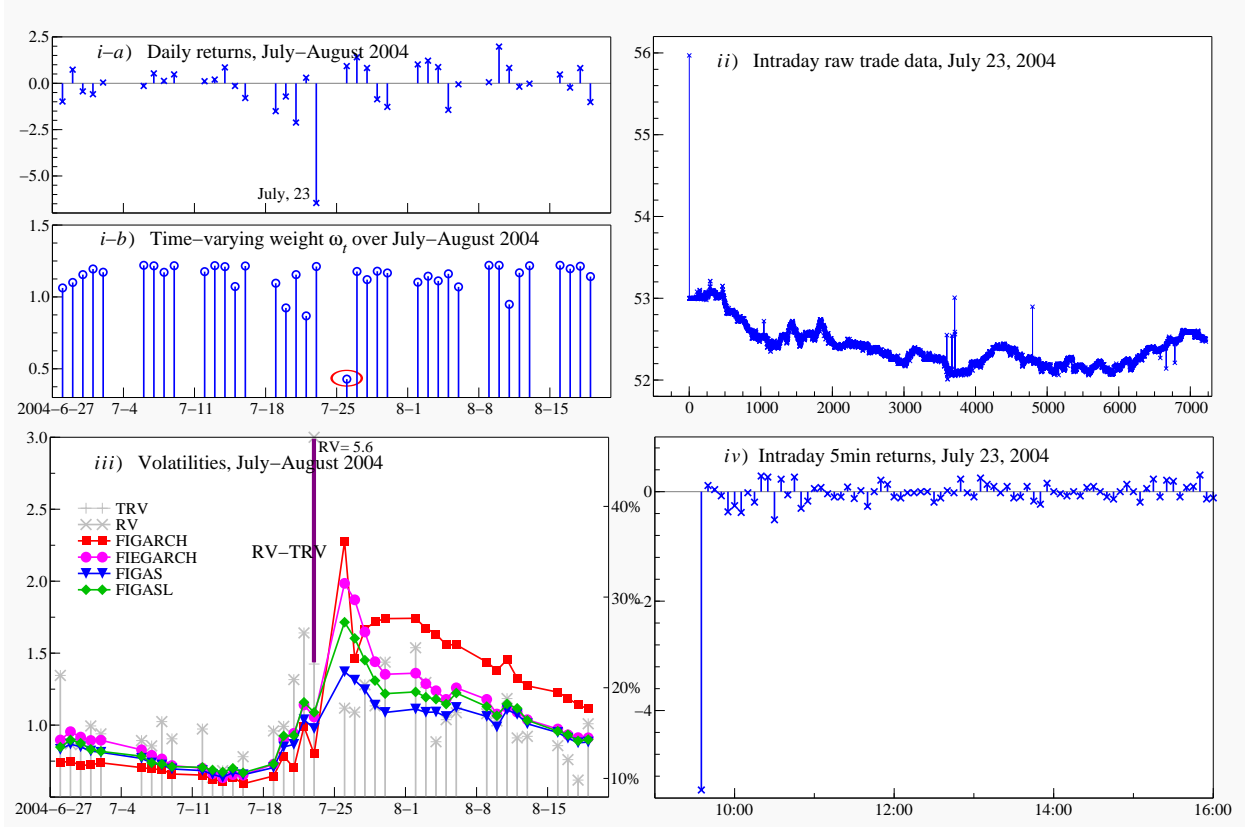
Figure 2: *Case study for Procter & Gamble (PG) equity returns*

The upper-left panel consists of a top panel containing the daily equity returns, and a bottom panel containing the weights ω_t of the FIGAS model. The upper-right panel contains the intraday time series plot of the price level. The lower-left panel presents the time series of the realised variance together with the model implied volatility estimates. The lower-right panel plots the intraday 5 minute returns.

In the upper-left panel of Figure 2 we present daily PG equity returns in the summer of 2004 together with the weights ω_t of (2.8) which are used for computing the score $\nabla_t^{(h)}$ in (2.7). The aberrant observation of July 23, 2004 is clearly visible and takes the value of approximately -6.5% : it is caused by the opening price of 55.97\$ and a closing price of 52.49\$ as depicted in the upper-right panel. This daily price increment was much larger (in absolute terms) than in adjacent days. From the high-frequency intraday data we observe that after the first trade, the remaining transactions over this day were relatively stable at a price level of around 52.50\$-53.00\$. Consequently, the first intraday return as well as the daily return are large. In the lower-left panel of Figure 2 we show the noticeable influence of this event on the volatility estimate. The substantial divergence between the standard realised variance measure (RV) and TRV on this day is noticeable. Also the model-based volatility estimates increase instantly and it takes some time before they return to their previous level or to TRV. However, the effect is more pronounced in the FIGARCH estimate than in its FIGAS counterpart. The FIGAS updating scheme downweights the extreme observation through the weight ω_t. The resulting volatility estimate is more robust. The volatility dynamics in the FIGAS+L specification are designed to respond relatively more to negative shocks compared to the FIGAS specification. We therefore notice a higher impact of the large return on the FIGAS+L compared to the FIGAS specification. However, the FIGAS+L volatility remains more robust than its FIGARCH and FIEGARCH counterparts.

4.3   Conditional dependence between returns on equities

Next we study the dynamic behavior of the dependence structure in pairs of equities. We compute standardized returns using the volatility estimates of a particular model and trans-

form the returns into uniform variables using the probability integral function of the Student's t density. Then a set of competing models are considered for pairs of such transformed series. We report estimation results for three pairs only: AXP/GE, GE/KO and KO/PG. Results for other pairs are available in the online appendix. Table 3 presents the parameter estimates and associated standard errors for four models that are introduced in Section 3.1: Fisher, GAS, FIBase and FIGAS. For each model, we consider Gaussian (G) and Student's t (t) copula functions.

In a preliminary analysis, we compute the GPH estimates of the fractional order of integration $d^{(g)}$ of the threshold realised correlation $TRCorr$. The GPH estimates are reported in the second column of Table 3 and they have values in the range from 0.35 to 0.50 which indicates long memory features in the correlation processes. However, the GPH estimates for realised correlations are smaller than those for realised volatilities. Within the model-based framework, the persistence of the correlation processes can be inferred from the estimates of the autoregressive coefficients $\beta^{(g)}$ that are in the range $0.970 - 0.997$ for the Fisher and GAS models. The lower bound of this range is smaller than the $\beta^{(h)}$ estimates for volatilities. It suggests that the degree of memory for the correlation processes may be different. In general, we find that the correlation series are persistent. The empirical results of Dias and Embrechts (2010) also indicate that high persistence of dependence processes for dynamic copulas is a common feature for financial returns of different sampling frequencies. Hafner and Manner (2010) report that copula dependence exhibits 'nearly integrated' behaviour.

The last two columns of Table 3 report the AIC and MAE criteria which may be used to select the 'best' model. From these criteria we conclude that the FIGAS-t specification is always best or second best for the three equity pairs. This finding also applies to the

Table 3: *Maximum likelihood estimates for conditional dependence models*

Pair	d_{TRCorr}	Model	$\delta^{(g)}/\mu^{(g)}$	$\alpha^{(g)}$	$\beta^{(g)}$	$d^{(g)}$	κ	AIC	MAE_{TRCorr}
AXP/GE	0.497	Fisher-G	0.007 (0.003)	0.049 (0.012)	0.980 (0.005)	-	-	-1220.05	1.000
		GAS-G	0.004 (0.003)	0.036 (0.008)	0.996 (0.002)	-	-	-1217.33	0.965
		FIBase-G	1.026 (0.152)	0.031 (0.043)	0.588 (0.537)	0.546 (0.073)	-	-1224.87	0.967
		FIGAS-G	0.950 (0.196)	0.045 (0.027)	0.408 (0.384)	0.714 (0.077)	-	-1221.64	0.941
		Fisher-t	0.005 (0.003)	0.045 (0.011)	0.982 (0.005)	-	9.942 (1.890)	-1253.73	0.990
		GAS-t	0.003 (0.003)	0.038 (0.009)	0.997 (0.002)	-	9.890 (1.912)	-1248.64	0.951
		FIBase-t	1.059 (0.145)	0.037 (0.042)	0.556 (0.559)	0.586 (0.072)	8.495 (1.27)	-1231.26	0.946
		FIGAS-t	0.955 (0.209)	0.074 (0.030)	0.251 (0.334)	0.711 (0.074)	9.969 (1.938)	-1252.60	0.934
GE/KO	0.462	Fisher-G	0.005 (0.002)	0.034 (0.009)	0.985 (0.004)	-	-	-753.75	1.000
		GAS-G	0.003 (0.002)	0.023 (0.005)	0.996 (0.002)	-	-	-755.65	0.991
		FIBase-G	1.204 (0.116)	0.215 (0.121)	0.220 (0.398)	0.318 (0.096)	-	-722.27	1.004
		FIGAS-G	0.817 (0.140)	0.057 (0.027)	0.351 (0.356)	0.653 (0.085)	-	-747.06	0.989
		Fisher-t	0.005 (0.002)	0.031 (0.009)	0.985 (0.005)	-	8.174 (1.361)	-799.50	1.001
		GAS-t	0.003 (0.002)	0.027 (0.008)	0.996 (0.003)	-	8.184 (1.352)	-802.58	0.985
		FIBase-t	0.947 (0.068)	0.203 (0.097)	0.055 (0.474)	0.298 (0.090)	7.960 (1.243)	-772.76	1.066
		FIGAS-t	0.804 (0.163)	0.062 (0.033)	0.405 (0.379)	0.652 (0.098)	8.084 (1.309)	-797.67	0.984
KO/PG	0.394	Fisher-G	0.011 (0.006)	0.041 (0.014)	0.979 (0.008)	-	-	-897.53	1.000
		GAS-G	0.024 (0.009)	0.049 (0.011)	0.978 (0.008)	-	-	-906.37	1.001
		FIBase-G	1.402 (0.126)	0.150 (0.129)	0.408 (0.502)	0.291 (0.082)	-	-872.77	0.993
		FIGAS-G	1.090 (0.065)	0.043 (0.018)	0.969 (0.029)	0.078 (0.232)	-	-905.64	1.005
		Fisher-t	0.014 (0.007)	0.048 (0.015)	0.975 (0.010)	-	7.106 (1.051)	-957.83	0.994
		GAS-t	0.030 (0.012)	0.066 (0.015)	0.971 (0.011)	-	7.500 (1.166)	-961.99	0.992
		FIBase-t	1.237 (0.105)	0.044 (0.023)	0.997 (0.002)	0.055 (0.096)	7.650 (1.142)	-955.85	1.079
		FIGAS-t	1.074 (0.079)	0.052 (0.018)	0.945 (0.045)	0.173 (0.207)	7.484 (1.156)	-961.45	0.990

Note: d_{TRCorr} denotes the estimate of the fractional difference parameter for the $TRCorr$ measure based on log-periodogram regression method of Geweke and Porter-Hudak (1983). AIC stands for Akaike Information Criterion. MAE^{TRCorr} stands for the mean of the absolute errors between the model-based correlation process and threshold realised correlation. Suffixes G and t stand for Gaussian and t copula, respectively. The Fisher and FIBase models are discussed in Section 3.1. The (FI)GAS model for dependence is presented in Section 2.2.

(unreported) other three pairs. The estimates of $d^{(g)}$ for the FIGAS specification have values in the range from -0.01 to 0.72 for all six pairs. For the three reported pairs we find

significant estimates of $d^{(g)}$ for the pairs AXP/GE and GE/KO. Such findings also hold for the FIBase model. For the AXP/GE pair, the $d^{(g)}$ estimate is even statistically different from 0.5; it suggests that the dependence measure is mean-reverting but is not covariance-stationary. These results also provide empirical evidence that the restriction adopted in Pafka and Mátyás (2001), i.e., imposing a common degree of long memory across volatilities and covariances, may be too restrictive.
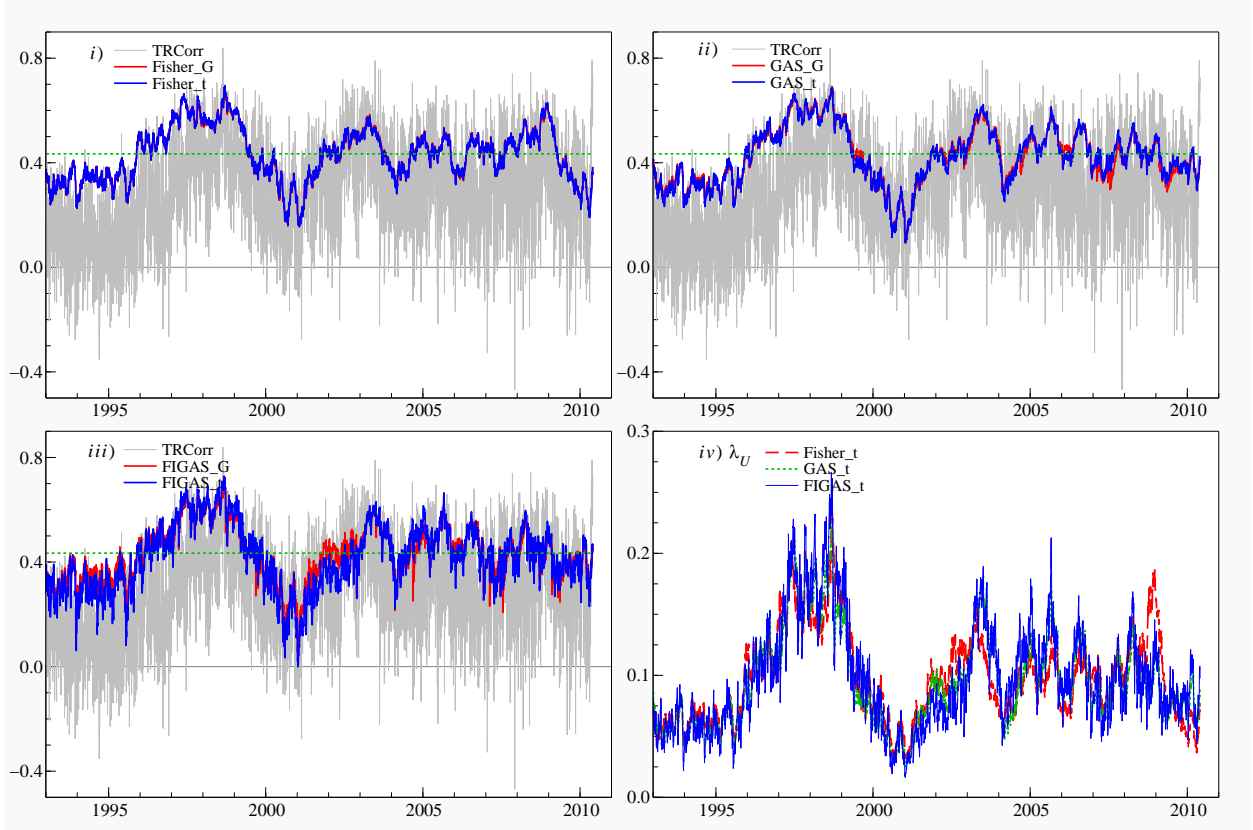

Figure 3: *Conditional dependence between returns on GE and KO equities*
The figure contains realised and model implied correlation measures. The dotted horizontal line is the estimate of ρ in the static bivariate t copula.

Figure 3 presents the time-varying correlations together with the threshold realised correlation ($TRCorr$) for the equity pair GE/KO. The realised correlation is much noisier than the model-based correlations. The GE/KO dependence oscilattes around the value obtained from the static copula (long-run mean) without any significant breaks over time. The Stu-

dent's t copula density generalizes the Gaussian copula density by allowing for non-zero dependence in the tails of the distribution. The correlation ρ_t and the degrees of freedom κ determine the tail dependence for the Student's t copula. This copula density is symmetric: the upper and lower tail dependence are equal and are given by

$$\lambda_{L,t} = \lambda_{U,t} = 2T_{\kappa+1}\left(-\sqrt{\kappa+1}\frac{\sqrt{1-\rho_t}}{\sqrt{1+\rho_t}}\right) \in [0,1]. \tag{4.3}$$

The lower-right panel of Figure 3 presents the implied time-varying tail dependence. It appears that the highest dependence between extreme events for this pair of equities was around 1997-1998.
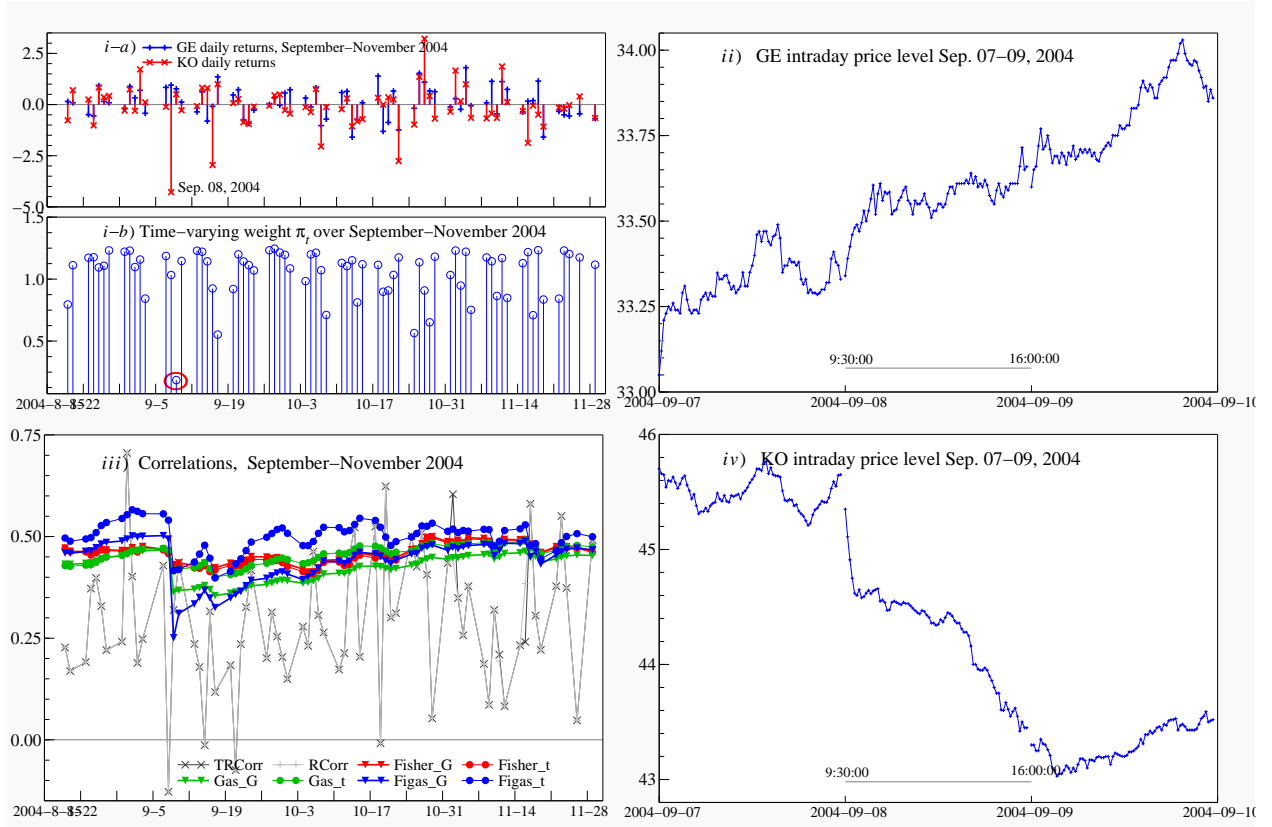

Figure 4: *Case study for the GE/KO equities*

The upper-left panel contains the daily equity returns for GE and KO (upper half), and the weights π_t of the FIGAS-t model (lower half). The upper-right and lower-right panels contains the intraday price dynamics for GE and KO, respectively. The lower-left panel presents the time series of the realised correlation together with the model implied correlation estimates.

The model-based correlations in Figure 3 look very similar. However, the correlations of different models can behave rather differently when extreme observations occur in the sample. An example is given by Figure 4 for GE/KO. The daily return of GE for September 8, 2004 was positive while the return for KO was substantially negative, approximately -4.5%. The upper and lower right-hand panels in the graph present the 5 minute prices of GE and KO over three consecutive days. During September 8, 2004, the two equity prices clearly moved in opposite directions while this is not the case for the surrounding days. The isolated patterns for this particular day has a large effect on model-based correlations as we observe in the lower-left panel of Figure 4. The correlations drop sharply downwards and slowly return to their previous levels afterwards. Given the long memory features of the models, the mean reversion is very gradual. The lower half of the upper-left panel presents the weights π_t based on the GAS and FIGAS models with the Student's t copula specification. The aberrant pattern on Sep 8, 2004 receives a smaller weight for both Student's t models. Consequently, these correlation estimates are more robust compared to their Gaussian counterparts. The values of the GAS-G and FIGAS-G specifications appear to be heavily affected by the occurrence of the isolated extreme return realizations. The FIGAS-t is much more robust.

5   Summary and conclusions

We have proposed a new observation driven fractionally integrated model for time-varying volatility and dependence in data with heavy tailed and long memory features. The new model extends the literature in several ways. First, correlations rather than covariances are modeled as time-varying processes. In contrast to Teyssière (1997) and Brunetti and

Gilbert (2000), our model lets covariances vary over time due to changes in both variances and correlations. Second, the correlation processes are modeled in a copula framework and the volatilities are modeled by means of marginal distributions. This approach enabled us to allow for different degrees of fractional integration in volatilities and in correlations. Third, by employing the generalized autoregressive score (GAS) framework of Creal, Koopman, and Lucas (2011), we have shown that the correlation and volatility estimates are more robust to incidental extreme observations when considering Student's t copula and marginal density functions, respectively. The GAS updating scheme introduces a weighting scheme that mitigates the effects of aberrant observations on correlation and volatility estimates. The robustness is particularly relevant in the context of a fractionally integrated process: each shock has a long-lasting impact due to its long memory dynamic property.

In a Monte Carlo study we have shown that the new model outperforms a number of recently proposed benchmark models for a variety of correlation processes. In an empirical study we have shown that the robustness property of the new model is a key distinguishing feature. All model-based volatility estimates have been benchmarked against non-parametric threshold realised variance and correlation estimates based on high-frequency returns. From the perspective of our dynamic copula approach, we have found that the degree of long memory in the individual volatilities is similar across equities. We have also found that the degree of long memory in correlations across equity pairs can be substantially different. These findings stress the need for flexible time series models that allow for cross-sectional differences in the dynamic behavior and persistence of correlations and volatilities.

References

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

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*(453), 42–55.

Ausin, M. C. and H. F. Lopes (2010). Time-varying joint distribution through copulas. *Computational Statistics and Data Analysis 54*(11), 2383–2399.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics 73*, 5–59.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 74*, 3–30.

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. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*(2), 253–280.

Bartram, S. M., S. J. Taylor, and Y.-H. Wang (2007). The Euro and European financial market dependence. *Journal of Banking & Finance 31*(5), 1461–1481.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics 21*, 79–109.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics 69*(3), 542–547.

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

Brunetti, C. and C. L. Gilbert (2000). Bivariate FIGARCH and fractional cointegration. *Journal of Empirical Finance 7*, 509–530.

Conrad, C., M. Karanasos, and N. Zeng (2011). Multivariate fractionally integrated APARCH modeling of stock market volatility: A multi-country study. *Journal of Empirical Finance 18*, 147–159.

Creal, D., S. J. Koopman, and A. Lucas (2011). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, Volume 26, forthcoming.

Demarta, S. and A. J. McNeil (2005). The t copula and related copulas. *International Statistical Review 73*(1), 111–129.

Dias, A. and P. Embrechts (2004). Dynamic copula models for multivariate high-frequency data in finance. Manuscript, ETH Zurich.

Dias, A. and P. Embrechts (2010). Modeling exchange rate dependence dynamics at different time horizons. *Journal of International Money and Finance 29*(8), 1687–1705.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica 50*, 987–1008.

Engle, R. F. (2002). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics 20*(3), 339–350.

Geweke, J. F. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis 4*(4), 221–238.

Glosten, L., R. Jagannathan, and D. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*, 1779–1801.

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*(1), 15–29.

Hafner, C. M. and H. Manner (2010). Dynamic stochastic copula models: estimation, inference and applications. Journal of Applied Econometrics, Forthcoming.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics 24*, 127–161.

Harvey, A. and T. Chakravarty (2008). Beta-t-(E)GARCH. Cambridge Working Papers in Economics 0840, Faculty of Economics, University of Cambridge.

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

Joe, H. and J. J. Xu (1996). The estimation method of inference functions for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia.

Jondeau, E. and M. Rockinger (2006). The Copula-GARCH model of conditional dependencies: An international stock market application. *Journal of International Money*

and Finance 25, 827–853.

Lange, K. L., R. J. A. Little, and J. M. G. Taylor (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association 84* (408), 881–896.

Maasoumi, E. and M. McAleer (2008). Realized volatility and long memory: An overview. *Econometric Reviews 27*(1-3), 1–9.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics 36* (2), 270–296.

Mancini, C. and F. Gobbi (2010). Identifying the brownian covariation from the co-jumps given discrete observations. Working paper, Dipartimento di Matematica, Universita' degli Studi di Firenze.

Maynard, A. and P. C. B. Phillips (2001). Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. *Journal of Applied Econometrics 16*, 671–708.

Nelson, D. (1992). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59* (2), 347–370.

Pafka, S. and L. Mátyás (2001). Multivariate diagonal FIGARCH: specification, estimation and application to modelling exchange rates volatility. Working Paper, Central European University. Economics Department.

Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. Wiley-Interscience: Wiley Series in Probability and Statistics.

Patton, A. J. (2002). Applications of copula theory in financial applciations. Unpublished Ph.D. dissertation. University of California, San Diego.

Patton, A. J. (2006). Modeling asymmetric exchange rate dependence. *International Economic Review 42*(2), 527–556.

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

Silvennoinen, A. and T. Teräsvirta (2009). Multivariate GARCH models. In T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 201–229. Springer-Verlag, Berlin.

Teyssière, G. (1997). Modelling exchange rates volatility with multivariate long-memory ARCH processes. Working paper, GREQAM 97b03, Universite d'Aix-Marseille III.

Tse, Y. K. (1998). The conditional heteroscedasticity of the yen-dollar exchange rate. *Journal of Applied Econometrics 13*(1), 49–55.