A Score-Driven Conditional Correlation Model for Noisy and Asynchronous Data: an Application to High-Frequency Covariance Dynamics

Giuseppe Buccheri[1], Giacomo Bormetti[2], Fulvio Corsi[3,4], and Fabrizio Lillo[2,5]

[1]University of Rome Tor Vergata, Italy

[2]University of Bologna, Italy

[3]University of Pisa, Italy

[4]City University of London, UK

[5]CADS, Human Technopole, Milan, Italy

Online appendix

1 Model extensions

In this section we describe several extensions of the model described in Section (2.2) of the paper. The first is a modification to the filtering algorithm of Section (2.2) aimed to provide robust covariance forecasts in the presence of jumps. We complement the results of the Monte-Carlo and empirical sections with those obtained by applying the alternative methodology. This also serves as a robustness check for the results obtained in the paper. We also discuss some alternative parameterizations of the covariances. The first two are aimed to facilitate the estimation of the model at high dimensions, whereas the latter accommodates for deterministic intraday volatility patterns.

1.1 A Robust update scheme

It is well-known that high-frequency prices are characterized by the presence of jumps. The filtering algorithm based on the Gaussian density is sensitive to outliers and might provide inaccurate forecasts when a jump occurs. To achieve robustness to fat-tails, we modify the Gaussian algorithm as follows. Instead of the normal conditional density in Eq. (11) of the paper, we consider the following Student-t conditional density:

$$\begin{aligned}\log p_\nu(Y_t|f_t,\mathcal{F}_{t-1},\Theta) = &\log\Gamma\left(\frac{\nu+n_t}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{n_t}{2}\log[(\nu-2)\pi] - \frac{1}{2}\log|F_t| \\ &- \frac{\nu+n_t}{2}\log\left(1+\frac{v_t'F_t^{-1}v_t}{\nu-2}\right)\end{aligned} \tag{1}$$

where v_t and F_t are defined as in Section (2). Inspired by Harvey and Luati (2014) and Buccheri et al. (2019), we re-write the update in Eq. (12) of the paper as:

$$a_t = a_{t-1} + K_{t-1}v_{t-1}^{(\mathrm{r})} \tag{2}$$

where $v_t^{(\mathrm{r})} = \frac{v_t(\nu+n_t)}{\nu-2+v_t'F_t^{-1}v_t}$ is proportional to the score of the Student-t density in Eq. (1) computed with respect to a_t. Furthermore, we re-write the score in Eq. (13) of the paper as:

$$\nabla_t = -\frac{1}{2}\left[\dot{F}_t'(\mathbb{I}_{n_t}\otimes F_t^{-1})\mathrm{vec}(\mathbb{I}_{n_t} - v_t^{(\mathrm{r})}v_t'F_t^{-1}) + 2\dot{v}_t'F_t^{-1}v_t^{(\mathrm{r})}\right] \tag{3}$$

To understand the rationale behind the new filtering algorithm, observe that the prediction error v_t coincides, up to a normalization, with the score of the Gaussian density in Eq. (11) of the paper computed with respect to a_t (see Buccheri et al. (2019) for a detailed discussion on the relation between the score and the Kalman filter recursions). We thus replace the Gaussian score v_t in the update of the latent efficient price with the score $v_t^{(\mathrm{r})}$ of the Student-t density in Eq. (1). Observe also that the expression in Eq. (3) coincides with the score of the Student-t density in Eq. (1) computed with respect to the time-varying parameter vector f_t.

The new filtering recursions, being based on the scores of the Student-t density, provide robust estimates of both the latent efficient prices and the covariances. If ν goes to infinity, $v_t^{(\mathrm{r})}$ reduces to v_t and we recover the filtering algorithm of Section (2.2) in the paper. In contrast, when ν is finite, the two filters behave differently. If an outlier occurs, the prediction error v_t is large, and therefore outliers have a large impact in the standard filtering algorithm described in the paper. On the other hand, due to the nonlinear dependence

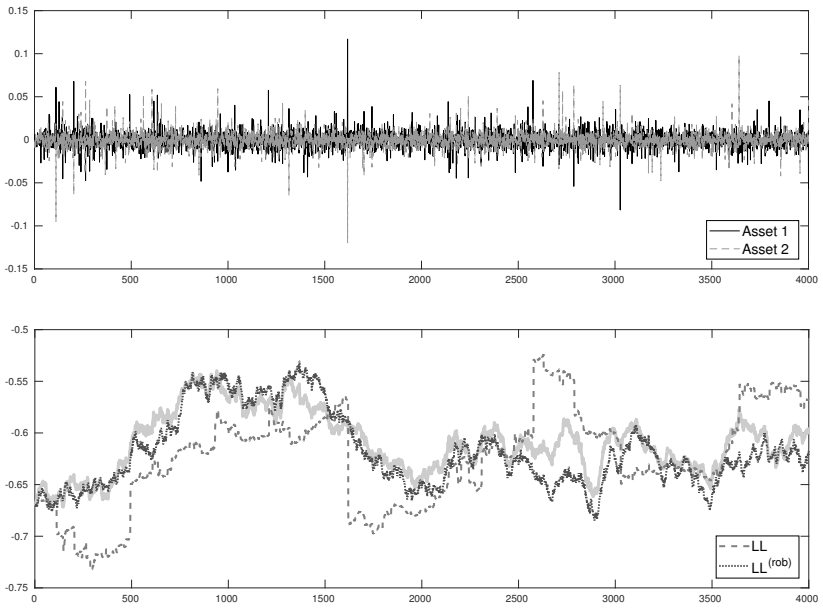

Figure 1: Up: log-returns simulated from a Student-t conditional density with $\nu = 3$ degrees of freedom. Bottom: simulated correlation (grey line), filtered correlation obtained from the local-level model with Gaussian score (dashed line), filtered correlation obtained from the local-level model with the robust update scheme (dotted line).

on v_t, $v_t^{(\mathrm{r})}$ remains finite when Y_t is large. Thus, in the new filtering algorithm, outliers have a little impact on the forecast of the time-varying parameters.

The proposed method only requires a minor modification to the algorithm described in Section (2.2). Indeed, one only needs to replace Eq. (12), (13) of the paper with Eq. (2), (3), respectively, and maximize with respect to the Student-t density in Eq. (1). However, it has the advantage of performing much better in the presence of fat-tails. Figure (1) shows one simulation of a bivariate t-GAS model (Creal et al. 2011) based on a Student-t conditional density with $\nu = 3$ degrees of freedom. As expected, the correlation filtered by the Gaussian local-level model is significantly affected by the presence of outliers. In contrast, the filtering algorithm based on the robust scores provides very accurate estimates of the simulated correlation. A similar result is obtained when comparing the estimates of the variances or, if observations are noisy, when comparing the filtered log-prices.

1.1.1 Additional Monte-Carlo results

In this section we perform the same Monte-Carlo experiment of Section (3.3) in the paper, but include the results coming from the estimation of the local-level model with robust update scheme described above. Figure (2) shows the out-of-sample average Frobenius losses reported in the paper, with in addition the average loss of the robust local-level model. The local-level model with Student-t scores behaves differently compared to the Gaussian local-level. On the figure on the left, we change the degree of sparsity of the data by varying the probability of missing values λ. We see that, for small λ, the performance of the robust model is much closer to that of the t-GAS, confirming that the proposed approach significantly improves over the

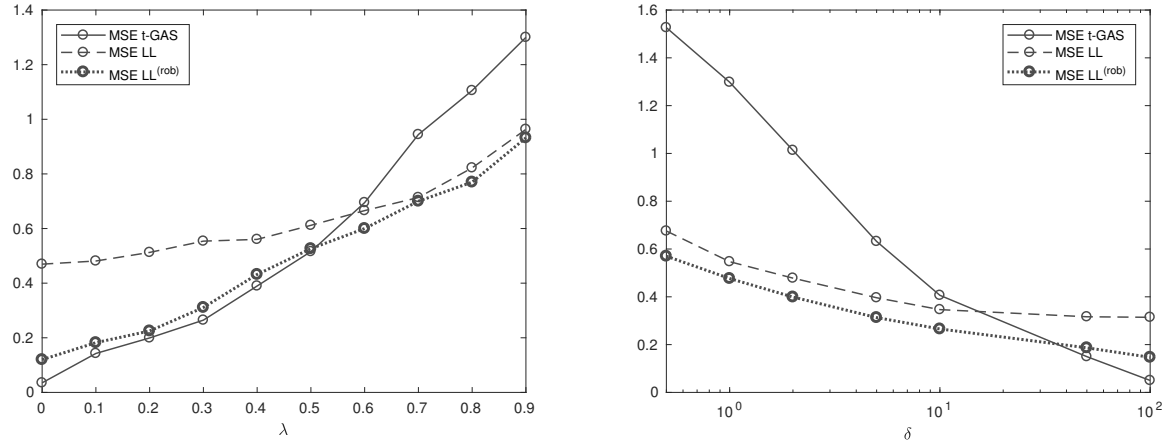

Figure 2: Left: average out-of-sample Frobenius losses, as a function of the average probability of missing values λ, of the t-GAS, the local-level and the local-level with robust update scheme. The DGP is a t-GAS with $\nu = 3$. Right: average out-of-sample Frobenius losses of the same models as a function of the signal-to-noise ratio δ. The DGP is a t-GAS with $\nu = 3$ contaminated by a Student-t distributed measurement error with $\nu_{err} = 3$ and we set $\lambda = 0$.

Gaussian algorithm when observations have fat-tails. On the right, we change the level of noise by varying the signal-to-noise ratio δ. We note that the performance of the robust local-level is closer to that of the t-GAS when δ is large, i.e. when estimation errors are small. When δ decreases, it rapidly improves over the t-GAS, in a similar fashion to the standard local-level model.

1.1.2 Additional empirical results

We denote by $\text{LL}_h^{(\text{rob})}$ and $\text{LL}_e^{(\text{rob})}$ the two robust local-level specifications with hyperspherical coordinates and equicorrelations, respectively. Table (1) shows the average and the standard deviation of the parameters estimated by the four local-level models and the t-GAS. We note that the parameter ν estimated by the $\text{LL}_h^{(\text{rob})}$ and $\text{LL}_e^{(\text{rob})}$ models is significantly lower than that of the t-GAS. Such difference is most likely due to data reduction and microstructure noise, which both affect the inference of the t-GAS parameters. It thus implies a higher degree of robustness to jumps for our local-level specifications.

Parameter	t-GAS	LL_h	$\mathrm{LL}_h^{(\mathrm{rob})}$	LL_e	$\mathrm{LL}_e^{(\mathrm{rob})}$
A^{h}	-	0.0323	0.0496	0.0179	0.0274
	-	(0.0137)	(0.0179)	(0.0117)	(0.0158)
A^{d}	0.1515	0.0226	0.0312	0.0210	0.0290
	(0.0373)	(0.0088)	(0.0116)	(0.0093)	(0.0115)
A^{r}	0.0232	0.0048	0.0053	0.0057	0.0045
	(0.0109)	(0.0018)	(0.0038)	(0.0029)	(0.0038)
ν	13.555	-	9.6123	-	8.2604
	(4.2204)	-	(2.3327)	-	(2.1200)

Table 1: Average of maximum likelihood estimates of parameters A^{h}, A^{d}, A^{r}, ν of t-GAS, LL_h, $\mathrm{LL}_h^{(\mathrm{rob})}$, LL_e and $\mathrm{LL}_e^{(\mathrm{rob})}$ models. The averages are computed over the whole sample of 251 business days. Standard deviations are indicated in parenthesis.

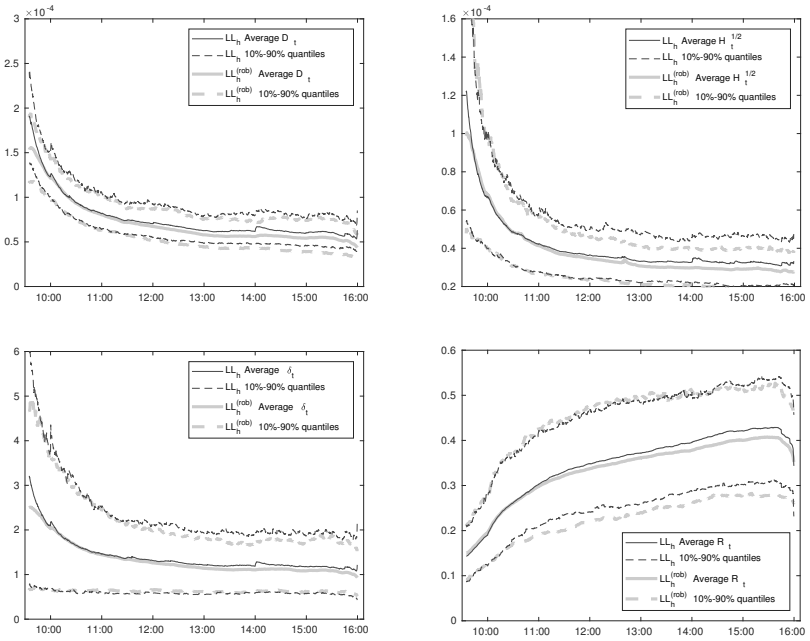

Figure 3: Average intraday patterns of the estimated LL_h and $\mathrm{LL}_h^{(\mathrm{rob})}$ time-varying parameters D_t, $H_t^{1/2}$, δ_t and R_t. For each $t = 1, \ldots, 23400$, we report 10% and 90% quantiles of the distribution (over days) of averages of D_t, $H_t^{1/2}$, δ_t and R_t (over assets or couples of assets).

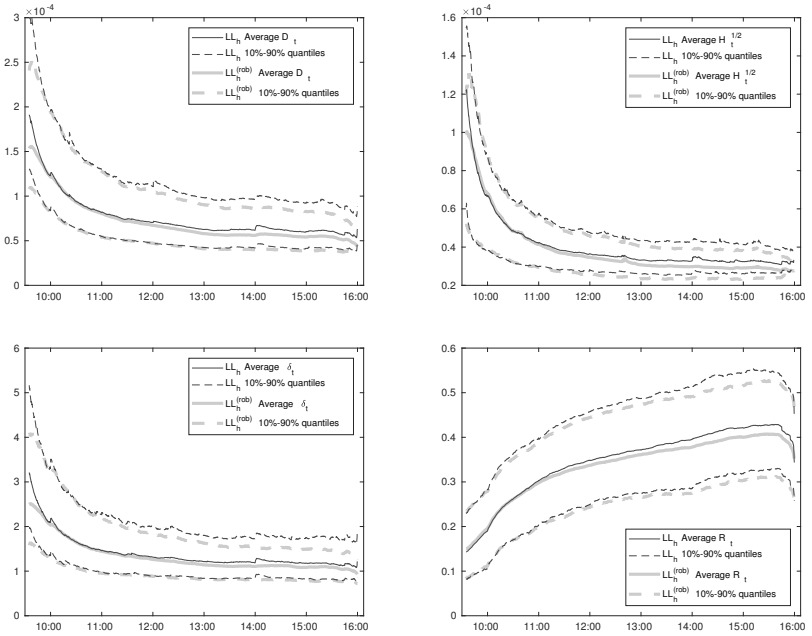

Figure 4: Average intraday patterns of the estimated LL_h and $\mathrm{LL}_h^{(\mathrm{rob})}$ time-varying parameters D_t, $H_t^{1/2}$, δ_t and R_t. For each $t = 1, \ldots, 23400$, we report 10% and 90% quantiles of the distribution (over assets or couples of assets) of averages of D_t, $H_t^{1/2}$, δ_t and R_t (over days).

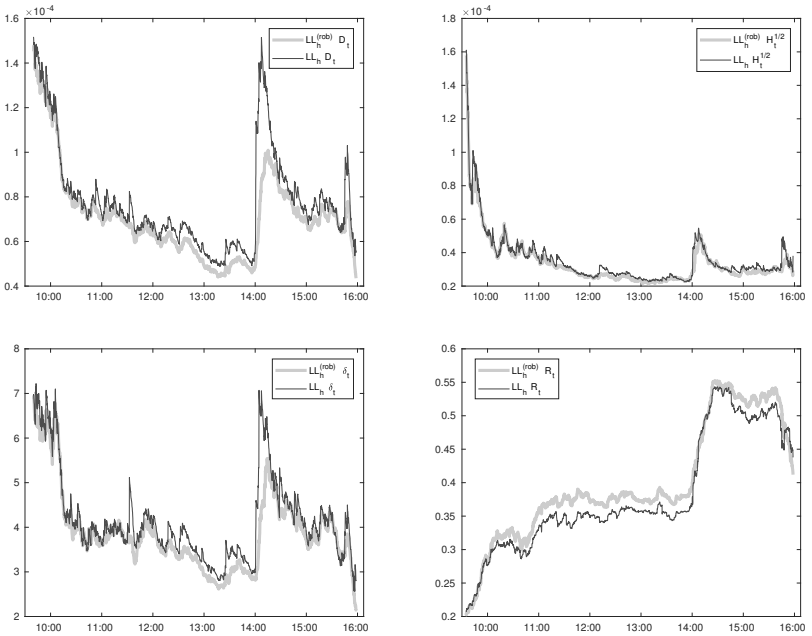

Figure 5: For both LL_h and $\mathrm{LL}_h^{(\mathrm{rob})}$, we plot $\tilde{d}_t^j$, $\tilde{h}_t^j$, $\tilde{\delta}_t^j$, $\tilde{\rho}_t^j$ computed for $j = 82$, corresponding to 30-04-2014

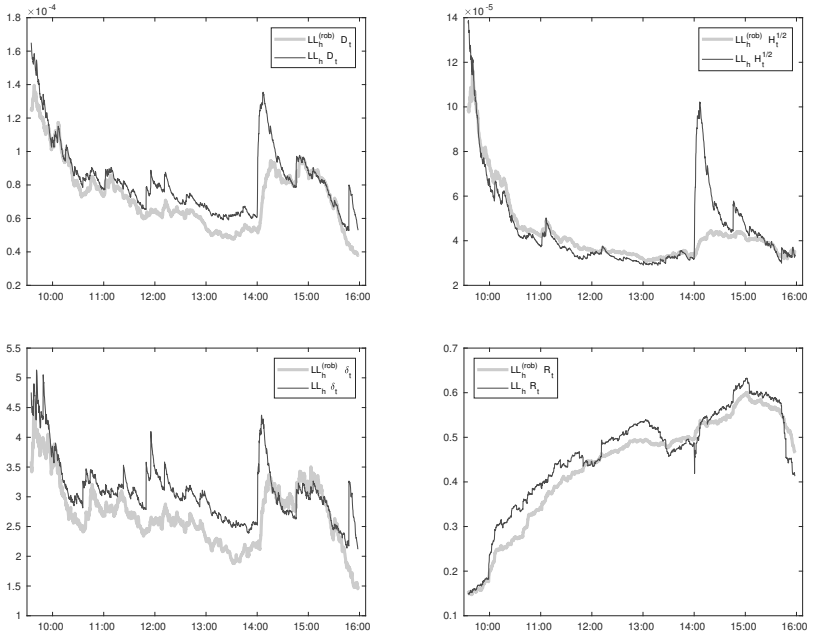

Figure 6: For both LL_h and $\text{LL}_h^{(\text{rob})}$, we plot $\tilde{d}_t^j$, $\tilde{h}_t^j$, $\tilde{\delta}_t^j$, $\tilde{\rho}_t^j$ computed for $j = 145$, corresponding to 30-07-2014

We report in Figures (3)-(6) the average intraday patterns of LL_h and LL_h^{rob} models and their filtered estimates during FOMC days. It is interesting to note that the average intraday patterns of the $\text{LL}_h^{(\text{rob})}$ model are similar, in levels, to the average intraday patterns of the LL_h model. However, the two models behave a little bit differently in the presence of jumps and other large movements in the log-prices. This is more evident when examining the dynamics of the time-varying parameters during the two FOMC days. We see that, at 14:00, the LL_h is characterized by a steep increase in volatilities and correlations, whereas the $\text{LL}_h^{(\text{rob})}$ provides a more gradual response. This is due to the different update rule used in the two models, which are based on Gaussian and Student-t scores. Despite this difference, both models rapidly respond to the release of information, suggesting that the increase of volatilities and correlations in the standard algorithm is not due to the Gaussian assumption for the conditional density.

	LL_h	$\text{LL}_h^{(\text{rob})}$	LL_e	$\text{LL}_e^{(\text{rob})}$	t-GAS	DCC	EWMA
	Ex-post Portfolio Variance						
Avg. variance	0.5428	0.5412	0.5501	0.5353	0.5895	0.6987	0.7023
N. of days in $\mathcal{M}_{90\%}$	170	181	165	185	98	23	0
Avg. p-value	0.6599	0.7118	0.6238	0.7300	0.2621	0.0061	0.0000
	99% VaR						
Avg. coverage	0.0198	0.0166	0.0202	0.0156	0.0016	0.4976	0.4812
N. of rejections of $\mathcal{H}_0^K$	67	56	82	60	198	250	250
Avg. p-value	0.2681	0.3161	0.2151	0.3047	0.0236	0.000	0.000
	95% VaR						
Avg. coverage	0.0566	0.0537	0.0605	0.0532	0.0038	0.4976	0.4923
N. of rejections of $\mathcal{H}_0^K$	48	33	52	45	239	250	250
Avg. p-value	0.2885	0.3446	0.2885	0.2948	0.0025	0.000	0.0000

Table 2: Top: Average variance ($\times 10^5$) of 1-minute GMV portfolios constructed through out-of-sample covariance forecasts of LL_h, $\text{LL}_h^{(\text{rob})}$, LL_e, $\text{LL}_e^{(\text{rob})}$, t-GAS and DCC models. We report the number of days in which the models are included in $\mathcal{M}_{90\%}$ and the average of the p-values of the MCS test. Middle: Average coverage of 1-minute VaR's at 99% CL obtained from the covariance forecasts of GMV portfolios. We report the number of rejections of the Kupiec test with a significance level of 5% and the average p-value. Down: As before, but VaR's are computed at 95% CL.

Table (2) shows the results of the out-of-sample experiment performed in Section (4.4) in the paper, with in addition the performance of the two robust local-level models. We first note that the portfolios based on $\text{LL}_h^{(\text{rob})}$ and $\text{LL}_e^{(\text{rob})}$ models have slightly lower ex-post variance compared to LL_h and LL_e portfolios, and are included in the MCS a larger number of times. When applied to high-frequency data, the robust update scheme thus improves over the standard Gaussian algorithm of Section (1.1). Similarly, we see that the $\text{LL}_h^{(\text{rob})}$ and $\text{LL}_e^{(\text{rob})}$ models provide average coverage rates that are closer to the nominal confidence level. This is not surprising, as the two Gaussian specifications naturally tend to underestimate the risk.

1.2 Reduced rank correlations

If n is large, the dimension of the vector f_t of time-varying parameters becomes huge and this can make the estimation process computationally demanding. In this case, one can use a reduced rank decomposition of the correlation matrix R_t in order to lower the number of time-varying parameters. We write R_t as:

$$R_t = Z_t' Z_t \tag{4}$$

where Z_t is now an $r \times n$ matrix with $r < n$. The number of free parameters of the reduced rank correlation matrix is $(r-1)(n-r/2)$. In the hyperspherical parameterization, the nonzero entries z_{ij} of Z_t read (see

e.g. Rapisarda et al. 2007):

$$z_{ij} = \begin{cases} 1 & i = j = 1 \\ \cos\theta_{1j} & i = 1, j > 1 \\ \cos\theta_{ij} \prod_{k=1}^{i-1} \sin\theta_{kj} & 1 < i < \min(r, j) \\ \prod_{k=1}^{r-1} \sin\theta_{kj} & i = \min(r, j) \end{cases} \tag{5}$$

Thus, the matrix Z_t has the following structure:

$$Z_t = \begin{pmatrix} 1 & c_{12} & c_{13} & \dots & c_{1r} & c_{1r+1} & \dots & c_{1n} \\ 0 & s_{12} & c_{23}s_{13} & \dots & c_{2r}s_{1r} & c_{2r+1}s_{1r+1} & \dots & c_{2n}s_{1n} \\ 0 & 0 & s_{23}s_{13} & \dots & c_{3r}s_{2r}s_{1r} & c_{3r+1}s_{2r+1}s_{1r+1} & \dots & c_{3n}s_{2n}s_{1n} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \prod_{k=1}^{r-1} s_{kr} & \prod_{k=1}^{r-1} s_{kr+1} & \dots & \prod_{k=1}^{r-1} s_{kn} \end{pmatrix}$$

where we neglect the time subscript t for ease of notation. The first r columns are parameterized by the $r(r-1)/2$ angles $\theta_{12}, \theta_{13}, \dots, \theta_{r-1r}$, while each of the remaining $n-r$ columns are parameterized by the $r-1$ angles $\theta_{1k}, \theta_{2k}, \dots, \theta_{r-1,k}$, $k = r+1, \dots, n$. The total number of parameters is therefore $k_r = r(r-1)/2 + (n-r)(r-1)$, which is equal to the number of free parameters of an $n \times n$ correlation matrix of rank r. Eq. (25) still holds, provided that the derivative of Z_t with respect to f_t is computed as:

$$\frac{\partial z_{ij}}{\partial \theta_{lm}} = \begin{cases} 0 & i > j, j \neq m, l \geq m, l > i \\ -z_{ij} \tan\theta_{ij} & i \neq \min(j, r), l = i \\ \frac{z_{ij}}{\tan\theta_{ij}} & l < i \end{cases} \tag{6}$$

1.3 Common factor structure in the covariances

Another alternative method to reduce the number of time-varying parameters is to assume a factor structure, in a similar fashion to Creal et al. (2011). We can write:

$$\log(\text{diag}(H_t)) = a_H + M_H f_t \tag{7}$$

$$\log(\text{diag}(D_t^2)) = a_D + M_D f_t \tag{8}$$

$$\phi_t = a_\phi + M_\phi f_t \tag{9}$$

where f_t is a vector of lower dimension and a_H, a_D, a_ϕ, M_H, M_D, M_ϕ are vectors and factor loading matrices of appropriate dimension and subject to identification restrictions. The structure of the filter remains the same, except that one needs to take into account the linear transformation (7), (8), (9) when computing the Jacobian. As discussed by Creal et al. (2011), the existence of a factor structure allows to decrease the number of time-varying parameters, but increases the number of static parameters in the maximum likelihood estimation.

1.4 Deterministic volatility patterns

The proposed approach allows the modeling of intraday deterministic patterns in D_t and H_t. We write the new matrices $D_t^{(s)}$ and $H_t^{(s)}$ as:

$$D_t^{(s)} = s_t D_t \tag{10}$$

$$H_t^{(s)} = s_t^2 H_t \tag{11}$$

where s_t is a common deterministic trend. The latter is set as in Andersen et al. (2012) and Bollerslev et al. (2016), namely:

$$s_t = c + ae^{-\alpha t} + be^{-\beta(1-t)} \tag{12}$$

where $a = 0.75$, $b = 0.25$, $c = 0.88929198$ and $\alpha = \beta = 10$. The filtering recursions remain unaltered, provided that one replaces D_t and H_t with $D_t^{(s)}$ and $H_t^{(s)}$, respectively, and multiplies their derivatives by s_t.

2 Computation of $\dot{v}_t$ and $\dot{F}_t$

We first introduce the notation. The $n \times n$ identity matrix is denoted as $\mathbb{I}_n$. We use $\otimes$ to denote the Kronecker product between two matrices. The operator vec$[\cdot]$, applied to an $m \times n$ matrix A, stacks the columns of A into an $mn \times 1$ vector. The operator diag$[\cdot]$ applied to an $n \times n$ matrix stacks its diagonal elements into an $n \times 1$ vector. When applied to an $n \times 1$ vector, it gives a diagonal $n \times n$ matrix with the elements of the vector in the main diagonal. We also introduce the commutation matrix C_{mn}, i.e. the $mn \times mn$ matrix such that $C_{mn}\text{vec}A = \text{vec}A'$ for every $m \times n$ matrix A. The derivative of an $m \times n$ matrix function $F(X)$ with respect to the $p \times q$ matrix X is defined as in Abadir and Magnus (2005), i.e. as the $mn \times pq$ matrix computed as $\partial\text{vec}(F(X))/\partial\text{vec}(X)'$.

Let us define $a_t = \text{E}_{t-1}[X_t]$ and $P_t = \text{Cov}_{t-1}[X_t]$. Due to asynchronous trading, Y_t is a vector with $n_t \leq n$ components. We define the $n_t \times n$ selection matrix Γ_t with ones in the columns corresponding to observed prices. We also define the prediction error $v_t = Y_t - \Gamma_t a_t$ and the prediction error variance $F_t = \Gamma_t(P_t + H_t)\Gamma_t'$. The Kalman filter recursions for the local-level model in Eq. (5), (6) of the paper are given by:

$$a_t = a_{t-1} + K_{t-1}v_{t-1} \qquad P_t = P_{t-1}(\mathbb{I}_n - K_{t-1}\Gamma_{t-1})' + Q_t \tag{13}$$

where $K_{t-1} = P_{t-1}\Gamma_{t-1}'F_{t-1}^{-1}$. If at time $t-1$ all the observations are missing, we set $a_t = a_{t-1}$ and $P_t = P_{t-1} + Q_t$, as discussed by Durbin and Koopman (2012). It is convenient to introduce the auxiliary vector of time-varying parameters:

$$\tilde{f}_t = \begin{pmatrix} \text{diag}(H_t) \\ \text{diag}(D_t^2) \\ \phi_t \end{pmatrix} \tag{14}$$

The latter is related to f_t by the following link-function:

$$\tilde{f}_t = \mathcal{L}(f_t) = \begin{pmatrix} \exp f_t^{(1)} \\ \vdots \\ \exp f_t^{(2n)} \\ f_t^{(2n+1)} \\ \vdots \\ f_t^{(k)} \end{pmatrix} \tag{15}$$

The Jacobian of the transformation is:

$$J_{\mathcal{L}} = \left(\frac{\partial \tilde{f}_t}{\partial f_t{}'} \right) = \begin{pmatrix} H_t & 0_{n\times n} & 0_{n\times q} \\ 0_{n\times n} & D_t^2 & 0_{n\times q} \\ 0_{q\times n} & 0_{q\times n} & \mathbb{I}_q \end{pmatrix} \tag{16}$$

Note that, using the chain rule, ∇_t and $\mathcal{I}_{t|t-1}$ can be expressed as:

$$\nabla_t = J_{\mathcal{L}} \tilde{\nabla}_t, \quad \mathcal{I}_{t|t-1} = J_{\mathcal{L}} \tilde{\mathcal{I}}_{t|t-1} J_{\mathcal{L}} \tag{17}$$

where:

$$\tilde{\nabla}_t = \left[\frac{\partial \log p(Y_t | \tilde{f}_t, \mathcal{F}_{t-1}, \Theta)}{\partial \tilde{f}_t'} \right]', \quad \tilde{\mathcal{I}}_{t|t-1} = \mathrm{E}[\tilde{\nabla}_t \tilde{\nabla}_t'] \tag{18}$$

which can be computed as in Eq. (13), (14) of the paper, but deriving with respect to $\tilde{f}_t$ rather than f_t. We thus focus on $\dot{v}_t = \partial v_t / \partial \tilde{f}_t'$ and $\dot{F}_t = \partial \mathrm{vec}(F_t) / \partial \tilde{f}_t'$. As a particular case of the general recursions appearing in Delle Monache et al. (2019), we obtain:

$$\dot{v}_t = 0 \tag{19}$$

$$\dot{F}_t = (\Gamma_t \otimes \Gamma_t)(\dot{P}_t + \dot{H}_t) \tag{20}$$

where:

$$\dot{P}_t = \dot{Q}_t \tag{21}$$

$$\dot{Q}_t = [(D_t R_t \otimes \mathbb{I}_n) + (\mathbb{I}_n \otimes D_t R_t)] \dot{D}_t + (D_t \otimes D_t) \dot{R}_t \tag{22}$$

Here $\dot{H}_t = \frac{\partial \mathrm{vec}(H_t)}{\partial \tilde{f}_t'}$, $\dot{D}_t = \frac{\partial \mathrm{vec}(D_t)}{\partial \tilde{f}_t'}$ are $n^2 \times k$ matrices given by:

$$\dot{H}_t = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ & & \vdots & & & \vdots & \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ & & \vdots & & & \vdots & \\ & & \vdots & & & \vdots & \\ \underbrace{0 & \dots & 0 & 1}_{n} & \underbrace{0 & \dots & 0}_{n+q} \end{pmatrix} \tag{23}$$

$$\dot{D}_t = \frac{1}{2}\begin{pmatrix} 0 & \dots & 0 & \frac{1}{D_{t,11}} & 0 & \dots & 0 & 0 & \dots & 0 \\ & \vdots & & & \vdots & & & & \vdots & \\ 0 & \dots & 0 & 0 & \frac{1}{D_{t,22}} & \dots & 0 & 0 & \dots & 0 \\ & \vdots & & & \vdots & & & & \vdots & \\ & \vdots & & & \vdots & & & & \vdots & \\ \underbrace{0 \quad \dots \quad 0}_{n} & & & \underbrace{0 \quad \dots\dots \quad \frac{1}{D_{t,nn}}}_{n} & & & & \underbrace{0 \quad \dots \quad 0}_{q} & & \end{pmatrix} \tag{24}$$

The computation of $\dot{R}_t$ depends on the parameterization. We distinguish the case where the hyperspherical coordinates are used and the case where the equicorrelation parameterization is used. In the first case, we have:

$$\dot{R}_t = [(Z_t' \otimes \mathbb{I}_n)C_{nn} + (\mathbb{I}_n \otimes Z_t')]\dot{Z}_t \tag{25}$$

The derivative of the element $Z_{ij,t}$ with respect to the hyperspherical angle $\theta_{lm,t}$ is given by:

$$\frac{\partial Z_{ij}}{\partial \theta_{lm}} = \begin{cases} 0 & i > j, j \neq m, l \geq m, l > i \\ -Z_{ij}\tan\theta_{ij} & i < j, l = i \\ \frac{Z_{ij}}{\tan\theta_{ij}} & i \leq j, l < i \end{cases} \tag{26}$$

Note that the time index was suppressed for ease of notation. In the second case we have:

$$\dot{R}_t = [0_{n^2 \times 2n}, \dot{\rho}_t \text{vec}(-\mathbb{I}_n + \mathcal{J}_n)] \tag{27}$$

where:

$$\dot{\rho}_t = \frac{1}{2}\left(1 + \frac{1}{n-1}\right)\frac{1}{\cosh^2\theta_t} \tag{28}$$

References

Abadir, K., Magnus, J., 2005. Matrix Algebra. Econometric Exercises, Cambridge University Press.

Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. Journal of Econometrics 169, 75 – 93. Recent Advances in Panel Data, Nonlinear and Nonparametric Models: A Festschrift in Honor of Peter C.B. Phillips.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics 192, 1 – 18.

Buccheri, G., Bormetti, G., Corsi, F., Lillo, F., 2019. Filtering and smoothing with score-driven models. Working paper .

Creal, D., Koopman, S.J., Lucas, A., 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. Journal of Business & Economic Statistics 29, 552–563.

Delle Monache, D., Petrella, I., Venditti, F., 2019. Price dividend ratio and long-run stock returns: a score driven state space model. Working paper .

Durbin, J., Koopman, S., 2012. Time Series Analysis by State Space Methods: Second Edition. Oxford Statistical Science Series, OUP Oxford.

Harvey, A., Luati, A., 2014. Filtering with heavy tails. Journal of the American Statistical Association 109, 1112–1122.

Rapisarda, F., Brigo, D., Mercurio, F., 2007. Parameterizing correlations: a geometric interpretation. IMA Journal of Management Mathematics 18, 55–73.