

# A Bayesian High-Frequency Estimator of

tion and similar papers at [core.ac.uk](http://core.ac.uk)

bro  
provided by

STEFANO PELUSO

*University of Lugano*

FULVIO CORSI

*Ca' Foscari University of Venice and City University London*

ANTONIETTA MIRA

*University of Lugano*

## ABSTRACT

A multivariate positive definite estimator of the integrated covariance matrix of noisy and asynchronously observed asset returns is proposed. We adopt a Bayesian Dynamic Linear Model where microstructure noise is interpreted as measurement error, and asynchronous trading as missing observations in an otherwise synchronous series. Missing observations are treated as any other parameter, as typical in a Bayesian framework. An augmented Gibbs algorithm is used since all full conditionals are available and its convergence and robustness are discussed. A realistic simulation study compares our estimator with existing alternatives, under different liquidity and microstructure noise conditions. The results suggest that our estimator is superior in terms of RMSE particularly under severe conditions, such as portfolios of assets with heterogeneous liquidity and high level of microstructure noise. The application to the empirical dataset of ten tick-by-tick stock price series confirms the simulation results. (JEL: C13; C51; C52; C58)

**KEYWORDS:** asynchronicity; data augmentation; Gibbs sampler; missing observations; realized covariance

---

The authors would like to thank Prof. Francesco Audrino for fruitful discussions and for the data used in the empirical application. We acknowledge the financial support of the Swiss National Science Foundation grant number 200021\_130394. Address correspondence to Stefano Peluso, University of Lugano, Swiss Finance Institute, Via Giuseppe Buffi 13, CH-6904 Lugano, Switzerland, or e-mail: [stefano.peluso@usi.ch](mailto:stefano.peluso@usi.ch).

doi:10.1093/jfinec/mbu017 Advanced Access publication July 2, 2014

© The Author, 2014. Published by Oxford University Press. All rights reserved.

For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Available intra-day prices can be used to improve the estimation of the covariance among several financial assets, so that even the covariation of asset prices within the day can be included in the inferential process. The two main concerns when dealing with several time series of Ultra-High Frequency (UHF) prices is that they are observed at different trading times and with microstructure noise. The first problem is known as asynchronicity of UHF asset prices and its effect on the estimation of the covariance was first identified by Epps (1979), who found that the correlation is biased toward zero as the sampling frequency increases when the realized covariance is computed on a regularized grid. The Realized Covariance estimator proposed by Hayashi and Yoshida (2005) (HY) is unaffected by this asynchronicity problem. The second feature of UHF asset prices (and in general of asset prices) is that they present a component due to the microstructure noise. In the literature, there are several proposed corrections to HY that make it robust to this noise. For example, Voev and Lunde (2007), Bibinger (2011), based on a multiscale subsampling correction of HY that improves the convergence properties of the estimator proposed by Palandri (2006) and the Two-Scales Realized Covariance (TSRC) estimator of Zhang (2011). A consistent QMLE-type estimator of the high-frequency covariance of two assets observed asynchronously with microstructure noise was introduced by Ait-Sahalia, Fan, and Xiu (2010).

When instead of the covariance between two assets we consider the covariance matrix of several assets, all estimators mentioned above, which successfully deal with the asynchronicity and noise in the bivariate case, do not guarantee a positive semi-definite estimator in the multivariate setting. Among the few works in the literature proposing a multivariate covariance estimator that preserves positivity, the most popular is the Multivariate Realized Kernel of Barndorff-Nielsen *et al.* (2011). They suggest to synchronize the high-frequency prices using a Refresh Time Scheme combined with a multivariate realized kernel to provide a consistent and positive semi-definite estimator of the covariance matrix. A drawback of their methodology is that the synchronization of the time series with the Refresh Time Scheme can cause a large loss of information if the involved assets are traded with very different liquidity. Furthermore, they need to tune a bandwidth parameter. Finally, their results are valid only asymptotically, when the mesh of trading intervals converges to zero and the number of observations goes to infinity. The Composite Kernel has been recently proposed in Lunde, Shephard, and Sheppard (2012) to solve these important issues, but it is again a bivariate estimator, then requiring some kind of projection to retain the positivity in a multivariate framework.

Motivated by the need of an unbiased, positive-semidefinite estimator of the multivariate integrated covariance matrix of asynchronous and noisy UHF asset prices, we cast the problem into a Bayesian framework and consider the asynchronous times series as synchronous series with missing observations, that are treated as any other parameter of the problem as typically done in a Bayesian framework. The flexibility of the Dynamic Linear Model we adopt allows us to easily treat the true latent price process, not affected by microstructure noise,

as additional parameters to be estimated. Since the joint posterior distribution of the parameters and the missing values is not standard, we use Markov chain Monte Carlo algorithms to obtain samples from it. In particular, the posterior covariance matrix is sampled through a Gibbs sampler from an Inverse Wishart distribution which naturally preserves its positivity. The simulation results show superior performance of the proposed estimator compared to the benchmark one, particularly for a high number of assets and dispersed missing probabilities. These performance improvements are mainly due to the high robustness of the proposed estimator to market microstructure noise and to its ability to fully incorporate all the relevant information available from the whole set of assets. In fact, one of the main advantages of our approach is to be able to reconstruct the unobserved dynamics for each asset by using the information contained in the observed dynamics of all the other assets. This type of multivariate signal extraction is particularly convenient for the less liquid series which can benefit more from the information contained in the dynamics of the more liquid assets.

We confirm the simulation results by applying our methodology to an empirical dataset composed of ten tick-by-tick series of stock prices from January to November 2007. The empirical analysis shows that the covariance matrix of the proposed estimator produces a vector of standardized returns closer to a multivariate standard normal than the ones obtained with alternative methods.

Recently, the quest for a realized covariance estimator which guarantees positive definiteness and robustness to both market microstructure and asynchronicity has spurred several research papers employing different methodologies. Mancino and Sanfelici (2011) and Park and Linton (2012) present a covariance estimators based on the Fourier domain techniques introduced by Malliavin and Mancino (2009) which avoid explicit time synchronizations.

A more similar approach to ours, but considered from the classical frequentist viewpoint, has been recently investigated in the concurrent works of Corsi, Peluso, and Audrino (2014) and Shephard and Xiu (2012). In the latter, the focus is more on the derivation of the theoretical asymptotic properties of the estimator showing that it converges at the optimal rate, while in Corsi, Peluso, and Audrino (2014) the reconstruction of the missing values from the noisy observed prices is performed by a Kalman smoother and Expectation Maximization (KEM) algorithm. Although similarities between EM-type algorithms and stochastic algorithms like the Gibbs sampler exists (see Tanner and Wong, 1987 and Meng and van Dyk, 1997), differently from the KEM estimator of Corsi, Peluso, and Audrino (2014), our Bayesian approach is more flexible to generalizations of the dynamics of the latent price process and to different distribution of microstructure noise (i.e., we can extend our framework to non-Gaussian latent log returns, to autocorrelated microstructure error or to other model features we believe empirically consistent). Furthermore, in our Bayesian estimator there is no additional computational cost in dealing with nondiagonal covariance matrices of the microstructure noise. Finally, the KEM methodology finds the MLE

of the integrated covariance matrix, while the Bayesian estimator produces samples, approximating the entire posterior distribution of any function of interest of the integrated covariance matrix. This could be of interest when some functional of the integrated covariance matrix is in use,<sup>1</sup> or when, for risk management purposes, having the quantiles of the distribution of the integrated covariance allows a precise quantification of the uncertainty in the covariance estimation.

We present our methodology in Section 1: in Section 1.1 we introduce the case of asynchronicity without noise, which is extended in Section 1.2 to noisy observations, where we propose the augmented Gibbs sampler. In Section 2, Monte Carlo simulation experiments are performed to compare our Bayesian estimator with alternative estimators presented in Section 2.1, with a simulation design introduced in Section 2.2 that reflect realist scenarios, in a bivariate (Section 2.3) and multivariate (Section 2.4) setting. Section 3 contains the application and comparison of the methodologies to the empirical dataset of ten stocks, with some Gaussianity tests and with a portfolio allocation exercise that measure the economic relevance of the alternative estimation procedures. We check convergence properties of the Gibbs sampler and its robustness to strong microstructure noise and high missings percentages in Sections 4.1 and 4.2, respectively. Section 5 concludes.

## 1 METHODOLOGY

### 1.1 Asynchronicity without Noise

We start by considering the model with asynchronous prices that are observed at different times within the day, but without being contaminated by microstructure noise. The simplified model is

$$d\mathbf{X}_t = \mu(\mathbf{X}_t, \theta)dt + \sqrt{\Sigma(\mathbf{X}_t, \theta)}d\mathbf{W}_t, \tag{1}$$

where, for some compact  $\Theta \subseteq \mathbb{R}^k$ ,  $\theta$  the unknown parameter vector,  $\mu : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^d$ ,  $\Sigma : \mathbf{X}^{-1}(\mathbb{R}^d) \times \Theta \rightarrow \mathbb{R}^{d \times d, +}$ ,  $\mathbf{X}_t$  is a  $d$ -dimensional log-price diffusion process,  $\mathbf{W}_t$  is a  $d$ -dimensional Brownian motion and  $\mathbb{R}^{d \times d, +} := \{M \in \mathbb{R}^{d \times d} : M > 0, \text{ symmetric}\}$ , that is the space of square,  $d \times d$  positive definite symmetric matrices. We assume that the drift and the diffusion functions satisfy the Lipschitz condition

$$\|\mu(x, \theta) - \mu(y, \theta)\| + \|\sqrt{\Sigma(x, \theta)} - \sqrt{\Sigma(y, \theta)}\| \leq C\|x - y\| \tag{2}$$

for some positive constant  $C$  and with  $\|\cdot\|$  indicating the Euclidean norm. We need this assumption to ensure the existence of a strong (unique), square integrable solution to (1) (Oksendal, 2002).

---

<sup>1</sup>As in Chiriac and Voev (2010) who focused on the *vech* of the upper triangular matrix in the Cholesky decomposition of the integrated multivariate covariance.

We work with the discretized version (time instants between adjacent observations  $\Delta t=1$ ) of (1) with constant diffusion coefficient and zero drift, so that the inference is directly conducted on  $\Sigma$ . Assuming a deterministic drift equal to zero is reasonable since a bounded variation drift is negligible for inference on  $\Sigma$ . Furthermore, we recognize that the volatility is a time-varying process, and we rely on the recent result of Xiu (2010) that shows that the QMLE of the volatility of a misspecified model with constant volatility remains consistent and optimal in terms of its rate of convergence under fairly general assumptions. Let us first consider the bivariate case, for  $i=1,2$

$$X_{i,t} = X_{i,t-1} + \epsilon_{i,t} \quad \epsilon_{i,t} \sim N(0, \sigma_i^2), \tag{3}$$

with  $\text{corr}(\epsilon_{1,t}, \epsilon_{2,t}) = \rho$ ,  $X_{i,t}$  is the log-price of asset  $i$  observed at time  $t$ ,  $\sigma_i$  is its volatility and  $\rho$  is the correlation coefficient.

We define  $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ , the missing and observed parts of  $\mathbf{X}$  respectively as  $\mathbf{X}^{miss}$  and  $\mathbf{X}^{obs}$  and partition the time interval  $[1, \dots, T]$  as  $[\mathbf{t}_i^{miss}, \mathbf{t}_i^{obs}]$  accordingly. The likelihood is

$$L(\Sigma, \mathbf{X}_{1:T}^{miss} | \mathbf{X}_{1:T}^{obs}) \propto \prod_{i=1}^2 \frac{1}{(1-\rho^2)\sigma_i^2} \exp\left(-\frac{1}{2} \sum_{t \in \mathbf{t}_i^{obs}} \frac{(X_{i,t} - m_{i,t})^2}{(1-\rho^2)\sigma_i^2}\right),$$

where  $X_{i,s:t}$  indicates the log price of asset  $i$ , from time  $s$  to time  $t$ , both extremes included,  $m_{i,t} := X_{i,t-1} - \rho \frac{\sigma_i}{\sigma_{-i}} (X_{-i,t} - X_{-i,t-1})$  and  $-i$  is the other asset. We assume a Jeffrey’s uninformative prior (Jeffreys, 1946) for  $\Sigma$ ,  $p(\Sigma) \sim |\Sigma|^{-3/2}$ , but an informative prior that incorporates in the analysis prior knowledge on the problem can easily be adopted. For example, in an empirical Bayesian approach, an Inverse Wishart prior for  $\Sigma$ , with parameters  $T_0$  and  $S_0$  being respectively, the sample size and the sum of squared observed returns, ignoring the missing data, would still retain the conjugacy of the problem.

The asynchronicity (or, in other words, the presence of missing observations) complicates the form of the likelihood, but, once we condition on missing observations, we can still use standard results from multivariate normal theory to derive the full conditional for  $\Sigma$ :

$$p(\Sigma | \mathbf{X}) \propto |\Sigma|^{-(T+3)/2} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1} \sum_{t=1}^T \epsilon_t \epsilon_t')\right) \propto IW\left(\sum_{t=1}^T (\mathbf{X}_t - \mathbf{X}_{t-1})(\mathbf{X}_t - \mathbf{X}_{t-1})', T\right)$$

and find that  $\phi(X_{i,t}^{miss} | \Sigma)$ , the full conditionals of  $X_{i,t}^{miss}$ ,  $i=1,2$ , are normal:

$$X_{i,t}^{miss} | X_{i,1:t-1}, X_{-i,1:t}, \Sigma \sim N\left(m_{i,t}, (1-\rho^2)\sigma_i^2\right),$$

where the subscript  $-i$  refer to the other asset. The extension of (3) to the multivariate case is straightforward and sampling the covariance matrix from an

Inverse Wishart assures that the resulting estimate is positive definite. The generic  $d$ -dimensional discretized model is, for  $i = 1, \dots, d$ :

$$X_{i,t} = X_{i,t-1} + \epsilon_{i,t} \quad \epsilon_{i,t} \sim N(0, \sigma_i^2), \tag{4}$$

with  $\text{corr}(\epsilon_{i,t}, \epsilon_{j,t}) = \rho_{ij}$ .  $\mathbf{X}$  and  $\Sigma$  are, respectively, a  $d \times T$  and a  $d \times d$  matrix. We write the likelihood as

$$L(\Sigma, \mathbf{X}_{1:T}^{miss} | \mathbf{X}_{1:T}^{obs}) \propto \prod_{i=1}^d v_i^{-1} \exp\left(-\frac{1}{2} \sum_{t \in \mathcal{I}_i^{obs}} v_i^{-1} (X_{i,t} - \tilde{m}_{i,t})^2\right),$$

where  $v_i := \sigma_i^2 - \Sigma_i \cdot \Sigma_{-i}^{-1} \Sigma_{-i}'$ ,  $\Sigma_{-i}$  is obtained from  $\Sigma$  by dropping the row and column corresponding to asset  $i$ , while  $\Sigma_i$  is the  $i$ -th row of  $\Sigma$  without its  $i$ -th element,  $\tilde{m}_{i,t} := X_{i,t-1} + \Sigma_i \cdot \Sigma_{-i}^{-1} (\mathbf{X}_{-i,t} - \mathbf{X}_{-i,t-1})$  and  $\mathbf{X}_{-i,s:t}$  is the matrix of assets log prices  $j$ ,  $\forall j \neq i$  and from time  $s$  to time  $t$ . The Jeffrey's uninformative prior for  $\Sigma$  is now  $p(\Sigma) \sim |\Sigma|^{-(d+1)/2}$  and its full conditional is

$$p(\Sigma | \mathbf{X}) \propto |\Sigma|^{-(T+d+1)/2} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1} \sum_{t=1}^T \epsilon_t \epsilon_t')\right) \propto IW\left(\sum_{t=1}^T (\mathbf{X}_t - \mathbf{X}_{t-1})(\mathbf{X}_t - \mathbf{X}_{t-1})', T\right).$$

The full conditional of  $X_{i,t}^{miss}$ ,  $\phi(X_{i,t}^{miss} | \Sigma)$ , is easily derived as:

$$X_{i,t}^{miss} | X_{i,1:t-1}, \mathbf{X}_{-i,1:t}, \Sigma \sim N(\tilde{m}_{i,t}, v_i).$$

To sample from the posterior distribution of  $X^{miss}$  and  $\Sigma$ , we use a Gibbs sampler and iteratively sample  $\{X^{miss}, \Sigma\}$  from their full conditionals. The algorithm at each iteration thus consists of the following two steps:

1. Draw a covariance matrix  $\Sigma$  from its full conditional, that is an Inverse Wishart distribution,  $IW(S(\mathbf{X}^{miss}), T)$ , with  $S(\mathbf{X}^{miss}) = \sum_{t=1}^T (\mathbf{X}_t - \mathbf{X}_{t-1})(\mathbf{X}_t - \mathbf{X}_{t-1})'$ .  $S$  is expressed as a function of  $\mathbf{X}^{miss}$  to highlight the dependence on the imputed missing log prices.
2. Impute the missing observations,  $\mathbf{X}^{miss}$ , by drawing from  $\phi(X_{i,t}^{miss} | \Sigma)$ ,  $\forall t \in \mathcal{I}_i^{miss}$ ,  $i = 1, \dots, d$ .

### 1.2 Asynchronicity and Noise

A step toward a more realistic model is made by introducing the microstructure noise, so that our model becomes, for  $t = 1, \dots, T$

$$\mathbf{Y}_t = \mathbf{X}_t + \sqrt{\Omega} d\mathbf{B}_t \tag{5}$$

$$d\mathbf{X}_t = \mu(\mathbf{X}_t, \theta) dt + \sqrt{\Sigma(\mathbf{X}_t, \theta)} d\mathbf{W}_t, \tag{6}$$

where  $\mu$ ,  $\Sigma$ ,  $\mathbf{X}_t$ , and  $\mathbf{W}_t$  are defined as in (1),  $\Omega$  is the constant variance of the microstructure noise term,  $Y_t$  is a  $d$ -dimensional observed (with noise) log-price diffusion process,  $\mathbf{B}_t$  is a  $d$ -dimensional Brownian motion,  $\mathbf{B}_t \perp \mathbf{W}_t$ . Assume that the Lipschitz condition (2) is satisfied. The discretized version of (5) and (6) with constant diffusion coefficient is

$$\mathbf{Y}_t = \mathbf{X}_t + \eta_t, \quad \eta_t \sim N(\mathbf{0}, \Omega), \tag{7}$$

$$\mathbf{X}_t = \mathbf{X}_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(\mathbf{0}, \Sigma), \tag{8}$$

where  $\mathbf{Y}_t$  is the observed log-price,  $\mathbf{X}_t$  is considered as the true latent log-price process,  $\eta_t$  is the microstructure noise and  $\Omega$  its covariance matrix,  $\epsilon$  is the true latent return with covariance  $\Sigma$ .  $\eta_t$  and  $\epsilon_t$  are assumed independent.

This is a linear state space model, consisting in the observation Equation (7) and the state Equation (8). In this particular form, is also known in the literature as local level model, or random walk plus noise model or steady forecasting model, extensively covered in Harvey (1989) and in West and Harrison (1997). Despite its simplicity, the local level model can be used to analyze real datasets in various settings and scenarios, as it has been pointed out by many authors, see for example Durbin (2004) or Triantafyllopoulos (2011). The model can also be viewed as a particular case of a Dynamic Linear Model (West and Harrison, 1997), DLM in short, characterized in its general form by  $\{A, C, R, Q\}$ , respectively the observation matrix, transition matrix, observation error variance matrix and transition error variance matrix, possibly time-varying. Then, our model is a time-invariant DLM with matrices  $\{I_d, I_d, \Omega, \Sigma\}$ .

Under this model, the observed log-return follows an MA(1) process (Ait-Sahalia, Fan, and Xiu, 2010), and we can still obtain the likelihood in a product form by noting that  $\{\epsilon_t | \mathbf{Y}_t, \mathbf{X}_{t-1}, \Omega\}_{t=1}^T$  are i.n.i.d (independent but not identically distributed) Gaussian random vectors with zero means and covariance matrix  $\tilde{\Sigma}_t = V_t^{t-1} + \Omega$ .  $V_t^{t-1}$  is the variance of the prediction error obtained through the Kalman filter. Hence, the likelihood is

$$L(\Sigma, \Omega, \mathbf{Y}_{1:T}^{miss}, \mathbf{X}_{1:T} | \mathbf{Y}_{1:T}^{obs}) \propto \prod_{i=1}^d \prod_{t \in \mathbf{t}^{obs}} \tilde{v}_{i,t}^{-1} \exp\left(-\frac{1}{2} \tilde{v}_{i,t}^{-1} (X_{i,t} - \tilde{m}_{i,t})^2\right),$$

with  $\tilde{v}_{i,t} := \tilde{\sigma}_{i,t}^2 - \tilde{\Sigma}_{i,t} \tilde{\Sigma}_{-i,t}^{-1} \tilde{\Sigma}'_{i,t}$  and  $\mathbf{Y}^{miss}$  and  $\mathbf{Y}^{obs}$  are the missing and observed parts of  $\mathbf{Y}$  corresponding to the partition  $[\mathbf{t}^{miss}, \mathbf{t}^{obs}]$ . If we assume uninformative Jeffrey’s priors for  $\Sigma$  and  $\Omega$ , the complete data joint posterior density of  $\Sigma$  and  $\Omega$  is proportional to

$$(|\Omega| |\Sigma|)^{-\frac{d+1}{2}} \prod_{t=1}^T N(\mathbf{Y}_t; \mathbf{X}_t, \Omega) N(\mathbf{X}_t; \mathbf{X}_{t-1}, \Sigma).$$

The Gibbs sampler can be used also in this more complicated setting since the full conditionals are available in standard form. In our simulation approach,

we augment the data twice, by considering both the missing observations and the latent process as additional parameters. Therefore, we implement the algorithm by iteratively sampling  $\{\mathbf{X}, \Sigma, \Omega, \mathbf{Y}^{miss}\}$  from their full conditional densities. The full conditionals of  $\Sigma$  and  $\Omega$  are still proportional to Inverse Wishart:  $p(\Sigma|\Omega, \mathbf{Y}_{1:t}, \mathbf{X}_{1:t}) \propto IW(SS_\Sigma, T)$  and  $p(\Omega|\Sigma, \mathbf{Y}_{1:t}, \mathbf{X}_{1:t}) \propto IW(SS_\Omega, T)$ , with  $SS_\Sigma = \sum_{t=1}^T (\mathbf{X}_t - \mathbf{X}_{t-1})(\mathbf{X}_t - \mathbf{X}_{t-1})'$  and  $SS_\Omega = \sum_{t=1}^T (\mathbf{Y}_t - \mathbf{X}_t)(\mathbf{Y}_t - \mathbf{X}_t)'$ .

To sample the missing observations, we partition  $\mathbf{Y}$  in  $[\mathbf{Y}^{miss}, \mathbf{Y}^{obs}]$  and the full conditional density of the missing observations are still available in a standard form since they are obtained as the conditional normal density that result from the joint density of observed and missing log-prices  $\mathbf{Y}_t | \mathbf{Y}_{1:t-1}, \mathbf{X}, \Sigma, \Omega \propto N(\mathbf{X}_t, \Omega)$ , that is  $Y_{i,t}^{miss} | Y_{i,1:t-1}, \mathbf{Y}_{-i,1:t}, \mathbf{X}, \Sigma, \Omega$ , distributed as  $N(\mathbf{X}_{i,t} + \Omega_i \Omega_{-i}^{-1} (\mathbf{Y}_{-i,t} - \mathbf{X}_{-i,t}), \omega_i^2 - \Omega_i \Omega_{-i}^{-1} \Omega_i')$ , where  $\mathbf{Y}_{-i,s:t}$  is the matrix of log-prices for assets  $j, \forall j \neq i$ , from time  $s$  to time  $t$  and, similarly,  $\mathbf{X}_{-i,s:t}$  is the matrix of latent log-prices sampled with the FFBS (Forward Filtering Backward Simulation) algorithm for assets  $j, \forall j \neq i$  and from time  $s$  to time  $t$ .  $\Omega_{-i}$  is obtained from  $\Omega$  by dropping the row and column corresponding to asset  $i$ , and  $\Omega_i$  is the  $i$ -th row of  $\Omega$  without its  $i$ -th element.

Finally, we extract the latent log-price by using the FFBS algorithm, a Kalman smoother in which the smoothing recursions are replaced by simulations of the latent process. Following Fruwirth-Schnatter (1994), we can write the distribution of  $\mathbf{X} | \mathbf{Y}, \Sigma, \Omega$  as

$$p(\mathbf{X} | \mathbf{Y}, \Sigma, \Omega) = \prod_{t=1}^T p(\mathbf{X}_t | \mathbf{X}_{t+1:T}, \mathbf{Y}), \tag{9}$$

where the last factor in the product is simply  $p(\mathbf{X}_T | \mathbf{Y})$ , that is, the filtering distribution of  $\mathbf{X}_T$ , which is  $N(\mathbf{X}_T^t, V_t^t)$ , with  $\mathbf{X}_T^t$  the filtered latent log-price and  $V_t^t$  its covariance matrix. In order to obtain a draw from the distribution on the left-hand side, one can start by drawing  $\mathbf{X}_T$  from  $N(\mathbf{X}_T^t, V_t^t)$  and then, for  $t = T - 1, T - 2, \dots, 1$ , recursively draw  $\mathbf{X}_t$  from  $p(\mathbf{X}_t | \mathbf{X}_{t+1:T}, \mathbf{Y})$ . It can be shown that  $p(\mathbf{X}_t | \mathbf{X}_{t+1:T}, \mathbf{Y}) = p(\mathbf{X}_t | \mathbf{X}_{t+1}, \mathbf{Y}_{1:t})$  and this distribution is  $N(\mathbf{X}_t^t + V_t^t (V_{t+1}^t)^{-1} (\mathbf{X}_{t+1} - \mathbf{X}_{t+1}^t), V_t^t - V_t^t (V_{t+1}^t)^{-1} V_t^t)$ , where  $\mathbf{X}_{t+1}^t$  is the predicted latent log-price.

Summarizing, the implemented Gibbs sampler executes the following steps at each iteration:

1. Draw the covariance matrix  $\Sigma$  from its full conditional, that is an Inverse Wishart distribution  $IW(SS_\Sigma, T)$ , with  $SS_\Sigma = \sum_{t=1}^T (\mathbf{X}_t - \mathbf{X}_{t-1})(\mathbf{X}_t - \mathbf{X}_{t-1})'$ .
2. Draw the covariance matrix  $\Omega$  from its full conditional, that is an Inverse Wishart distribution  $IW(SS_\Omega, T)$ , with  $SS_\Omega = \sum_{t=1}^T (\mathbf{Y}_t - \mathbf{X}_t)(\mathbf{Y}_t - \mathbf{X}_t)'$ .
3. Impute, for  $i = 1, \dots, d$  and  $t \in \mathbf{t}_i^{miss}$ , the missing observations  $\mathbf{Y}^{miss}$  by drawing from  $N(\mathbf{X}_{i,t} + \Omega_i \Omega_{-i}^{-1} (\mathbf{Y}_{-i,t} - \mathbf{X}_{-i,t}), \omega_i^2 - \Omega_i \Omega_{-i}^{-1} \Omega_i')$ , where the dependence on  $i$  of  $\Omega_i$  and  $\Omega_{-i}$  has been suppressed to simplify notation.



4. Apply the FFBS algorithm to the DLM  $\{I_d, I_d, \Omega, \Sigma\}$  to extract the latent process  $\mathbf{X}$  from its full conditional  $\prod_{t=1}^T N(m_t, W_t)$ , where we have defined  $m_t \equiv \mathbf{X}_t^t + V_t^t(V_{t+1}^t)^{-1}(\mathbf{X}_{t+1}^t - \mathbf{X}_{t+1}^t)$  and  $W_t \equiv V_t^t - V_t^t(V_{t+1}^t)^{-1}V_t^t$ .

## 2 SIMULATION STUDY

### 2.1 Competing Estimators

In this section, we compare the performance of our Gibbs estimator with other estimators available in the literature. The first alternative is proposed by Ait-Sahalia, Fan, and Xiu (2010) (AFX), who estimate the covariance as a function of variances after synchronizing the asset returns. They use the Refresh Time Scheme introduced by Barndorff-Nielsen *et al.* (2011), which consists in aligning the returns on an irregular time grid by selecting those ticks at which all the assets have been traded at least once in the interval. This scheme includes the largest amount of data among all the Generalized Synchronization Schemes as defined in Ait-Sahalia, Fan, and Xiu (2010), but the loss of information still strongly depends on the presence of illiquid assets since several observations for the more liquid assets are neglected within each grid interval. After the synchronization, they estimate the covariance by applying the QMLE estimator suggested in Ait-Sahalia, Mykland, and Zhang (2005) to the identity  $\text{Cov}(X_1, X_2) = \frac{1}{4}(\text{Var}(X_1 + X_2) - \text{Var}(X_1 - X_2))$ , valid for any random variables  $X_1$  and  $X_2$ . It is fair to note that the polarization identity results in an increased variance when volatilities of the components are not equal (higher differences lead to higher losses by polarization), so that it could be interesting in a future work to see if the bivariate QMLE in the version by Shephard and Xiu (2012) can be a better benchmark.

The second estimator we include in the comparison study is the Composite Realized Kernel of Lunde, Shephard, and Sheppard (2012) (CompK). CompK synchronizes the high-frequency prices using a Refresh Time Scheme combined with a pairwise realized kernel. Compared to the Multivariate Realized Kernel of Barndorff-Nielsen *et al.* (2011), the use of a bivariate synchronization of the data (through the Refresh Time Scheme) increases the accuracy of the covariance estimation (since a smaller amount of data needs to be discarded) but makes the estimator no longer positive semi-definite by construction. For its implementation, we follow the suggestions contained in Lunde, Shephard, and Sheppard (2012) choosing a jittering parameter equal to 2 and a Parzen kernel function as weight function for the realized autocovariances.

The bandwidth is set to its optimal value by using the true  $\Sigma$  and  $\Omega$  (that cannot be observed in reality) and thus favoring the CompK methodology. For more details, we refer the reader to the original paper.

The third estimator is Bibinger (2011) multiscale estimator (MultiSC). This estimator is the multiscale extension of the estimator proposed by Hayashi and Yoshida (2005) (HY), that is the cross-product of all returns with

at least a partial overlapping. HY is robust to the asynchronicity but not to the microstructure noise. The multiscale extension is performed by first, adopting the pseudo-aggregation of Palandri (2006) and then choosing the optimal number of scales and optimal weights for the subsample estimators according to Bibinger (2011). We further report the results of the HY estimator.

An estimator that guarantees a positive semidefinite multivariate covariance matrix is essential in many applications. To enforce it when using AFX, CompK, and MultiSc in a multivariate setting, we project the resulting estimated matrix into the space of positive semi-definite matrices, by minimizing different notions of distances. In addition to be an *ad hoc* procedure, these types of projections suffer from the problem of loosing the financial interpretation of the covariance between the assets since, as noted in Frigessi *et al.* (2010), some entries of the covariance matrix, upon projection, can dramatically change.

We now describe the three types of projections we use to render all the estimators positive definite. The first is the Eigenvalue clipping projection built on the Marcenko–Pastur theorem (Marčenko and Pastur, 1967) from Random Matrix theory, and recently employed in Hautsch, Kyj, and Oomen (2012). Random Matrix theory suggests that all the significant eigenvalues of the estimated covariance has to be higher than the specific threshold  $\lambda_M = (1 - \frac{\lambda_{(1)}}{d})(1 + \sqrt{(\frac{d}{N})})^2$ , where  $\lambda_{(1)}$  is the highest eigenvalue,  $d$  is the number of assets and  $N$  is the sample size used to compute the covariance. If we define  $k = \sum_{i=1}^d I(\lambda_{(i)} \geq \lambda_M)$  and  $I$  is the indicator function, all the eigenvalues below  $\lambda_M$  are converted to  $\lambda_{(i)}^* = \frac{\sum_{i=k+1}^d \max(0, \lambda_{(i)})}{d-k}$ ,  $i = k+1, \dots, d$ .

The second is the so-called Factor projection where only a fixed number  $k$  of largest eigenvalues is kept.

To the estimated covariance matrix  $\Sigma$ , it corresponds an estimated correlation matrix  $R$  that can be decomposed, through Spectral Decomposition, as  $LDL'$ , where  $L$  is the matrix with eigenvectors as columns,  $D$  is a diagonal matrix with the eigenvalues ordered from largest to smallest  $\lambda_{(i)}$ ,  $i = 1, \dots, d$ , on the diagonal.  $\lambda_{(i)}$  is converted to  $\lambda_{(i)}^* = \lambda_{(i)}$ ,  $i = 1, \dots, k$  and to  $\lambda_{(i)}^* = 0$  otherwise. Finally, the covariance matrix is reconstructed by considering the converted eigenvalues and only the eigenvectors corresponding to the nonzero eigenvalues.

Finally, in the Shrinkage projection the optimal shrinking matrix  $E$  of Ledoit and Wolf (2003) is used to shrink the estimated correlation matrix  $R$  as  $(1-a)R + aE$ , where  $a$  is chosen to be the minimum weight for which the shrunked  $R$  becomes positive.

## 2.2 Data-Generating Process

The data-generating process is a multifactor extension of the stochastic volatility Heston (1993) model with two volatility factors having different persistence. In addition, a deterministic intraday volatility component is added to the stochastic

process in order to reproduce the observed U-shape intraday pattern. For  $i = 1, \dots, d$ ,  $t \in [0, T]$  and  $j \in \{a, b\}$ :

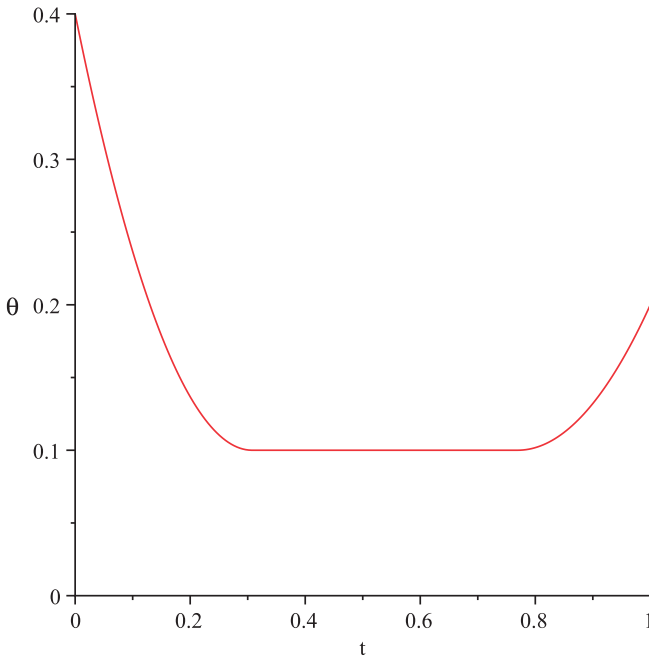
$$\begin{aligned}
 dX_{i,t} &= \sigma_{i,t}^a dW_{i,t}^a + \sigma_{i,t}^b dW_{i,t}^b, \\
 d(\sigma_{i,t}^j)^2 &= k^j \{\theta_t^j - (\sigma_{i,t}^j)^2\} + s^j \sigma_{i,t}^j dB_{i,t}^j, \\
 \theta_t^j &= \begin{cases} (h_1^j - 1)\bar{\theta}^j \left(\frac{t}{\alpha^j T} - 1\right)^2 + \bar{\theta}^j & t \in [0, \alpha^j T] \\ \bar{\theta}^j & t \in [\alpha^j T, (1 - \beta^j)T] \\ (h_2^j - 1)\bar{\theta}^j \left(\frac{t}{\beta^j T} - \frac{1 - \beta^j}{\beta^j}\right)^2 + \bar{\theta}^j & t \in [(1 - \beta^j)T, T], \end{cases}
 \end{aligned}$$

where  $E(dW_{i,t}^j dW_{k,t}^j) = \rho_{ik}^j dt$ ,  $E(dW_{i,t}^j dB_{k,t}^j) = \delta_{ik} \pi_{i,t}^j dt$ ,  $E(dW_{i,t}^a dW_{k,t}^b) = 0$  and  $E(dB_{i,t}^a dB_{k,t}^b) = 0 \forall i \neq k$ ,  $\alpha_j, \beta_j \in [0, 1]$ , with  $\alpha_j + \beta_j \leq 1$ , and  $h_1^j, h_2^j \in \mathbb{R}$ .  $X_{i,t}$  is the log latent price of asset  $i$  at time  $t$ ,  $W_{i,t}^a, W_{i,t}^b, B_t^a$  and  $B_t^b$  are Brownian motions with correlation structure defined above,  $k^j$  and  $s^j$  are, respectively, the speed of the mean reversion and the volatility of the variance process  $(\sigma_{i,t}^j)^2$ ,  $\theta^j$  captures the deterministic intraday pattern of  $\bar{\theta}^j$ , the asymptotic mean of  $(\sigma_{i,t}^j)^2$ . In particular, following for example Ellickson *et al.* (2011), we split the trading day into three parts: the first  $\alpha^j$  part of the trading day is the a region of price discovery, where  $\theta^j$  starting from values around  $h_1^j \bar{\theta}^j$ , decreases to  $\bar{\theta}^j$ , around which it stays for the following  $1 - \alpha^j - \beta^j$  part of the day, until raising again in the final  $\beta^j$  part of the trading hours, up to  $h_2^j \bar{\theta}^j$ .

The two-factors Heston model, also known as Double Heston model, was introduced in Christoffersen, Heston, and Jacobs (2009), who model the variance process with two uncorrelated processes. The addition of a second factor improves the quality of fit on empirical data compared to the original Heston model while computing time remains comparable. Empirical studies show that adding a third process, however, does not lead to further significant improvements. To introduce the well known U-shape intraday volatility pattern in the model, we substitute the asymptotic means of the volatility processes with the deterministic time-varying process  $(\theta^j)_{t \in [0,1]}$  that appears as in Figure 1. Fixing  $\alpha^j = \frac{4}{13}$ ,  $\beta^j = \frac{3}{13}$ ,  $h_1^j = 4$  and  $h_2^j = 2$  allows us to capture the stylized facts of a high variance at the beginning of the trading day (up to 4 times the midday level) that quickly decreases until stabilizing within 2 hours, with a slight increase again toward the end of the day (up to two times the midday level). Furthermore, the time  $t$  observed vector of log prices  $Y_t$  is supposed to be

$$Y_t = X_t + T_t,$$

where  $T_t \sim t_d(x; \Omega, v) \propto \left(1 + \frac{1}{v} x' \Omega^{-1} x\right)^{-\frac{v+d}{2}}$  a  $d$ -dimensional scaled Student's  $t$ -distribution with scale matrix  $\Omega$  and  $v$  degrees of freedom that explains large outlier returns, hardly justifiable by Gaussian disturbances. Finally, to reproduce



**Figure 1** Volatility factor deterministic intraday trend.

the asynchronicity of intraday log asset prices, we randomly cancel observations  $Y_{i,t}$ . Since trading volumes and volatility are inversely related, we negatively relate the number of missing observations to the volatility level.

We use an Euler discretization scheme to generate the data (see Authier and Possamaï, 2010 for a discussion on alternative discretization schemes of the Double Heston model). The first observation for the variance process is drawn from a Gamma distribution centered in the mean variance. All the codes have been written in Matlab 7.11.0 (R2010b) and run (possibly in parallel) with Intel(R) Xeon(R) CPU X7460 @ 2.66 GHz. In Table 1, we report the computational cost in seconds of one iteration of our Gibbs sampler, as function of the sample size  $T$  used for the estimation of the integrated covariance matrix and of the number of assets  $d$ . The algorithm CPU time is  $O\left(T\sqrt{\frac{d(d+1)}{2}}\right)$ : there is a linear computational cost in the sample size  $T$ , and a square root cost in  $\frac{d(d+1)}{2}$ , the number of parameters in the covariance matrix.

In all the following simulations and in the subsequent empirical analysis, the frequency at which the synchronized latent process is reconstructed is chosen to be of one second. Obviously, our methodology can be applied with a finer or coarser grid, depending on the characteristics of the available data. When performing the Monte Carlo analysis with a time-varying covariance matrix, the estimated

**Table 1** CPU time in seconds per Gibbs iteration, as function of the sample size  $T$  and number of assets  $d$ .

d	2	4	10	50	100
T					
23,400/4	1.25	1.45	1.7	6.7	24.6
23,400/2	2.5	3	3.5	13.4	49
23,400	5	5.6	6.8	27	98
23,400 · 2	10	11.25	13.5	54	196
23,400 · 10	50	56	67	270	980

(constant) integrated covariance matrices are computed at the daily horizon and compared with the realized covariance matrix of the latent process for that day. Hence, each day the covariance estimators are compared with a realized measure of the latent process which is different for every day.

### 2.3 Bivariate Case

We first run a bivariate simulation study with: sample size of  $T = 23,400$ , of which  $1/3$  and  $1/2$  are removed completely at random from the whole generated sample, for the first and second asset, respectively, and two starting log prices of  $\log(100)$  and  $\log(40)$ . The true  $\Sigma$  and  $\Omega$  matrices, (the same for the two volatility factors) are, respectively,

$$\Sigma = \begin{bmatrix} 0.16 & 0.06 \\ 0.06 & 0.09 \end{bmatrix}, \quad \Omega = \begin{bmatrix} 0.08 & 0 \\ 0 & 0.04 \end{bmatrix}.$$

The other parameters  $\{\bar{\theta}^j, k^j, s^j, \rho_{12}^j, \pi_1^j, \pi_2^j\}$  are, respectively,  $\{0.1, 0.179, 0.28, 0, -0.834, -0.834\}$  and  $\{0.1, 1.303, 0.28, 0, -0.834, -0.834\}$  for the two volatility factors, as estimated by Christoffersen, Heston, and Jacobs (2009), on S&P500 call options in the period 1990–2004. For each compared estimator, we generate  $M = 100$  matrices of prices and we run our Gibbs sampler for each generated sample of prices for 5000 steps, after 5000 initial iterations of burn-in. The starting point (of the MCMC simulation) of the missing values is the local mean ignoring the missing data up to ten ticks before and after the missing trade. To speed up the convergence of the Markov chain, we need to be careful about the initial values for the covariance parameters: We initialize the sampler from the pairwise Hayashi and Yoshida (2005) estimate of the covariance of the partially observed noisy returns series for the off-diagonal terms. For the variances, we use the Two Time Scale Estimator of Zhang, Myckland, and Ait-Sahalia (2005). Then, we interpret the return variability not captured by the Two Time Scale Estimator as the contribution coming from the microstructure noise. We stress the fact that different starting values affect the convergence speed of the Markov chain, but not

**Table 2** Simulation results: bivariate case.  $M = 100$  Monte Carlo estimates for each compared estimator are computed and the mean is reported. The Gibbs sampler runs for 5000 iterations, plus 5000 of burn-in. The RMSE is reported in parenthesis. RCov5min and RCov10min refer to the realized covariance matrix computed with, respectively, 5-minute and 10-minute returns

	Asset 1	Asset 2	Asset 1–2		
			No Lead	Lead1	Lead2
RCov5min	0.0539 (0.0097)	0.0301 (0.0051)	0.0186 (0.0051)	0.0202 (0.0063)	0.0202 (0.0064)
RCov10min	0.0536 (0.0138)	0.0295 (0.0069)	0.0188 (0.0073)	0.0202 (0.0081)	0.0216 (0.0073)
AFX	0.0534 (0.0063)	0.0298 (0.0036)	0.0199 (0.0028)	0.0200 (0.0026)	0.0165 (0.0038)
CompK	0.0560 (0.0073)	0.0314 (0.0040)	0.0189 (0.0046)	0.0143 (0.0061)	0.0108 (0.0093)
MultiSc	0.0514 (0.0103)	0.0289 (0.0049)	0.0186 (0.0047)	0.0198 (0.0054)	0.0201 (0.0050)
HY	0.0532 (0.0063)	0.0296 (0.0036)	0.0204 (0.0066)	0.0155 (0.0057)	0.0106 (0.0098)
Gibbs	0.0535 (0.0062)	0.0299 (0.0035)	0.0199 (0.0027)	0.0195 (0.0023)	0.0152 (0.0050)
True value	0.0532	0.0298	0.0199	0.0199	0.0199

the validity of the results, since the chain is independent on the chosen starting values, once it has reached stationarity.

The results of this first simulation are reported in Table 2 where the Gibbs estimator is shown to perform slightly better than all other competitors in terms of RMSE, still all methods return fairly precise estimates of the covariance matrix. Thus, with only two assets, no significant differences are observed among the estimation procedures under comparison. Notable exceptions are the realized measures estimated through low-frequency returns at 5 and 10 minutes (reported in line 1 and 2 of Table 2, respectively), which suffer from higher estimation variability.

In order to test the robustness of the estimators to the presence of noncontemporaneous covariation, we further stress the simulation setting by adding delayed covariation in the true latent asset prices: the so-called lead-lag effect empirically observed by, for instance, Griffin and Oomen (2009). To this aim we let the more liquid asset (asset 1) *lead* the other one designing two different settings: in the scenario *Lead1*, a portion of 30% and 20% of the covariation is shifted, respectively, by one and two seconds, while in *Lead2* setting, the covariation is shifted over a longer time span of 10 seconds with a proportion of  $(10-i+1)\%$  shifted by  $i$  seconds, for  $i = 1, \dots, 10$ . The results are reported in the last two columns of Table 2. In *Lead1*, CompK and HY already show a relevant bias since they fail to fully capture

the noncontemporaneous portion (50%) of the total covariance. These estimators, together with realized covariance estimators at 5 and 10 minutes, present the worst RMSE. On the other hand, MultiSC, AFX, and Gibbs seem more robust in terms of bias-variance trade-off in presence of lead-lag effects. In the more challenging *Lead2*, there is a further slight deterioration in the RMSE of the low-frequency estimators, and an important bias is visible for CompK and HY. MultiSc is still robust to the lead-lag effect, while some distortions start appearing in AFX and Gibbs. The results are quite favorable to our Bayesian estimator, since it remains among the best alternatives. In particular, it is interesting to note how the Gibbs estimator turns out to be much more robust than its competing tick-by-tick estimator, that is the HY one. For our Bayesian estimator, both a preprocessing and a postprocessing remedy to eliminate the bias induced by the lead-lag effect are subjects of current research: the first one investigates the use of a lower frequency for the reconstructed synchronized series; the second one, inspired by the covariance correction proposed by Scholes and Williams (1977), adds lead-lag cross-product terms to the standard sample covariance to capture the noncontemporaneous portion of covariation.

## 2.4 Multivariate Case

For the multivariate simulation analysis, we estimate the covariance matrix for a portfolio of ten assets. Our Gibbs estimator is naturally extended to the ten-dimensional case, without any modification. If the covariance matrices obtained with the pointwise procedures are not positive semi-definite, we project them onto the space of positive semi-definite matrices through the procedures described above. The true covariance matrices of the data-generating process are  $\Sigma$ , reported in Table 3, and  $\Omega = \text{diag}([0.08, 0.04, 0.02, 0.05, 0.1, 0.06, 0.1, 0.15, 0.15, 0.08])$ . The simulation is initialized from  $P_0 = \log([100, 40, 60, 80, 40, 20, 90, 30, 50, 60])$ , with probabilities of missing observations equal to  $v = 0.5978 - 1.087 \cdot \text{diag}(\Sigma)$ , so that trading volumes and volatility levels are inversely related. The probability of missing values ranges from 25% for the most liquid asset, to 50% for the least liquid asset. We note that, as the dimension of the portfolio increases, it becomes more and more difficult to obtain, through pairwise inferences, an estimated covariance matrix that retains positivity.

We iterate Gibbs sampler for 5000 (burn-in) plus 5000 steps. As synthetic measure of the performance we choose the Frobenius norm of the matrix difference  $\Delta$  between the estimated and generated realized covariance matrix. In particular, we define our norm as  $\|\Delta_{i,j}\| = \|(\hat{\Sigma}_i^j - \Sigma)\|_F$ , where  $\|\cdot\|_F$  is the traditional Frobenius norm, and  $\hat{\Sigma}_i^j$  is the  $j$ -th covariance matrix estimated with methodology  $i$ , and  $j = 1, \dots, M$ . We further define  $\hat{E}[\|\Delta_i\|] := \frac{1}{M} \sum_{j=1}^M \|\Delta_{i,j}\|$  and  $\hat{\sigma}[\|\Delta_i\|] := \left(\frac{1}{M-1} \sum_{j=1}^M (\|\Delta_{i,j}\| - \hat{E}[\|\Delta_i\|])^2\right)^{1/2}$  as the estimated expected value and standard deviation of the Frobenius distance for methodology  $i$ . The results are summarized

**Table 3** 10-dimensional  $\Sigma$  matrix used to generate simulated data

---


$$\Sigma = \begin{bmatrix} 0.16 & 0.06 & -0.01 & 0.10 & 0.12 & 0.04 & 0.06 & 0.09 & 0.09 & 0.04 \\ 0.06 & 0.09 & 0.04 & 0.05 & 0.05 & 0.02 & 0.04 & 0.03 & 0.05 & 0.10 \\ -0.01 & 0.04 & 0.12 & 0.03 & 0.08 & 0.05 & 0.09 & 0.05 & 0.11 & 0.04 \\ 0.10 & 0.05 & 0.03 & 0.11 & 0.11 & 0.04 & 0.06 & 0.11 & 0.13 & 0.04 \\ 0.12 & 0.05 & 0.08 & 0.11 & 0.25 & 0.06 & 0.09 & 0.07 & 0.16 & 0.03 \\ 0.04 & 0.02 & 0.05 & 0.04 & 0.06 & 0.16 & 0.10 & 0.07 & 0.14 & 0.06 \\ 0.06 & 0.04 & 0.09 & 0.06 & 0.09 & 0.10 & 0.18 & 0.07 & 0.16 & 0.05 \\ 0.09 & 0.03 & 0.05 & 0.11 & 0.07 & 0.07 & 0.07 & 0.32 & 0.11 & 0.05 \\ 0.09 & 0.05 & 0.11 & 0.13 & 0.16 & 0.14 & 0.16 & 0.11 & 0.28 & 0.05 \\ 0.04 & 0.10 & 0.04 & 0.04 & 0.03 & 0.06 & 0.05 & 0.05 & 0.05 & 0.24 \end{bmatrix},$$


---

in Table 4 and Figure 2a. They clearly show that the proposed multivariate Bayesian estimator is the most accurate measure of the true covariance matrix, followed, in order, by AFX, CompK, HY, and MultiSc. In Figure 3a, we compare graphically the kernel density of the Frobenius distances for all methodologies observing, again, that the Gibbs estimator outperforms the other competitor measures. The kernel density estimate is based on a Gaussian kernel function, using the optimal window parameter (width)  $\hat{a} \left( \frac{4}{3M} \right)^{1/5}$ , where  $\hat{a} = \text{med}(|\Delta_i| - \text{med}(|\Delta|)) / 0.6745$  is a robust estimate of the standard deviation of the Frobenius distances. The density is evaluated at 100 equally spaced points that cover the range of the data. For the projection methodologies, the factor approximation method induces an important bias in the estimation, while there is no clear ordering between shrinking and clipping methods.

Furthermore, following Andersen *et al.* (2003), in Figures 4 and 5, we use as simple graphical diagnostic the *volatility signature plot* and the *correlation signature plot* for two randomly chosen stocks, which show the pattern of bias injected in realized covariances as underlying returns are sampled progressively more frequently. The key insight is that microstructure bias will likely manifest itself as sampling frequency increases by distorting the average realized volatility. Consequently, a plot of average realized volatility against sampling frequency may help reveal the severity of microstructure bias as sampling frequency increases. A comparison with the estimated covariance matrix provided by the competing methodology can reveal if they are robust to microstructure noise. For readability, we plot only the *clip* projections of the pairwise estimators, that is the best projection in this case. What we observe is that all the methodologies considered are robust to the microstructure noise, except for HY in the correlation estimation, expected to be downward biased.

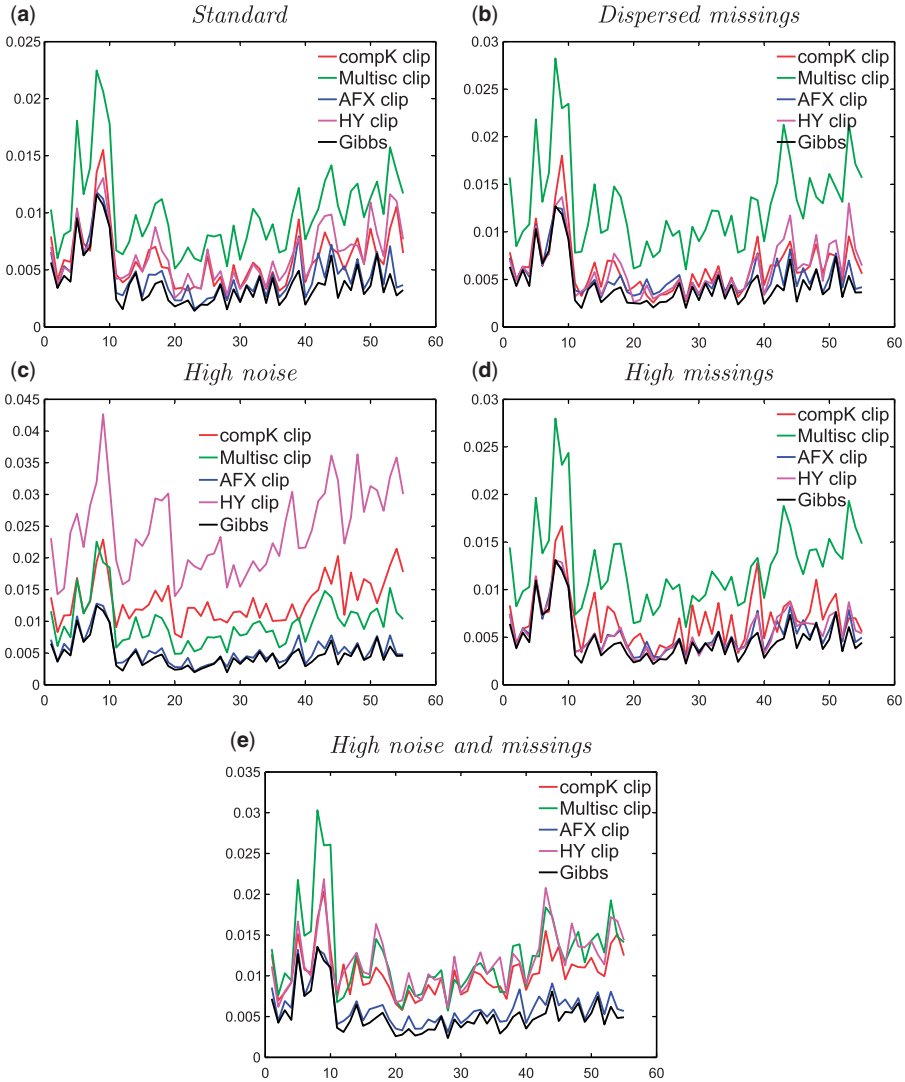


**Table 4** Estimated expected values and standard deviations of the Frobenius distances between the estimated and the true realized covariance matrix

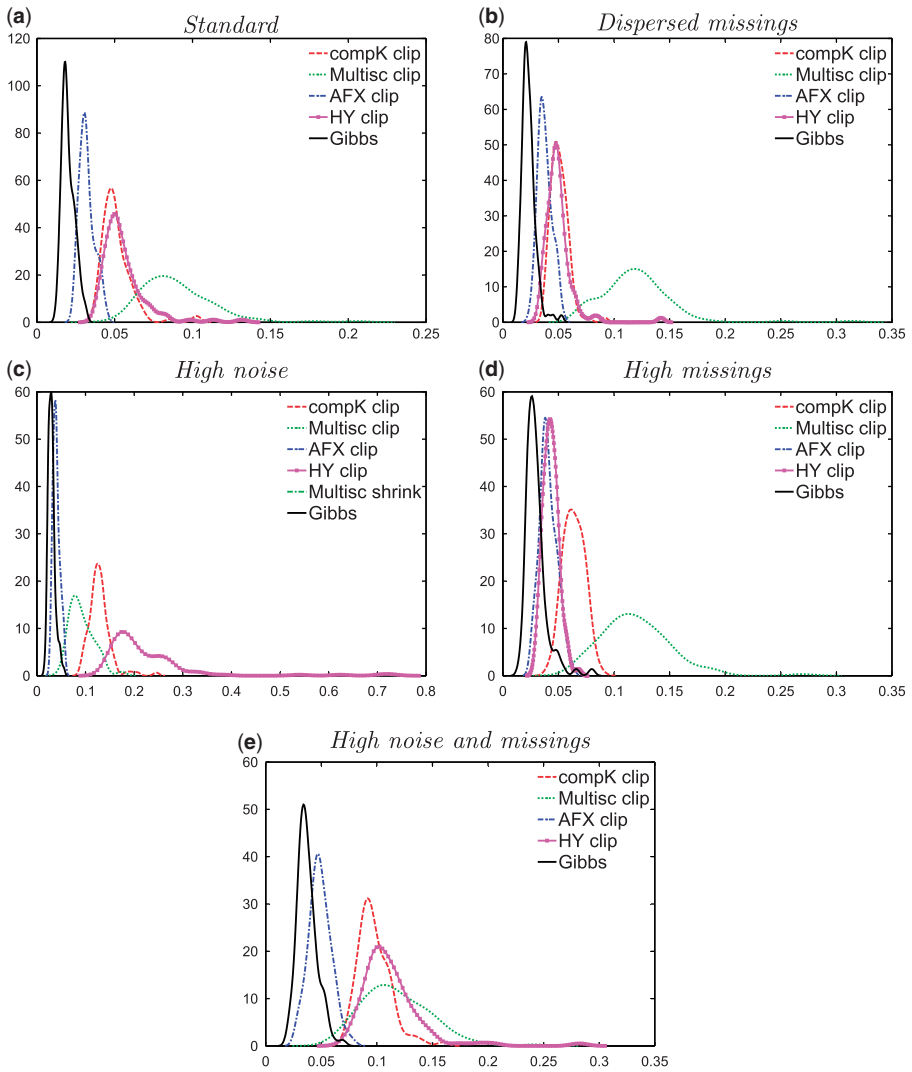
	AFX			CompK			Multisc			HY			
	Gibbs	Clipping	Factor	Shrink	Clipping	Factor	Shrink	Clipping	Factor	Shrink	Clipping	Factor	Shrink
<i>(a) Standard</i>													
$\hat{E}[\ \Delta_i\ ]$	0.0201	0.0320	0.1313	0.0240	0.0513	0.1329	0.0486	0.0902	0.1505	0.0864	0.0555	0.1339	0.0502
$\hat{\sigma}[\ \Delta_i\ ]$	0.0040	0.0048	0.0132	0.0056	0.0107	0.0138	0.0099	0.0242	0.0200	0.0249	0.0143	0.0140	0.0111
<i>(b) Dispersed missings</i>													
$\hat{E}[\ \Delta_i\ ]$	0.0232	0.0378	0.1325	0.0314	0.0510	0.1324	0.0483	0.1202	0.1663	0.1163	0.0497	0.1338	0.0454
$\hat{\sigma}[\ \Delta_i\ ]$	0.0062	0.0063	0.0144	0.0064	0.0082	0.0134	0.0082	0.0344	0.0292	0.0342	0.0132	0.0143	0.0107
<i>(c) High noise</i>													
$\hat{E}[\ \Delta_i\ ]$	0.0302	0.0401	0.1327	0.0340	0.1289	0.1626	0.0952	0.0923	0.1520	0.0887	0.2259	0.2365	0.1168
$\hat{\sigma}[\ \Delta_i\ ]$	0.0070	0.0069	0.0134	0.0071	0.0228	0.0213	0.0132	0.0250	0.0204	0.0254	0.1009	0.0989	0.0238
<i>(d) High missings</i>													
$\hat{E}[\ \Delta_i\ ]$	0.0298	0.0408	0.1332	0.0350	0.0638	0.1383	0.0621	0.1196	0.1661	0.1160	0.0432	0.1325	0.0388
$\hat{\sigma}[\ \Delta_i\ ]$	0.0096	0.0078	0.0147	0.0080	0.0093	0.0154	0.0095	0.0319	0.0269	0.0319	0.0069	0.0145	0.0070
<i>(e) High noise missings</i>													
$\hat{E}[\ \Delta_i\ ]$	0.0373	0.0489	0.1341	0.0442	0.0977	0.1470	0.0865	0.1173	0.1638	0.1139	0.1125	0.1546	0.0846
$\hat{\sigma}[\ \Delta_i\ ]$	0.0087	0.0099	0.0152	0.0099	0.0148	0.0170	0.0127	0.0304	0.0259	0.0304	0.0280	0.0254	0.0120

$\hat{E}[\|\Delta_i\|] := \frac{1}{M} \sum_{j=1}^M \|\Delta_{i,j}\|$  and  $\hat{\sigma}[\|\Delta_i\|] := \left( \frac{1}{M-1} \sum_{j=1}^M (\|\Delta_{i,j}\| - \hat{E}[\|\Delta_i\|])^2 \right)^{1/2}$ , with  $M=100$  and 10 assets. Define missing probabilities  $v=0.5978-1.087 \cdot \text{diag}(\Sigma)$  and noise matrix  $\Omega = \text{diag}(0.08, 0.04, 0.02, 0.05, 0.1, 0.06, 0.1, 0.15, 0.15, 0.08)$

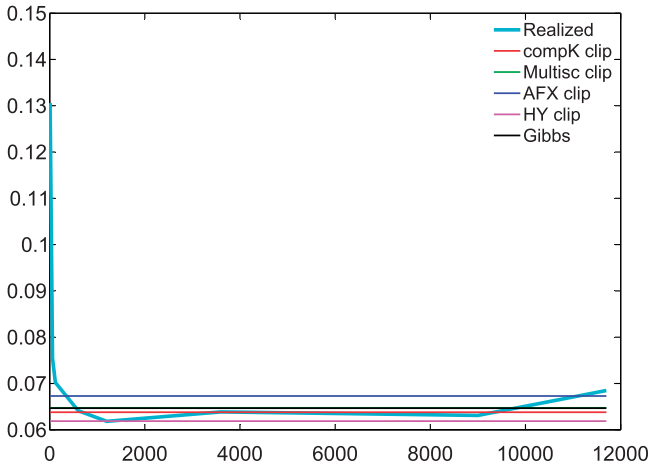
(a) *Standard*: missing probabilities  $v$  and noise matrix  $\Omega$ .  
 (b) *Dispersed missings*: more dispersed missing probabilities 1.2130-3.4783·diag( $\Sigma$ ) and noise matrix  $\Omega$ .  
 (c) *High noise*: missing probabilities  $v$  and noise matrix  $\Omega+0.35I$ .  
 (d) *High missings*: missing probabilities  $v+0.35$  and noise matrix  $\Omega$ .  
 (e) *High noise and missings*: missing probabilities  $v+0.35$  and noise matrix  $\Omega+0.35I$ .



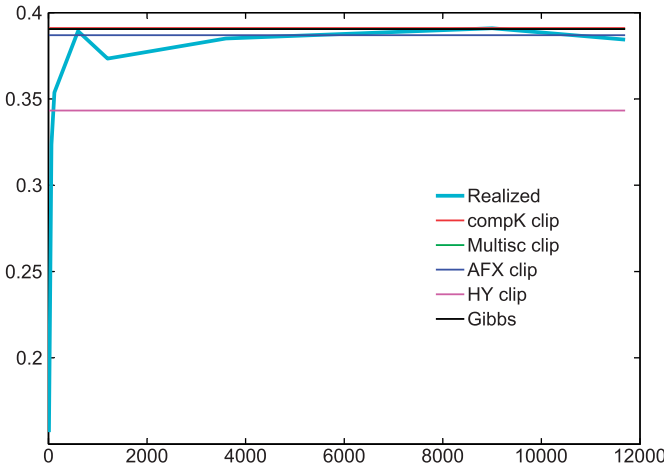
**Figure 2** Simulated root mean squared errors.  $M=100$  Monte Carlo estimates for each estimator are computed. The Gibbs sampler runs for 5000 iterations, plus 5000 of burn-in. The  $x$ -axis is the index for the fifty-five parameters of the covariance matrix, starting from the ten variances. Define missing probabilities  $v=0.5978-1.087 \cdot \text{diag}(\Sigma)$  and noise matrix  $\Omega = \text{diag}(0.08, 0.04, 0.02, 0.05, 0.1, 0.06, 0.1, 0.15, 0.15, 0.08)$  (a) *Standard*: missing probabilities  $v$  and noise matrix  $\Omega$ . (b) *Dispersed missings*: more dispersed missing probabilities  $1.2130-3.4783 \cdot \text{diag}(\Sigma)$  and noise matrix  $\Omega$ . (c) *High noise*: missing probabilities  $v$  and noise matrix  $\Omega+0.35I$ . (d) *High missings*: missing probabilities  $v+0.35$  and noise matrix  $\Omega$ . (e) *High noise and missings*: missing probabilities  $v+0.35$  and noise matrix  $\Omega+0.35I$ .



**Figure 3** Kernel density estimates of the Frobenius distances  $\|\Delta_{i,j}\| = \|(\hat{\Sigma}_i^j - \Sigma)\|_F$ ,  $\hat{\Sigma}_i^j$  is the  $j$ -th covariance matrix estimated with methodology  $i$  and  $j=1, \dots, M=100$  and  $\Sigma$  is the true covariance matrix. Define missing probabilities  $v=0.5978 - 1.087 \cdot \text{diag}(\Sigma)$  and noise matrix  $\Omega = \text{diag}([0.08, 0.04, 0.02, 0.05, 0.1, 0.06, 0.1, 0.15, 0.15, 0.08])$ . (a) *Standard*: missing probabilities  $v$  and noise matrix  $\Omega$ . (b) *Dispersed missings*: more dispersed missing probabilities  $1.2130 - 3.4783 \cdot \text{diag}(\Sigma)$  and noise matrix  $\Omega$ . (c) *High noise*: missing probabilities  $v$  and noise matrix  $\Omega + 0.35I$ . (d) *High missings*: missing probabilities  $v + 0.35$  and noise matrix  $\Omega$ . (e) *High noise and missings*: missing probabilities  $v + 0.35$  and noise matrix  $\Omega + 0.35I$ .



**Figure 4** Signature plot for one variance. The line called *Realized* is the value of the variance from the Realized Covariance computed over returns synchronized at the frequency in seconds given by the *x*-axis.



**Figure 5** Signature plot for one correlation. The line called *Realized* is the value of the correlation from the Realized Covariance computed over returns synchronized at the frequency in seconds given by the *x*-axis.

We expect the performance of the Bayesian estimator to improve relative to the other estimators in the presence of a more dispersed probabilities of missing (i.e., when assets have very different liquidity profiles), since in this case our Bayesian estimator capture multivariate information that is not available to the pairwise methodologies. To validate this expectation, we repeat the simulations

for ten assets, holding everything as in the previous simulation setting, but with more dispersed missing probabilities  $1.2130 - 3.4783 \cdot \text{diag}(\Sigma)$ , still inversely proportional to  $\Sigma$ , but in the range [10%, 90%]. The results in Figure 2b confirm our intuition: particularly affected by the changed scenario are AFX and MultiSc, but the ordering of the estimators remains as in the base case, with the Gibbs methodology performing best. The relative norms are summarized in Table 4 and in Figure 3b the relative kernel densities estimates are plotted. The explanation is that information contained in the co-movement of more liquid assets is incorporated in the estimation of covariances of less liquid asset only with a truly multivariate approach, which can then significantly improve the covariance estimation accuracy in portfolios containing assets with very different liquidity profiles.

### 3 EMPIRICAL APPLICATION

#### 3.1 Dataset and Gaussianity Tests

Our dataset consists of NYSE TAQ tick-by-tick data from 3 January 2007 to 21 November 2007 for ten U.S. stocks: Alcoa Inc., Citigroup Inc., Hasbro Inc., Harley Davidson, Intel Corp., Microsoft Corp., Nike Inc., Pfizer Inc., Exxon Mobil Corp., and Tektronix. The average missing probability per day ranges from 0.67 to 0.96, respectively, for Citigroup and Tektronix, whose time series (in log scale) are reported in Figure 6, for the first 1000 seconds of one randomly selected trading day. The correlation between these two stocks, estimated through the competing methodologies described above, is shown in Figure 7. Not all projection procedures are reported for brevity, but are available under request.

Clark (1973) and Ane and Geman (2000) theoretically argue that, for an underlying continuous-time diffusion process, rescaling the log-return by an appropriate measure of the market activity, recovers the standard Gaussian distribution. Andersen *et al.* (2000, 2001b, 2003), Andersen *et al.* (2001b), and Andersen, Bollerslev, and Dobrev (2007) indeed show that, when daily returns are standardized by the corresponding daily realized volatility, the resulting distribution is nearly Gaussian. Analogously, in a multivariate context, rescaling the daily return vector by the appropriate integrated covariance measure should recover the multivariate standard Gaussian distribution. Theoretically, the presence of jumps could induce deviations from the normality of standardized returns. However, since all the considered measures are equally affected by the presence of jumps, we assume that the estimator whose covariance matrix produces a vector of standardized returns closer to a multivariate standard normal will be the preferred one.

Following this intuition, we compare the different covariance matrix estimates through the use of two tests. In particular, we compute the 10-dimensional vector of

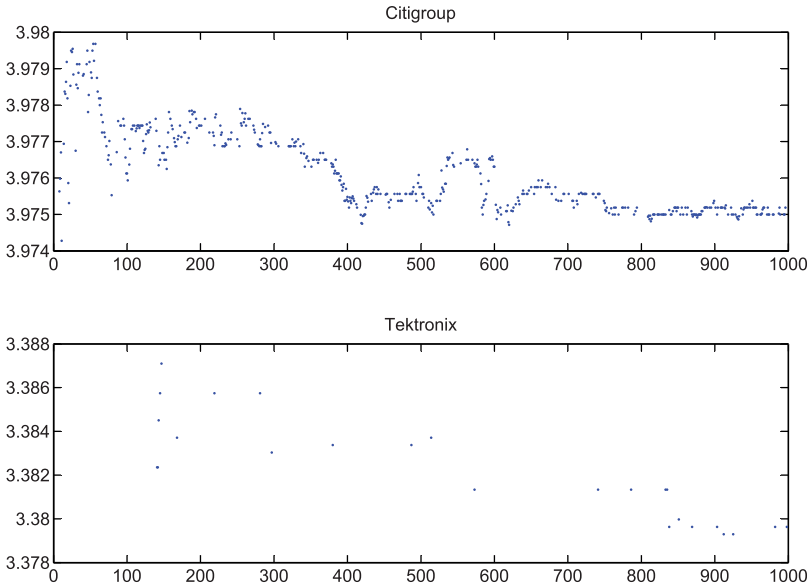


Figure 6 First 1000 seconds of 11 May 2007 for Citigroup (above) and Tektronix (below) log prices.

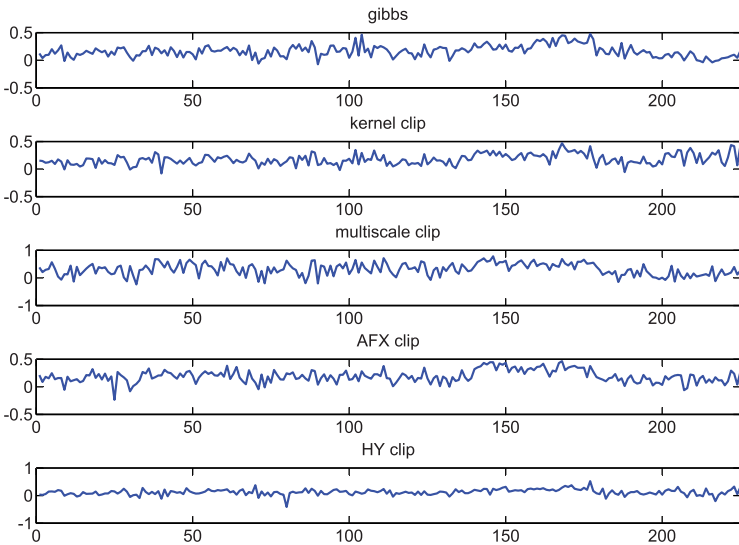


Figure 7 Estimated correlation between Citigroup and Tektronix.

**Table 5** Empirical tests: in the first column, all the estimation methodologies with associated projection procedures are reported. In the second column, we have the test based on moments comparison, while in the last column the  $p$ -value for the  $\chi^2$  test. See the text for details on the estimation methodologies, the projection procedures, and the empirical tests

Method and proj	Moments test		$\chi^2$ test $p$ -val
	Mean dist	Cov dist	
Gibbs	0.1186	1.4517	0.7135
CompK clipping	0.3969	22.9628	0.2696
CompK factor	0.3084	4.9784	0.2798
CompK shrink	1.0130	264.0134	0.1629
Multisc clipping	0.2176	3.1529	0.3193
Multisc factor	0.2331	3.9824	0.2598
Multisc shrink	0.2223	4.0357	0.2834
AFX clipping	0.1637	1.4658	0.5606
AFX factor	0.1546	2.3255	0.4870
AFX shrink	0.1637	4.3677	0.5535
HY clipping	0.2133	6.9434	0.5844
HY factor	0.2652	23.1247	0.5335
HY shrink	0.4801	54.3585	0.5406

open-to-close daily returns for each day, and we standardize its demeaned version using the estimated covariance matrix. Then,

- in the first test we conduct a comparison based on the difference between the sample and hypothesized moments of the standardized returns. In particular, we take the mean over the days of the standardized returns and we compute the Euclidean distance between this obtained 10-dimensional vector and the null vector of same dimension. In the same way, we compute the sample covariance matrix of the standardized returns and we compute the Frobenius distance between this obtained  $10 \times 10$  matrix and the identity matrix of same dimensions. We expect the *best* methodology to report the lowest distances.
- In the second test we compute the probability of observing more extreme data, given the null hypothesis of multivariate standard gaussianity of the standardized returns. In particular, under the null, the vector of daily returns is an observation from the multivariate standard Gaussian distribution, or, equivalently, its squared Euclidean norm is an observation from a  $\chi^2$  distribution with 10 degrees of freedom, denoted  $\chi_{10}^2$ . Then, the area under the  $\chi_{10}^2$  density on the right of the observed squared Euclidean norm is a sort of  $p$ -value, and we expect the *best* methodology to show a higher  $p$ -value, suggesting more confidence in the null hypothesis.

The results for the tests above are reported in Table 5, where it is clear that our Bayesian estimator is superior to alternative methodologies, and for any projection

procedures. In the table, we limit the analysis of the first test to the first two moments, but results not reported show the same conclusions for higher moments.

The signature plots averaged all over the sample period show that the estimates of all the methodologies under comparison are not centered on the low-frequency synchronized realized covariance. We claim that this effect is not due to the market microstructure noise or asynchronicity since, as also shown in the simulation analysis, all the considered estimator are robust to these type of effects, but that are due to other intrinsic characteristics of the empirical data. In particular, as already noted by Griffin and Oomen (2009) this might be due to a delay in the manifestation of the correlation among assets. This effect is more likely to be observed in a dataset where assets have very different trading volumes, like in our case, since assets which are more liquid, tend to lead the behavior of other less traded assets, but with some delay due to the absence of market activity. To confirm our intuition, we plot in Figure 8 the signature plot of the correlations among the latent price processes reconstructed from our Bayesian methodology, that is on the latent prices obtained after synchronization and microstructure noise filtering. The fact that, as the frequency decreases, we still observe an increase in the realized covariance estimation on synchronized and noise-free data seems to suggest that such behavior might be induced by the presence of a lead-lag correlation structures among the assets. In fact, comparing it with the equivalent signature plot computed using the same reconstructed latent price process but from the simulated data (using the data-generating process employed in the previous section), we observe that in contrast to the signature plot of the empirical data, the one on simulated data is virtually horizontal. This comparison suggests that the upward slope observed on empirical data is more a feature of the data rather than the methodology which is (as extensively shown in the previous section) perfectly able to cope with a wide range of market frictions and imperfections: market microstructure noise, asynchronicity, stochastic volatility, intraday pattern, and heavy tails in the return distributions.

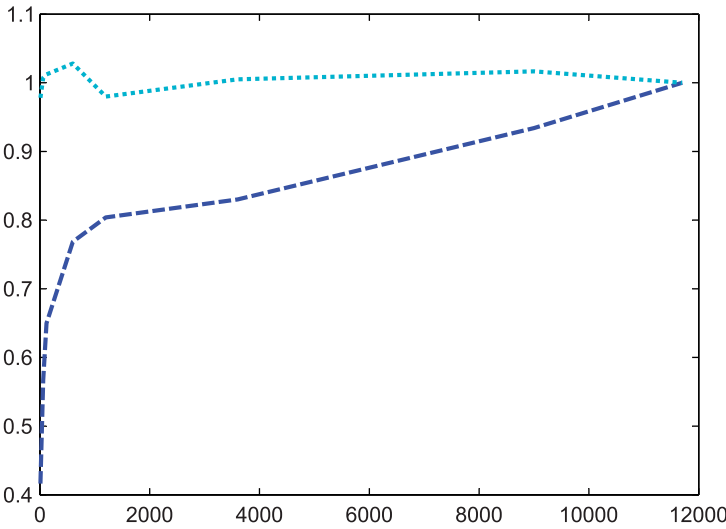
### 3.2 Portfolio Allocation Analysis

We also evaluate the economic benefit of the proposed covariance estimators, employing the methodology suggested by West, Edison, and Cho (1993) and Fleming, Kirby, and Ostdiek (2001, 2003), which compares the utility level obtained by different covariance estimators in the context of portfolio allocation. In particular, we consider a risk-averse investor who uses a conditional mean-variance optimization rule to allocate funds across the stocks in our dataset, with a daily rebalance.

Define  $R_t^f$  as the risk-free rate,  $\mu_t^i = E_t[R_{t+1}^i]$  as the expected value at time  $t$  of the stock  $i$  return at time  $t+1$  and  $\mu_t^p = E_t[R_{t+1}^p]$  as the target expected return of the portfolio. The investor solves the quadratic program

$$\min_{w_t} w_t' \Sigma_t w_t$$





**Figure 8** Signature plot for the averaged correlation on the latent prices reconstructed through our Bayesian methodology: comparison between simulated and empirical data. The dashed line is the normalized signature plot of the correlation from empirical synchronized returns, at the frequency in seconds given by the x-axis. The dotted line is the equivalent plot for simulated data.

subject to

$$w_t' \mu_t + (1 - w_t' \mathbf{1}_d) R_t^f = \mu_p,$$

where  $w_t$  is the vector of portfolio weights,  $\Sigma_t$  is the covariance matrix estimated at time  $t$  with tick prices observed in day  $t - 1$ , and  $\mathbf{1}_d$  is a  $d$ -dimensional vector of ones. Assuming absence of predictability in the expected returns of the single stocks at the daily level, we assume  $\mu_t^i = \mu_t = R_t^f + 0.1 \forall i, t$ . Fixing the expected stock returns equal to the risk-free rate plus a constant spread in a mean-variance framework, is equivalent to following a volatility-timing strategy where the weights vary only with  $\Sigma_t$ . The vector of weights that minimize the portfolio variance is

$$w_t = \frac{(\mu_p - R_t^f) \Sigma_t^{-1} (\mu_t - R_t^f) \mathbf{1}_{10}}{(\mu_t - R_t^f) \mathbf{1}_{10} \Sigma_t^{-1} \mathbf{1}_{10}' (\mu_t - R_t^f)'}$$

Following Engle and Colacito (2006); Audrino and Trojani (2006); Corsi and Audrino (2012), the economic differences between the covariance estimates is computed as the AU measure, interpreted as the fee that the investor would be willing to pay to switch from one covariance measure to the other. The AU is expressed as:

$$AU = \frac{\lambda}{2} \frac{1}{T} \sum_{i=1}^T (R_t^p - \bar{R}^p)^2 \tag{10}$$

**Table 6** Annualized fees (expressed in basis points) that an investor following a volatility timing strategy would be willing to pay to employ the Bayesian integrated covariance estimator in place of the alternative estimators. The portfolio weights are obtained minimizing the conditional variance of a portfolio containing the ten stocks in our database and one-month U.S. Treasury bonds

Target (%)	Method	$\lambda=2$	$\lambda=7$	$\lambda=10$
0.06	Kernel	0.8	2.9	4.1
	Multiscale	0.2	0.8	1.2
	AFX	0.1	0.4	0.5
	HY	0.3	1.1	1.6
0.10	Kernel	9.3	32.4	46.3
	Multiscale	2.1	7.3	10.5
	AFX	-0.6	-1.9	-2.8
	HY	1.5	5.3	7.5
0.15	Kernel	31.8	111.3	159.0
	Multiscale	6.9	24.3	34.7
	AFX	-3.3	-11.6	-16.6
	HY	2.7	9.5	13.6

where

$$R_t^p = R_t^f + w'_{t-1}(R_t - R_t^f \mathbf{1}_{10})$$

is the return of the portfolio constructed at time  $t-1$ ,  $\bar{R}^p$  is the sample mean of portfolio returns, and  $\lambda$  is a coefficient of risk aversion.

We compute AU in (10) for the covariances estimated by the different methodologies in our dataset. As risk-free return we use the U.S. one-month rate, but similar results are obtained for the U.S. three-month rate. The results are reported in Table 3.2, for risk-aversion parameter  $\lambda$  equal to 2, 7, and 10, and for target portfolio expected return  $\mu_p$  of 6%, 10%, and 15%. The figures show that an investor following a volatility timing strategy would almost always be willing to pay a positive annual fee to switch to the Bayesian estimator of the integrated covariance. The exception (but for modest amounts) is AFX in case of large target portfolio returns and high-risk aversion; we expect a favorable result for the Gibbs estimator even in this scenario for the estimation of larger covariance matrices, where the projection of a pairwise methodologies can be much more distorting. As a general trend, we note that the economic gains become greater for larger values of  $\mu_p$  and  $\lambda$ .

#### 4 MCMC CONVERGENCE ISSUES AND ROBUSTNESS

Our model with noise is a nonstationary latent Gaussian Bayesian model with Gaussian response variables and we use an MCMC approach for inference. It is

well known that MCMC tends to exhibit poor performance when applied to these models. The first reason is that the different points of the latent process  $\mathbf{X}$  are strongly dependent on each other. Second, the latent process and  $\Sigma$  are also strongly dependent, especially in large sample settings as ours. This is known in the literature as *Roberts–Stramer critique* (Roberts and Stramer, 2001) and is formalized by noting that

$$\text{plim}_{\Delta t \rightarrow 0} \sum_{t=1}^T (\mathbf{X}_t - \mathbf{X}_{t-1})' (\mathbf{X}_t - \mathbf{X}_{t-1}) = \Sigma. \tag{11}$$

This asymptotic relationship between  $\Sigma$  and  $\mathbf{X}$  causes, in the limit, the Markov Chain to be reducible, that is unable to escape from the current value.

A common approach to overcome the strong posterior dependence *within* the latent process, is to sample the whole process  $\mathbf{X}$  jointly. This is what we do in (9) by using the FFBS algorithm of Fruwirth-Schnatter (1994) and the simulation study results suggest that our Gibbs estimator is not severely affected by the Roberts–Stramer reducibility problem. This happens thanks to the double augmentation scheme, since the sampling of the missing values of the noisy observed process adds a disturbance element to the sampling of the latent process, breaking the deterministic relation between  $\Sigma$  and  $\mathbf{X}$ . Our doubly augmented algorithm naturally complements the traditional Gibbs sampler that does not sample  $\mathbf{Y}^{\text{miss}}$ : the convergence speed of the traditional Gibbs sampler increases when the signal-to-noise ratio of the problem is higher (Roberts and Sahu, 1997), while in our augmented Gibbs sampler, as the observational noise increases, sampling of  $\mathbf{X}$  will be more “disturbed” by sampling of  $\mathbf{Y}^{\text{miss}}$ , improving the performance of the Bayesian estimator.

We investigate the robustness of our methodology in two ways: we increase the noise, everything else being fixed, and look at simulation results to validate the hypothesis that with higher noise, the Gibbs sampler works relatively better. Then, to test the robustness to a finer grid, that is to a higher number of points between observations in the latent process, we increase the missing percentage, holding everything else constant.

With an increase in noise,  $\mathbf{Y}^{\text{miss}}$  introduces an higher “disturbance” to the deterministic relation (11) between the latent process and the covariance matrix. Since the reducibility problem is attenuated, it is natural to expect that the Bayesian estimation improves relative to the alternative methodologies. In Figure 2c, we report the results for the different methodologies compared, with noise variance  $\Omega_1 = \Omega + 0.35I$ , and we indeed note that there is an improvement in favor of the Bayesian approach. As expected, HY is severely distorted, since not robust to microstructure noise, and only AFX resists close after the Gibbs method. The estimated expected values and standard deviations of the Frobenius distances are reported in Table 4. Figure 2c compares the RMSEs and components of the methodologies and Figure 3c plots the Frobenius distance kernel densities.

The increase in missing percentage has the effect of increasing the number of points of the latent process to be estimated, severing the reducibility problem. Still, with missing probabilities equal to the missing percentages of the standard multivariate case plus 0.35 for each assets (missing percentages from 60% to 85%), the simulation study shows superior results for the Bayesian estimator as reported in Figures 2d, 3d, and in Table 4, with HY resisting better, since it does not use synchronization schemes that are heavily affected by high missing values. We thus conclude that the other estimation procedures deteriorate as or more rapidly than ours when the market conditions become more severe.

Finally, we add simultaneously more variance noise and more missings and we report the results in Figures 2e, 3e, and in Table 4, confirming the higher robustness of the Bayesian methodology to more extreme market conditions.

## 5 CONCLUSIONS

In this article, we study the problem of estimation of the multivariate integrated covariance matrix of noisy and asynchronous observations. The Dynamic Linear Model is the setting chosen to deal with presence of microstructure noise in the data, and we treat the asynchronous time series as synchronous series with missing observations. The Bayesian approach allows us to deal with missing observations (asynchronicity) by treating them as additional parameters of the problem. An augmented Gibbs algorithm is implemented to sample the covariance matrix, the observational error variance matrix, the latent process, and the missing observations of the noisy process. Our MCMC estimator is positive definite by construction and we compare it with several alternatives available in the literature. A simulation study suggests that our estimator is superior in terms of RMSE in a 10-dimensional setting, even with dispersed and high missing percentages and with high noise. This suggests that our Bayesian estimator performs better in normal and severe conditions, as with portfolios of assets with heterogeneous liquidity profiles, or particularly illiquid, or when there is a high level of microstructure noise in the market. Since noise variances are usually estimated to be rather small in large samples, the outperformance of the Bayesian estimator is less crucial when large amount of data are available, still it is indicative of good small sample performances. The application of our methodology to an empirical dataset composed of ten tick-by-tick series of stock prices seems to confirm simulation results by showing that the covariance matrix of the proposed estimator produces a vector of standardized returns closer to a multivariate standard normal than the ones obtained with alternative methods, and that there is a concrete economic advantage in using our Bayesian estimator.

As possible extension, our methodology could be applied to factors of much larger portfolios. Furthermore, we could extend the simulation algorithm to adaptive MCMC samplers (Haario, Saksman, and Tamminen, 2001) or to Particle MCMC methods (Andrieu, Doucet, and Holenstein, 2010) to face the problem of

lost conjugacy of the full conditionals in the case of nonlinear and nonnormal measurement and transition equations.

### APPENDIX A: FINITENESS OF THE POSTERIOR

We check, first without missings and then with missings, that the posterior distribution of  $\Sigma$  is proper. Without missings, proving  $\int p(\Sigma|Y)d\Sigma < \infty$  is equivalent to prove

$$p(Y) = \int p(Y|\Sigma, \Omega, X)p(\Sigma, \Omega, X)d\Sigma d\Omega dX < \infty$$

when  $p(Y|\Sigma, \Omega, X) = \prod_{t=1}^T N_d(y_t; x_t, \Omega) < \infty$  by the assumed state space model and  $\Sigma, \Omega,$  and  $X$  are a priori independent. We denote with  $N_d(y_t; x_t, \Omega)$  the  $d$ -dimensional Gaussian distribution, with location  $x_t$  and scale matrix  $\Omega$ . A priori,  $p(\Sigma) \propto |\Sigma|^{-\frac{d+1}{2}}, p(\Omega) \propto |\Omega|^{-\frac{d+1}{2}},$  and  $p(X) \propto c,$  so that their joint distribution  $p(\Sigma, \Omega, X) \propto (|\Sigma||\Omega|)^{-\frac{d+1}{2}}$  is finite. For positive-definite  $\Sigma$  and  $\Omega,$

$$p(Y) = \int p(Y|\Sigma, \Omega, X)p(\Sigma, \Omega, X)d\Sigma d\Omega dX \tag{A1}$$

$$= \int \prod_{t=1}^T N_d(y_t; x_t, \Omega)(|\Sigma||\Omega|)^{-\frac{d+1}{2}} d\Sigma d\Omega dX \tag{A2}$$

$$\propto \int |\Omega|^{-T/2} \exp\left\{-\frac{1}{2} \sum_{t=1}^T (y_t - x_t)' \Omega^{-1} (y_t - x_t)\right\} (|\Sigma||\Omega|)^{-\frac{d+1}{2}} d\Sigma d\Omega dX$$

$$= \int |\Omega|^{-\frac{T+d+1}{2}} \exp\left\{-\frac{1}{2} \sum_{t=1}^T (y_t - x_t)' \Omega^{-1} (y_t - x_t)\right\} d\Omega dX \int |\Sigma|^{-\frac{d+1}{2}} d\Sigma$$

$$\leq \left(\min_{i \in \{1, \dots, d\}} (\lambda_{\Sigma, i})\right)^{-\frac{d(d+1)}{2}} \int |\Omega|^{-\frac{T+d+1}{2}} \exp\left\{-\frac{1}{2} \sum_{t=1}^T (y_t - x_t)' \Omega^{-1} (y_t - x_t)\right\} d\Omega dX$$

$$= \left(\min_{i \in \{1, \dots, d\}} (\lambda_{\Sigma, i})\right)^{-\frac{d(d+1)}{2}} \int |\Omega|^{-\frac{d+1}{2}} \int_{\mathbb{R}^{dT}} |\Omega|^{-T/2} \exp\left\{-\frac{1}{2} \sum_{t=1}^T (y_t - x_t)' \Omega^{-1} (y_t - x_t)\right\} dX d\Omega$$

$$\propto \left(\min_{i \in \{1, \dots, d\}} (\lambda_{\Sigma, i})\right)^{-\frac{d(d+1)}{2}} \int |\Omega|^{-\frac{d+1}{2}} d\Omega$$

$$\leq \left(\min_{i \in \{1, \dots, d\}} (\lambda_{\Sigma, i})\right)^{-\frac{d(d+1)}{2}} \left(\min_{i \in \{1, \dots, d\}} (\lambda_{\Omega, i})\right)^{-\frac{d(d+1)}{2}} < \infty,$$

where  $\lambda_{A, i}$  is the  $i$ -th eigenvalue of the matrix  $A$ . We have used the inequality  $|A| = \prod_{i=1}^d \lambda_{A, i} \geq (\min_{i \in \{1, \dots, d\}} (\lambda_{A, i}))^d$ . Note that for singular  $\Sigma,$  the result does not hold since  $(\min_{i \in \{1, \dots, d\}} (\lambda_{\Sigma, i})) = 0$ . The same is true for singular  $\Omega$ .

When missings are introduced, the relation to prove becomes

$$p(Y^o) = \int p(Y^o | \Sigma, \Omega, X, Y^m) p(\Sigma, \Omega, X, Y^m) d\Sigma d\Omega dXdY^m d\Sigma < \infty,$$

where  $p(\Sigma, \Omega, X, Y^m) \propto (|\Sigma||\Omega|)^{-\frac{d+1}{2}}$  and

$$p(Y^o | \Sigma, \Omega, X, Y^m) \propto \prod_{i=1}^d \prod_{t \in \mathcal{I}_i^o} v_{i,t}^{-1} \exp \left\{ -\frac{1}{2} v_{i,t}^{-1} (x_{i,t} - m_{i,t})^2 \right\}$$

with  $v_{i,t} := \tilde{\sigma}_{i,t}^2 - \tilde{\Sigma}_{i,t} \Sigma_{-i,t}^{-1} \tilde{\Sigma}'_{i,t}$ ,  $\tilde{\Sigma}_t = V_t^{t-1} + \Omega$ ,  $m_{i,t} := x_{i,t-1} + \tilde{\Sigma}_{i,t} \Sigma_{-i,t}^{-1} (x_{i,t} - x_{i,t-1})$  and  $V_t^{t-1}$  is the prediction error variance at time  $t$  obtained through Kalman iterations. Computations can be strongly simplified thanks to the relation  $p(Y^o) = \int p(Y^o, Y^m) dY^m = \int p(Y) dY^m$ :

$$\begin{aligned} p(Y^o) &= \int \prod_{t=1}^T N_d(y_t; x_t, \Omega) (|\Sigma||\Omega|)^{-\frac{d+1}{2}} d\Sigma d\Omega dXdY^m \\ &\leq \left( \min_{i \in \{1, \dots, d\}} (\lambda_{\Sigma, i}) \right)^{-\frac{d(d+1)}{2}} \int |\Omega|^{-\frac{d+1}{2}} d\Omega dY^m \\ &\leq \left( \min_{i \in \{1, \dots, d\}} (\lambda_{\Sigma, i}) \right)^{-\frac{d(d+1)}{2}} \left( \min_{i \in \{1, \dots, d\}} (\lambda_{\Omega, i}) \right)^{-\frac{d(d+1)}{2}} < \infty. \end{aligned}$$

Similar calculations, not reported for brevity, show that the posterior distribution of  $\Omega$  is also proper.

Received June 14, 2012; revised April 4, 2014; accepted May 16, 2014.

**REFERENCES**

Ait-Sahalia, Y., J. Fan, and D. Xiu. 2010. High-Frequency Covariance Estimates with Noisy and Asynchronous Financial Data. *Journal of the American Statistical Association* 105: 1504–1517.

Ait-Sahalia, Y., P. Mykland, and L. Zhang, 2005. How Often to Sample a Continuous-Time Process in the presence of Market Microstructure Noise. *The Review of Financial Studies* 18: 351–416.

Andersen, T. G., T. Bollerslev, F. Diebold, and H. Ebens. 2001a. The Distribution of Realized Stock Return Volatility. *Journal of Financial Economics* 61: 43–76.

Andersen, T. G., T. Bollerslev, F. Diebold, and P. Labys. 2001b. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96: 42–55.

- Andersen, T. G., T. Bollerslev, F. Diebold, and P. Labys. 2000. Great Realizations. *Risk* 13: 105–108.
- Andersen, T. G., T. Bollerslev, F. Diebold, and P. Labys. 2003. Modeling and Forecasting Realized Volatility. *Econometrica* 71: 579–625.
- Andersen, T. G., T. Bollerslev, and D. Dobrev. 2007. No-Arbitrage Semi-Martingale Restrictions for Continuous-Time Volatility Models Subject to Leverage Effects, Jumps and i.i.d. Noise: Theory and Testable Distributional Implications. *Journal of Financial Econometrics* 138: 125–180.
- Andrieu, C., A. Doucet, and R. Holenstein. 2010. Particle Markov Chain Monte Carlo Methods. *Journal of Royal Statistical Society B* 72: 269–342.
- Ane, T., and H. Geman. 2000. Order Flow, Transaction Clock, and Normality of Asset Returns. *The Journal of Finance* 55: 2259–2284.
- Audrino, F., and F. Trojani. 2006. A General Multivariate Threshold GARCH Model with Dynamic Conditional Correlations. *Journal of Business and Economic Statistics* 29: 138–149.
- Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard. 2011. Multivariate Realized Kernels: Consistent Positive Semi-Definite Estimator of the Covariation of Equity Prices With Noise and Non-Synchronous Trading. *Journal of Econometrics* 162: 149–169.
- Bibinger, M. 2011. Efficient Covariance Estimation for Asynchronous Noisy High-Frequency Data. *Scandinavian Journal of Statistics* 38: 23–45.
- Chiriac, R., and V. Voev. 2010. Modelling and Forecasting Multivariate Realized Volatility. *Journal of Applied Econometrics* 26: 922–947.
- Christoffersen, P., S. Heston, and K. Jacobs. 2009. The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work So Well. *Management Science* 55: 1914–1932.
- Clark, P. 1973. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica* 41: 135–155.
- Corsi, F., and F. Audrino, 2012. Realized Covariance Tick-by-Tick in Presence of Rounded Time Stamps and General Microstructure Effects. *Journal of Financial Econometrics* 10: 591–616.
- Corsi, F., S. Peluso, and F. Audrino. 2014. Missing in Asynchronicity: A Kalman-em Approach for Multivariate Realized Covariance Estimation. *Journal of Applied Econometrics*. Forthcoming.
- Durbin, P. 2004. *Introduction to State Space Time Series Analysis*. Cambridge: Cambridge University Press.
- Ellickson, B., B. Hood, T. Liu, D. Whang, and P. Zhou. 2011. Stocks in the Short run. Available at SSRN 1868544.
- Engle, R., and R. Colacito. 2006. Testing and Valuing Dynamic Correlations for Asset Allocation. *Journal of Business and Economic Statistics* 24: 238–253.
- Epps, T. 1979. Comovements in Stock Prices in the Very Short Run. *Journal of the American Statistical Association* 74: 291–296.
- Fleming, J., C. Kirby, and B. OstDiek. 2001. The Economic Value of Volatility Timing. *Journal of Finance* 56: 329–352.

- Fleming, J., C. Kirby, and B. Ostdiek. 2003. The Economic Value of Volatility Timing Using Realized Volatility. *Journal of Financial Economics* 67: 473–509.
- Frigessi, A., A. Loland, A. Pievatolo, and F. Ruggieri. 2010. Statistical Rehabilitation of Improper Correlation Matrices. *Quantitative Finance*. iFirst.
- Fruwirth-Schnatter, S. 1994. Data Augmentation and Dynamic Linear Models. *Journal of Time Series Analysis* 15: 183–202.
- Gauthier, P., and D. Possamai. 2010. Efficient Simulation of the Double Heston Model. *preprint SSRN*.
- Griffin, J., and R. Oomen. 2009. Covariance Measurement in the Presence of Non-Synchronous Trading and Market Microstructure Noise. *Journal of Econometrics* 160: 58–68.
- Haario, H., E. Saksman, and J. Tamminen. 2001. An Adaptive Metropolis Algorithm. *Bernoulli* 7: 223–242.
- Harvey, A. 1989. *Forecasting Structural Time Series Model and the Kalman Filter*. Cambridge: Cambridge University Press.
- Hautsch, N., L. M. Kyj, and R. C. Oomen. 2012. A Blocking and Regularization Approach to High-Dimensional Realized Covariance Estimation. *Journal of Applied Econometrics* 27: 625–645.
- Hayashi, T., and N. Yoshida. 2005. On Covariance Estimation of Non-Synchronously Observed Diffusion Processes. *Bernoulli* 11: 359–379.
- Heston, S. 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies* 6: 327–343.
- Jeffreys, H. 1946. An Invariant Form of the Prior Probability in Estimation Problems. *Proceedings of Royal Statistical Society, Series A* 186: 453–461.
- Ledoit, O., and M. Wolf. 2003. Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance* 10: 603–621.
- Lunde, A., N. Shephard, and K. Sheppard. 2012. Econometric Analysis of Vast Covariance Matrices using Composite Realized Kernels. *Unpublished paper: Department of Economics, University of Oxford*.
- Malliavin, P., and M. E. Mancino. 2009. A Fourier Transform Method for Nonparametric Estimation of Multivariate Volatility. *The Annals of Statistics* 37: 1983–2010.
- Mancino, M. E., and S. Sanfelici. 2011. Estimating Covariance Via Fourier Method in the Presence of Asynchronous Trading and Microstructure Noise. *Journal of financial econometrics* 9: 367–408.
- Marčenko, V. A., and L. A. Pastur. 1967. Distribution of Eigenvalues for Some Sets of Random Matrices. *Sbornik: Mathematics* 1: 457–483.
- Meng, X., and D. van Dyk. 1997. The EM Algorithm, An Old Folk-Song Sung to a fast New Tune. *Journal of Royal Statistical Society. Series B* 59: 511–567.
- Oksendal, B. 2002. *Stochastic Differential Equations. An Introduction with Applications*. New York: Springer.



- Palandri, A. 2006. Consistent Realized Covariance for Asynchronous Observations Contaminated by Market Microstructure Noise. Technical report, University of Copenhagen.
- Park, S., and O. Linton. 2012. Estimating the Quadratic Covariation Matrix for an Asynchronously Observed Continuous Time Signal Masked by Additive Noise. *FMG Discussion Papers* 703.
- Roberts, G., and S. Sahu. 1997. Updating Schemes, Correlation Structure, Blocking and Parameterization for the Gibbs Sampler. *Journal of Royal Statistical Society B* 59: 291–317.
- Roberts, G., and O. Stramer, 2001. On Inference for Partially-Observed Nonlinear Diffusion Models Using the Metropolis-Hastings Algorithm. *Biometrika* 88: 603–621.
- Scholes, M., and J. Williams. 1977. Estimating Betas from Nonsynchronous Data. *Journal of Financial Economics* 5: 181–212.
- Shephard, N., and D. Xiu. 2012. Econometric Analysis of Multivariate Realised qml: Efficient Positive Semi-Definite Estimators of the Covariation of Equity Prices. *Chicago Booth Research Paper*.
- Tanner, M., and W. Wong. 1987. The Calculation of Posterior Distributions by Data Augmentation (with discussion). *Journal of American Statistical Association* 82: 528–550.
- Triantafyllopoulos, K. 2011. Real-Time Covariance Estimation for the Local Level Model. *Journal of Time Series Analysis* 32: 93–107.
- Voev, V., and A. Lunde. 2007. Integrated Covariance Estimation using High Frequency Data in the Presence of Noise. *Journal of Financial Econometrics* 5: 68–104.
- West, K., H. Edison, and D. Cho. 1993. A Utility-Based Comparison of Some Models of Exchange Rate Volatility. *Journal of International Economics* 35: 23–45.
- West, M., and P. Harrison. 1997. *Bayesian Forecasting and Dynamic Models*. New York: Springer.
- Xiu, D. 2010. Quasi-Maximum Likelihood Estimation of Volatility With High Frequency Data. *Journal of Econometrics* 159: 235–250.
- Zhang, L., P. Mykland, and Y. Ait-Sahalia. 2005. A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data. *Journal of the American Statistical Association* 100: 1394–1411.
- Zhang, L. 2011. Estimating Covariation: Epps Effect, Microstructure Noise. *Journal of Econometrics* 160: 33–47.