A bivariate fractional stochastic volatility model

Ranieri Dugo* Giacomo Giorgio† Paolo Pigato‡

Abstract

Motivated by empirical evidence from the joint behavior of realized volatility time series, we propose to model the joint dynamics of log-volatilities using a bivariate fractional Ornstein-Uhlenbeck (2fOU) process. This is a mean reverting Gaussian process with fractal features living in $\mathbb{R}^2$. It is the solution of the Langevin equation with the multivariate fractional Brownian motion, in the sense of Amblard et al. (2012), as driving term. This model is a multivariate version of the Rough Fractional Stochastic Volatility (RFSV) model proposed by Gatheral, Jaisson, and Rosenbaum (2018).

We discuss the main features of the process and propose different estimation procedures to identify its parameters. The first is a two-step method that takes as given the parameters governing the univariate marginals, whereas the second method identifies all the parameters at once. Regarding the first one, the estimation of the univariate process is well documented in the literature, but it often presents a bias in the speed of mean-reversion parameter, an issue that we also try to overcome. We derive the asymptotic properties of the estimators and compare the performance of the finite sample behavior to the asymptotic theory with Monte Carlo experiments.

Finally, an empirical investigation is carried out on 7 realized volatility time series overlapping in time and available on a long sampling period. Our results show how realized volatility time series are strongly correlated and present different degrees of asymmetry in their cross-covariance structure, which can be linked to what are known as spillover effects. These features can all be well captured by our model. Moreover, in accordance with the existing literature, we observe behaviors close to nonstationarity and roughness in the trajectories. A forecasting exercise provides further evidence of the consistency of our model with observed data.

1 Introduction

Over the past few years, modeling realized volatility with log-normal fractional processes in continuous time has gained some popularity. The most common adopted specification is the one in which the log-volatility process is the solution to the Langevin equation driven by fractional Brownian motion B^H (Mandelbrot and

*Department of Economics and Finance, University of Rome Tor Vergata, dugoranieri@gmail.com
†Department of Mathematics, University of Rome Tor Vergata
‡Department of Economics and Finance, University of Rome Tor Vergata

Van Ness 1968)

$$dY_t = \alpha(\mu - Y_t)dt + \nu dB_t^H.$$

The process Y is known as the fractional Ornstein-Uhlenbeck process and has been studied extensively by Cheridito, Kawaguchi, and Maejima (2003). The first time that a similar version of such process has appeared in volatility modeling was due to Comte and Renault (1998). The authors employed such specification to allow for long-memory in continuous-time stochastic volatility modeling.

Gatheral, Jaisson, and Rosenbaum (2018) adopt the specification above in their Rough Fractional Stochastic Volatility model in which they argue for rough trajectories, meaning $H < \frac{1}{2}$, and a very slow mean reverting behavior of Y due to $\alpha \to 0$, thus justifying the fact that log-volatility behaves locally as a fractional Brownian motion. Wang, Xiao, and Yu (2021) propose an estimation procedure for the fractional Ornstein-Uhlenbeck process with the method of moments, develop the related asymptotic theory of the estimators, and perform a forecasting exercise. In forecasting, the model under study is able to outperform all the considered benchmarks for a number of volatility time series of US stock indices. Bolko et al. (2022) develop a Generalized Method of Moments estimation procedure utilizing the moments of integrated variance as orthogonality conditions, making use for the first time of the GMM procedure for fractional stochastic volatility models. The authors also consider how microstructure noise affects estimation.

Some other related papers do not use the fractional Ornstein-Uhlenbeck process but fractional processes of similar nature. Bennedsen et al. (2022) introduce a class of models for log-volatility that can, but do not have to, be Gaussian, according to the specification. Such models allow to decouple the short-time property of the trajectories and the long-time memory structure of the process, which were previously determined together by H. The authors still find rough trajectories and good forecasting performance.

Starting from Bayer et al. (2016), rough volatility models have also been widely used in option pricing.

Motivated by previously mentioned literature and the joint dynamics of realized volatility time series, we propose an extension to the bivariate setting of the fractional Ornstein-Uhlenbeck process for volatility modeling, develop an estimation procedure, and perform extensive empirical and forecasting studies.

As shown in Figure 1, volatility time series are characterized by different self-similarity coefficients. These coefficients are related to the regularity of the trajectories as well as the memory structure of the processes. With regard to the memory structure of the process, it is customary in univariate models to look at the decay of the auto-covariance, but in our bivariate setting, it also enters in the picture how the cross-covariances decay for large lags.

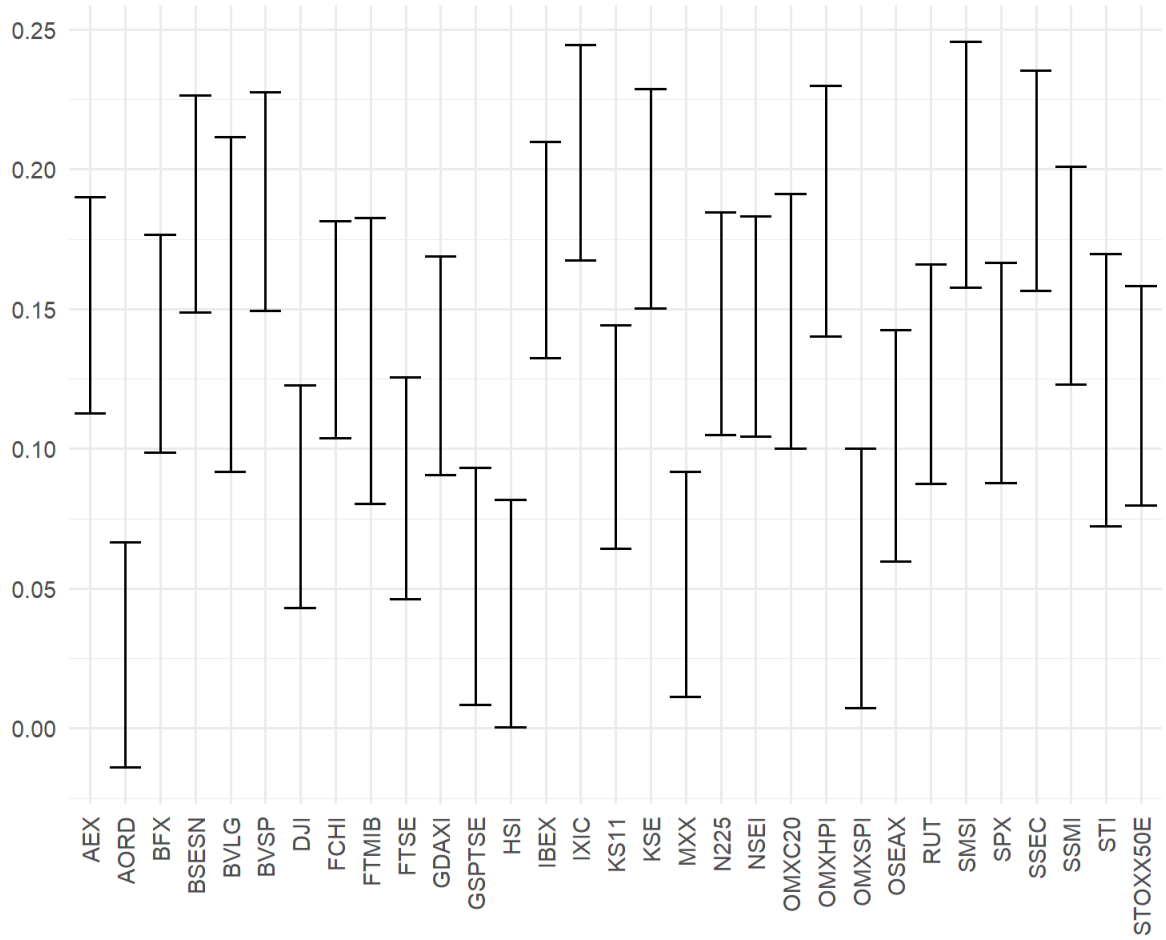

Figure 1: Estimates of H across realized volatility time series using the COF estimator.

In order to allow for different self-similarity coefficients in the single components of our bivariate model, we decide to use as a driving term in the Langevin equation the fractional Brownian motion in the sense of Amblard et al. (2010). Doing so, we are able to decouple for each marginal component the regularity of the trajectory as well as its memory structure, as captured by the autocovariance. However, when looking at the bivariate model, both self-similarity coefficients play together a role in determining the behavior of the cross-covariance function.

After introducing the model, we outline possible estimation strategies that rely on the method of moments. In particular, one estimation strategy relies on a two-step procedure in which the parameters of the unviariate marginals are first estimated according to some methodology and then the parameters ruling the joint dynamics are determined by the inversion of the cross-covariance function. Another estimation strategy attempts the joint identification of all parameters of the model by employing the Generalized Method of Moments procedure. For both procedures we develop an asymptotic theory. Both procedures seem to perform well in finite sample according to our Monte Carlo experiments.

Using these estimators, we perform an extensive empirical analysis on all pairs of realized volatility time series of 7 major equity indices obtained from the Realized library of the Oxford-Man institute. In line with Gatheral, Jaisson, and Rosenbaum (2018), we find that log-volatility mean-reverts slowly and can be well approximated locally by a version of the fractional Ornstein-Uhlenbeck process taht mean-reverts very weakly, or similarly by the multivariate fractional Brownian motion, which is non-stationary altogether. In addition, we find strong empirical support in favor of our model in the fit of the theoretical cross-covariances obtained with the estimated parameters to the empirical ones.

Finally, we perform a forecasting exercise based on standard results for Gaussian vectors. The forecasting is based both on the almost non-stationary bivariate fractional Ornstein-Uhlenbeck process and the bivariate fractional Brownian motion. We find that our model has a forecasting accuracy in line with the Heterogeneous autoregressive model by Corsi (2009).

The article is structured as follows: Section 2 introduces the model starting from the definition of the multivariate fractional Brownian motion; Section 3 contains the identification procedures for the parameters and their asymptotic theories; in Section 4 we test these estimation procedures on simulation; Section 5 contains the empirical analysis consisting of the estimation of the model on realized volatility time series and evaluation of the fit of the so-obtained theoretical cross-covariance function; the forecasting exercise is carried out in Section 6; Section 7 concludes. Some calculations useful for the proofs of the asymptotic theory are in Appendix A. Throughout the paper, we use $\xrightarrow{p}$, $\xrightarrow{d}$, and $\overset{d}{=}$ to denote convergence in probability, convergence in distribution, and equivalence in distribution, respectively.

2 The Model

The proposed model for the logarithm of realized volatility is a continuous-time model described by a multivariate stochastic differential equation (SDE) driven by the fractional Brownian motion. The constituents of this framework are the Langevin equation, giving the volatility model the property of mean-reversion, long advocated since Fouque, Papanicolaou, and Sircar (2000), and a random noise driving the dynamics. The noise of our choice is the multivariate fractional Brownian motion in the sense of Amblard et al. (2010). The resulting log-volatility dynamics is Gaussian (empirically valid hypothesis according to Andersen et al. 2001) and fractional. The latter is an important property, attributed to volatility time series since Ding and Granger (1996).

The resulting model is a generalization of the fractional Ornstein-Uhlenbeck process by Cheridito, Kawaguchi, and Maejima (2003), adopted in the seminal paper by Gatheral, Jaisson, and Rosenbaum (2018) for rough volatility modeling.

2.1 The multivariate fractional Brownian motion

The multivariate fractional Brownian motion introduced by Amblard et al. (2010) is a vector-valued Gaussian process $B \in \mathbb{R}^d$, with cross-covariance function

$$\begin{aligned} E\left(B_s^{H_i} B_t^{H_j}\right) = & \frac{\sigma_i \sigma_j}{2} \{ \left(\rho_{i,j} + \eta_{i,j} \text{sign}(t)\right) |s|^{H_i + H_j} \\ & + \left(\rho_{i,j} - \eta_{i,j} \text{sign}(s)\right) |t|^{H_i + H_j} \\ & - \left(\rho_{i,j} - \eta_{i,j} \text{sign}(t-s)\right) |t-s|^{H_i + H_j} \}, \end{aligned} \tag{1}$$

where $H_i + H_j \neq 1$ (alternative formulation for $H_i + H_j = 1$ available in Appendix 2), $\rho_{ij} \in [-1, 1]$, $\rho_{ij} = \rho_{ji} = \text{Cor}(B_1^{H_i}, B_1^{H_j})$, $\eta_{ij} \in \mathbb{R}$, $\eta_{ij} = -\eta_{ji}$. While ρ_{ij} is a contemporaneuos correlation coefficient, η_{ij} determines the time-reversibility of the process. By time reversibility we mean the property that $(B_t^H)_{t \in \mathbb{R}} \overset{d}{=} (B_{-t}^H)_{t \in \mathbb{R}}$ $\forall i, j = 1, \ldots, d$, which happens only for $\eta_{i,j} = 0$.
The parameters of the model need to satisfy the constraint

$$C_{ij} = \frac{\Gamma H + 1^2}{\Gamma(2H_1 + 1)\Gamma 2H_2 + 1} \frac{\rho_{12}^2 \sin^2\left(\frac{\pi}{2}(H_1 + H_2)\right) + \eta_{12}^2 \cos^2\left(\frac{\pi}{2}(H_1 + H_2)\right)}{\sin \pi H_1 \sin \pi H_2} \leq 1, \forall i, j = 1, \ldots, d \tag{2}$$

in order to have a semipositive definite covariance matrix (Equation (1)). In the following, we will refer to the Expression (2) as the Coherency constraint.
When this constraint is satisfied, the process is well defined and enjoys stationary increments. However, the value of each increment of the process depends on its entire history of realizations. Moreover, the correlation decay of such increments is a power law in the time lag as time lag goes to infinity. The power law can be integrable or not, depending on the value of $H_1 + H_2$. These properties are carried on to the Ornstein-Uhlenbeck dynamics, as we will see them in the following section.

2.1 The multivariate fractional Ornstein-Uhlenbeck process

Let us now consider a multivariate fractional Brownian motion $(B_t^H)_{t \geq 0}$ with dimension $d = 2$. The bivariate fractional Ornstein-Uhlenbeck (2fOU) process is defined as the vector $Y_t = (Y_t^1, Y_t^2) \in \mathbb{R}^2$ whose elements solve pathwise the Langevin equation

$$dY_t^i = \alpha_i(\mu_i - Y_t^i)dt + \nu_i dB_t^{H,i}, \qquad i = 1, 2 \tag{3}$$

where $\mu_i \in \mathbb{R}$ is the long-term mean, $\nu_i > 0$ is the diffusion coefficient, $H_i \in (0, 1)$ the Hurst coefficient of the $i - th$, $i = 1, 2$ marginal one dimensional component of the bivariate fractional Brownian motion, and $\alpha_i > 0$ is the speed of mean-reversion coefficient.

Also, the integral with respect to the fractional Brownian motion has to be intended as a Riemann-Stieltjes integral (see Cheridito, Kawaguchi, and Maejima 2003).
The stationary solution to (3) is given by

$$Y_t^i = \mu_i(1 - e^{-\alpha_i t}) + \nu_i \int_\infty^t e^{-\alpha_i(t-s)} dB_s^{H,i}, \tag{4}$$

.

2.1.1 Properties

The 2fOU process is Gaussian and stationary. It admits time reversibility, i.e. $(Y_t)_{t\in\mathbb{R}} \overset{d}{=} (Y_{-t})_{t\in\mathbb{R}}$, when $\eta_{12} := \eta = 0$ and $\alpha_1 = \alpha_2$, conditions that imply symmetry in the cross-covariance function (ccf) of the process.
The expected value of the process at time t is

$$\mathbb{E}\left[(Y_t^1, Y_t^2)\right] = (\mu_1(1 - e^{-\alpha_1 t}), \mu_2(1 - e^{-\alpha_2 t})). \tag{5}$$

Its lagged cross-covariance function is

$$\mathrm{Cov}\left(Y_t^i, Y_{t+k}^j\right) = e^{-\alpha_j k}\mathrm{Cov}\left(Y_t^i, Y_t^j\right) + \nu_1\nu_2 e^{-\alpha_j k} H(H-1)\frac{\rho - \eta}{2} \int_0^s e^{\alpha_j v}\left(\int_{-\infty}^0 e^{\alpha_i u}(v-u)^{H-2} du\right) dv \tag{6}$$

where the contemporeaneous cross-covariance, or simply covariance, is given by

$$\mathrm{Cov}\left(Y_t^i, Y_t^j\right) = \frac{\Gamma(H+1)\nu_1\nu_2}{2(\alpha_1 + \alpha_2)}\left(\left(\alpha_1^{1-H} + \alpha_2^{1-H}\right)\rho + \left(\alpha_2^{1-H} - \alpha_1^{1-H}\right)\eta\right). \tag{7}$$

Two asymptotic properties of the cross-covariance function that will be useful for our later purposes (see Dugo, Giorgio, and Pigato 2024 for further detail) are given in the following.
For brevity, let us denote in the following $\gamma_{i,j}(k) := \mathrm{Cov}\left(Y_t^i, Y_{t+k}^j\right)$.

Proposition 1 (Large-lag asymptotics)**.** *Let $H_1, H_2 \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$ and $k \in \mathbb{R}$. Then for fixed $t \in \mathbb{R}$, when $k \to \infty$*

$$\gamma_{i,j}(k) \sim k^{H_1+H_2-2}. \tag{8}$$

where $f(k) \sim g(k)$ means $\frac{f(k)}{g(k)} \to C,\ C \in \mathbb{R}_{\setminus\{0\}}$.

Remark 1. *if $\rho \neq 0$ or $\eta \neq 0$ and $\rho \neq \eta$,*

- $H_1 + H_2 > 1 \implies \sum_{k=0}^{\infty} \gamma_{Y^1,Y^2}(k) = \infty$ *(long-range interdependence);*

- $H_1 + H_2 < 1 \implies \sum_{k=0}^{\infty} |\gamma_{Y^1,Y^2}(k)| < \infty$ *(short-range interdependence).*

For further details on the topic of long and short-range dependance refer to Beran et al. (2013).

Proposition 2 (Small-lag asymptotics)**.** *For all $t \in \mathbb{R}$, as $k \to 0$*

$$\gamma_{i,j}(k) = \gamma_{i,j}(0) - \frac{\rho - \eta}{2}\nu_1\nu_2 k^{H_1+H_2} + o(k), \tag{9}$$

where $o(k)$ means that the additional terms $f(k)$ satisfy $\lim_{k\to 0} \frac{f(k)}{k} = 0$.

Proposition 3. *For all $t \in \mathbb{R}$ and $k \in \mathbb{R}$, as $(\alpha_i, \alpha_j) \to (0, 0)$,*

$$\gamma^{\alpha}_{i,j}(k) - \gamma^{\alpha}_{i,j}(0) \to -\frac{\rho - \eta}{2}\nu_1\nu_2 k^{H_1+H_2}, \tag{10}$$

Proof. The previous result can be easily seen by taking the aforementioned limit in (6) and substituting

$$\int_0^s \left(\int_{-\infty}^0 (v-u)^{H-2} du \right) dv = -\frac{k^H}{H(H-1)}.$$

It is interesting to notice that when $\alpha_i, \quad i = 1, 2$ is very small, $\alpha_i << \frac{1}{T}$, the bivariate fractional Ornstein-Uhlenbeck process behaves locally as a bivariate fractional Brownian motion on $[0, T]$. This equivalence is established component-wise in Gatheral, Jaisson, and Rosenbaum (2018). We report the adjusted statement below.

Proposition 4. *Let $B^H = (B^{H_1}, B^{H_2})$ be a bivariate fractional Brownian motion in the sense of Amblard et al. (2010), and $Y^H = (Y^{H_1}, Y^{H_2})$ be a bivariate fractional Ornstein-Uhlenbeck process with parameters $(0, 0) < H = (H_1, H_2) < (1, 1)$, $\alpha = (\alpha_1, \alpha_2) > (0, 0)$, and $\nu = (\nu_1, \nu_2) > (0, 0)$. As $\alpha = (\alpha_1, \alpha_2) \to (0, 0)$*

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left|\left| Y^H_t - Y^H_0 - \nu \odot B^H_t \right|\right| \right] \to 0,$$

where $||\cdot||$ represent the L^2 norm and $\odot$ indicates the Hadamard product.

Proof. The statement follows immediately from the triangle inequality of the Euclidean norm and the proof for the univariate components in Gatheral, Jaisson, and Rosenbaum (2018).

Figure 2 displays sample paths of a bivariate fractional Ornstein-Uhlenbeck process. For different values of H, growing from top to bottom, we have decreasing roughness in the trajectories. See the caption for details regarding the parameters.

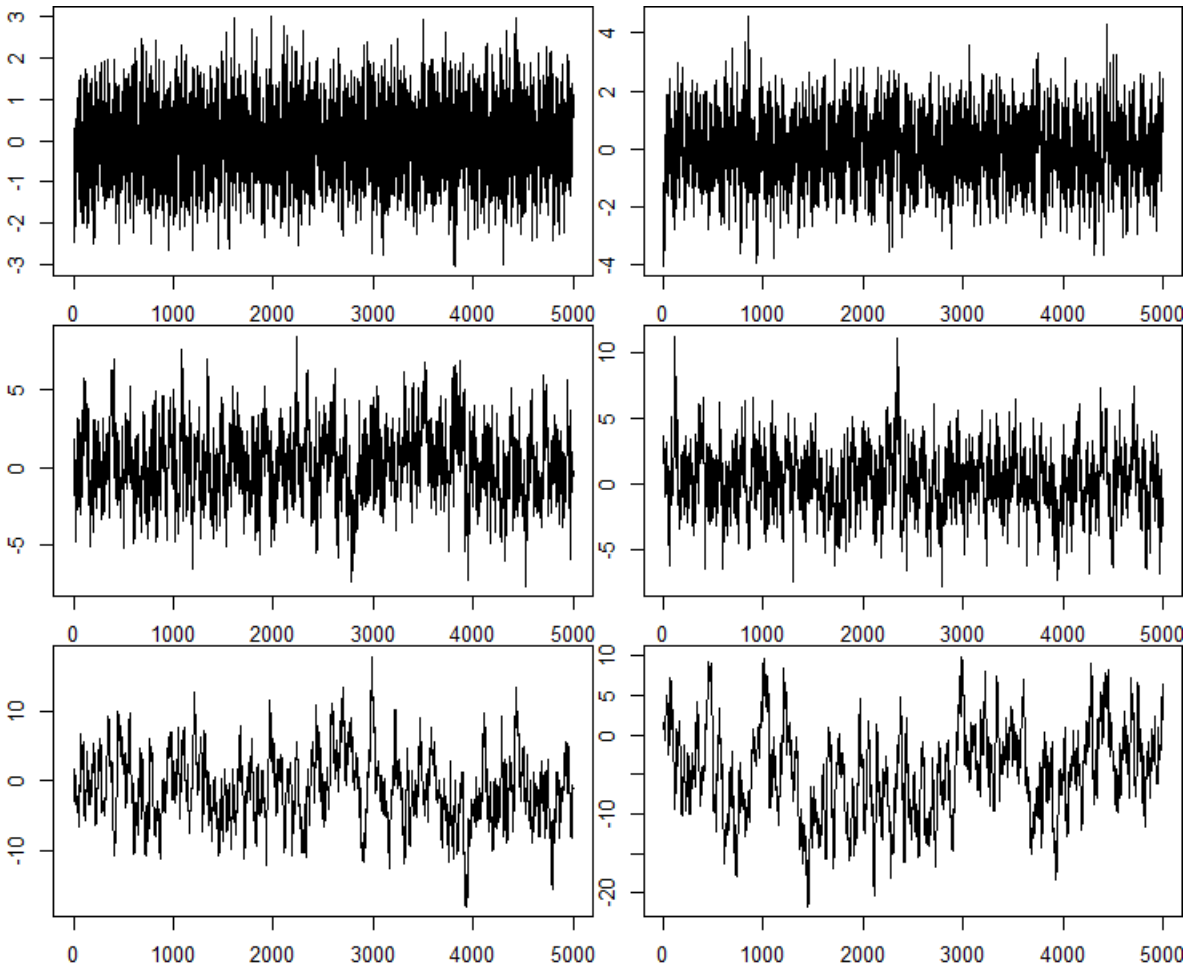

Figure 2: $H_1 = 0.1,\ H_2 = 0.2$ (top); $H_1 = 0.5,\ H_2 = 0.5$ (middle); $H_1 = 0.8,\ H_2 = 0.9$ (bottom)

3 Estimation procedure

In order to estimate our model we rely on the Generalized Method of Moments (GMM). The estimation procedure entails identifying 8 parameters govering the marginal distributions ($\alpha_i,\ \nu_i,\ \mu_i,\ H_i,\ i = 1, 2$) and 2 parameters $\rho$and η that weave the two elements into the multivariate dynamics.

First, we estimate the parameters of the univariate marginals adopting the estimators introduced by Wang, Xiao, and Yu (2021). Then we obtain ρ and η by solving an exactly determined system of two equations and two unknowns. The two equations are given by the cross-rellation function at an arbitrary lag taken with positive and negative sign (see Dugo, Giorgio, and Pigato 2024). The cross-correlations are not observed directly but can be estimated by simple averages from the data.

Then, we use the set of estimates obtained as explained above as starting point for a more general GMM estimation routine in which the system is highly overdetermined. Being the system overdetermined, GMM

relies on a numerical minimization routine of a loss function in order to retrieve the parameters. The orthogonality conditions for the GMM routine correspond to the elements of the cross-covariance function for all lags up to a certain maximum plus some spare ones. A similar approach was adopted by Bolko et al. (2022).

3.1 Two-step estimation

3.1.1 Exact relationship for covariances

Since the cross-correlation function is linear in ρ and η, we can invert it at lags k and $-k$ for given values of $\alpha_i,\ \nu_i,\ H_i,\ i = 1, 2$, and derive explicit expressions for ρ and η,

$$\begin{aligned}\rho =& A_1(k)\mathrm{Cor}(Y_t^1, Y_t^2) + A_2(k)\mathrm{Cor}(Y_{t+k}^1, Y_t^2) + A_3(k)\mathrm{Cor}(Y_t^1, Y_{t+k}^2)\\ \eta =& A_4(k)\mathrm{Cor}(Y_t^1, Y_t^2) + A_5(k)\mathrm{Cor}(Y_{t+k}^1, Y_t^2) + A_6(k)\mathrm{Cor}(Y_t^1, Y_{t+k}^2).\end{aligned} \tag{11}$$

Remark 2. *In the former expressions, $A_i,\ i = 1, \ldots, 6$ are explicit functions of the parameters governing the univariate marginals.*

Assuming Y a zero mean process (without loss of generality), and substituting empirical cross-correlations in place of theoretical ones, we obtain the estimators

$$\begin{aligned}\hat{\rho}_n =& A_1(k)\frac{\frac{1}{n}\sum_{i=1}^{n} Y_i^1 Y_i^2}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i^1)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i^2)^2}} + A_2(k)\frac{\frac{1}{n-k}\sum_{i=1}^{n-k} Y_{i+k}^1 Y_i^2}{\sqrt{\frac{1}{n-k}\sum_{i=1}^{n-k}(Y_{i+k}^1)^2}\sqrt{\frac{1}{n-k}\sum_{i=1}^{n-k}(Y_i^2)^2}}\\ &+ A_3(k)\frac{\frac{1}{n-k}\sum_{i=1}^{n-k} Y_i^1 Y_{i+k}^2}{\sqrt{\frac{1}{n-k}\sum_{i=1}^{n-k}(Y_i^1)^2}\sqrt{\frac{1}{n-k}\sum_{i=1}^{n-k}(Y_{i+k}^2)^2}}\\ \hat{\eta} =& A_4(k)\frac{\frac{1}{n}\sum_{i=1}^{n} Y_i^1 Y_i^2}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i^1)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i^2)^2}} + A_5(k)\frac{\frac{1}{n-k}\sum_{i=1}^{n-k} Y_{i+k}^1 Y_i^2}{\sqrt{\frac{1}{n-k}\sum_{i=1}^{n-k}(Y_{i+k}^1)^2}\sqrt{\frac{1}{n-k}\sum_{i=1}^{n-k}(Y_i^2)^2}}\\ &+ A_6(k)\frac{\frac{1}{n-k}\sum_{i=1}^{n-k} Y_i^1 Y_{i+k}^2}{\sqrt{\frac{1}{n-k}\sum_{i=1}^{n-k}(Y_i^1)^2}\sqrt{\frac{1}{n-k}\sum_{i=1}^{n-k}(Y_{i+k}^2)^2}}.\end{aligned} \tag{12}$$

Due to the ergodicity of the process the estimators defined in (12) possess all the desirable properties.

Theorem 1. *(i) Asymptotic unbiasedness*

$$E\big[(\hat{\rho}_n, \hat{\eta}_n)\big] \to (\rho, \eta) \qquad \text{as } n \to \infty;$$

(ii) Consistency

$$(\hat{\rho}_n, \hat{\eta}_n) \xrightarrow{p} (\rho, \eta) \qquad \text{as } n \to \infty;$$

(iii) Asymptotic normality (CLT): if $H_1 + H_2 < \frac{3}{4}$,

$$\sqrt{n}\big(\hat{\rho}_n - \rho,\ \hat{\eta}_n - \eta\big) \xrightarrow{d} N(0, \Sigma)$$

where

$$\Sigma_{ij} = \sum_{s=0}^{\infty} Cov[G_1(\tilde{Y}_t), G_2(\tilde{Y}_{t+s})] + \sum_{s=0}^{\infty} Cov[G_1(\tilde{Y}_{t+s}), G_2(\tilde{Y}_t)],$$

$i, j = 1, 2$, $G_1, G_2 : \mathbb{R}^3 \to \mathbb{R}$ *homogeneous polynomials defining* ρ *and* η *in terms of the CCF, and* $\tilde{Y}_t = (Y_t^1, Y_t^2, Y_{t+k}^1, Y_{t+k}^2)$*;*

(iv) Asymptotic non-normality (NCLT): if $H_1 \vee H_2 > \frac{3}{4}$,

$$n^{2-2H_1+H_2}\big(\hat{\rho}_n - \rho,\ \hat{\eta}_n - \eta\big) \xrightarrow{d} (Z_1, Z_2),$$

where the distribution of $Z = (Z_1, Z_2)$ *is not clear yet [analogy with Rosenblatt in the univariate case].*

See Dugo, Giorgio, and Pigato (2024) for details and proofs.

Standard errors associated to $\hat{\rho}$ and $\hat{\eta}$

The asymptotic standard errors of the estimators above are difficult to calculate. However, it is possible to obtain finite sample approximations of the standard errors of the estimators $\hat{\rho}$ and $\hat{\eta}$ above by bootstrap. In particular, when $H_1 \vee H_2 < \frac{3}{4}$, Arcones (1994) establishes validity of the Moving Block Bootstrap (Lahiri 1993). Whereas $H_1 \vee H_2 \geq \frac{3}{4}$, subsamplig techniques (Hall, Jing, and Lahiri 1998) could be used to approximate the limiting distribution as well as calculate t-statistics.

3.1.2 Short-lag relationship for covariances

Estimating the speed of mean reversion parameter of an Ornstein-Uhlenbeck process is known in the literature to be very challenging. In many cases the obtained estimates are biased. See for example Phillips and Yu (2009) for the Markovian case, and Wang, Xiao, and Yu (2021) for the non-Markovian one.
The parameters $\alpha_i,\ i = 1, 2$, play in turn a key role in our model in estimating ρ and η for two reasons.
First, the $\alpha_i,\ i = 1, 2$ enter directly the cross-correlation function thus making ρ and η functionally dependent on them.
Second, the inversion of the ccf and isolation of ρ and η requires numeric integrations that see $\alpha_i,\ i = 1, 2$ at the exponents. When $\alpha_i,\ i = 1, 2$ are increasingly small, precision might be lost.

For these reasons, statistical on the one hand, and numerical on the other one, inference on ρ and η might be invalidated by the $\alpha_i,\ i = 1, 2$ from the first step. To avoid this problem, we exploit the asymptotic relationship for the cross-covariance function stated in Proposition 2, which looking at very small lags,

makes the drift become negligible, and makes the functional dependence of ρ and η on $\alpha_i,\ i=1,2$ disappear altoghether in such (asymptotic) regime.

In practice, inverting the two conditions (identified by the choice of i and j or equivalently considering $-k$ and k) of the small-lag relation for an arbitrary value of k,

$$\gamma_{i,j}(k) = \gamma_{i,j}(0) - \frac{\rho - \eta}{2}\nu_1\nu_2 k^{H_1+H_2} + O(k^{H_1+H_2}), \qquad as\ k \to 0 \tag{13}$$

leads to

$$\begin{aligned} \rho &= \frac{2\mathrm{Cov}[Y_t^1, Y_t^2] - \mathrm{Cov}[Y_{t+k}^1, Y_t^2] - \mathrm{Cov}[Y_t^1, Y_{t+k}^2]}{\nu_1\nu_2 k^{H_1+H_2}} + O(k^{H_1+H_2}) \\ \eta &= \frac{\mathrm{Cov}[Y_t^1, Y_{t+k}^2] - \mathrm{Cov}[Y_{t+k}^1, Y_t^2]}{\nu_1\nu_2 k^{H_1+H_2}} + O(k^{H_1+H_2}), \end{aligned} \tag{14}$$

which, after plugging-in empirical covariances and leaving the remainder $O(k^{H_1+H_2})$ out, deliver two new estimators $\tilde{\rho},$ and $\tilde{\eta}$.

Remark 3. *These estimators do not depend on $\alpha_i,\ i=1,2$.*

Theorem 2. *Suppose to observe $(Y_t)_{t\in[0,T_n]}$, n times at time increments of size $\Delta_n = \frac{T_n}{n}$ on an equally spaced grid. As $\Delta_n \to 0$, we have*

(i) Unbiasedness:

$$\mathbb{E}\big[(\hat{\rho}_n, \hat{\eta}_n)\big] \to (\rho, \eta) \qquad as\ n\ \to \infty;$$

(ii) Consistency for ρ:

$$\hat{\rho} \xrightarrow{p} \rho \qquad as\ n\Delta_n \to \infty;$$

(iii) Asymptotic normality for ρ: if $H_1 + H_2 < \frac{3}{4}$,

$$\sqrt{n}\big(\hat{\rho}_n - \rho\big) \xrightarrow{d} N(0, \sigma^2) \qquad as\ n\Delta_n^2\ \to 0\ and\ n\Delta_n \to \infty,$$

where σ is a positive constant.

See Dugo, Giorgio, and Pigato (2024) for details and proofs.

3.2 Generalized method of moments

Given a discrete sample of observations of the bivariate fractional Ornstein-Uhlenbeck process $(Y_n)_{n=1,\dots,N} = \big(Y_n^1, Y_n^2\big)_{n=1,\dots,N}$, the full set of parameters $\theta = (\alpha_1, \alpha_2, \nu_1, \nu_2, \rho, \eta, H_1, H_2, \mu_1, \mu_2)$ can be obtained via the

Generalized Method of Moments (Hansen 1982). Such procedure relies on a set of orthogonality conditions which we state in the following. Let us define the functions

- $g_0^i(\theta) = \mathbb{E}_\theta \left[Y_t^i\right] = \mu_i, \quad i = 1, 2;$
- $g_k^{ii}(\theta) = \mathbb{E}_\theta \left[(Y_t^i)^2\right] = \gamma_{ii}(0) + \mu_i^2, \quad i = 1, 2;$
- $g_k^{ii}(\theta) = \mathbb{E}_\theta \left[Y_{t+k}^i Y_t^i\right] = \gamma_{ii}(k) + \mu_i^2, \quad i = 1, 2;$
- $g_0^{ij}(\theta) = \mathbb{E}_\theta \left[Y_t^i Y_t^j\right] = \gamma_{ij}(0) + \mu_i \mu_j;$
- $g_k^{ij}(\theta) = \mathbb{E}_\theta \left[Y_{t+k}^i Y_t^j\right] = \gamma_{ij}(k) + \mu_i \mu_j;$
- $g_k^{ji}(\theta) = \mathbb{E}_\theta \left[Y_t^i Y_{t+k}^j\right] = \gamma_{ji}(k) + \mu_i \mu_j;$

where $\mathbb{E}_\theta$ indicates the expectation with respect to the law induced by θ, $k \in \mathcal{L} \subset \mathbb{N}_{\setminus\{0\}}$ is an index of lags different than zero, and $\gamma_{ij}(k)$ is the cross-covariance function at lag k defined exactly in (6) or given asymptotically in Proposition 2.

Let us collect all the values of the moments on which we rely in a column vector

$$G(\theta) = \left(g_0^i(\theta), g_0^j(\theta), g_0^{ii}(\theta), g_0^{jj}(\theta), g_0^{ij}(\theta), \left(g_k^{ii}(\theta), g_k^{jj}(\theta), g_k^{ij}(\theta), g_k^{ji}(\theta)\right)_{k\in\mathcal{L}}\right)' \in \mathbb{R}^{5+4L} \tag{15}$$

where $L = \#\mathcal{L}$ indicates the cardinality of the set $\mathcal{L}$. Given $n \in \mathbb{N}$, we define the vector

$$\hat{Y}_n = \left(Y_n^1, Y_n^2, (Y_n^1)^2, (Y_n^2)^2, Y_n^1 Y_n^2, \left(Y_{n+k}^1 Y_n^1, Y_n^2 Y_{n+k}^2, Y_{n+k}^1 Y_n^2, Y_n^1 Y_{n+k}^2\right)_{k\in\mathcal{L}}\right)', \tag{16}$$

the random function

$$\hat{m}_N(\theta) = \frac{1}{N}\sum_{n=1}^{N} \hat{Y}_n - G(\theta), \tag{17}$$

and introduce the GMM estimator $\hat{\theta}_N$ as

$$\begin{aligned} \hat{\theta}_N &= \arg\min_\theta \mathcal{M}(\theta) \\ &= \arg\min_\theta \hat{m}_N(\theta)' W_N \hat{m}_N(\theta) \end{aligned} \tag{18}$$

where W_n is a $(5+4L)\times(5+4L)$ symmetric positive definite weighting matrix.
In order to state consistency and asymptotic normality, we rely on the following.

Assumptions

(i) $k \to \gamma_{i,j}(k) \in \mathbb{R}$ is a continuous function $\forall k \in \mathbb{R}$;

(ii) $k \to \gamma_{i,j}(k)$ is twice differentiable $\forall k \in \mathbb{R}$ and its gradient $\nabla\gamma_{i,j}(k)$ and Hessian $\nabla^2\gamma_{i,j}(k)$ exist, are continuous, and bounded (see Appendix A);

(iii) $W\hat{m}(\theta) = 0$ if and only if $\theta = \theta_0$;

(iv) under the measure induced by the true parameters' value θ_0, the random weight-matrix W_N converges in probability to a non-random matrix W, as $N \to \infty$.

Theorem 3.

(I) under the assumptions (i)-(iii) above,

$$\hat{\theta}_N \xrightarrow{p} \theta_0 \qquad \text{as } N \to \infty,$$

(II) when $H_1 + H_2 < \frac{3}{2}$ and assumption (iv) above holds,

$$\sqrt{n}\left(\hat{\theta}_N - \theta\right) \xrightarrow{d} N(0, \Sigma)$$

where $\Sigma = (D'WD)^{-1} D'W\Gamma WD (D'WD)^{-1}$, D is the Jacobian matrix of $\hat{m}_N(\theta)$ with rows given by $[D]_p = -\nabla\gamma_{i,j}(k), p = 1, \ldots, 5 + L$, and Γ is the covariance matrix of the (vector-valued) random function $\hat{m}_N(\theta)$.

Proof of (I)

By definition of the GMM estimator, $\hat{\theta}$ satisfies the first order condition

$$\frac{\mathrm{d}\mathcal{M}(\theta)}{\mathrm{d}\theta}|_{\theta=\hat{\theta}} = 0. \tag{19}$$

Using a first-order Taylor expansion with Lagrange form of the remainder around θ_0, we have

$$\frac{\mathrm{d}\mathcal{M}(\theta)}{\mathrm{d}\theta}|_{\theta=\hat{\theta}} = \frac{\mathrm{d}\mathcal{M}(\theta)}{\mathrm{d}\theta}|_{\theta=\theta_0} + \frac{\mathrm{d}^2\mathcal{M}(\theta)}{\mathrm{d}\theta^2}|_{\theta=\bar{\theta}}\left(\hat{\theta} - \theta_0\right) = 0 \tag{20}$$

where $\bar{\theta}$ lies between θ_0 and $\hat{\theta}$.
Solving for $\hat{\theta} - \theta_0$, we get

$$\hat{\theta} - \theta_0 = -\left(\frac{\mathrm{d}^2\mathcal{M}(\theta)}{\mathrm{d}\theta^2}\right)^{-1}|_{\theta=\bar{\theta}} \frac{\mathrm{d}\mathcal{M}(\theta)}{\mathrm{d}\theta}|_{\theta=\theta_0}. \tag{21}$$

Given assumption (ii),we can explicitly calculate the terms in Equation (21), as

$$\frac{\mathrm{d}\mathcal{M}(\theta)}{\mathrm{d}\theta} = -2D'W_N\hat{m}_N(\theta) \tag{22}$$

and

$$\frac{\mathrm{d}^2\mathcal{M}(\theta)}{\mathrm{d}\theta^2} = 2D'W_N D + o\left(\hat{m}_N(\theta)\right) \tag{23}$$

where $o\left(\hat{m}_N(\theta)\right)$ represents a term containing the second derivatives of $\hat{m}_N(\theta)$ times $\hat{m}_N(\theta)$ itself.
Dugo, Giorgio, and Pigato (2024) show that

$$\frac{1}{N}\sum_{n=1}^{N}\hat{Y}_n \xrightarrow{p} G(\theta) \quad as\ N \to \infty \tag{24}$$

which, in our terms, is equivalent to

$$\hat{m}_N(\theta) \xrightarrow{p} 0. \tag{25}$$

which, in turn, together with the boundedness of the second order derivatives, allows us to conclude that

$$\frac{\mathrm{d}\mathcal{M}(\theta)}{\mathrm{d}\theta} \xrightarrow{p} 0 \tag{26}$$

and

$$\frac{\mathrm{d}^2\mathcal{M}(\theta)}{\mathrm{d}\theta^2} \xrightarrow{p} 2D'WD, \tag{27}$$

where D is the Jacobian matrix of $\hat{m}_N(\theta)$ calculated at the true value of the parameters and W is the limiting weights matrix. Equations (26) and (27) imply together that $\bar{\theta} \xrightarrow{p} \theta_0$ and $\hat{\theta} \xrightarrow{p} \theta_0$, thus concluding the proof of (I). ■

Proof of (II)

In order to establish the asymptotic distributional properties of $\hat{\theta}$, we need to understand the limit distribution of the term in (8).
We know from Theorem 4 in Arcones (1994), that when $H_1 \wedge H_2 < \frac{3}{2}$, $\forall k$,

$$\sqrt{n}\left(\frac{1}{N}\sum_{i=1}^{N}Y_{n+k}^{i}Y_{n}^{j} - \gamma_{i,j}(k)\right) \xrightarrow{d} N(0,\sigma^2). \tag{28}$$

Using Equation (28) we can establish the limit in distribution of the vector $\hat{m}_N(\theta)$, which basically stacks together the difference of a number of empirical ccfs minus their theoretical counterpart, as

$$\sqrt{n}\hat{m}_N(\theta) \xrightarrow{d} N(0, \Gamma). \tag{29}$$

Under technical mixing conditions on the vector $\hat{m}_N(\theta)$, Γ can be estimated consistently using Newey and West (1987) HAC-type estimators (see Bolko et al. 2022).
Therefore we have

$$\sqrt{n}\frac{\mathrm{d}\mathcal{M}(\theta)}{\mathrm{d}\theta} = -\sqrt{n}2D'W_N\hat{m}_N(\theta) \xrightarrow{d} N(0, 4D'W\Gamma WD) \tag{30}$$

which together with Equation (21), Equation (27) and Slutsky's Theorem establishes (II). ■

Remark 4. *Notice that*

$$(D'WD)^{-1}D'W\Gamma WD(D'WD)^{-1} - (D'\Gamma^{-1}D)^{-1} \geq 0, \tag{31}$$

suggests that higher efficiency for $\hat{\theta}$ could be achieved by choosing $W = \Gamma^{-1}$, which delivers the lowest asymptotic variance $\Sigma^* = (D'\Gamma^{-1}D)^{-1}$ for our estimator.

The reasoning behind this choice follows the optimal weighting in generalised linear models (GLS). See Hayashi (2011) for details.
Finally, assumption (iii) simply tells us that the parameters obtained in this way are indeed the true ones.

4 Simulation study

In this section we discuss numerical restults supporting the statistical theory presented so far.

4.1 Simulation procedure

We present in what follows an exact and an approximate scheme for simulating the bivariate fractional Ornstein-Uhlenbeck process.

4.1.1 Exact simulation

The exact simulation can be performed with general methods for stationary Gaussian processes, such as the Cholesky decomposition of the covariance matrix, which can be constructed starting from Equation (6). The triangular matrix obtained by decomposition can be transformed into trajectories of the process by multiplying it to arrays of independently and identically distributed standard Gaussian random variables.

This methodology is powerful due to the exactness of the law of the simulated trajectories. However, it has a main limitation, namely it cannot be used to generate very long paths (large N) and/or set large mean-reversion parameters $\alpha_i, \quad i = 1, 2$. The reason lies in the numerical evaluation of the integral in the expression of the ccf in Equation (6) for choices of the parameters N and $\alpha_i,\ i = 1, 2$ that lead to integrating large exponentials.

4.1.2 Approximate simulation

Another natural approach to simulate the bivariate fractional Ornstein-Uhlenbeck process is the Euler-Maruyama method. It is a commonly used numerical scheme for solving stochastic differential equations, which can be used with great flexibility and efficiency. The methodology consists in approximating the solution to the Langevin equation at discrete time steps by using the Euler method for the deterministic part and adding a random noise term to account for the stochasticity.
In practice, the noise term is a previously obtained fractional Gaussian noise $\left(B^{H_j}_{i\Delta} - B^{H_j}_{(i-1)\Delta}\right)^N_{i=1}$, which can be simulated efficiently according to a suitably designed circular embedding method (available in Section 5 of Amblard et al. 2010) and whose implementation is available either on the website of the authors in R, or, upon request to us, coded in C using the basic routines provided by Galassi et al. (2002).

The Euler approximation of the model $Y^H_n = (Y^{H_1}_n, Y^{H_2}_n)$ over the interval $((n-1)\Delta, n\Delta)$ takes the form

$$Y^{H_i}_{n\Delta} = Y_{(n-1)\Delta} + \alpha_i \left(\mu_i - Y^{H_i}_{(n-1)\Delta}\right)\Delta + \nu_i \left(B^{H_i}_{n\Delta} - B^{H_i}_{(n-1)\Delta}\right), \qquad i = 1, 2$$

where $\mu_j \in \mathbb{R},\ \alpha_j \in \mathbb{R}^+,\ \nu_j \in \mathbb{R}^+$ are the parameters of the Langevin equation and $(B^{H_j}_{i\Delta} - B^{H_j}_{(i-1)\Delta})$ are the simulated increments of the i_{th} component of the multivariate fractional Brownian Motion. It is worth noticing that the constant diffusion coefficient of the fractional Ornstein-Uhlenbeck process leads to the equivalence of the Euler-Murayama scheme and the refined version obtained by Milstein (1974), which results in higher efficiency.

The Euler-Maruyama method is relatively simple to implement, but it remains the drawback of being approximate as a consequence of the discretization in time of a continuous process. In order to reduce the error induced by such discretization, a finer grid $\Delta' = \frac{\Delta}{S}$ is used to obtain a sample of trajectories of length $N' = NS$, of which only N equally spaced observations at distance Δ are eventually retained. This approach was adopted in a context similar to ours by Wang, Xiao, and Yu (2021) and Bolko et al. (2022) and is helpful at better approximating the exact distribution of the process at the relevant sampling times.

The order of convergence of the Euler-Maruyama scheme is well established in the case of diffusion processes driven by standard Brownian motion (see Kloeden et al. 1992). However, for stochastic differential equations (SDEs) driven by fractional Brownian motion, it is still an ongoing matter of research. Useful results are provided in Mishura (2008), Hu, Liu, and Nualart (2016), and Liu and Tindel (2019).

An alternative methodology would be solving the Langevin equation over the interval $(n\Delta, (n-1)\Delta)$ with $Y^{H_j}_{(n-1)\Delta}$ as initial condition, so to get

$$Y^{H_i}_{n\Delta} = \mu_i + (Y^{H_i}_{(n-1)\Delta} - \mu_i)e^{-\alpha_i \Delta} + \nu_i \int_{(n-1)\Delta}^{n\Delta} e^{-\alpha_i(n\Delta - t)} dB^{H_j}_t, \quad i = 1, 2,$$

and approximating the stochastic integral $\int_{(i-1)\Delta}^{i\Delta} e^{-\alpha_j(i\Delta - t)} dB^{H_j}_t \approx e^{-\alpha_j \Delta/2}(B^{H_j}_{i\Delta} - B^{H_j}_{(i-1)\Delta})$ (see Bolko et al. 2022).

In the following, we adopt the Euler-Maruyama method with N and S varying according to the different setups of the experiments.

4.2 Two-step procedure

In the first place, we verify on simulated trajectories that the estimators defined in Equation 12 satisfy the properties stated in Theorem 1.

4.2.1 Estimators based on the exact ccf relationship

Numerical verification of CLT (iii)

Figure 3 considers the case $H_1 + H_2 < \frac{3}{2}$. In particular, it shows different aspects that verify asymptotic unbiasedness, consistency and asymptotic normality. The exercise was carried out on a sample of size $M = 10^4$ made of trajectories of length $N = 400$. Such trajectories were obtained with the exact simulation method.
In particular, from left to right, for $\theta = \rho, \eta$, we have:

(i) the density of the estimates $\hat{\theta}$ for a growing sample size, which is detailed in the legend. In red, it is superposed a Gaussian density centered at ρ and having standard deviation equal to the estimated standard error for $\hat{\theta}_N$ when N = 400. This picture shows that, for given parameters, $\hat{\theta}$ is normally distributed as far as the estimation is carried out on a sample of size at least N = 400;

(ii) the log of the absolute mean estimation error $\log \frac{1}{N} \sum_{n=1}^{N} \left| \hat{\theta}_n - \theta \right|$ against $\log n$. The angular coefficient of the line verifies the speed of convergence in the central limit theorem because

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} Z$$

$$\log \mathbb{E}\left[\hat{\theta}_n - \theta\right] \to \log \mathbb{E}\big[Z\big] - \frac{1}{2} \log n;$$

(iii) the superposed density of the rescaled estimation error $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$, which appear to be distributed as a normal random variable for every sample length N, as demonstrated by the superposed standard normal in red, show that the central limit theorem holds and that the scaling $\sqrt{n}$ is correct.

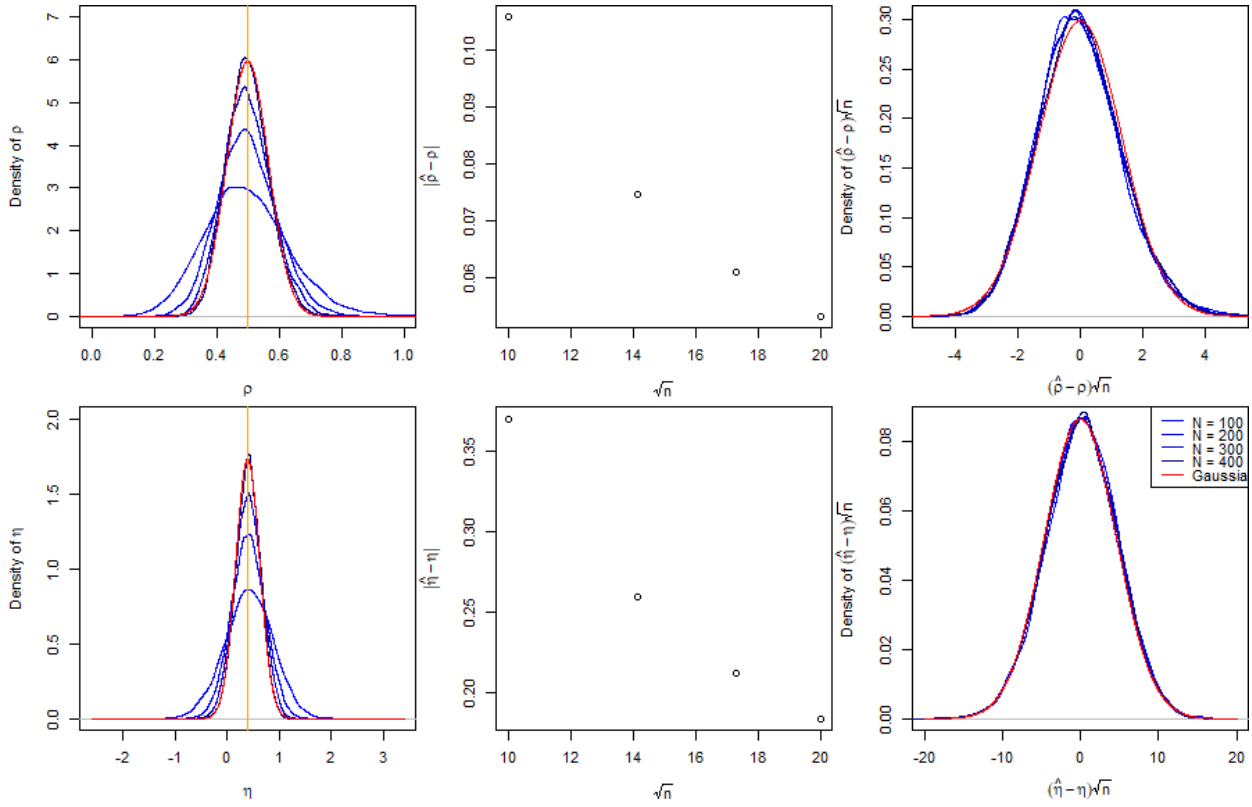

Figure 3: (top) ρ, (bottom) η - parameter values $\rho = 0.5,\ \eta = 0.4,\ H_1 = 0.3,\ H_2 = 0.4,\ \alpha_1 = \alpha_2 = 0.5,\ \sigma_1 = \sigma_2 = 1,\ N = 400,\ M = 10^4$

Numerical verification of NCLT (iv)

In Figure 4, we show the behavior of the estimator $\hat{\theta}$ under $H_1 + H_2 > \frac{3}{2}$. Again we use data sampled with an exact simulation scheme and the same size/length as before. With a big difference with respect to before, despite having unbiasedness, our estimator has now a skewed distribution and the rate of convergence to the limiting one is now different from $\frac{1}{2}$. Figure 3 shows unscaled (left) and rescaled (right) estimation errors with the aim to show the shape of the limiting distribution. Please do not consider the plot in the middle, because it is wrong.

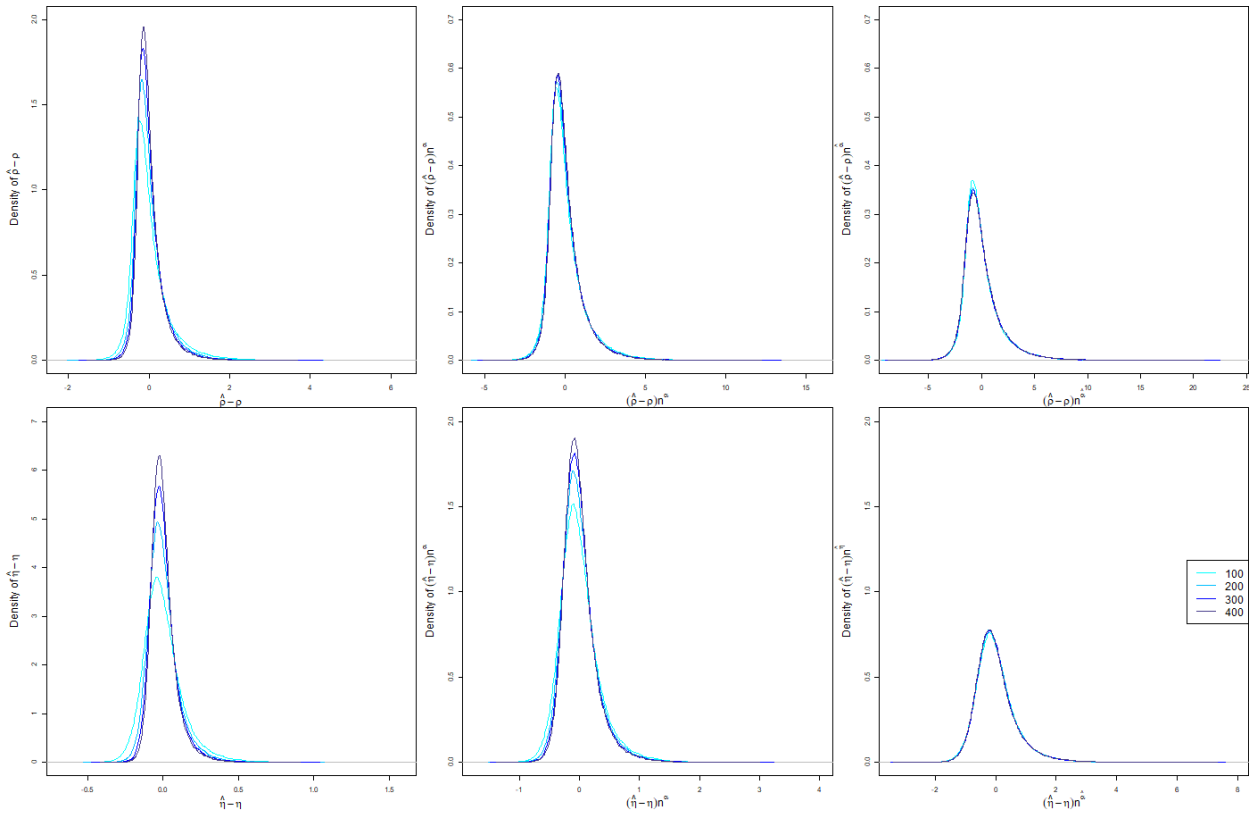

Figure 4: (top) ρ, (bottom) η - parameter values $\rho = 0.5,\ \eta = 0,\ H_1 = 0.8,\ H_2 = 0.9,\ \alpha_1 = \alpha_2 = 0.5,\ \sigma_1 = \sigma_2 = 1,\ N = 400,\ M = 10^4$

Rate of convergence in CLT (iii) and NCLT (iv)

Figure 5 shows a numerical verification of the rate of convergence of the estimation error for varying values of $H_1 + H_2$. To obtain these values, we relied on simulated time series obtained with the approximate method so as to be longer than what we can produce with the exact method. In particular, the simulated series were characterized by a length of $N = 20000$ time steps within which $S = 1024$ further substeps were simulated and then filtered out. The latter to reduce discretization error. In the figure, we have the theoretical rates as broken lines and those estimated in simulation represented by crosses. The two seem satisfactorily close to each other.

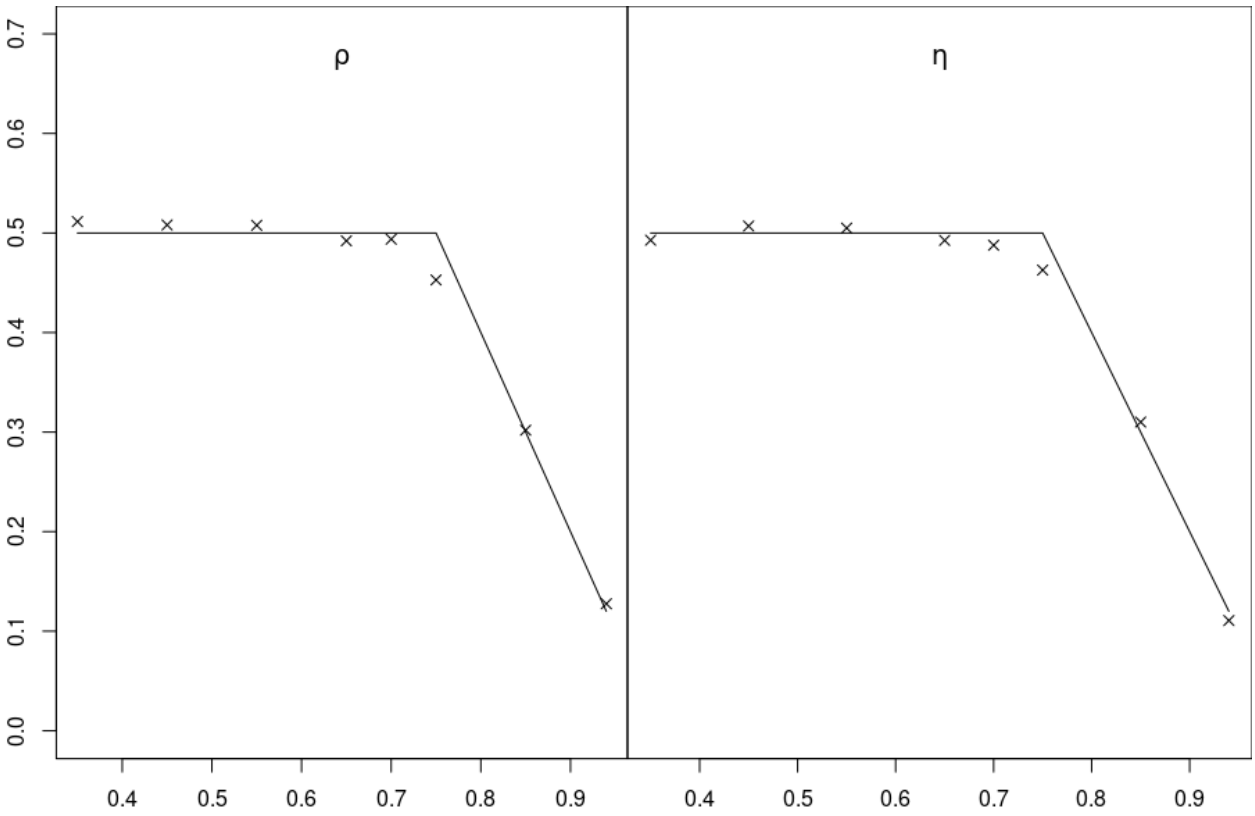

Figure 5: Rate of convergence of the estimates to true parameters on numerical experiment - N = 20000, substeps = 1024, H1 = H2 - 0.1 shown on the x-axis

The graphics show that the rate of convergence set out by points (iii) and (iv) of Theorem 2 do indeed hold.

4.2.2 Estimators based on the asymptotic ccf relationship

In this section, we show how the estimators $\tilde{\rho}$ and $\tilde{\eta}$ based on the small-lag asymptotic ccf relationship outperform those based on the exact relationship $\hat{\rho}$ and $\hat{\eta}$ when α_i $i = 1, 2$ become very small. What shown is exactly the reason that motivated us in deriving a second set of estimators based on the small-lag asymptotic ccf relationship and their qualities.

Comparison of the estimators for ρ

Figure 6 shows the densities of the estimation error for rho $\sqrt{n}\,(\hat{\rho} - \rho)$. From left to right the value of $\alpha_1 = \alpha_2$ decreases, whereas from top to bottom we have the distinction of using either the standard estimator $\hat{\rho}$ or the small-lag regime one $\tilde{\rho}$.

We can see that the "asymptotic density" (approximated for large N = 400) goes off target for the exact estimator when $\alpha_i,\ i = 1, 2$ becomes small. The opposite happens for $\tilde{\rho}$, where the error density almost coincide with a standard normal as long as $N = 400$ and $\alpha_i = 0.01,\ i = 1, 2$. We guess, this is a consequence

of the numerical issues in integration and such a situation could well be encountered in practice due to poor estimations of $\alpha_i,\ i = 1, 2$. Another explanation can be given by a statistical point of view. Because $\alpha_i\ i = 1, 2$ always multiplies t and therefore we might have different influences of the parameters on the rates of convergence of the two estimators. Therefore, if we believe that our $\alpha_i,\ i = 1, 2$ is very small, we should be using the small-lag regime estimator $\tilde{\rho}$.

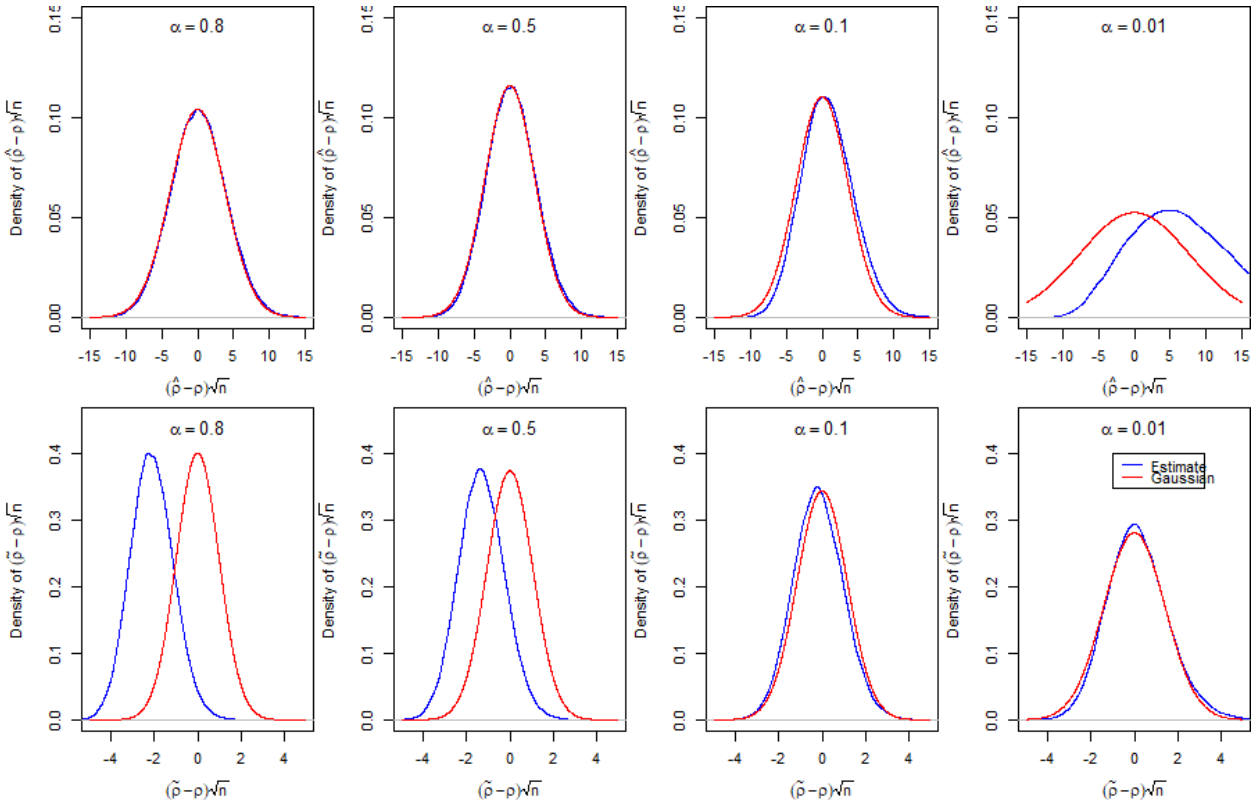

Figure 6: Density of re-scaled estimation errors: (top) based on exact CCF, (bottom) based on small-time CCF - parameter values $\rho = 0.5,\ \eta = 0.4,\ H_1 = 0.3,\ H_2 = 0.4,\ \nu_1 = \nu_2 = 1$

Comparison of the estimators for η

Figure 7 shows the same concept from Figure 6 for $\hat{\eta}$ and $\tilde{\eta}$. Again we observe that the quality of the estimator based on the exact ccf relationship $\hat{\eta}$ (top row) deteriorates as $\alpha = \alpha_i = \alpha_j$ gets smaller. This is shown by the limit distribution going off target with respect to the Gaussian one. On the other hand, the estimation error of $\tilde{\eta}$, based on the small-lag regime, seems to converge to the desired Gaussian as $\alpha = \alpha_i = \alpha_j$ gets smaller.

Notice that with respect to the previous figure we are taking $\alpha = \alpha_i = \alpha_j$ one order of magnitude smaller than before in each column, suggesting a different behavior in $\tilde{\eta}$ than $\tilde{\rho}$. This case indeed has to be treated with care because of the lack of underlying analytical theory.

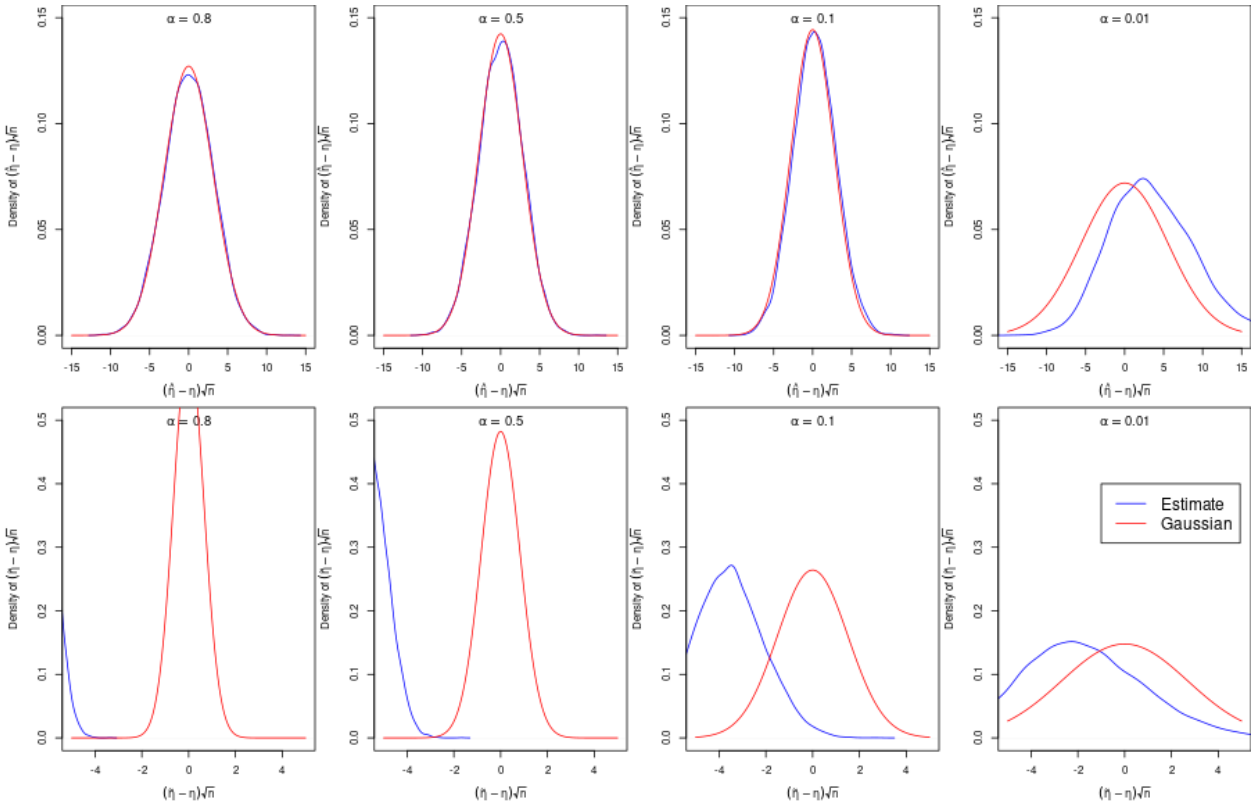

Figure 7: Density of re-scaled estimation errors: (top) based on exact CCF, (bottom) based on small-time CCF - parameter values $\rho = 0.5,\ \eta = 0.4,\ H_1 = 0.3,\ H_2 = 0.4,\ \nu_1 = \nu_2 = 1$

4.3 Full GMM routine

In this paragraph we check on simulation the quality of the GMM estimation. The algorithm employed for indentification uses the conditions with exact auto-covariances and cross-covariances at lag 0 through 5 and 20 and 50 (as chosen in Bolko et al. 2022). The initial conditions for the GMM routine are the estimates obtained with the two-step procedure detailed above.

For this purpose, we perform five experiments and report the results in the following tables. In particular, the tables report the true value of the parameters, the estimates obtained according to the two-step method under the name Point 2S, and the estimates obtained with the GMM routine that uses the two-step values as starting point, under the name Point GMM. Similar names are used for reporting the standard errors.

The estimation experiments are performed on a sample of $M = 2000$ replications of the process for each set of parameters, where the length of the trajectories was set to $N = 5000$ (comparable to the time series used in the empirical study in Section 5). Point estimates and standard errors, which are shown in a smaller font, are obtained as a simple average over the sample estimates.

The simulations underlying this experiment relied on the approximate routine with $S = 512$ substeps within

each two continguous observations of the process.

Table 1: Results of experiment 1 (refer to main text for interpretation)

	ρ	η	a_1	a_2	v_1	v_2	H_1	H_2	μ_1	μ_2
True value	0.500	0.030	0.100	0.200	1.000	2.000	0.600	0.700	0.000	0.000
Point 2S	0.493	0.086	0.097	0.175	0.992	1.915	0.592	0.672	-0.014	-0.031
StdErr 2S	0.076	0.073	0.012	0.018	0.016	0.039	0.019	0.018	0.329	0.782
Point GMM	0.499	0.033	0.104	0.204	0.999	1.992	0.599	0.695	-0.013	-0.028
StdErr GMM	0.011	0.082	0.014	0.025	0.015	0.059	0.016	0.023	0.325	0.745

Table 2: Results of experiment 2 (refer to main text for interpretation)

	ρ	η	a_1	a_2	v_1	v_2	H_1	H_2	μ_1	μ_2
True value	0.500	0.300	0.100	0.200	1.000	2.000	0.600	0.700	0.000	0.000
Point 2S	0.493	0.411	0.097	0.174	0.992	1.914	0.593	0.671	0.001	-0.014
StdErr 2S	0.076	0.090	0.012	0.018	0.016	0.039	0.019	0.018	0.328	0.765
Point GMM	0.499	0.305	0.104	0.207	0.999	1.999	0.600	0.698	0.001	-0.009
StdErr GMM	0.011	0.082	0.012	0.026	0.015	0.061	0.015	0.023	0.325	0.731

Table 3: Results of experiment 3 (refer to main text for interpretation)

	ρ	η	a_1	a_2	v_1	v_2	H_1	H_2	μ_1	μ_2
True value	0.500	0.300	0.100	0.200	1.000	2.000	0.100	0.200	0.000	0.000
Point 2S	0.505	0.288	0.100	0.178	0.999	1.988	0.100	0.189	0.000	0.001
StdErr 2S	0.014	0.028	0.041	0.034	0.012	0.023	0.022	0.022	0.009	0.018
Point GMM	0.500	0.303	0.104	0.205	0.998	1.998	0.101	0.202	0.000	0.000
StdErr GMM	0.013	0.031	0.016	0.025	0.012	0.023	0.008	0.017	0.010	0.020

Table 4: Results of experiment 4 (refer to main text for interpretation)

	ρ	η	a_1	a_2	v_1	v_2	H_1	H_2	μ_1	μ_2
True value	0.500	0.030	0.100	0.200	1.000	2.000	0.100	0.200	0.000	0.000
Point 2S	0.504	0.022	0.098	0.177	0.999	1.987	0.099	0.190	0.000	0.001
StdErr 2S	0.014	0.022	0.040	0.033	0.011	0.022	0.021	0.021	0.009	0.018
Point GMM	0.499	0.025	0.109	0.204	0.997	1.996	0.104	0.202	0.000	0.001
StdErr GMM	0.011	0.033	0.026	0.024	0.011	0.023	0.013	0.016	0.010	0.020

Table 5: Results of experiment 5 (refer to main text for interpretation)

	ρ	η	a_1	a_2	v_1	v_2	H_1	H_2	μ_1	μ_2
True value	0.500	0.030	0.100	0.200	1.000	2.000	0.050	0.100	0.000	0.000
Point 2S	0.508	0.025	0.098	0.173	1.000	1.994	0.049	0.093	0.000	0.000
StdErr 2S	0.018	0.018	0.067	0.056	0.012	0.023	0.022	0.022	0.009	0.017
Point GMM	0.501	0.025	0.117	0.206	1.000	1.998	0.054	0.103	0.000	0.001
StdErr GMM	0.012	0.023	0.045	0.033	0.014	0.024	0.016	0.014	0.010	0.018

In all cases, reported from Table 1 to Table 5, we can see that the average values of the estimated parameters are much closer to the true values when employing GMM rather than the 2-step procedure, whereas the standard errors do not change considerably for varying methodology. This is especially true for $\alpha,\ i = 1, 2$, which is initially downward biased (especially α_2 as a consequence of being larger), and η which is estimated correctly only after applying GMM. Also GMM slightly fails at centering the true value of eta when we have very rough trajectories and small values of the parameter, cases in which it still delivers and important improvement.

In addition, to understand the accuracy of the asymptotic theory, Figure 8 displays kernel smoothed densities of the statistics $\frac{\left(\hat{H}_1 - H_1\right)}{\hat{se}(\hat{H}_1)}$ and $\frac{(\hat{\rho}-\rho)}{\hat{se}(\hat{\rho})}$, where the standard errors at the denominator are estimates extracted from the matrix Σ^* defined in Remark 4, against the density of a standard $N(0, 1)$. The proximity of the curves confirms the results in Theorem 3 and we believe the same to hold for the rest of the parameters.

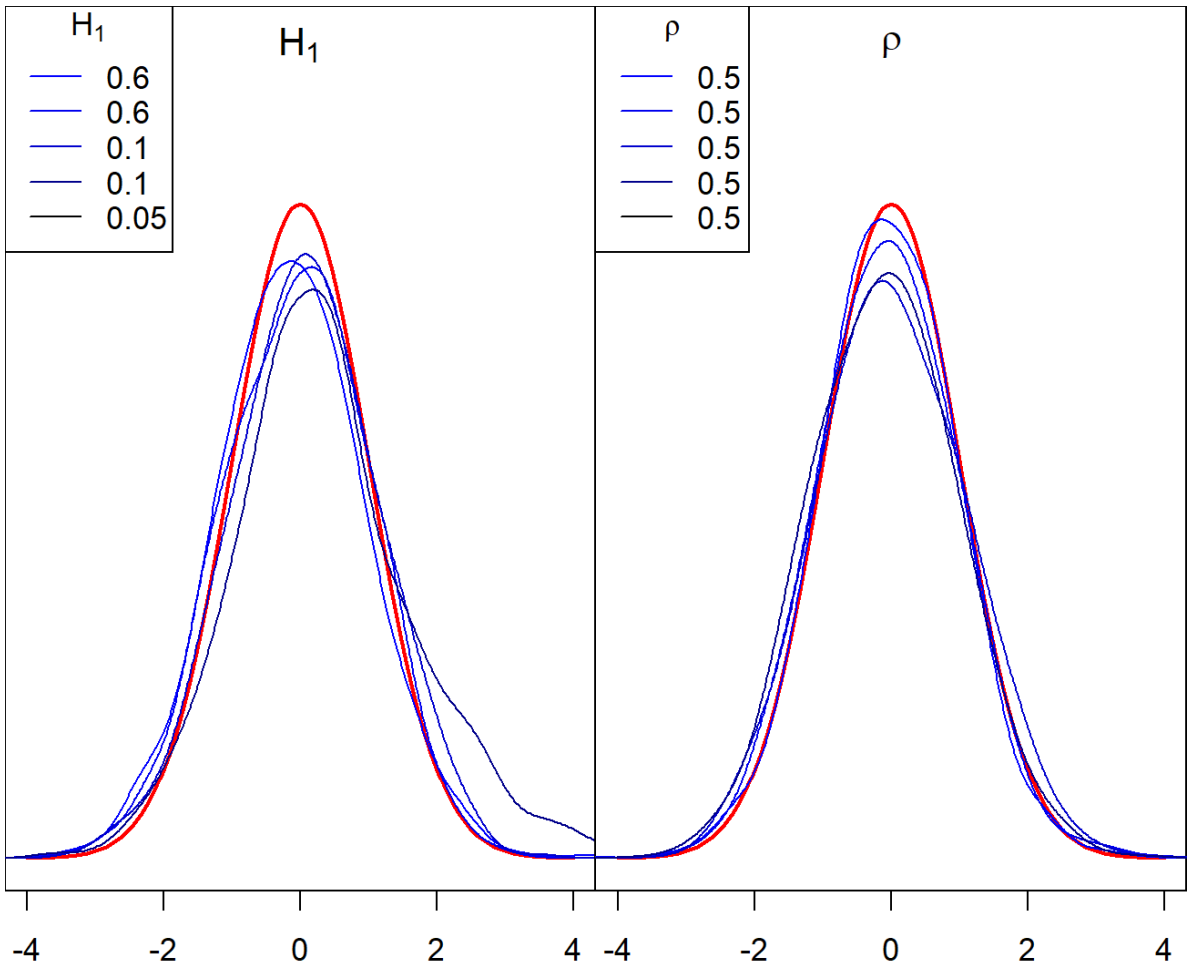

Figure 8: Densities of standardized estimation errors for variables H1 and a1 in experiments 1 through 5.

5 Empirical results

In this section, the bivariate log-normal fractional stochastic volatility model introduced in Sections 2 is estimated on real data.

5.1 Data considerations

The daily data under analysis cover a comprehensive selection of realized volatility time series of indices constructed from high frequency observations. We downloaded the latest version of the Oxford-Man Institute's "realized library". The website provides, among others, daily realized volatility times series computed at a 5-min frequency for thirty-one leading stock indexes covering major financial markets. Before proceeding with the analysis, we chose 7 realized volatility time series overlapping in time and available on a long

sampling period and of notable importance, leaving aside all of the rest. The chosen ones are the symbols: IXIC (NASDAQ), DJI (Dow Jones Industrial Average), SPX (Standard & Poor's 500), STOXX50E (Euro Stoxx 50), FTSE (Financial Times Stock Exchange), IBEX (IBEX 35 - Madrid), and GDAXI (DAX 30 - Frankfurt). Starting from these 7, we obtained bivariate realized volatility time series by taking all their possible combinations, resulting in a sample of 21 bivariate series. Theoretically, we could consider a full-fledged multivariate model with 7 components and perform GMM jointly. However, we leave this idea for future work. Now, for each time series we consider the first 5000 overlapping observations for estimation and ignore the rest. This results in variable sampling periods as 5000 trading days from 2000-01-03 would end up to 2019-01-28. However, missing data are spread heterogeneously across time series, resulting in slightly different periods. With this choice, we leave outside of the sample approximately two years of data, one of which contains the Covid pandemic. Later we will include these data in the out-of-sample forecasting.
Additionally, we discarded days in which realized volatility appeared identically equal to zero.

5.2 Evidence of weak mean-reversion

The parameter $\alpha_i,\ i = 1, 2$, which is well know in the literature to be very small and often estimated with a bias (see for example Wang, Xiao, and Yu 2021 and references therein), creates us some problems in the estimation of the bivariate parameters ρ and η and using the model therefater.
We attempted estimation of the full model containing the $\alpha_i,\ i = 1, 2$ but obtained values that seemed more often than not inconsistent with the observed cross-covariance structure of the data.
Furthermore, given that $\alpha_i,\ i = 1, 2$ is well known in the literature to be very small, see for example Gatheral, Jaisson, and Rosenbaum (2018), at least for the time-horizon which we look at (i.e. the volatility process mean reverts slowly), we decided to carry out the empirical study leveraging on the small-lag asymptotic relationship in Proposition 2.
First, we estimate $H_i,\ i = 1, 2$ using the semi-parametric estimator indotruced by Lang and Roueff (2001) and test the asymptotic relationship on the data.
We report results related to the symbols IXIC, SPX, and STOXX50E but the same analysis is similar for the rest of the indices in our sample.
Table 6 contains such estimates. Notice that they vary slightly, for the same index, when paired with a different series due to the fact that the samples vary slighly, mainly due to missing values in different positions.

Table 6: Estimates of H on the three symbols of interest using change of frequency estimator

symbol	H_1	H_2
IXICSPX	0.223	0.127
IXICSPX	0.021	0.021
IXICSTOXX50E	0.225	0.099
IXICSTOXX50E	0.021	0.021
SPXSTOXX50E	0.132	0.098
SPXSTOXX50E	0.021	0.021

Using the estimates from Table 6, we can verify the asymptotic relationship in Proposition 2.
Figure 9 to 11 plot the empirical counterpart of the relationship in Proposition 2. In particular, on the y axis we have empirical estimates of the cross-covariance function for lags 0 to 50 and on the x axis we have a suitable rescaled version of the lag, where the scaling is obtained with the estimate of H in Table 6. Red lines represent a linear fit on the points.

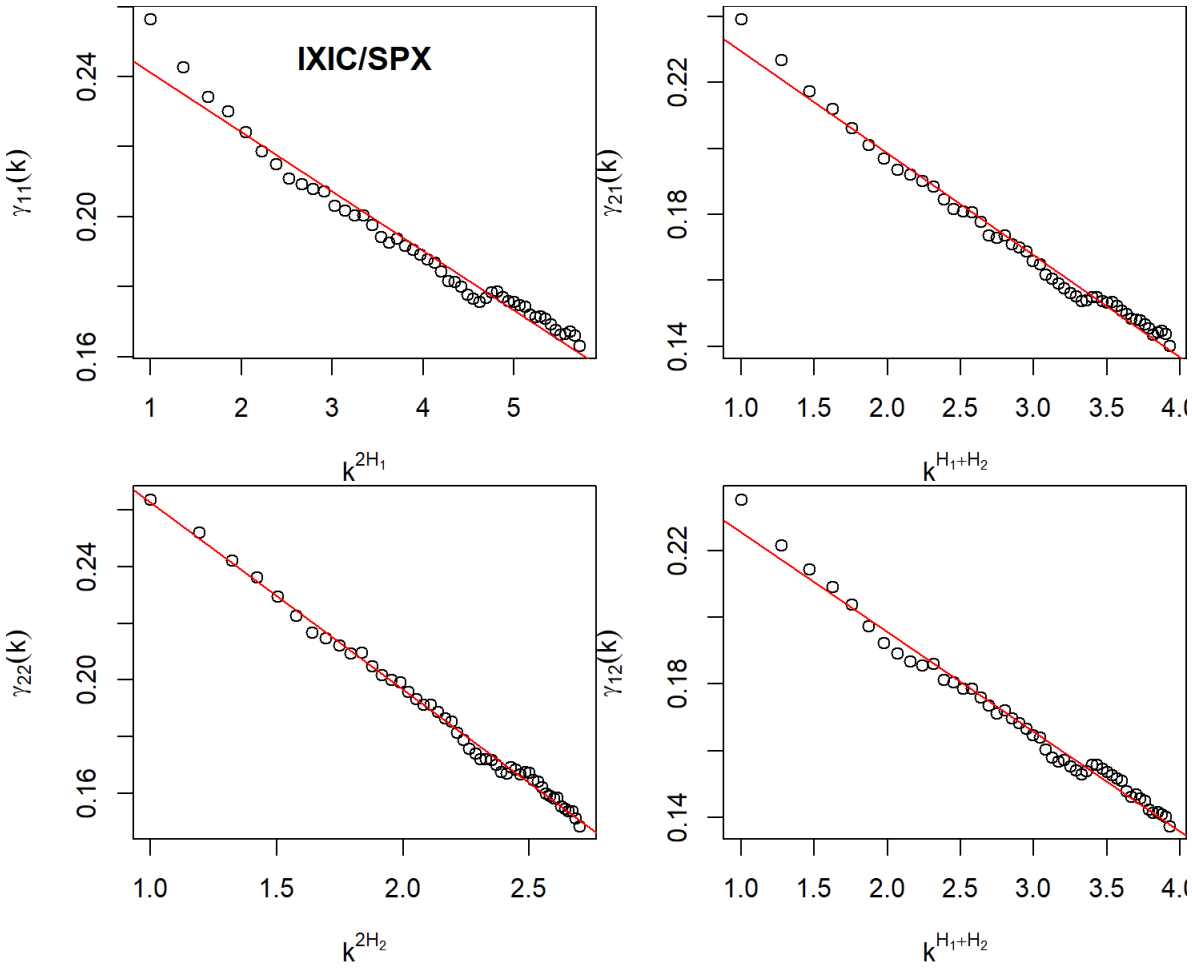

Figure 9: Cross-covariance function vs. suitable power of the lag for IXIC/SPX.

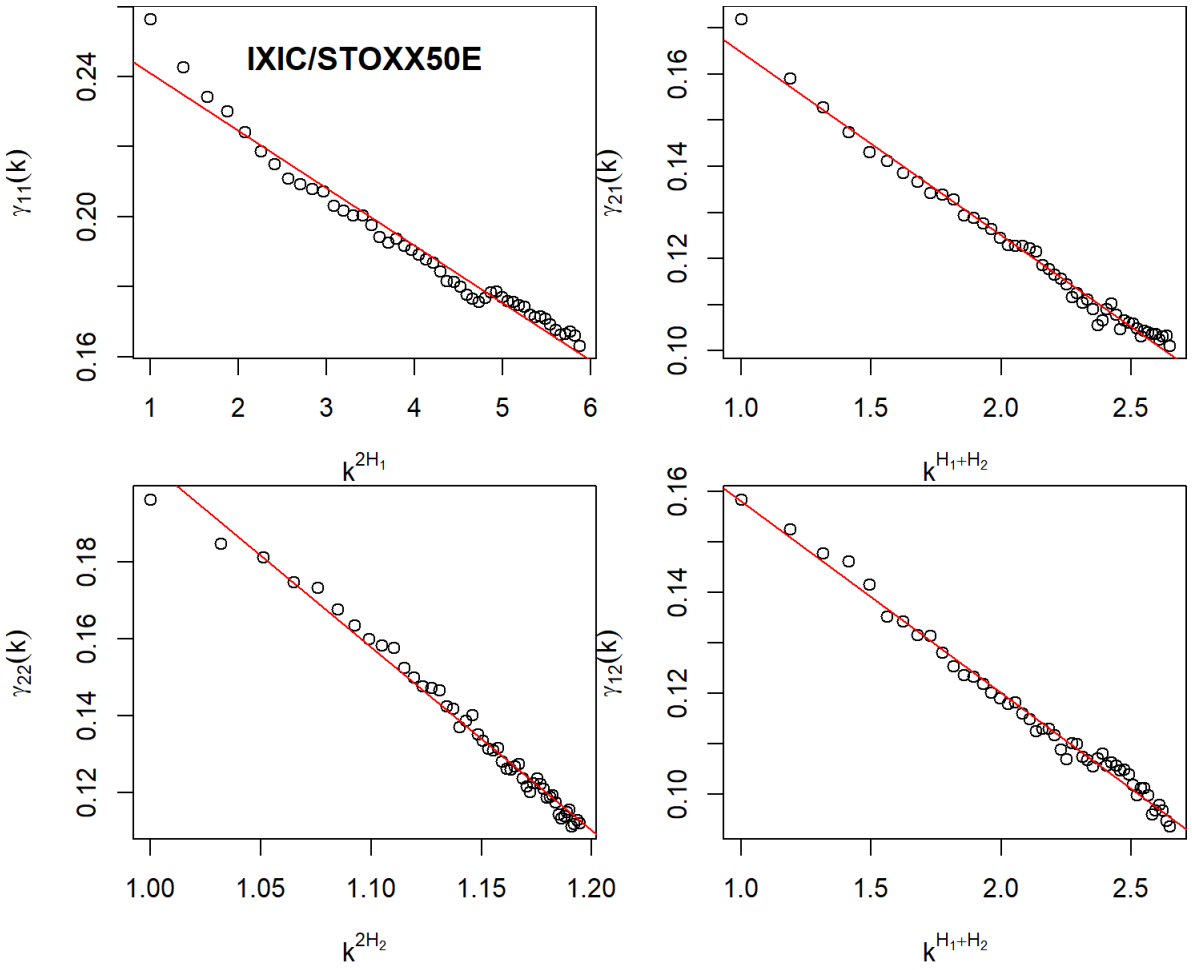

Figure 10: Cross-covariance function vs. suitable power of the lag for IXIC/STOXX50E.

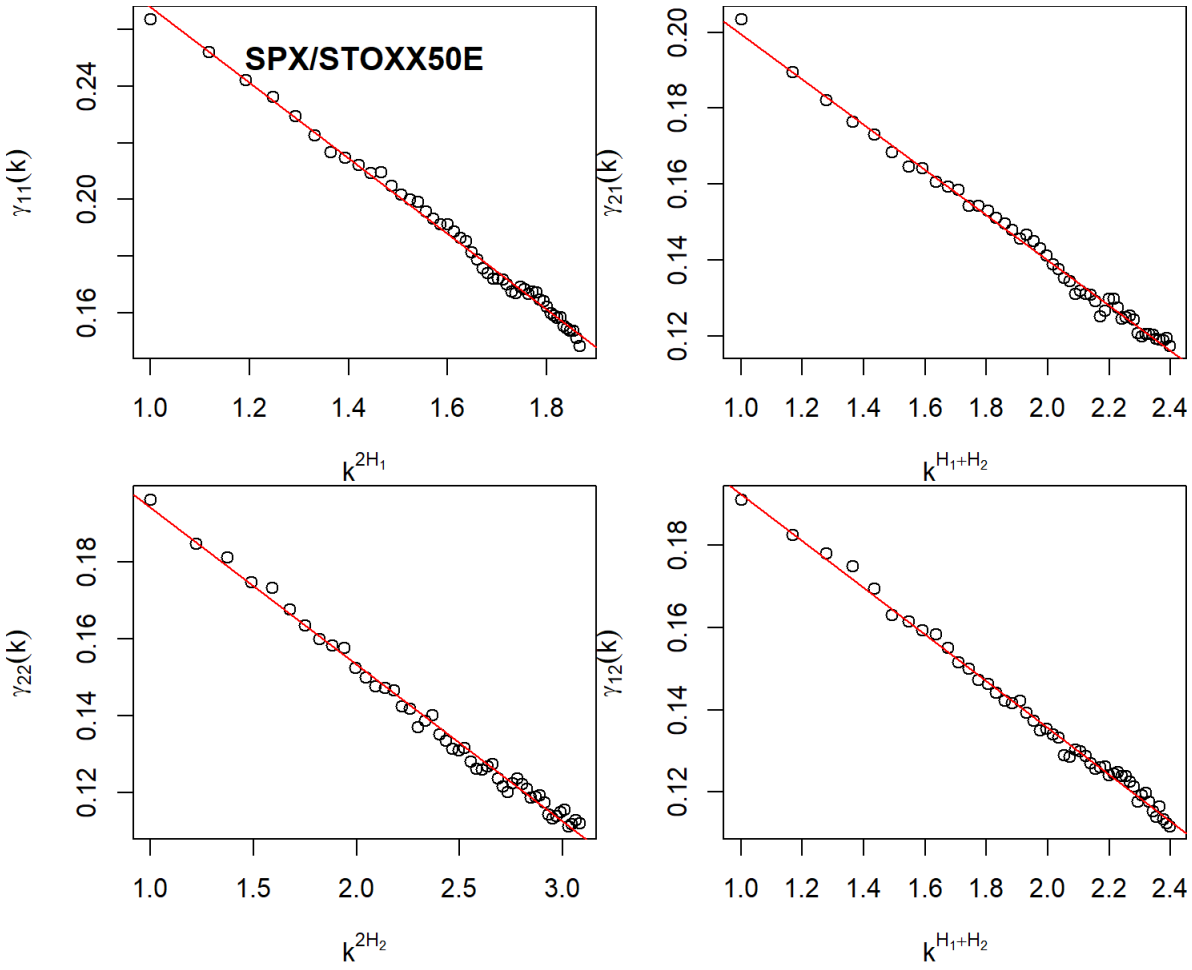

Figure 11: Cross-covariance function vs. suitable power of the lag for SPX/STOXX50E.

From the figures above we can see that the asymptotic linear relationship is indeed satisfied in the observed data. This gives us the ground on which we will exploit such relationship up to lag 50 in estimation.
It is also worth mentioning that the above relationship also highlights the almost non stationarity in the data, as was already mentioned in Proposition 4. This is because the linearity for k large implies that the mean reversion is weak, i.e. $\alpha,\ i = 1, 2$ is small, relative to the same time horizon.

5.3 Estimation results

In the following, we report estimation results obtained using the exact relationship with the two-step procedure, using the asymptotic relationship and the two-step procedure, and afterwards those obtained using the asymptotic relationship with GMM. The reasons behind using GMM are twofolds: it allows to estimate all the parameters at once, and it also allows us obtain standard errors.

Notice that when we use the asymptotic relationship we do not consider the α_{ij} but we do consider the variance and covariance of the process at lag 0 in the estimation procedure.

5.3.1 Two-step procedure and exact cross-correlation function

Table 7 reports estimates obtained with the two-step procedure in which we obtain the parameters govering the univariate distributions according to Wang, Xiao, and Yu (2021) and estimate ρ and η with the inversion of the exact cross-correlation function for a given lag as discussed in Section 3.1. The lag that we chose for this procedure was $k = 1$. Standard errors are obtained directly according to Wang, Xiao, and Yu (2021) for the univariate parameters and by using the Moving Block Bootstrap introduced by Lahiri (1993), which is proven to be valid in our case where $H_1 \wedge H_2 < \frac{3}{4}$, for ρ and η. The authors of the bootstrap methodology suggest a block size of length $B = CN^{\frac{1}{3}}$ where N is the length of the observed trajectory. In our case, we set $C = 70$ as it resulted the value around which the estimates of the standard errors were the most stable. The last column report the coherency constraint from Equation (2) which demonstrates the admissibility of our estimates.
The rows alternate between point estimates (larger font size) and standard errors (smaller font size). Standard errors for the coherency constraint do not have any meaning in the table.

Table 7: Empirical estimates on RV time series via 2-step method

symbol	ρ	η	a_1	a_2	v_1	v_2	H_1	H_2	μ_1	μ_2	**cohe**
IXICDJI	0.576	0.013	2.787	0.012	1.033	0.570	0.224	0.081	2.518	2.431	0.425
IXICDJI	0.062	0.036	0.612	0.048	0.119	0.068	0.021	0.022	0.036	2.992	0.301
IXICSPX	0.690	0.006	2.723	0.207	1.025	0.689	0.223	0.127	2.519	2.418	0.514
IXICSPX	0.096	0.039	0.606	0.185	0.118	0.081	0.021	0.021	0.037	0.246	0.349
IXICSTOXX50E	0.309	0.103	2.867	0.197	1.041	0.651	0.225	0.099	2.518	2.685	0.153
IXICSTOXX50E	0.066	0.034	0.620	0.188	0.119	0.077	0.021	0.021	0.036	0.223	0.274
IXICFTSE	0.254	0.120	2.663	0.006	1.037	0.510	0.225	0.072	2.527	2.540	0.165
IXICFTSE	0.030	0.066	0.598	0.034	0.119	0.061	0.021	0.022	0.038	5.318	0.971
IXICIBEX	0.314	0.091	2.209	0.422	0.983	0.626	0.215	0.148	2.525	2.705	0.122
IXICIBEX	0.061	0.050	0.550	0.258	0.113	0.073	0.021	0.021	0.043	0.116	0.570
IXICGDAXI	0.344	0.082	2.342	0.078	0.999	0.572	0.218	0.117	2.527	2.706	0.152
IXICGDAXI	0.048	0.030	0.564	0.115	0.115	0.068	0.021	0.021	0.041	0.524	0.212
DJISPX	0.930	-0.007	0.010	0.187	0.565	0.682	0.079	0.125	2.432	2.419	0.910

Table 7: Empirical estimates on RV time series via 2-step method *(continued)*

symbol	ρ	η	a_1	a_2	v_1	v_2	H_1	H_2	μ_1	μ_2	**cohe**
DJISPX	0.218	0.011	0.044	0.176	0.067	0.080	0.022	0.021	3.521	0.268	0.074
DJISTOXX50E	0.339	0.081	0.015	0.161	0.577	0.641	0.083	0.096	2.432	2.685	0.194
DJISTOXX50E	0.102	0.017	0.053	0.171	0.069	0.076	0.022	0.021	2.471	0.266	0.074
DJIFTSE	0.355	0.094	0.007	0.007	0.561	0.513	0.079	0.073	2.442	2.540	0.278
DJIFTSE	0.142	0.052	0.036	0.036	0.067	0.061	0.022	0.022	5.169	4.783	0.603
DJIIBEX	0.306	0.074	0.008	0.358	0.565	0.613	0.080	0.144	2.440	2.705	0.146
DJIIBEX	0.091	0.022	0.039	0.239	0.067	0.072	0.022	0.021	4.447	0.133	0.114
DJIGDAXI	0.362	0.054	0.010	0.057	0.569	0.558	0.081	0.113	2.442	2.706	0.166
DJIGDAXI	0.048	0.040	0.043	0.100	0.068	0.066	0.022	0.021	3.794	0.686	0.348
SPXSTOXX50E	0.354	0.096	0.270	0.195	0.710	0.650	0.132	0.098	2.417	2.684	0.194
SPXSTOXX50E	0.085	0.019	0.210	0.187	0.084	0.077	0.021	0.021	0.197	0.226	0.090
SPXFTSE	0.357	0.109	0.143	0.006	0.675	0.512	0.124	0.072	2.428	2.540	0.263
SPXFTSE	0.143	0.048	0.155	0.035	0.080	0.061	0.021	0.022	0.344	5.172	0.534
SPXIBEX	0.329	0.087	0.183	0.372	0.691	0.617	0.128	0.145	2.427	2.705	0.145
SPXIBEX	0.115	0.020	0.174	0.243	0.081	0.072	0.021	0.021	0.278	0.129	0.107
SPXGDAXI	0.374	0.072	0.226	0.084	0.705	0.575	0.131	0.118	2.428	2.706	0.171
SPXGDAXI	0.057	0.041	0.193	0.119	0.083	0.068	0.021	0.021	0.233	0.494	0.382
STOXX50EFTSE	0.637	0.014	0.379	0.016	0.682	0.528	0.117	0.078	2.687	2.528	0.425
STOXX50EFTSE	0.121	0.025	0.254	0.055	0.081	0.063	0.021	0.022	0.129	2.090	0.151
STOXX50EIBEX	0.668	-0.010	0.448	0.969	0.677	0.705	0.127	0.165	2.690	2.699	0.454
STOXX50EIBEX	0.180	0.025	0.273	0.383	0.080	0.082	0.021	0.021	0.111	0.060	0.176
STOXX50EGDAXI	0.802	-0.024	0.649	0.199	0.708	0.628	0.128	0.127	2.694	2.702	0.646
STOXX50EGDAXI	0.222	0.024	0.328	0.182	0.083	0.074	0.021	0.021	0.081	0.233	0.181
FTSEIBEX	0.395	-0.004	0.052	1.298	0.562	0.743	0.089	0.176	2.527	2.696	0.174
FTSEIBEX	0.073	0.035	0.098	0.439	0.067	0.086	0.022	0.021	0.708	0.049	0.282
FTSEGDAXI	0.508	-0.021	0.059	0.227	0.565	0.631	0.091	0.137	2.529	2.698	0.271
FTSEGDAXI	0.082	0.040	0.104	0.192	0.067	0.074	0.022	0.021	0.629	0.211	0.360

Table 7: Empirical estimates on RV time series via 2-step method *(continued)*

symbol	ρ	η	a_1	a_2	v_1	v_2	H_1	H_2	μ_1	μ_2	**cohe**
IBEXGDAXI	0.636	-0.013	1.310	0.213	0.743	0.627	0.177	0.133	2.700	2.699	0.413
IBEXGDAXI	0.126	0.026	0.440	0.187	0.086	0.074	0.021	0.021	0.048	0.221	0.165

In the context of ensuring positive semi definiteness of the covariance matrix, we observe that the values of the coherency constraint in the last columns are below 1 in all the cases, despite not imposing such constraint in estimation.

Regarding regularity of trajectories, we have that the values estimated for H_1 and H_2 are close to or below 0.2, which is far from the critical value of $\frac{1}{2}$, determining roughness in volatility time series as already found in Gatheral, Jaisson, and Rosenbaum (2018), Bolko et al. (2022), and Wang, Xiao, and Yu (2021), among others.

With emphasis on the univariate distributions, it is worth noticing that the $\alpha_i,\ i = 1, 2$ are estimated quite small, especially for indices like DJI and FTSE. These very small values seem hard to estimate and have an important impact on the standard errors of μ_i in the asymptotic theory that we adopt. Another issue with $\alpha_i,\ i = 1, 2$ is their instability with respect to the sampling period. We can see that slight variation in the sample, due to matching the joint time series missing data points, which can result in slight changes between different bivariate series containing common components, might result in quite different $\alpha_i,\ i = 1, 2$. Behavior that is less pronounced in other parameters' values.

In analyzing the dataset, particular attention should be given to ρ and η, which contribute to the cross-correlation structure of our bivariate model.

Regarding the estimates of ρ, which represent the contemporaneous correlation of the underlying fractional Brownian motions, we observe a wide range of values ranging between 0.314 for the couple IXIC/IBEX and 0.926 in the case of DJI/SPX, values that are consistent with what we would expect from economic considerations. All the values of ρ are positive, and according to the standard errors obtained by bootstrap as detailed above, they all appear statistically significant at the 5% confidence level.

The values of η appear to be both positive and negative and possess a range that goes from -0.024 for STOXX50E/GDAXI to 0.120 for IXIC/FTSE. It is hard to interpret the meaning of the η because it contributes to the asymmetry in the ccf together with all the other parameters. Significance for η could only be established in 7 out of 21 cases.

In order to show the reasons behind our choice not to use the exact ccf in estimation, however, we plot the empirical and theoretical exact ccf where the theoretical one is evaluated with parameters taken from Table

7. Again, we look at the three indices above but the remaining ones are available upon request.

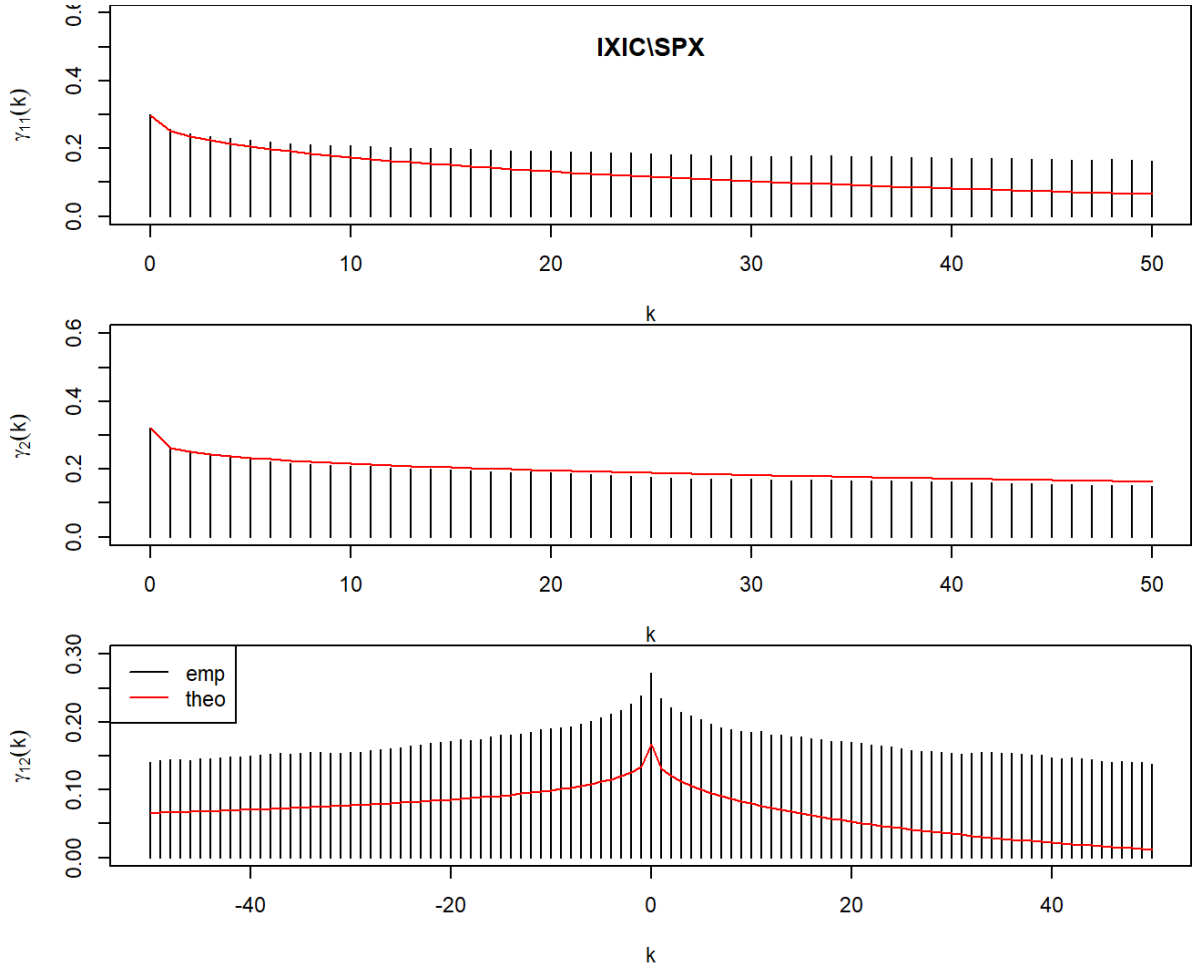

Figure 12: IXIC/SPX.

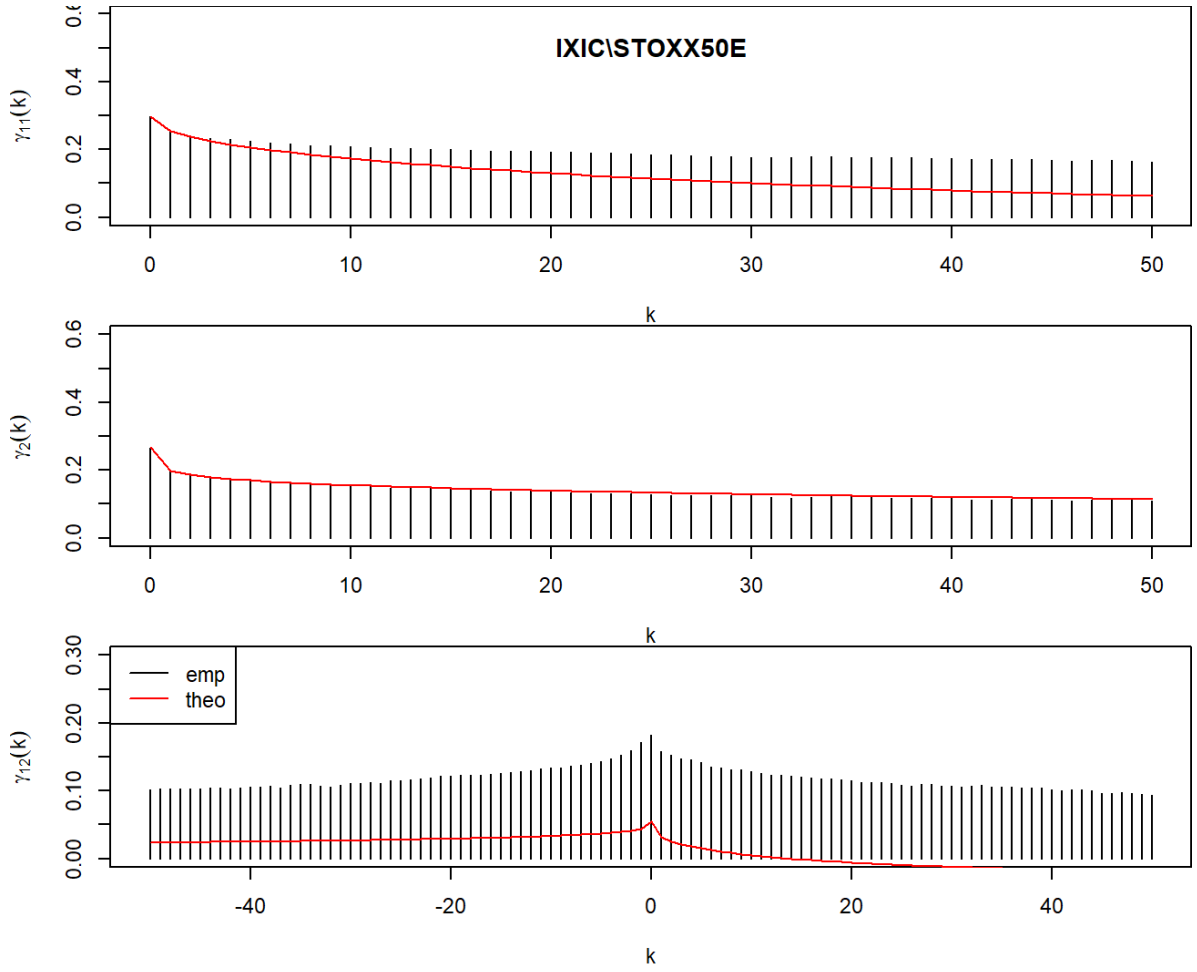

Figure 13: IXIC/STOXX50E.

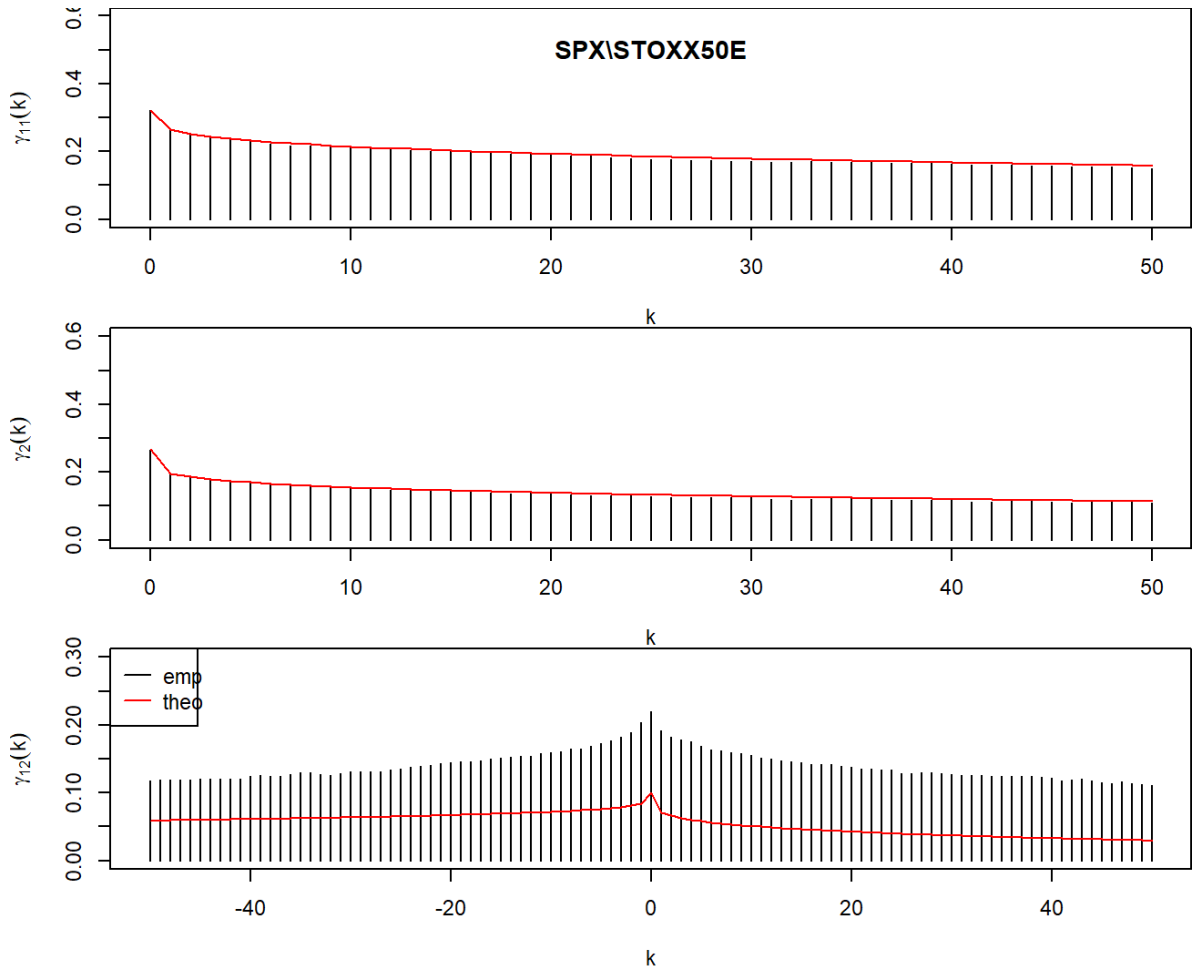

Figure 14: for SPX/STOXX50E.

Figures 12-14 show the difficulty in our two-step method based on the exact relationship in replicating the ccf estimated on real data. In one out of three cases (IXIC), the autocovariance function could not be well replicated theoretically with the exact relationship and the estimates at hand. In all three cases, instead, the full model estimates are not able at all at replicating the empirical cross-covariance, probably due to some parameters estimated poorly in the first stage.

Things already improve when we substitute the formula for the exact cross-covariance with the asymptotic one in the estimation of the bivariate parameters and afterwards in calculating and plotting the theoretical cross-covariance against the empirical one.

5.3.2 Two-step procedure and asymptotic cross-covariance function

In this paragraph we report the estimates of ρ and η obtained with the estimators based on the asymptotic small-lag cross-covariance relationship, which we name $\tilde{\rho}$ and $\tilde{\eta}$. The first step for the parameters govering

the marginal distributions is the one in Section 5.3.1. .

Table 7 reports estimates obtained for lag $k = 10$ because it seemed the one that delivered the best estimates in terms of cross-covariance fit, however, we will see that this method would not always work well if we changed the chosen lag. With regard to the previous methodology (table), using the asymptotic relationship of the cross-covariance function, which does not account for the $\alpha_i, \quad i = 1, 2$ always delivers higher ρs. In one case, unfortunately ρ is estimated above 1 and the coherency constraint is consequently not satisfied. The ηs on the other hand appear more moderate in magnitude than before.

Table 8: Empirical estimates on RV time series via 2-step method and asymptotic relationship

symbol	ρ	η	**cohe**
IXICDJI	0.709	0.024	0.645
IXICSPX	0.753	0.021	0.614
IXICSTOXX50E	0.462	0.026	0.253
IXICFTSE	0.482	0.041	0.323
IXICIBEX	0.436	0.029	0.199
IXICGDAXI	0.502	0.040	0.282
DJISPX	1.093	-0.003	1.256
DJISTOXX50E	0.616	0.019	0.386
DJIFTSE	0.688	0.029	0.488
DJIIBEX	0.572	0.024	0.360
DJIGDAXI	0.659	0.042	0.464
SPXSTOXX50E	0.590	0.020	0.359
SPXFTSE	0.669	0.034	0.492
SPXIBEX	0.564	0.024	0.322
SPXGDAXI	0.632	0.040	0.410
STOXX50EFTSE	0.734	0.003	0.561
STOXX50EIBEX	0.681	-0.017	0.473
STOXX50EGDAXI	0.794	-0.001	0.631
FTSEIBEX	0.520	-0.006	0.303
FTSEGDAXI	0.648	0.007	0.437
IBEXGDAXI	0.632	0.023	0.409

Figures 15-17 show the fit of the theoretical asymptotic auto-covariance and cross-covariance functions to the empirical ones represented by the histograms. We notice in the first place that the marginal fit of the auto-covariances (top) is broadly in line with the one obtained with the exact relationship in the previous paragraph. IXIC, which is characterized by one of the highest α in the entire dataset, stands out to have a bad marginal fit.

When we look at the bivariate fit of the cross-covariance (bottom), we see that the estimates reported in Table 8 together with the asymptotic formula do indeed improve the fit to the empirical cross-covariances. However, a drawback that is now highlighted is that changing the value of k can dramatically change the estimates of ρ and η and in turn worsen or improve the fit (notice the difference between the red and blue lines).

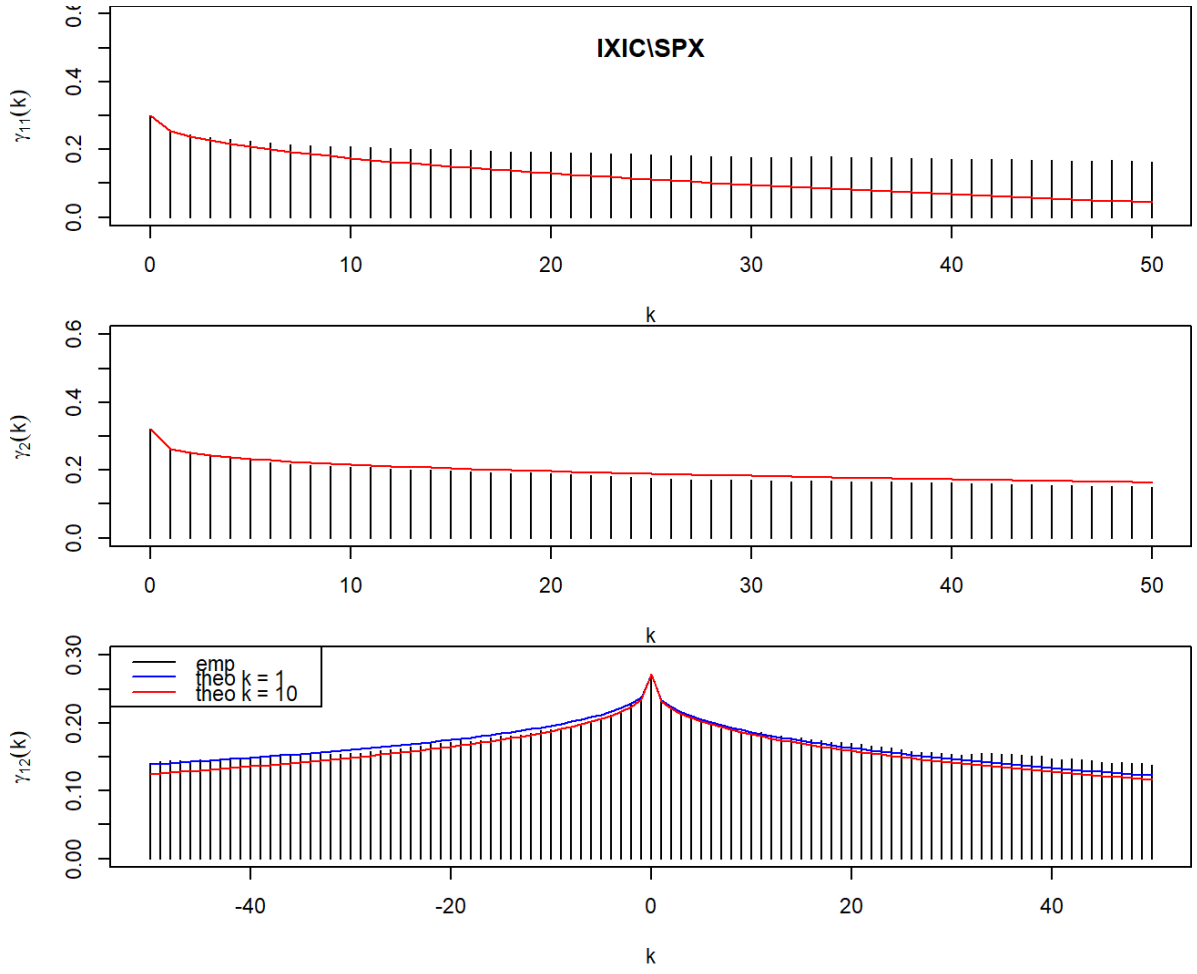

Figure 15: Empirical (histogram) and asymptotic theoretical (red and blue lines) cross-covariance function for IXIC/SPX.

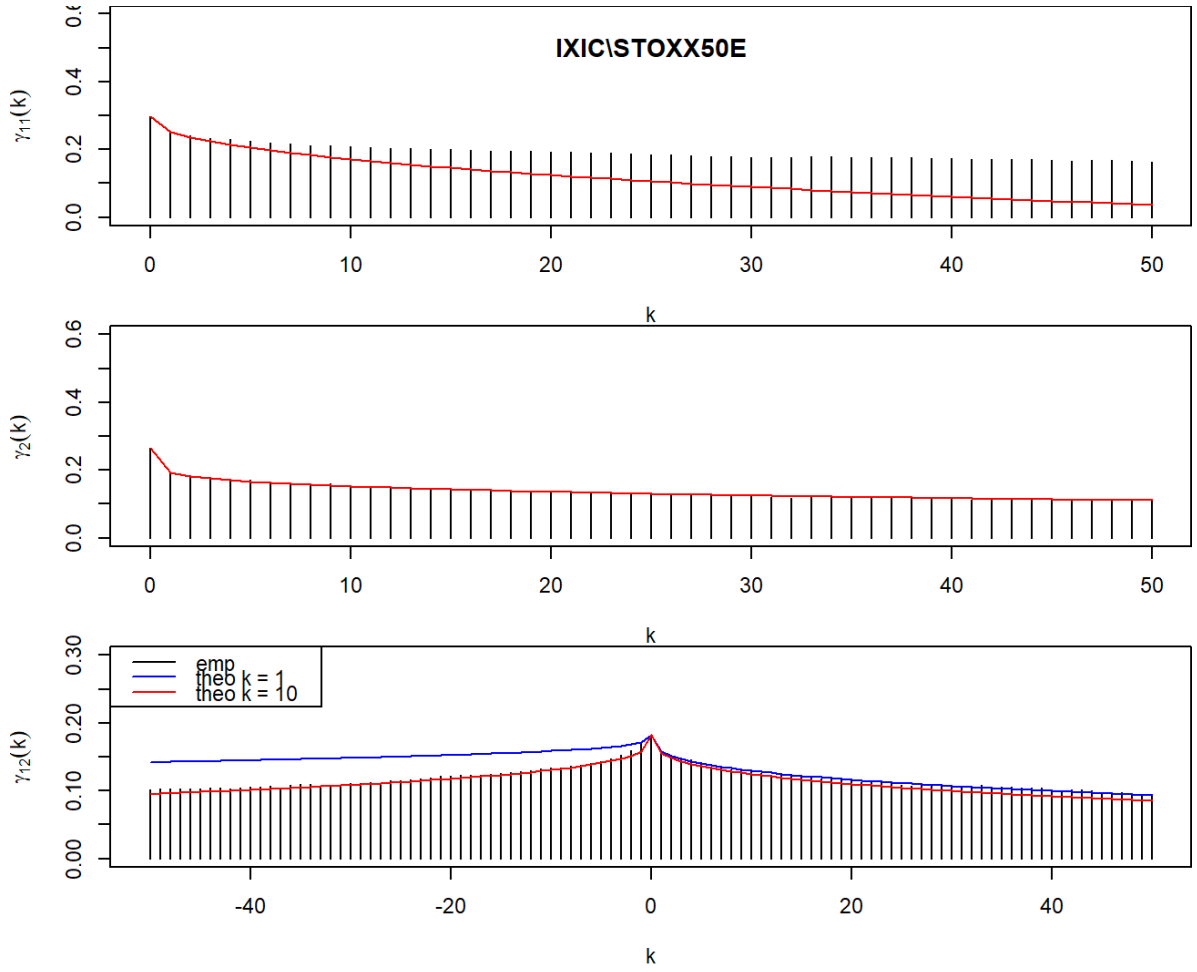

Figure 16: Empirical (histogram) and asymptotic theoretical (red and blue lines) cross-covariance function for IXIC/STOXX50E.

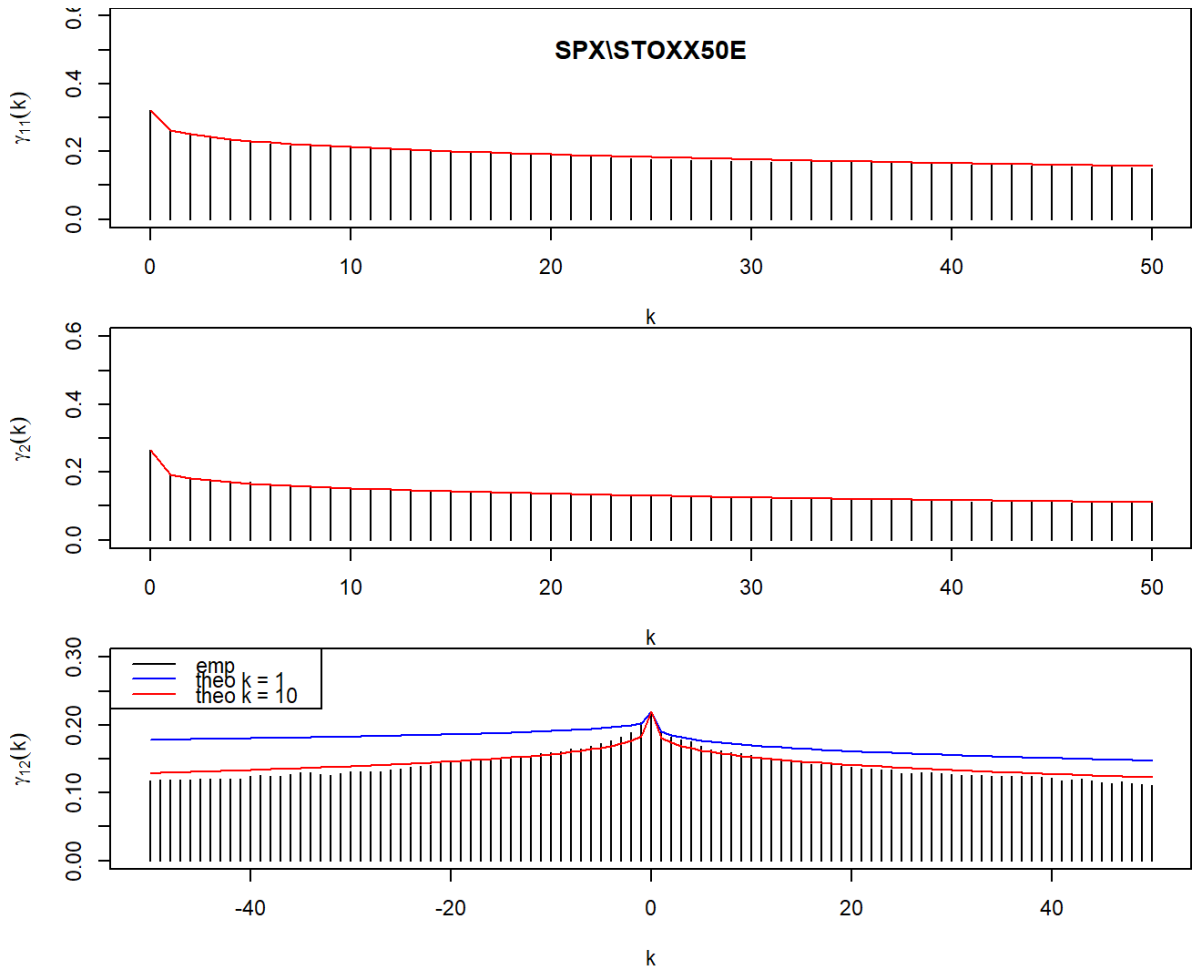

Figure 17: Empirical (histogram) and asymptotic theoretical (red and blue lines) cross-covariance function for SPX/STOXX50E.

In order to avoid the change in performance arising when choosing different lags for the cross-covariance to be inverted at, and simultaneously try to improve the fits of both auto-covariances and cross-covariances, we turn to GMM.

5.3.3 Generalized method of moments (GMM)

Technically speaking, the GMM procedure that we implemeneted exploits as orthogonality conditions the distances between the asymptotic cross-covariance (CCF) function $\tilde{\gamma}_{i,j}(l)$ and its empirical counterpart for lags $l = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50$ and $(i, j) = (1, 1), (1, 2), (2, 1), (2, 2)$. The number of lags was chosen based on the consideration that low-order lags are the most informative and some distant lags are necessary to recover the memory structure (see Bolko et al. 2022). The parameters to be estimated now do not include $\alpha_i, \quad i = 1, 2$ but, on the other hand, they include the values of the variances and cross-covariance at lag

0. The starting values of the optimization routine were given by the estimates obtained by the two-step procedure with $\tilde{\rho}$ and $\tilde{\eta}$ as bivariate parameters. The starting values are reported in Table 7, with ρ and η as in Table 8. Regarding variances and cross-covariance at lag 0, the starting values were the simple sample average estimators.

We attempted a proper GMM routine with the inverse of the covariance matrix of the moment conditions as weighting matrix of the loss function to be optimized but we obtained inferior results that adopting an identity matrix. One reason to be investigated might be that the moment structure shows very high degree of autocorrelation and cross-correlations which the standard Newey-West HAC procedure could not deal with. We therefore opted for a simple minimum distance estimation in which the weighing matrix was given by the identity matrix, thus giving equal weights to auto-covariances and cross-covariances at all lags included.

Remark 5. *Notice that the lags considered do not include 0, i.e. the empirical variance and covariance at lag 0 are not included as orthogonal conditions to be matched by the empirical ones. The reason lies in the fact that the asymptotic relationship shows the highest error with respect to the empirical counterparts at such lags (see Figures 9-11). In addition, in the model in Bolko et al. (2022), the effect of microstructure noise appears in the moment at lag zero of the integrated variance.*

Table 9 reports point estimates and standard errors for parameters of the bivariate fractional volatility model obtained via the Generalized Method of Moments (GMM) for each bivariate time series in our selection . As in Table 6, each symbol has two rows: one for point estimates (larger font) and one for standard errors (smaller font).

Table 9: Empirical estimates on RV time series via GMM

symbol	ρ	η	$\mathbf{Var}_1$	$\mathbf{Var}_2$	**Cov**	v_1	v_2	H_1	H_2	μ_1	μ_2	**cohe**
IXICDJI	0.787	0.274	0.023	0.342	0.260	0.566	0.676	0.141	0.073	2.512	2.425	0.690
IXICDJI	0.324	0.080	0.107	0.671	0.241	0.312	0.685	0.542	0.434	0.071	0.066	0.107
IXICSPX	0.801	0.266	0.031	0.360	0.261	0.562	0.737	0.145	0.082	2.515	2.417	0.703
IXICSPX	0.176	0.207	0.109	0.530	0.207	0.313	0.428	0.515	0.351	0.071	0.069	0.033
IXICSTOXX50E	0.535	0.259	0.053	0.303	0.172	0.621	0.697	0.198	0.053	2.512	2.686	0.452
IXICSTOXX50E	1.500	0.296	0.279	1.380	0.108	0.282	1.702	0.388	0.491	0.071	0.058	2.298
IXICFTSE	0.722	0.288	-0.008	0.243	0.208	0.731	0.740	0.224	0.214	2.518	2.532	0.522
IXICFTSE	0.463	0.700	0.076	0.036	0.034	0.313	0.440	0.467	0.269	0.071	0.062	0.236
IXICIBEX	0.493	0.316	0.033	0.238	0.162	0.741	0.673	0.177	0.159	2.515	2.701	0.247
IXICIBEX	0.438	1.055	0.100	0.091	0.061	0.276	0.274	0.635	0.288	0.073	0.058	0.243

Table 9: Empirical estimates on RV time series via GMM *(continued)*

symbol	ρ	η	$\mathbf{Var}_1$	$\mathbf{Var}_2$	**Cov**	v_1	v_2	H_1	H_2	μ_1	μ_2	**cohe**
IXICGDAXI	0.509	0.308	0.033	0.280	0.210	0.931	0.581	0.195	0.122	2.525	2.696	0.277
IXICGDAXI	0.470	1.356	0.099	0.185	0.070	0.218	0.152	0.416	0.450	0.072	0.068	0.222
DJISPX	0.990	0.310	0.001	0.334	0.319	0.662	0.688	0.090	0.096	2.429	2.416	0.981
DJISPX	0.831	0.582	0.010	0.339	0.360	0.324	0.207	0.383	0.349	0.063	0.065	0.692
DJISTOXX50E	0.622	0.302	0.027	0.269	0.220	0.611	0.665	0.101	0.093	2.425	2.683	0.395
DJISTOXX50E	0.450	0.926	0.098	0.352	0.257	0.325	0.239	0.700	0.361	0.064	0.057	0.255
DJIFTSE	0.770	0.316	0.021	0.244	0.240	0.828	0.736	0.163	0.211	2.437	2.532	0.604
DJIFTSE	0.234	1.219	0.049	0.036	0.050	0.290	0.423	0.366	0.258	0.067	0.061	0.058
DJIIBEX	0.650	0.329	0.005	0.236	0.182	0.827	0.675	0.125	0.203	2.439	2.698	0.448
DJIIBEX	0.218	1.562	0.066	0.043	0.080	0.282	0.379	0.521	0.267	0.067	0.057	0.055
DJIGDAXI	0.732	0.323	0.012	0.279	0.249	0.797	0.673	0.139	0.199	2.437	2.694	0.554
DJIGDAXI	0.193	1.403	0.049	0.043	0.071	0.285	0.407	0.442	0.293	0.069	0.068	0.040
SPXSTOXX50E	0.584	0.320	0.030	0.274	0.224	0.700	0.658	0.140	0.067	2.411	2.686	0.398
SPXSTOXX50E	0.545	1.329	0.129	0.794	0.218	0.274	0.912	0.513	0.457	0.068	0.058	0.299
SPXFTSE	0.727	0.330	0.030	0.256	0.250	0.898	0.741	0.204	0.179	2.422	2.532	0.532
SPXFTSE	0.329	2.109	0.050	0.060	0.046	0.290	0.351	0.345	0.265	0.071	0.062	0.112
SPXIBEX	0.562	0.335	0.016	0.231	0.185	0.726	0.658	0.136	0.157	2.421	2.702	0.318
SPXIBEX	0.267	2.258	0.067	0.092	0.106	0.325	0.273	0.755	0.298	0.071	0.057	0.111
SPXGDAXI	0.633	0.334	0.021	0.280	0.252	0.733	0.607	0.163	0.130	2.418	2.697	0.408
SPXGDAXI	0.354	2.314	0.065	0.156	0.107	0.319	0.195	0.635	0.407	0.072	0.068	0.139
STOXX50EFTSE	0.684	0.272	0.003	0.249	0.216	0.708	0.565	0.109	0.083	2.687	2.525	0.477
STOXX50EFTSE	0.303	1.680	0.024	0.421	0.307	0.286	0.433	0.622	0.483	0.060	0.059	0.095
STOXX50EIBEX	0.634	0.268	-0.012	0.230	0.187	0.662	0.642	0.089	0.167	2.691	2.695	0.442
STOXX50EIBEX	0.202	1.540	0.055	0.065	0.143	0.313	0.265	0.694	0.253	0.058	0.056	0.061
STOXX50EGDAXI	0.742	0.268	-0.017	0.284	0.248	0.616	0.542	0.124	0.083	2.687	2.694	0.575
STOXX50EGDAXI	0.174	1.465	0.075	0.421	0.247	0.305	0.445	0.617	0.534	0.060	0.067	0.031
FTSEIBEX	0.494	0.254	-0.002	0.224	0.174	0.562	0.493	0.077	0.189	2.522	2.688	0.295

Table 9: Empirical estimates on RV time series via GMM
(continued)

symbol	ρ	η	$\mathbf{Var}_1$	$\mathbf{Var}_2$	**Cov**	v_1	v_2	H_1	H_2	μ_1	μ_2	**cohe**
FTSEIBEX	0.216	2.060	0.071	0.046	0.121	0.357	0.417	1.326	0.435	0.060	0.057	-0.009
FTSEGDAXI	0.626	0.255	-0.011	0.283	0.217	0.578	0.531	0.088	0.081	2.525	2.691	0.395
FTSEGDAXI	1.193	1.896	0.070	0.488	0.414	0.329	0.555	1.058	0.614	0.063	0.071	-1.901
IBEXGDAXI	0.564	0.241	-0.007	0.284	0.190	0.503	0.504	0.107	0.074	2.693	2.690	0.330
IBEXGDAXI	0.678	1.901	0.070	0.600	0.317	0.319	0.818	1.239	0.704	0.057	0.071	-0.014

After optimizing, again, with the goal to fit the asymptotic auto-covariance and cross-covariance functions, the coherency constraint is satisfied in all the cases.
Also, as already mentioned, the $\alpha_i,\ i = 1, 2$ are not quantities that have been estimated but we have instead the variances and covariance at lag 0.
The algorithm seems to suggest rougher trajectories than the previous method as both H_1 and H_2 are now smaller in value.

The values of ρ are now estimated quite high, in line with those from Table 8, with the difference that the ceiling value of 1 is never exceeded. It now ranges between 0.99 for DJI/SPX and 0.49 for IXIC/IBEX. The extreme values seem still valid heuristically since we have the highest contemporaneous correlation of the underlying fractional Brownian motions for two American indices and the lowest for two indices from different countries (US and EU).

Regarding the remaining parameters of the model, namely η, $\nu_i \quad i = 1, 2$, and $H_i \quad i = 1, 2$ we do not see any specific pattern when comparing to previous estimates. The values vary more or less in the same range as before, with the only difference that the values now change considerably when pairing a time series with different ones.

The drawback of this method appears to be twofolds: (i) the estimated standard errors appear extremely high, making most of the parameters strangely indistinguishable from zero, and (ii) univariate parameter values are inconsistent across different bivariate time seires of which they are component. The first drawback is known to affect these kinds of low-efficiency estimation methods (especially when the weighting matrix is given by the identity one), despite it seems here more problematic than elsewhere. The second issue arises from the fact that the values assigned to the parameters governing the marginal distributions change in order to fit simultaneously auto-covariances and cross-covariances, therefore chaning for each bivariate combination.

However, these estimates, in conjunction with the small-lag asymptotic formula for the cross-covariance, are able to fit properly the empirical auto-covariances and cross-covariances as shown in the next plots.

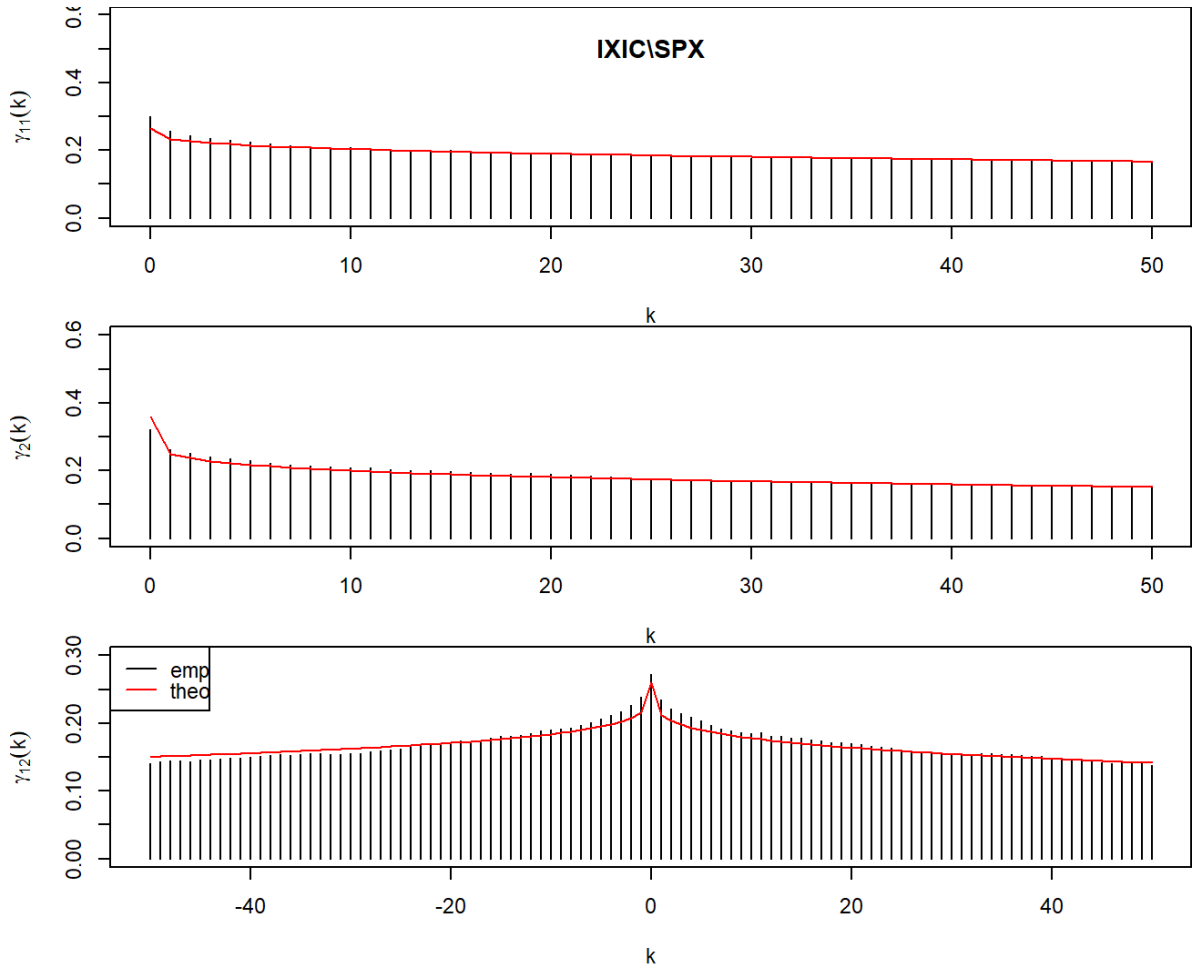

Figure 18: Empirical (histogram) and asymptotic theoretical (red line) cross-covariance function for IXIC/SPX.

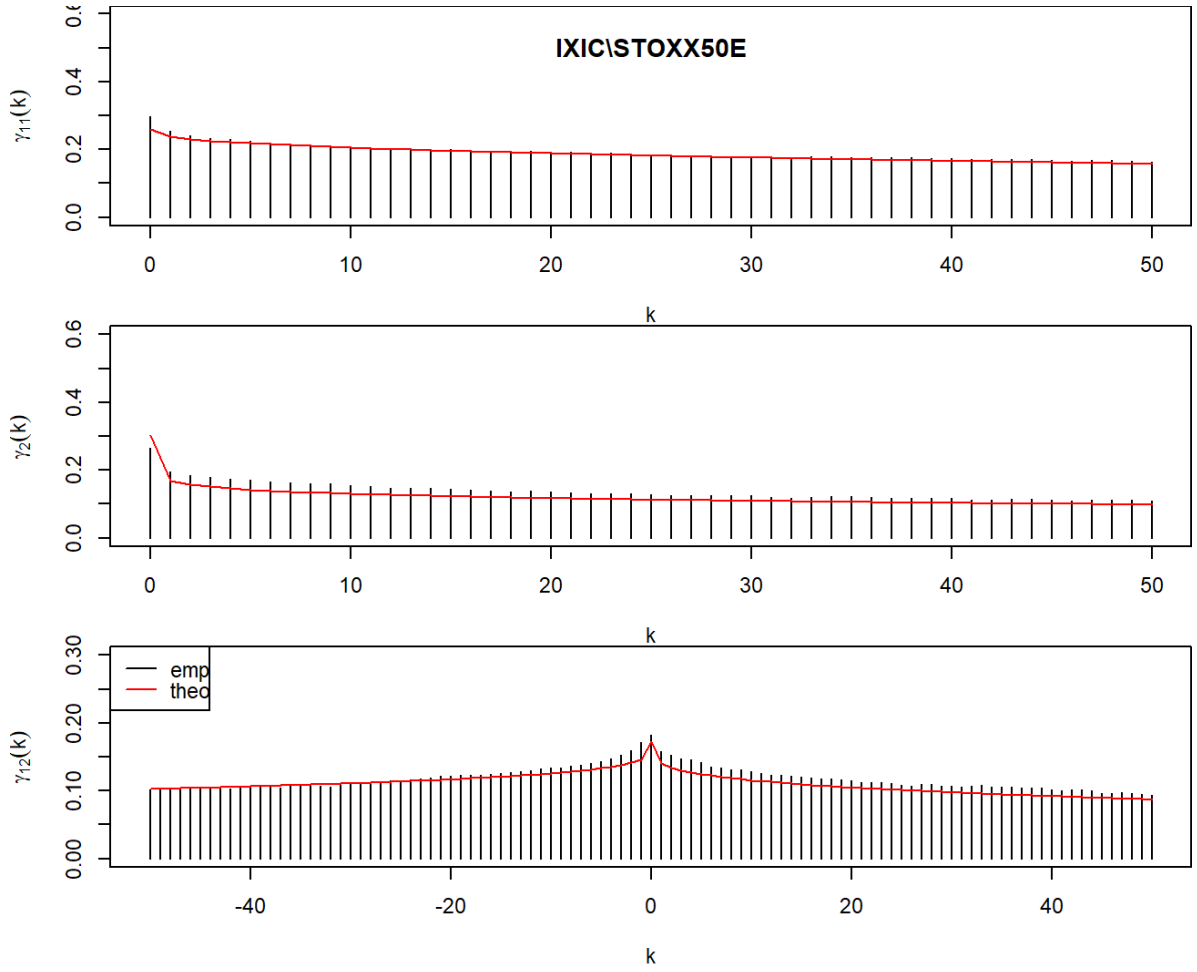

Figure 19: Empirical (histogram) and asymptotic theoretical (red line) cross-covariance function for IXIC/STOXX50E.

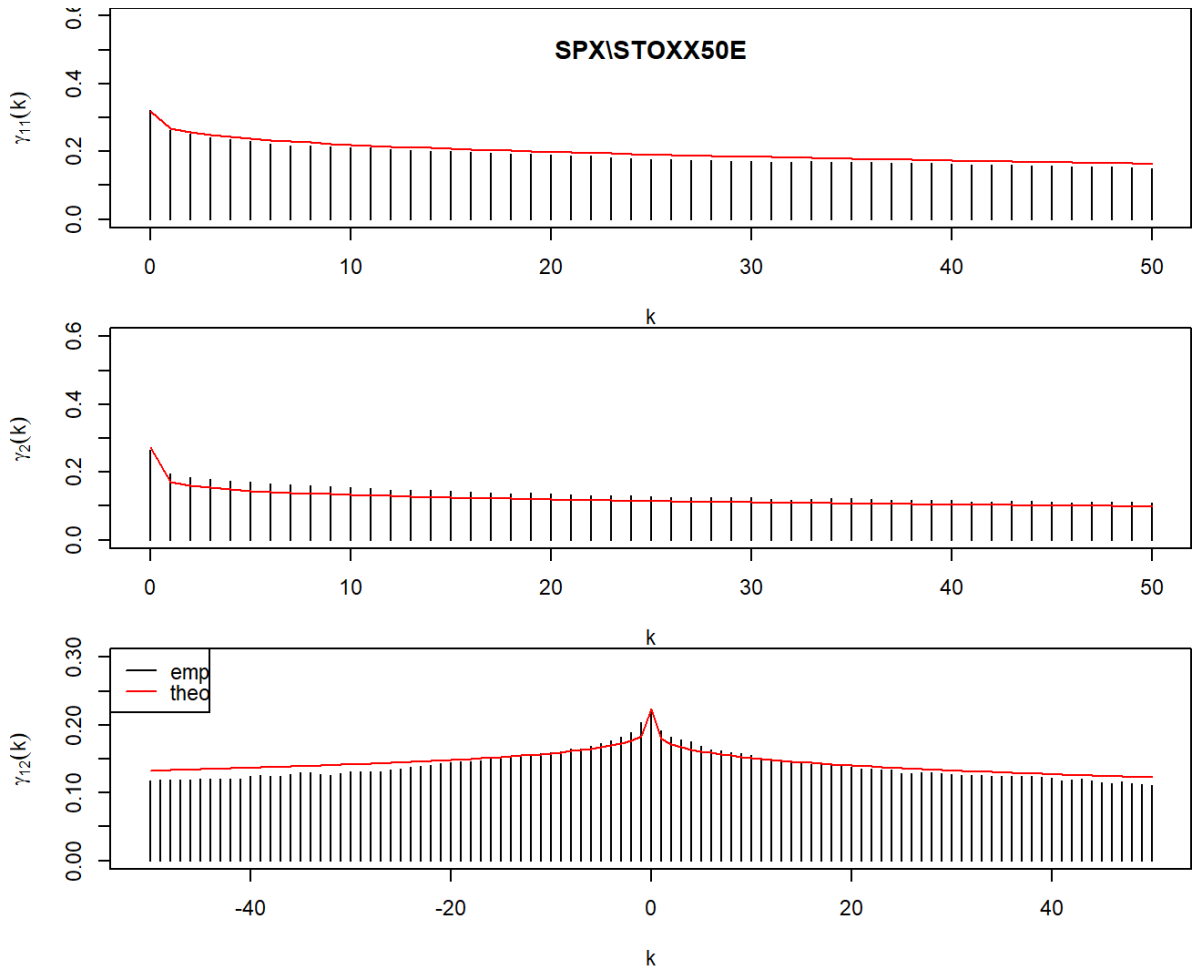

Figure 20: Empirical (histogram) and asymptotic theoretical (red line) cross-covariance function for SPX/STOXX50E.

Figures 18 to 20 show us that our proposed model is able to indeed fit the empirical auto-covariance and cross-covariance profiles reasonably well for up to 50 lags when the estimation considers only a bunch of them. Both univariate and bivariate profiles appear to line up well on top of the empirical estimates with only negligible errors remaining.

To easily visualize the correlation between the underlying Brownian drivers across volatility indices, we can recur to an undirected force graph as the one shown in Figure 21. The plot shows the indices under analysis as nodes of the graph. The length of the corresponding edges of the graph are inversely proportional to the ρs characterizing their bivariate relationships, which values are also superposed to the edge.
It is easy to see that realized volatilities of major US indices are more correlated among themselves than European ones and that the DJI realized volatility is the most connected to its European counterparties.

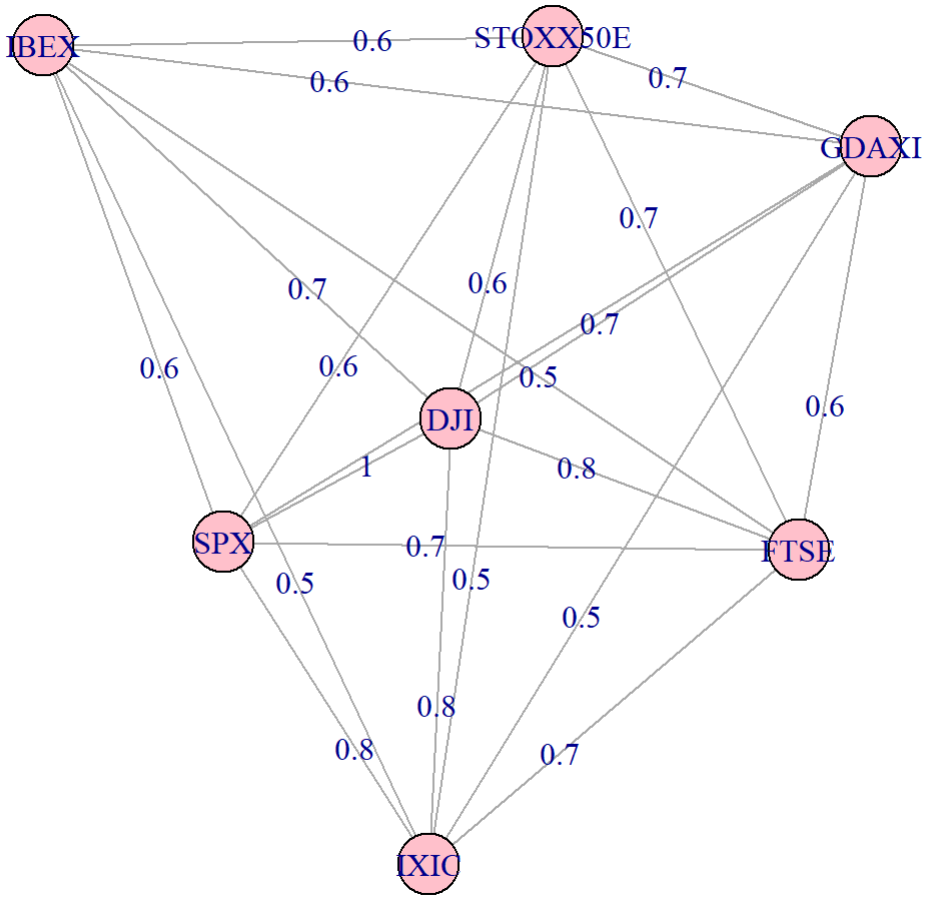

Figure 21: Empirical (histogram) and asymptotic theoretical (red line) cross-covariance function for SPX/STOXX50E.

5.3.4 Comparison between the two-step procedure and the GMM method

In general, it is hard to determine which estimation procedure works best on real data.

However, considering that the estimates obtained with the GMM method appear to fit the cross covariance function better and the fact that this method avoids dealing with the critical $\alpha_i,\ i = 1, 2$ or the choice of a lag k, we deem the GMM method as an appropriate choice despite the not completely stable estimates across series as well as the huge standard errors. Therefore we will employ this estimation technique in the forecasting scheme of the next section.

6 Forecasting

In this section we try to forecast future observations of the realized log-volatility time series using the GMM estimation routine detailed above. In addition, due to the close to non stationary behavior of the log-volatility time series, we will pair our GMM estimates to two closely related methods when forecasting which will become clear later: using the barely stationary bivariate fractional Ornstein-Uhlenbeck process or using the non-stationary fractional Brownian Motion.

6.1 Methodology

The forecasting exercise is carried out with standard formulae for Gaussian random variables adapted from Bennedsen et al. (2022) (Section 5.1). This approach, which is based on discrete time observations and matrix algebra, is even deemed optimal, and therefore better than a proper conditional expectation that assumes continuous time and requires discretization, according to Wang, Yu, and Zhang (2024) for the fractional Brownian motion. In addition, we conjecture that their argument remains valid for the case of the 2fOU with almost non-stationary behavior.

For a zero-mean Gaussian random vector $(x^1_{t+h}, x^1_t, x^1_{t-1}, \dots, x^1_{t-m+1}, x^2_{t+h}, x^2_t, x^2_{t-1}, \dots, x^2_{t-m+1})^T$, the distribution of $(x^1_{t+h}, x^2_{t+h})^T$ conditionally on its present and past observations

$$a = [x^1_t, x^1_{t-1}, \dots, x^1_{t-m+1}, x^2_t, x^2_{t-1}, \dots, x^2_{t-m+1}]^T \in \mathbb{R}^{2m}, \tag{32}$$

is given in closed form as

$$(x^1_{t+h}, x^2_{t+h})^T | a \ \overset{d}{=}\ \mathcal{N}(\mu, \xi^2), \tag{33}$$

where

$$\mu = \Gamma_{12}\Gamma_{22}^{-1} a, \tag{34}$$

and

$$\xi^2 = Var((x^1_t, x^2_t))(1 - \Gamma_{12}\Gamma 22^{-1}\Gamma_{21}) \tag{35}$$

with $\Gamma_{21} = \Gamma_{12}^T$.

In particular, Γ_{22} is the covariance matrix of the vector a,

$$\Gamma_{22} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ (\Lambda_{12})^T & \Lambda_{22} \end{bmatrix} \tag{36}$$

where, denoting for brevity $\text{Cov}(X,Y) := \gamma(X,Y)$, we can decompose the blocks of Γ_{22}, due to stationarity, as $(\Lambda_{11})_{i,j} = \gamma\left(x^1_{|i-j|}, x^1_0\right)$, $(\Lambda_{22})_{i,j} = \gamma\left(x^2_{|i-j|}, x^2_0\right)$

$$\begin{aligned}
\Lambda_{12} &= \begin{bmatrix} \gamma(x^1_t, x^2_t) & \gamma(x^1_t, x^2_{t-1}) & \dots & \gamma(x^1_t, x^2_{t-m}) \\ \gamma(x^1_{t-1}, x^2_t) & \gamma(x^1_{t-1}, x^2_{t-1}) & \dots & \gamma(x^1_{t-1}, x^2_{t-m}) \\ \dots & & & \\ \gamma(x^1_{t-m}, x^2_t) & \gamma(x^1_{t-m}, x^2_{t-1}) & \dots & \gamma(x^1_{t-m}, x^2_{t-m}) \end{bmatrix} \\
&= \begin{bmatrix} \gamma(x^1_0, x^2_0) & \gamma(x^1_0, x^2_{-1}) & \dots & \gamma(x^1_0, x^2_{-m}) \\ \gamma(x^1_{-1}, x^2_0) & \gamma(x^1_{-1}, x^2_{-1}) & \dots & \gamma(x^1_{-1}, x^2_{-m}) \\ \dots & & & \\ \gamma(x^1_{-m}, x^2_0) & \gamma(x^1_{-m}, x^2_{-1}) & \dots & \gamma(x^1_{-m}, x^2_{-m}) \end{bmatrix} \\
&= \begin{bmatrix} \gamma(x^1_0, x^2_0) & \gamma(x^1_1, x^2_0) & \dots & \gamma(x^1_m, x^2_0) \\ \gamma(x^1_0, x^2_1) & \gamma(x^1_0, x^2_0) & \dots & \gamma(x^1_{m-1}, x^2_0) \\ \dots & & & \\ \gamma(x^1_0, x^2_m) & \gamma(x^1_0, x^2_{m-1}) & \dots & \gamma(x^1_0, x^2_0) \end{bmatrix} \in \mathbb{R}^{2(m+1)\times 2(m+1)}
\end{aligned} \tag{37}$$

and

$$\begin{aligned}
\Gamma_{12} &= \begin{bmatrix} \gamma(x^1_{t+h}, x^1_t) & \dots & \gamma(x^1_{t+h}, x^1_{t-m}) & \gamma(x^1_{t+h}, x^2_t) & \dots & \gamma(x^1_{t+h}, x^2_{t-m}) \\ \gamma(x^2_{t+h}, x^1_t) & \dots & \gamma(x^2_{t+h}, x^1_{t-m}) & \gamma(x^2_{t+h}, x^2_t) & \dots & \gamma(x^2_{t+h}, x^2_{t-m}) \end{bmatrix} \\
&= \begin{bmatrix} \gamma(x^1_h, x^1_0) & \dots & \gamma(x^1_{h+m}, x^1_0) & \gamma(x^1_h, x^2_0) & \dots & \gamma(x^2_{h+m}, x^1_0) \\ \gamma(x^2_h, x^1_0) & \dots & \gamma(x^2_{h+m}, x^1_0) & \gamma(x^2_h, x^2_0) & \dots & \gamma(x^2_{h+m}, x^2_0) \end{bmatrix} \in \mathbb{R}^{2\times 2(m+1)}.
\end{aligned} \tag{38}$$

We can easily interpret $x_t = \log v_t - \mu$ being represented by the bivariate fractional Ornstein-Uhlenbeck process and implement the previous machinery in forecasting.

Therefore, the mechanism boils down to obtaining some weights $w = \Gamma_{12}\Gamma_{22}^{-1}$ that allow the computation of the conditional expectation of future realizations of the process. The flexibility is in the fact that Γ_{12} and Γ_{22} could be constructed for any Gaussian process. This allows us to obtain forecasts where the matrices are built either using the small-lag asymptotic relationship for cross-covariances (and hence the name barely stationary) of the 2fOU or the cross-covariance structure of a non stationary Brownian motion. According to how volatility time series seem to behave, the two should deliver similar results.

Additionally, we would like to compare our forecasts to a benchmark. The choice falls on the bivariate version of the famous HAR (see Corsi 2009) which appeared first in Bubak, Kocenda, and Zikes (2011). In particular, the vector HAR (VHAR) is specified by the equations

$$\log \mathbf{v}_t = \beta_0 + \boldsymbol{\beta}_1 \log \mathbf{v}_{(t-1|t-5)} + \boldsymbol{\beta}_2 \log \mathbf{v}_{(t-1|t-22)} + \boldsymbol{\epsilon}_t \tag{39}$$

where the $\boldsymbol{\beta}$ are square matrices of the coefficients (in our case 2x2), $\log \mathbf{v_t}$ is a bivariate observation of log-vol, and

$$\log \mathbf{v}_{(t-1|t-k)} = \frac{1}{k} \sum_{j=1}^{k} \log \mathbf{v}_{t-j}. \tag{40}$$

The VHAR is estimated by OLS.

Since our forecasting period covers part of the Covid pandemic, the metric that we chose to compare our forecasts is the median absolute error, which is supposed to be robust to extreme events.

6.2 Results

In this section, we report the results related to the forecasting exercise performed using a rolling window scheme. The rolling window scheme encompasses that the information used for estimation shifts ahead one-step at a time while we move the forecasting target one-step in the future.
In particular, estimation is always carried out on a sample of 4000 observations and the forecast is performed for the observation in place $4000 + h$, where h is the forecasting period, which ranges in the set of one day, one week and two weeks (in terms of trading days). For example, the first iteration for the one day ahead forecast will be using the first 4000 non-missing observations and the forecast will be attempted for day 4001. The orthogonal projection predicated in (32) and (34) is performed on a year of past data, that is on 252 observations taken at the end of the estimation sample. This choice is robust towards other choices of the projection space and is also due to computational feasibility. Accounting for missing values in the time series under analysis, the out-of-sample period consists of around 1500 observations, one quarter of which happen to be sampled during the Covid pandemic.

The following tables report the ratios of median absolute forecasting errors between the model under consideration at the numerator (2fOU or 2fBM) and the 2HAR at the denominator for all the bivariate series under analaysis. Forecasting was performed according to the methodology outlined above.
The number in the header refers to the component of the bivariate time series.

1-day ahead

Table 10: 1-day ahead rolling window forecasts

value	fbm1	fbm2	fou1	fou2
IXICDJI	**0.98**	**0.98**	**0.96**	1
IXICSPX	**0.98**	**0.99**	**0.99**	1.01
IXICSTOXX50E	**0.98**	**0.96**	1.01	1.03
IXICFTSE	**0.99**	1	1.01	1.02

value	fbm1	fbm2	fou1	fou2
IXICIBEX	**0.96**	1.02	**0.98**	1.03
IXICGDAXI	**0.97**	1	**0.98**	1.02
DJISPX	1.14	1.06	1.13	1.07
DJISTOXX50E	**0.98**	1.01	**0.98**	1.02
DJIFTSE	**0.97**	**0.99**	**0.97**	**0.99**
DJIIBEX	1.01	1.02	1.02	1.03
DJIGDAXI	1	**0.99**	1.01	1.01
SPXSTOXX50E	1	**0.97**	**0.99**	**0.97**
SPXFTSE	**0.98**	**0.99**	**0.98**	1
SPXIBEX	**0.99**	1.03	1	1.04
SPXGDAXI	**0.98**	1.01	1	1.02
STOXX50EFTSE	**0.97**	1	**0.97**	1
STOXX50EIBEX	**0.98**	1.03	1.01	1.03
STOXX50EGDAXI	**0.98**	**0.98**	**0.98**	**0.98**
FTSEIBEX	1.01	1.03	1.04	1.05
FTSEGDAXI	1.03	1	1.03	1
IBEXGDAXI	**0.99**	**0.99**	**0.99**	1

5-day ahead

Table 11: 5-day ahead rolling window forecasts

value	fbm1	fbm2	fou1	fou2
IXICDJI	**0.99**	1	**0.99**	1.03
IXICSPX	1	**0.97**	1	**0.99**
IXICSTOXX50E	1	1.01	1.06	1.11
IXICFTSE	1.02	**0.97**	1.03	1.06
IXICIBEX	1	**0.98**	1.01	**0.99**
IXICGDAXI	1.01	**0.99**	1.04	1.05
DJISPX	1.11	1.06	1.13	1.09
DJISTOXX50E	**0.94**	1.02	**0.96**	1
DJIFTSE	**0.96**	1	**0.97**	**0.99**
DJIIBEX	**0.99**	**0.99**	**0.99**	**0.98**
DJIGDAXI	1	**0.98**	**0.99**	**0.98**

value	fbm1	fbm2	fou1	fou2
SPXSTOXX50E	**0.97**	**0.99**	**0.97**	**0.99**
SPXFTSE	**0.95**	**0.99**	**0.96**	**0.98**
SPXIBEX	**0.96**	**0.98**	**0.96**	**0.99**
SPXGDAXI	**0.97**	**0.98**	**0.98**	1
STOXX50EFTSE	**0.98**	**0.97**	**0.98**	**0.96**
STOXX50EIBEX	**0.99**	**0.97**	1.04	**0.98**
STOXX50EGDAXI	**0.97**	1	**0.98**	1
FTSEIBEX	**0.98**	1	1.01	1.03
FTSEGDAXI	**0.98**	**0.99**	1	**0.99**
IBEXGDAXI	**0.97**	1.01	1.03	1.02

10-day ahead

Table 12: 10-day ahead rolling window forecasts

value	fbm1	fbm2	fou1	fou2
IXICDJI	1	**0.98**	**0.98**	1
IXICSPX	**0.99**	**0.97**	**0.98**	1.01
IXICSTOXX50E	**0.99**	**0.98**	1.03	1.1
IXICFTSE	**0.99**	**0.98**	1.01	1.04
IXICIBEX	1.01	**0.99**	1.03	**0.99**
IXICGDAXI	1	1	1.05	1.09
DJISPX	1.04	1.04	1.06	1.08
DJISTOXX50E	1.01	**0.98**	1.02	**0.98**
DJIFTSE	**0.96**	**0.98**	**0.96**	**0.96**
DJIIBEX	**0.99**	**0.98**	**0.99**	**0.98**
DJIGDAXI	1.01	1.01	1.02	1.02
SPXSTOXX50E	**0.96**	**0.96**	**0.97**	**0.95**
SPXFTSE	**0.93**	1	**0.92**	**0.98**
SPXIBEX	**0.96**	**0.97**	**0.97**	1.04
SPXGDAXI	**0.96**	1.01	**0.98**	1.02
STOXX50EFTSE	**0.95**	**0.98**	**0.95**	**0.98**
STOXX50EIBEX	**0.97**	**0.97**	**0.98**	**0.96**
STOXX50EGDAXI	**0.97**	1.03	**0.97**	1.03

value	fbm1	fbm2	fou1	fou2
FTSEIBEX	1.01	**0.99**	1.04	1.01
FTSEGDAXI	1.01	1.02	1.03	1.04
IBEXGDAXI	**0.97**	1.02	1.04	1.04

Tables 10 - 12 show that our model compares favourably against the benchmark. In many cases we did get an advantage in modeling the bivariate time series with the bivariate fractional volatility model, in some others we did not.

In particular, it seemed to be the simple bivariate fractional Brownian motion to perform best between the two proposed models. Despite the fact that the two models seemed to behave similarly. The results are however quite weak as the ratio is usually very close to one and when performing the Diebold-Mariano test (Diebold and Mariano 2002), we found most often than not the difference in forecasting accuracy to be insignificant.

However, we can notice that some indices gain advantage more than others from our modeling technique. In particular, IXIC and STOXX50E seem to benefit the most from both 2fOU and 2fBM modeling perspectives. There might be reasons behind the performance on such indices and not others, related to spillover effects that our model exploits better than the other.

7 Conclusion

Recently, the idea that volatility is rough was introduced by Gatheral, Jaisson, and Rosenbaum (2018) with their Rough Fractional Stochastic Volatility model. The idea they put forward was that realized volatility could be modeled by a simple fractional Brownian motion due to its monofractal scaling behavior. In addition to give volatility the stationary property, they characterized it with a mean reverting fractional process which behaved locally as a fractional Brownian motion.

Building on their idea, we proposed a bivariate version of the model. In order to allow for different self-similarity coefficients across the different components we used the multivariate fractional Brownian motion proposed by Amblard et al. (2010) as driving term in the Langevin equation. This allowed us to decouple the regularity of the trajectories and the decay of the auto-covariance of different components of the bivariate model and introduced a novel parameter η, which drives the time-reversibility of the underlying fractional Brownian motion.

In order to identify the model, we introduced different estimation procedures based on the method of moments. Our generalized moment estimators was able to infer the parameters ρ and η after taking as given those governing the univariate marginal distributions. We proposed this kind of estimator in a general framework and in a specific one which arises when the process mean reverts slowly. We named the latter small-lag

estimator because of its ability in identifying the parameters when the lag of the cross-covariance given the mean-reversion speed is too small to observe any meaningful mean-reverting behavior. In order to estimate all the parameters at once, we proposed another estimation procedure that relied on the Genrealized Method of Moments. In the latter case, the estimates that we obtained with the generalized moment estimator were given as starting values to the optimizer. It was worth noticing that the newly estimated self-similarity coefficients H appeared well below 0.5, confirming roughness in the trajectories.
We derived asymptotic theories for all sets of estimators and validated it successfully with Monte Carlo simulations.

After we estimated the model on real data, we could confirm the results in Gatheral, Jaisson, and Rosenbaum (2018) that the volatility time series mean revert slowly also from a bivariate perspective, where we looked at cross-covariances. This characteristic allowed us to employ the small-lag estimators in identification of the model and the small-lag asymptotic relationship of the cross-covariances to assess the goodness of fit to the data.
Interestingly, given the consistency of our model with the data, we can conclude that cross-covariances of volatility time series have a power law-like decay in the limit as the lag $k \to \infty$. In fact, it can be shown that as $k \to \infty$

$$\mathrm{Cov}(\sigma_t^1, \sigma_{t+k}^2) = \exp(\mu_1 + \mu_2) \exp\left(\frac{\mathrm{Var}(X_t^1) + \mathrm{Var}(X_{t+k}^2)}{2}\right) \left(\mathrm{Cov}\left(X_t^1, X_{t+k}^2\right) + o\left(\mathrm{Cov}\left(X_t^1, X_{t+k}^2\right)\right)\right), \quad (41)$$

meaning that the asymptotic behavior of the cross-covariance of the volatility is similar to the cross-covariance of its logarithm, whose behavior was already established in Proposition 1.

Finally, we conducted a forecasting exercise which further strengthened the belief of consistency of our model with realized volatiltiy time series. The forecasting exploited standard results for vectors of Gaussian random variables which allowed us to use the barely stationary or non-stationary specifications of our model. When confronted with a bivariate version of the famous HAR by Corsi (2009), the model behaved as well as its benchmark, thus confirming its ability in modeling log-volatilities.

Appendix A

In order to establish the asymptotic properties of the GMM estimator, we used the fact that the matrix of the moment conditions admits a bounded gradient. In our case, the former is strictly related to the bounded differentiability of the cross-covariance function in equation (1) with respect to each of the parameters.

Moreover, explicit formulas for the derivatives wirh respect to the parameters are of interest because they allow for faster convergence to the optimal solutions in many of the optimization routines that are variations of the gradient method, since they allow to avoid numerical differentiations which are well known to slow

down the convergence of the algorithms.

In this appendix we explicitly calculate such derivatives and show that are continuous. Continuity in the parameters ensures us boundedness in a neighbourhood of the solution, which is the space of interest of the optimization problem.

Differentiation of the integral function

Let us begin by stating the following result from Dugo, Giorgio, and Pigato (2024) for the integral appearing in Equation (2).

$$\begin{aligned} I(\alpha_i, \alpha_j, H) =& \int_0^s e^{\alpha_j v} \left(\int_{-\infty}^0 e^{\alpha_i u} (v-u)^{H-2} du \right) dv \\ =& \frac{1}{\alpha_i + \alpha_j} \int_0^s y^{H-2} \left(e^{\alpha_j y} - e^{-\alpha_i y} \right) dy \\ &+ \frac{e^{(\alpha_i+\alpha_j)s} - 1}{\alpha_i + \alpha_j} \int_s^\infty y^{H-2} e^{-\alpha_i y} dy. \end{aligned}$$

From the previous formulation and the Beppo Levi's (or Monotone Convergence) Theorem, we get the following derivatives.

$$\begin{aligned} \frac{\partial I}{\partial \alpha_i} =& -\frac{1}{(\alpha_i + \alpha_j)^2} \int_0^s y^{H-2} \left(e^{\alpha_j y} - e^{-\alpha_i y} \right) dy \\ &+ \frac{1}{\alpha_i + \alpha_j} \int_0^s y^{H-1} e^{-\alpha_i y} dy \\ &+ \frac{s e^{(\alpha_i+\alpha_j)s}}{\alpha_i + \alpha_j} \int_s^\infty y^{H-2} e^{-\alpha_i y} dy \\ &- \frac{e^{(\alpha_i+\alpha_j)s} - 1}{(\alpha_i + \alpha_j)^2} \int_s^\infty y^{H-2} e^{-\alpha_i y} dy \\ &- \frac{e^{(\alpha_i+\alpha_j)s} - 1}{\alpha_i + \alpha_j} \int_s^\infty y^{H-1} e^{-\alpha_i y} dy \\ \frac{\partial I}{\partial \alpha_j} =& -\frac{1}{(\alpha_i + \alpha_j)^2} \int_0^s y^{H-2} \left(e^{\alpha_j y} - e^{-\alpha_i y} \right) dy \\ &+ \frac{1}{\alpha_i + \alpha_j} \int_0^s y^{H-1} e^{\alpha_j y} dy \\ &+ \frac{s e^{(\alpha_i+\alpha_j)s}}{\alpha_i + \alpha_j} \int_s^\infty y^{H-2} e^{-\alpha_i y} dy \\ &- \frac{e^{(\alpha_i+\alpha_j)s} - 1}{(\alpha_i + \alpha_j)^2} \int_s^\infty y^{H-2} e^{-\alpha_i y} dy \end{aligned}$$

$$\begin{aligned} \frac{\partial I}{\partial H} =& \frac{1}{\alpha_i + \alpha_j} \int_0^s \log(y) y^{H-2} \left(e^{\alpha_j y} - e^{-\alpha_i y} \right) dy \\ &+ \frac{e^{(\alpha_i+\alpha_j)s} - 1}{\alpha_i + \alpha_j} \int_s^\infty \log(y) y^{H-2} e^{-\alpha_i y} dy. \end{aligned}$$

Differentiation of the covariance (cross-covariance at lag 0)

We are now going to calculate the gradient of

$$\gamma_0 = \frac{\Gamma(H+1)\nu_1\nu_2}{2(\alpha_1+\alpha_2)}\left(\left(\alpha_1^{1-H}+\alpha_2^{1-H}\right)\rho + \left(\alpha_2^{1-H}-\alpha_1^{1-H}\right)\eta\right)$$

as follows

$$\frac{\partial\gamma_0}{\partial\alpha_i} = \frac{\Gamma(H+1)\nu_1\nu_2}{2(\alpha_1+\alpha_2)}\left(-\frac{1}{\alpha_i+\alpha_j}\left(\left(\alpha_1^{1-H}+\alpha_2^{1-H}\right)\rho + \left(\alpha_2^{1-H}-\alpha_1^{1-H}\right)\eta\right) + (1-H)\alpha_i^{-H}(\rho-\eta)\right)$$

$$\frac{\partial\gamma_0}{\partial\alpha_j} = \frac{\Gamma(H+1)\nu_1\nu_2}{2(\alpha_1+\alpha_2)}\left(-\frac{1}{\alpha_i+\alpha_j}\left(\left(\alpha_1^{1-H}+\alpha_2^{1-H}\right)\rho + \left(\alpha_2^{1-H}-\alpha_1^{1-H}\right)\eta\right) + (1-H)\alpha_j^{-H}(\rho+\eta)\right)$$

$$\begin{aligned}\frac{\partial\gamma_0}{\partial H} = {} & \frac{\Gamma'(H+1)\nu_1\nu_2}{2\left(\alpha_i+\alpha_j\right)}\left(-\frac{1}{\alpha_i+\alpha_j}\left(\left(\alpha_1^{1-H}+\alpha_2^{1-H}\right)\rho + \left(\alpha_2^{1-H}-\alpha_1^{1-H}\right)\eta\right) + (1-H)\alpha_i^{-H}(\rho-\eta)\right) \\ & - \frac{\Gamma(H+1)\nu_1\nu_2}{2\left(\alpha_i+\alpha_j\right)}\left((\alpha_i^{1-H}\log\alpha_i + \alpha_j^{1-H}\log\alpha_j)\rho + (\alpha_j^{1-H}\log\alpha_j - \alpha_i^{1-H}\log\alpha_i)\eta\right)\end{aligned}$$

$$\frac{\partial\gamma_0}{\partial\rho} = \frac{\Gamma(H+1)\nu_1\nu_2}{2\left(\alpha_i+\alpha_j\right)}(\alpha_i^{1-H}+\alpha_2^{1-H})$$

$$\frac{\partial\gamma_0}{\partial\eta} = \frac{\Gamma(H+1)\nu_1\nu_2}{2\left(\alpha_i+\alpha_j\right)}(\alpha_j^{1-H}-\alpha_i^{1-H}).$$

Differentiation of the cross-covariance function

At this point, leveraging on previous derivatives and using the chain rule, it is easy to calculate the derivatives of the cross-covariance function $\gamma_{ij}(k)$ introduced in (1).

$$\frac{\partial\gamma_{ij}(k)}{\partial\alpha_j} = -se^{-\alpha_j s}\gamma_0 + e^{-\alpha_j s}\frac{\partial\gamma_0}{\partial\alpha_j} - s\nu_1\nu_2 e^{-\alpha_j s}H(H-1)\frac{\rho-\eta}{2}\frac{\partial I}{\partial\alpha_j}$$

$$\frac{\partial\gamma_{ij}(k)}{\partial\alpha_i} = e^{-\alpha_j s}\frac{\partial\gamma_0}{\partial\alpha_i} + \nu_1\nu_2 e^{-\alpha_j s}H(H-1)\frac{\rho-\eta}{2}\frac{\partial I}{\partial\alpha_i}$$

$$\frac{\partial\gamma_{ij}(k)}{\partial H} = e^{-\alpha_j s}\frac{\partial\gamma_0}{\partial H} + \nu_1\nu_2 e^{-\alpha_j s}H\frac{\rho-\eta}{2}I(\alpha_1,\alpha_2,H) + \nu_1\nu_2 e^{-\alpha_j s}H(H-1)\frac{\rho-\eta}{2}\frac{\partial I}{\partial H}$$

$$\frac{\partial\gamma_{ij}(k)}{\partial\nu_i} = e^{-\alpha_j s}\frac{\partial\gamma_0}{\partial\nu_i} + v_j e^{-\alpha_j s}H(H-1)\frac{\rho-\eta}{2}I(\alpha_1,\alpha_2,H)$$

$$\frac{\partial\gamma_{ij}(k)}{\partial\rho} = e^{\alpha_j s}\frac{\partial\gamma_0}{\partial\rho} + \frac{\nu_1\nu_2}{2}e^{-\alpha_j s}H(H-1)I(\alpha_1,\alpha_2,H)$$

$$\frac{\partial\gamma_{ij}(k)}{\partial\eta} = e^{\alpha_j s}\frac{\partial\gamma_0}{\partial\eta} - \frac{\nu_1\nu_2}{2}e^{-\alpha_j s}H(H-1)I(\alpha_1,\alpha_2,H).$$

Appendix B

When $H_i + H_j = 1$, the cross-covariance of the multivariate fractional Brownian motion takes the form

$$E\left(B_s^{H_i} B_t^{H_j}\right) = \frac{\sigma_i \sigma_j}{2}\left\{\rho(|s| + |t| - |s-t|) + \eta(t\log|t| - s\log|s| - (t-s)\log|t-s|\right\}$$

where $\rho_{ij} \in [-1, 1],\ \rho_{ij} = \rho_{ji} = \mathrm{Cor}(B_1^{H_i}, B_1^{H_j})$, $\eta_{ij} \in \mathbb{R},$ and $\eta_{ij} = -\eta_{ji}$.

References

Amblard, Pierre-Olivier, Jean-François Coeurjolly, Frédéric Lavancier, and Anne Philippe. 2010. "Basic Properties of the Multivariate Fractional Brownian Motion." *arXiv Preprint arXiv:1007.0828*.

Andersen, Torben G, Tim Bollerslev, Francis X Diebold, and Heiko Ebens. 2001. "The Distribution of Realized Stock Return Volatility." *Journal of Financial Economics* 61 (1): 43–76.

Arcones, Miguel A. 1994. "Limit Theorems for Nonlinear Functionals of a Stationary Gaussian Sequence of Vectors." *The Annals of Probability*, 2242–74.

Bayer, Christian, Peter K Friz, Sebastian Riedel, and John Schoenmakers. 2016. "From Rough Path Estimates to Multilevel Monte Carlo." *SIAM Journal on Numerical Analysis* 54 (3): 1449–83.

Bennedsen, Mikkel, Asger Lunde, Pakkanen, and Mikko S. 2022. "Decoupling the Short-and Long-Term Behavior of Stochastic Volatility." *Journal of Financial Econometrics* 20 (5): 961–1006.

Beran, Jan, Yuanhua Feng, Sucharita Ghosh, and Rafal Kulik. 2013. "Long-Memory Processes." *Long-Mem. Process.*

Bolko, Anine E, Kim Christensen, Mikko S Pakkanen, and Bezirgen Veliyev. 2022. "A GMM Approach to Estimate the Roughness of Stochastic Volatility." *Journal of Econometrics*.

Bubak, Vit, Evzen Kocenda, and Filip Zikes. 2011. "Volatility Transmission in Emerging European Foreign Exchange Markets." *Journal of Banking & Finance* 35 (11): 2829–41.

Cheridito, Patrick, Hideyuki Kawaguchi, and Makoto Maejima. 2003. "Fractional Ornstein-Uhlenbeck Processes."

Comte, Fabienne, and Eric Renault. 1998. "Long Memory in Continuous-Time Stochastic Volatility Models." *Mathematical Finance* 8 (4): 291–323.

Corsi, Fulvio. 2009. "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics* 7 (2): 174–96.

Diebold, Francis X, and Robert S Mariano. 2002. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 20 (1): 134–44.

Ding, Zhuanxin, and Clive WJ Granger. 1996. "Modeling Volatility Persistence of Speculative Returns: A New Approach." *Journal of Econometrics* 73 (1): 185–215.

Dugo, Ranieri, Giacomo Giorgio, and Paolo Pigato. 2024. "The Bivariate Fractional Ornstein-Uhlenbeck Process." *University of Rome Tor Vergata, Manuscript*.

Fouque, Jean-Pierre, George Papanicolaou, and K Ronnie Sircar. 2000. "Mean-Reverting Stochastic Volatility." *International Journal of Theoretical and Applied Finance* 3 (01): 101–42.

Galassi, Mark, Jim Davies, James Theiler, Brian Gough, Gerard Jungman, Patrick Alken, Michael Booth, Fabrice Rossi, and Rhys Ulerich. 2002. *GNU Scientific Library*. Network Theory Limited Godalming.

Gatheral, Jim, Thibault Jaisson, and Mathieu Rosenbaum. 2018. "Volatility Is Rough." *Quantitative Finance* 18 (6): 933–49.

Hall, Peter, Bing-Yi Jing, and Soumendra Nath Lahiri. 1998. "On the Sampling Window Method for Long-Range Dependent Data." *Statistica Sinica*, 1189–1204.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica: Journal of the Econometric Society*, 1029–54.

Hayashi, Fumio. 2011. *Econometrics*. Princeton University Press.

Hu, Yaozhong, Yanghui Liu, and David Nualart. 2016. "Rate of Convergence and Asymptotic Error Distribution of Euler Approximation Schemes for Fractional Diffusions."

Kloeden, Peter E, Eckhard Platen, Peter E Kloeden, and Eckhard Platen. 1992. *Stochastic Differential Equations*. Springer.

Lahiri, Soumendra Nath. 1993. "On the Moving Block Bootstrap Under Long Range Dependence." *Statistics & Probability Letters* 18 (5): 405–13.

Lang, Gabriel, and François Roueff. 2001. "Semi-Parametric Estimation of the hölder Exponent of a Stationary Gaussian Process with Minimax Rates." *Statistical Inference for Stochastic Processes* 4: 283–306.

Liu, Yanghui, and Samy Tindel. 2019. "FIRST-ORDER EULER SCHEME FOR SDES DRIVEN BY FRACTIONAL BROWNIAN MOTIONS." *The Annals of Applied Probability* 29 (2): 758–826.

Mandelbrot, Benoit B, and John W Van Ness. 1968. "Fractional Brownian Motions, Fractional Noises and Applications." *SIAM Review* 10 (4): 422–37.

Milstein, GN. 1974. "Approximate Integration of Stochastic Differential Equations, Theor." *Probab. Appl* 19: 557–62.

Mishura, Yuliya. 2008. *Stochastic Calculus for Fractional Brownian Motion and Related Processes*. Vol. 1929. Springer Science & Business Media.

Newey, Whitney K, and Kenneth D West. 1987. "Hypothesis Testing with Efficient Method of Moments Estimation." *International Economic Review*, 777–87.

Phillips, Peter CB, and Jun Yu. 2009. "A Two-Stage Realized Volatility Approach to Estimation of Diffusion Processes with Discrete Data." *Journal of Econometrics* 150 (2): 139–50.

Wang, Xiaohu, Weilin Xiao, and Jun Yu. 2021. "Modeling and Forecasting Realized Volatility with the Fractional Ornstein–Uhlenbeck Process." *Journal of Econometrics*.

Wang, Xiaohu, Jun Yu, and Chen Zhang. 2024. "On the Optimal Forecast with the Fractional Brownian Motion." *Quantitative Finance*, 1–10.