



# On the limiting spectral distribution of the covariance matrices of time-lagged processes

Christian Y. Robert<sup>a</sup>, Mathieu Rosenbaum<sup>b,\*</sup>

<sup>a</sup> CREST and ENSAE Paris Tech, Timbre J120, 3 Avenue Pierre Larousse, 92245 Malakoff Cedex, France

<sup>b</sup> CMAP-École Polytechnique Paris, UMR CNRS 7641, 91128 Palaiseau Cedex, France

## ARTICLE INFO

### Article history:

Received 5 January 2010

Available online 25 June 2010

### AMS subject classifications:

15A52

60F05

62F12

62H12

62M05

### Keywords:

Eigenvalues of covariance matrices

Lagged processes

Random matrix theory

Time lag estimation

Adaptive estimation

## ABSTRACT

We consider two continuous-time Gaussian processes, one being partially correlated to a time-lagged version of the other. We first give the limiting spectral distribution for the covariance matrices of the increments of the processes when the span between two observations tends to zero. Then, we derive the limiting distribution of the eigenvalues of the sample covariance matrices. This result is obtained when the number of paths of the processes is asymptotically proportional to the number of observations for each single path. As an application, we use the second moment of this distribution together with auxiliary volatility and correlation estimates to construct an adaptive estimator of the time lag between the two processes. Finally, we provide an asymptotic theory for our estimation procedure.

© 2010 Elsevier Inc. All rights reserved.

## 1. Introduction

Due to the rapid development of data acquisition and storage, statisticians encounter datasets for which the sample size and the number of variables are large. Understanding the sample covariance matrices issued from these datasets is of great importance in various areas, including for example statistical analysis (principal component analysis, classification...) and finance (risk management, portfolio allocation...). However, the estimation strategies that are used when the number of variables is fixed, and the sample size goes to infinity, fail when one considers that both the number of variables and the sample size are large. For samples of independent Gaussian random variables, a remarkable result by Marcenko and Pastur [7] describes the limiting behaviour of the distribution of the eigenvalues of the sample covariance matrices when the ratio of the number of variables and the sample size goes to a finite limit. This result was refined in the case of non-diagonal covariance matrices by Silverstein [9].

In this paper, we consider two continuous-time processes. The first process is a continuous Gaussian process with independent increments. The second process is a linear combination of the first one delayed by a constant  $\theta$  and of another independent Gaussian process. This situation typically appears in financial markets when two assets share fundamental economic characteristics. Their prices rarely react simultaneously to information: one price reacts before the other, resulting in a so-called “lead-lag” relationship. This has been well documented for various financial instruments, in particular indices and associated futures. For example, the case of S&P 500 Index and S&P 500 Futures is studied in [4,6]. Various possible explanations for this delay have been proposed including non-synchronous trading, transaction costs, asymmetric

\* Corresponding author.

E-mail addresses: [chrobert@ensae.fr](mailto:chrobert@ensae.fr) (C.Y. Robert), [mathieu.rosenbaum@polytechnique.edu](mailto:mathieu.rosenbaum@polytechnique.edu) (M. Rosenbaum).

information and market architectures. In practice, a lead-lag relationship can be viewed as a market inefficiency. Therefore, the knowledge of the value of this time delay may imply arbitrage profits.

We first prove the existence of an asymptotic distribution for the eigenvalues of the covariance matrices built from the increments of the two processes. Moreover, we characterise this limiting law. This is done when the span between two observations tends to zero. Then, based on the observations of sample paths of the processes, we consider the asymptotic distribution of the eigenvalues of the sample covariance matrices. This distribution is showed to satisfy a Marcenko–Pastur type equation. As an application, we use the second moment of this distribution together with auxiliary volatility and correlation estimates to construct an adaptive estimator of the time lag between the two processes. Finally, we provide an asymptotic theory for our estimation procedure.

The paper is organised as follows. We prove in Section 2 the existence of a limiting spectral distribution for the covariance matrices when the increments of the second process are shifted in a convenient way. We also derive the limiting distribution of the eigenvalues of the sample covariance matrices when the number of paths of the processes is asymptotically proportional to the number of observations for each single path. Section 3 is devoted to the construction and properties of the estimator of the time lag between the two processes. The proofs are gathered in a last section.

## 2. Limiting spectral distributions

### 2.1. Assumptions

On a filtered probability space  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ , we consider two processes  $(X_t)_{t \in [0,1]}$  and  $(Y_t)_{t \in [0,1]}$  defined by

$$X_t - X_0 = \int_0^t K_{s+\theta} dW_{s+\theta}, \quad (2.1)$$

$$Y_t - Y_0 = \rho \int_0^{t \wedge \bar{\theta}} K_s d\tilde{W}_s + \rho \int_{\bar{\theta}}^{t \vee \bar{\theta}} K_s dW_s + \int_0^t L_s dW'_s, \quad (2.2)$$

where  $(W_t, \tilde{W}_t, W'_t)_{t \in [0,1]}$  is a three-dimensional Brownian motion,  $\rho > 0$ ,  $0 < \theta < \bar{\theta} < 1$ , with  $\bar{\theta}$  a known constant and  $K_s$  and  $L_s$  are deterministic, twice continuously differentiable, positive functions. For a financial interpretation,  $[0, 1]$  can be seen as one trading day,  $X_t$  as the log-price of the “leader” asset and  $Y_t$  as the log-price of the “lagger” asset. The parameter  $\theta$  is referred to as the lead-lag parameter and measures the time lag between the processes. Note that  $[\theta, 1]$  is the time interval when the lead-lag relationship is in force. Thus, we have for  $s \in [\theta, 1]$ ,

$$dY_s = \rho dX_{s-\theta} + L_s dW'_s.$$

We assume that we observe  $m + 1$  equidistant values for each process, that is  $(X_{i/m}, Y_{i/m})$ , for  $i = 0, \dots, m$ . We write  $m = p \lfloor p^a \rfloor$ , where  $p$  is a positive integer and  $a > 0$ . The need for this parametrization will be clear after Section 2.3. Roughly speaking, in a financial context, the order of magnitude of the number of observed days will be  $p$  and we require the order of magnitude of the number of daily data being slightly bigger than  $p$ . In particular,  $p$  will be the parameter driving the asymptotics.

### 2.2. The covariance matrices

We denote the increments of a process  $(V_t)_{t \in [0,1]}$  on the  $\lfloor p^a \rfloor$  grids with mesh  $1/p$ , for  $i = 1, \dots, p-1$ ,  $l = 0, \dots, \lfloor p^a \rfloor - 1$  by

$$\Delta^{(l,p)} V_i = V_{i/p+l/m} - V_{(i-1)/p+l/m},$$

and for  $i = p$ ,  $l = 0, \dots, \lfloor p^a \rfloor - 1$  by

$$\Delta^{(l,p)} V_p = (V_1 - V_{1-1/p+l/m}) + (V_{l/m} - V_0).$$

The random variables  $\Delta^{(0,p)} X_i$  and  $\Delta^{(l,p)} Y_i$  are centered Gaussian with respective variance  $v_{i,0}^X$  and  $v_{i,l}^Y$  satisfying for  $i = 1, \dots, p-1$

$$v_{i,0}^X = \int_{(i-1)/p}^{i/p} K_{s+\theta}^2 ds, \quad v_{i,l}^Y = \int_{(i-1)/p+l/m}^{i/p+l/m} (\rho^2 K_s^2 + L_s^2) ds,$$

and for  $i = p$

$$v_{p,0}^X = \int_{1-1/p}^1 K_{s+\theta}^2 ds, \quad v_{p,l}^Y = \int_{1-1/p+l/m}^1 (\rho^2 K_s^2 + L_s^2) ds + \int_0^{l/m} (\rho^2 K_s^2 + L_s^2) ds.$$

Now, we set

$$Z^{(l,p)} = p^{1/2} (\Delta^{(0,p)} X_1, \dots, \Delta^{(0,p)} X_p, \Delta^{(l,p)} Y_1, \dots, \Delta^{(l,p)} Y_p)^\top$$

where  $^\top$  denotes the transpose operator.

Let  $\lfloor \theta \rfloor_p = \lfloor p\theta \rfloor / p$  be the largest number of the form  $k/p$ ,  $k \in \mathbb{N}$ , smaller than  $\theta$ . The vector  $Z^{(l,p)}$  is a Gaussian vector of size  $2p$  with 5-diagonal covariance matrix  $\Sigma_{(l,p)}$  satisfying

- for  $l = 0, \dots, \lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor$

$$\begin{cases} 1 \leq i \leq p, 1 \leq j \leq p, i = j & (\Sigma_{(l,p)})_{i,j} = pv_{i,0}^X \\ p+1 \leq i \leq 2p, p+1 \leq j \leq 2p, i = j & (\Sigma_{(l,p)})_{i,j} = pv_{j-p,l}^Y \\ 1 \leq i \leq p, p+1 \leq j \leq 2p, j-p = i + p\lfloor \theta \rfloor_p & (\Sigma_{(l,p)})_{i,j} = pv_{i,l,1}^{XY} \\ 1 \leq i \leq p, p+1 \leq j \leq 2p, j-p = i + p\lfloor \theta \rfloor_p + 1 & (\Sigma_{(l,p)})_{i,j} = pv_{i,l,2}^{XY} \\ p+1 \leq i \leq 2p, 1 \leq j \leq p, i-p = j + p\lfloor \theta \rfloor_p & (\Sigma_{(l,p)})_{i,j} = pv_{j,l,1}^{XY} \\ p+1 \leq i \leq 2p, 1 \leq j \leq p, i-p = j + p\lfloor \theta \rfloor_p + 1 & (\Sigma_{(l,p)})_{i,j} = pv_{j,l,2}^{XY} \end{cases}$$

with for  $i = 1, \dots, p(1 - \lfloor \theta \rfloor_p) - 1$

$$v_{i,l,1}^{XY} = \rho \int_{(i-1)/p}^{i/p - (\theta - \lfloor \theta \rfloor_p) + l/m} K_{s+\theta}^2 ds \quad \text{and} \quad v_{i,l,2}^{XY} = \rho \int_{i/p - (\theta - \lfloor \theta \rfloor_p) + l/m}^{i/p} K_{s+\theta}^2 ds,$$

and for  $i = p(1 - \lfloor \theta \rfloor_p)$

$$v_{p(1-\lfloor \theta \rfloor_p),l,1}^{XY} = \rho \int_{1-\lfloor \theta \rfloor_p - 1/p}^{1-\theta} K_{s+\theta}^2 ds,$$

all the other terms are equal to zero,

- for  $l = \lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor + 1, \dots, \lfloor p^a \rfloor - 1$

$$\begin{cases} 1 \leq i \leq p, 1 \leq j \leq p, i = j & (\Sigma_{(l,p)})_{i,j} = pv_{i,0}^X \\ p+1 \leq i \leq 2p, p+1 \leq j \leq 2p, i = j & (\Sigma_{(l,p)})_{i,j} = pv_{j-p,l}^Y \\ 1 \leq i \leq p, p+1 \leq j \leq 2p, j-p = i + \lfloor p\theta \rfloor - 1 & (\Sigma_{(l,p)})_{i,j} = pv_{i,l,1}^{XY} \\ 1 \leq i \leq p, p+1 \leq j \leq 2p, j-p = i + \lfloor p\theta \rfloor & (\Sigma_{(l,p)})_{i,j} = pv_{i,l,2}^{XY} \\ p+1 \leq i \leq 2p, 1 \leq j \leq p, i-p = j + \lfloor p\theta \rfloor - 1 & (\Sigma_{(l,p)})_{i,j} = pv_{j,l,1}^{XY} \\ p+1 \leq i \leq 2p, 1 \leq j \leq p, i-p = j + \lfloor p\theta \rfloor & (\Sigma_{(l,p)})_{i,j} = pv_{j,l,2}^{XY} \end{cases}$$

with for  $i = 1, \dots, p(1 - \lfloor \theta \rfloor_p) - 1$

$$v_{i,l,1}^{XY} = \rho \int_{(i-1)/p}^{(i-1)/p - (\theta - \lfloor \theta \rfloor_p) + l/m} K_{s+\theta}^2 ds \quad \text{and} \quad v_{i,l,2}^{XY} = \rho \int_{(i-1)/p - (\theta - \lfloor \theta \rfloor_p) + l/m}^{i/p} K_{s+\theta}^2 ds,$$

and for  $i = p(1 - \lfloor \theta \rfloor_p)$

$$v_{p(1-\lfloor \theta \rfloor_p),l,1}^{XY} = \rho \int_{1-\lfloor \theta \rfloor_p - 1/p}^{1-\theta - 1/p + l/m} K_{s+\theta}^2 ds \quad \text{and} \quad v_{p(1-\lfloor \theta \rfloor_p),l,2}^{XY} = \rho \int_{1-\theta - 1/p + l/m}^{1-\theta} K_{s+\theta}^2 ds,$$

and for  $i = p(1 - \lfloor \theta \rfloor_p) + 1$

$$v_{p(1-\lfloor \theta \rfloor_p)+1,l,1}^{XY} = 0,$$

all the other terms are equal to zero.

We call  $H_{l,p}$  the eigenvalue distribution of the covariance matrix  $\Sigma_{(l,p)}$ , that is the distribution that puts mass  $1/2p$  at each of the eigenvalues of this matrix. We are interested in the limiting distribution of  $H_{l,p}$  as  $p$  tends to infinity. The existence of a limiting distribution depends a priori on the choice of the sequence  $l = l_p$ . If

$$l_p = l_p^* = m(\theta - \lfloor \theta \rfloor_p),$$

then the covariance matrix becomes 3-diagonal. In this case, the limiting distribution can be quite easily obtained. However,  $l_p^*$  is only a target value since it is not an integer in general and it depends on the unknown parameter  $\theta$ . In our asymptotics, we will consider given sequences  $l_p^0$ , independent from everything else, so that we are close to the target value  $l_p^*$ . More precisely,  $(l_p^0)_{p \geq 1}$  will satisfy

$$\varepsilon_p := \frac{|l_p^0 - l_p^*|}{m} = o(1/p), \quad \text{as } p \rightarrow \infty. \quad (2.3)$$

In fact, we will also prove that such a sequence  $(l_p^0)_{p \geq 1}$  can be obtained from data, in an adaptive way, see Section 3.3.

Let us denote the  $k$ -th moment of  $H_{p,p}^0$  by  $\mu_{p,p}^{(k)}$

$$\mu_{p,p}^{(k)} := \frac{1}{2p} \text{Tr}(\Sigma_{(I_p^0, p)}^k) = \int_0^\infty \lambda^k dH_{p,p}^0(\lambda).$$

We have the following proposition.

**Proposition 2.1.** For any sequence  $(I_p^0)_{p \geq 1}$  satisfying (2.3), the distribution  $H_{p,p}^0$  converges weakly as  $p$  tends to infinity to a unique distribution  $H_\infty$ . The distribution  $H_\infty$  is characterised by the sequence of its moments  $(\mu^{(k)})_{k \geq 1}$ . Moreover, for any positive integer  $k$ , we have

$$\mu_{p,p}^{(k)} \rightarrow \mu^{(k)} := \int_0^\infty \lambda^k dH_\infty(\lambda) = \frac{1}{2} \int_0^1 P^{(k)}(t) dt + \frac{1}{2} \int_0^1 Q^{(k)}(t) dt, \quad \text{as } p \rightarrow \infty,$$

where  $P^{(k)}$  and  $Q^{(k)}$  are defined by the following recursion:

$$\begin{aligned} P^{(1)}(t) &= K_{t+\theta}^2, \\ Q^{(1)}(t) &= (\rho^2 K_{t+\theta}^2 + L_{t+\theta}^2) 1_{\{0 \leq t \leq 1-\theta\}} + (\rho^2 K_{t-(1-\theta)}^2 + L_{t-(1-\theta)}^2) 1_{\{1-\theta \leq t \leq 1\}}, \\ R^{(1)}(t) &= \rho K_{t+\theta}^2 1_{\{t \leq 1-\theta\}}, \end{aligned}$$

and for  $k \geq 2$ ,

$$\begin{aligned} P^{(k)}(t) &= P^{(1)}(t)P^{(k-1)}(t) + R^{(1)}(t)R^{(k-1)}(t) 1_{\{t \leq 1-\theta\}}, \\ Q^{(k)}(t) &= Q^{(1)}(t)Q^{(k-1)}(t) + R^{(1)}(t)R^{(k-1)}(t) 1_{\{t \leq 1-\theta\}}, \\ R^{(k)}(t) &= R^{(1)}(t)P^{(k-1)}(t) + Q^{(1)}(t)R^{(k-1)}(t) 1_{\{t \leq 1-\theta\}}. \end{aligned}$$

### 2.3. The sample covariance matrices

We now assume that we have a  $n$ -sample of the random vector  $Z^{(l,p)} : (Z_d^{(l,p)})_{d=1,\dots,n}$ . For financial interpretation, it means we have  $n$  days of observations of the two considered assets. Let  $Z^{(l)}$  be the matrix in  $\mathbb{R}^{2p \times n}$  defined by

$$Z^{(l)} = (Z_1^{(l,p)}, \dots, Z_n^{(l,p)}).$$

We consider the empirical counterpart of  $\Sigma_{(l,p)}$  given by

$$S_{(l,p)} = \frac{Z^{(l)}(Z^{(l)})^\top}{n}.$$

We denote by  $F_{l,p}$  the (random) eigenvalue distribution of  $S_{(l,p)}$ . Recall that a distribution  $G$  is characterised by its Stieltjes transform  $m_G$  defined by

$$m_G(z) = \int \frac{dG}{x - z}, \quad \text{for } z \in \mathbb{C}^+,$$

and set for  $\gamma > 0$

$$v_{\gamma,G}(z) = -(1-\gamma) \frac{1}{z} + \gamma m_G(z).$$

Let  $\gamma_p = 2p/n$ . The following proposition is derived from the Marcenko–Pastur result, see [9,11].

**Proposition 2.2.** Assume that  $\gamma_p \rightarrow \gamma > 0$  as  $p \rightarrow \infty$ . For any sequence  $(I_p^0)_{p \geq 1}$  satisfying (2.3), almost surely,  $F_{p,p}^0$  converges weakly to a unique (non-random) distribution  $F_\infty$ . This distribution is characterised by the Marcenko–Pastur type equation

$$-\frac{1}{v_{\gamma,F_\infty}(z)} = z - \gamma \int_0^{+\infty} \frac{\lambda dH_\infty(\lambda)}{1 + \lambda v_{\gamma,F_\infty}(z)}. \quad (2.4)$$

Furthermore, for any positive integer  $k$ , we have

$$\frac{1}{2p} \text{Tr}(\Sigma_{(I_p^0, p)}^k) = \int_0^{+\infty} \lambda^k dF_{p,p}^0(\lambda) \xrightarrow[p \rightarrow +\infty]{L^2} m^{(k)} = \int_0^{+\infty} \lambda^k dF_\infty(\lambda).$$

Note that Eq. (1.2) in [8] (see also [11]) gives a relation between the moments  $m^{(k)}$  and  $\mu^{(k)}$

$$m^{(k)} = \sum_{w=1}^k \gamma^{k-w} \sum_{\substack{n_1 + \dots + n_w = k-w+1 \\ n_1 + 2n_2 + \dots + wn_w = k}} \frac{k!}{n_1! \dots n_w!} (\mu^{(1)})^{n_1} \dots (\mu^{(w)})^{n_w}.$$

### 3. Application: estimation of the lead-lag parameter

In this section we propose an estimator of the lead-lag parameter  $\theta$ . This estimator can be thought of as a simple moment-based estimator in a random matrix context. Note that other methods have already been introduced for estimating a lead-lag parameter. For example, a three stage least square estimation strategy is used in [6], a cross correlation based technique is developed in [4] and a method based on the Hayashi–Yoshida estimator is proposed in [5]. Here our aim is not to prove any kind of optimality of our random matrix based method. We just show that it is a quite easy and elegant way to derive a limit theory for an estimator, which seems to be intricate for other procedures, see [5].

#### 3.1. Construction of the estimator

Let  $F_{(\gamma_p, H_{p,p}^0)}$  be the distribution function with Stieltjes transform satisfying Marcenko–Pastur Equation (2.4) associated to the value  $\gamma_p$  and to the distribution  $H_{p,p}^0$ . We denote by  $m_{p,p}^{(k)}$  its moment of order  $k$ . In particular, we have

$$m_{p,p}^{(2)} = \gamma_p \left( \mu_{p,p}^{(1)} \right)^2 + \mu_{p,p}^{(2)}.$$

This moment relation will be the cornerstone of the building of our estimator. Indeed, we have the following proposition.

**Proposition 3.1.** For any sequence  $(l_p^0)_{p \geq 1}$  satisfying (2.3), as  $p \rightarrow +\infty$ ,

$$\mu_{p,p}^{(1)} = \mu^{(1)} = \frac{1}{2} \int_0^1 (K_{s+\theta}^2 + (\rho^2 K_s^2 + L_s^2)) ds$$

and

$$\begin{aligned} \mu_{p,p}^{(2)} &= \frac{1}{2} \int_0^1 K_{s+\theta}^4 ds + \frac{1}{2} \int_0^1 (\rho^2 K_s^2 + L_s^2)^2 ds + \int_0^{1-\theta} \rho^2 K_{s+\theta}^4 ds - 2\rho^2 p\varepsilon_p(1-p\varepsilon_p) \int_0^{1-\theta} K_{s+\theta}^4 ds \\ &\quad - \frac{1}{p} (1-p(\theta - \lfloor \theta \rfloor_p)) p(\theta - \lfloor \theta \rfloor_p) \left[ \rho^2 K_1^2 + \frac{1}{2} ((\rho^2 K_0^2 + L_0^2) - (\rho^2 K_1^2 + L_1^2))^2 \right] + o(1/p). \end{aligned}$$

Remark that  $\mu_{p,p}^{(1)}$  does not depend on  $p$  and that its expression is obvious since it is equal to the trace of  $(1/2p) \Sigma_{(l_p^0, p)}$ .

Now define  $\theta_p$  such that

$$\begin{aligned} \int_0^{1-\theta_p} \rho^2 K_{s+\theta}^4 ds &= \int_0^{1-\theta} \rho^2 K_{s+\theta}^4 ds - 2\rho^2 p\varepsilon_p(1-p\varepsilon_p) \int_0^{1-\theta} K_{s+\theta}^4 ds \\ &\quad - \frac{1}{p} (1-p(\theta - \lfloor \theta \rfloor_p)) p(\theta - \lfloor \theta \rfloor_p) \left[ \rho^2 K_1^2 + \frac{1}{2} ((\rho^2 K_0^2 + L_0^2) - (\rho^2 K_1^2 + L_1^2))^2 \right]. \end{aligned}$$

We have that  $\theta_p$  tends to  $\theta$  under condition (2.3). Thus, the idea is to estimate  $m_{p,p}^{(2)}$  in the following way

$$\hat{m}_{p,p}^{(2)} = \frac{1}{2p} \text{Tr} \left( S_{(l_p^0, p)}^2 \right)$$

and to use estimators  $\hat{\mu}_p^{(1)}$  and  $\hat{\mu}_p^{(2)}(t)$  (see below) of  $\mu^{(1)}$  and of the function

$$\mu^{(2)}(t) = \frac{1}{2} \int_0^1 (K_{s+\theta}^4 + (\rho^2 K_s^2 + L_s^2)^2) ds + \rho^2 \int_0^{(1-t)} K_{s+\theta}^4 ds.$$

Hence,  $t$  will be close to  $\theta_p$  and  $\theta$  when

$$\hat{A}_p(t) = \hat{m}_{p,p}^{(2)} - \gamma_p (\hat{\mu}_p^{(1)})^2 - \hat{\mu}_p^{(2)}(t)$$

is close to zero. Eventually, note that it is more convenient to use finite distance moments than asymptotic moments. Indeed, it enables us to get more accurate associated empirical quantities, see [1] for details.

We now explain the way  $\hat{\mu}_p^{(1)}$  and  $\hat{\mu}_p^{(2)}(t)$  are built. For  $0 \leq t \leq 1$ , let

$$V_t^X = \int_0^t K_{s+\theta}^2 ds, \quad V_t^Y = \int_0^t (\rho^2 K_s^2 + L_s^2) ds, \quad Q_t^X = \int_0^t K_{s+\theta}^4 ds, \quad Q_t^Y = \int_0^t (\rho^2 K_s^2 + L_s^2)^2 ds.$$

The estimators  $\hat{\mu}_p^{(1)}$  and  $\hat{\mu}_p^{(2)}(t)$  are based on estimates of  $\rho$  and of the preceding quantities. Estimators  $\hat{V}_t^X$ ,  $\hat{V}_t^Y$ ,  $\hat{Q}_t^X$  and  $\hat{Q}_t^Y$  can be constructed using the grid with mesh  $1/m$ :

$$\hat{V}_t^X = \frac{1}{n} \sum_{d=1}^n \sum_{1 \leq j \leq \lfloor mt \rfloor} (\Delta^{(0,m)} X_j^{(d)})^2, \quad \hat{V}_t^Y = \frac{1}{n} \sum_{d=1}^n \sum_{1 \leq j \leq \lfloor mt \rfloor} (\Delta^{(0,m)} Y_j^{(d)})^2,$$

and

$$\hat{Q}_t^X = \frac{m}{3n} \sum_{d=1}^n \sum_{1 \leq j \leq \lfloor mt \rfloor} (\Delta^{(0,m)} X_j^{(d)})^4, \quad \hat{Q}_t^Y = \frac{m}{3n} \sum_{d=1}^n \sum_{1 \leq j \leq \lfloor mt \rfloor} (\Delta^{(0,m)} Y_j^{(d)})^4,$$

where the notations  $X^{(d)}$  and  $Y^{(d)}$  corresponds to the  $d$ -th sample of  $X$  and  $Y$ .

In order to estimate the coefficient  $\rho$ , we compute for each day the covariations between the values of  $X$  and the lagged values of  $Y$  on the grid with mesh  $1/m$ . These covariations will always be close to zero except for the two lag values corresponding to the interval of the grid with mesh  $1/m$  where  $\theta$  lies (or only one if  $\theta$  is a value of the grid with mesh  $1/m$ ). Let for  $k \leq m\bar{\theta}$ ,

$$\Gamma_k^{(d)} = \sum_{1 \leq j \leq m(1-\bar{\theta})} \Delta^{(0,m)} X_j^{(d)} \Delta^{(0,m)} Y_{j+k}^{(d)}.$$

We define the estimator of the correlation coefficient  $\rho$  by

$$\hat{\rho} = \frac{1}{n\hat{V}_{1-\bar{\theta}}^X} \sum_{d=1}^n \sum_{0 \leq k \leq m\bar{\theta}} \Gamma_k^{(d)} 1_{\left\{|\Gamma_k^{(d)}| \geq m^{\frac{4-b}{4+2b}}\right\}},$$

for some  $b > 4 + 6/a$ .

We now set

$$\hat{\mu}_p^{(1)} = \frac{1}{2}(\hat{V}_1^X + \hat{V}_1^Y), \quad \hat{\mu}_p^{(2)}(t) = \frac{1}{2}(\hat{Q}_1^X + \hat{Q}_1^Y) + \hat{\rho}^2 \hat{Q}_{1-t}^X$$

and we finally define  $\hat{\theta}_p$  by

$$\hat{\theta}_p = \min\{\theta_k \in \{k/m : k = 1, \dots, m\}, \hat{A}_p(\theta_k) \geq 0\}.$$

### 3.2. Asymptotic theory

We first give a result for the estimators of  $\mu^{(1)}$  and  $\mu^{(2)}(t)$ .

**Proposition 3.2.** As  $p \rightarrow \infty$ ,

$$p \left( (\hat{\mu}_p^{(1)})^2 - (\mu^{(1)})^2 \right) \xrightarrow{\mathbb{P}} 0$$

and

$$p \left( \hat{\mu}_p^{(2)}(t) - \mu^{(2)}(t) \right) \xrightarrow{ucp} 0,$$

where *ucp* denotes uniform convergence in probability over compact sets included in  $[0, 1]$ .

The asymptotic theory will be given using  $\theta_p$  and not  $\theta$ . This is due to the fact that an asymptotic bias appears since we do not take into account some unknown quantities in the expression of  $\mu_{p,p}^{(2)}$ . We have the following theorem.

**Theorem 3.3.** Assume that  $p|\gamma - \gamma_p| \rightarrow 0$  as  $p \rightarrow +\infty$ . For any sequence  $(l_p^0)_{p \geq 1}$  satisfying (2.3), as  $p \rightarrow +\infty$ ,

$$p(\hat{\theta}_p - \theta_p)$$

weakly converges to a non-degenerate Gaussian random variable.

Note that the bias and variance of the limit Gaussian variable can be derived from Theorem 1.1 in [1], see Section 4.4.

### 3.3. Adaptive choice of the sequence $l_p^0$

We now explain how the sequence  $l_p$  can be constructed from the data. For technical reasons, we assume in this section that  $\theta$  is a rational number. In this case, the set

$$\{p(\theta - \lfloor \theta \rfloor_p), p \geq 1\}$$

contains zero and is finite (the situation becomes quite different in the irrational case, see [10]). This will be an important point in the proof of the next result.

We consider a  $n'$ -sample  $(Z_k^{(l,p)})_{k=1,\dots,n'}$  where  $n' = n + \lfloor n/\log n \rfloor$ . We split this sample into two pieces of size  $n_1 = \lfloor n/\log n \rfloor$  and  $n_2 = n$ . We construct a sequence  $(l_p^e)_{p \geq 1}$  on the first piece of the sample and use it as a sequence  $(l_p^o)_{p \geq 1}$  for the second piece of the sample. Let  $Z_{k,i}^{(l,p)}$  be the  $i$ -th element of  $Z_k^{(l,p)}$ . We define for  $1 \leq i \leq p$  and  $p+1 \leq j \leq 2p$

$$s_{i,j}^{(l,p)} = \frac{1}{n_1} \sum_{k=1}^{n_1} Z_{k,i}^{(l,p)} Z_{k,j}^{(l,p)}, \quad T_{i,j}^{(l,p)}(s) = \mathbb{I}_{\{|s_{i,j}^{(l,p)}| > s\}}.$$

The quantity  $s_{i,j}^{(l,p)}$  is the empirical covariance between the  $i$ -th increment of  $X$  and the  $(j - p)$ -th increment of  $Y$ . The idea is to choose  $l_p$  so that being close to a tri-diagonal case, which is a case with minimum number of non-zero coefficients in the covariance matrix. In this way,  $l_p$  should be close to  $l_p^*$ . However, this strategy works only when  $\theta$  is not close to a value of the grid with mesh  $1/p$ . Indeed, in the case where  $\theta$  is close to a value of the grid with mesh  $1/p$ ,  $l_p^*$  is either close to zero or to  $\lfloor p^a \rfloor$ . Thus, both a choice of  $l_p$  close to zero and a choice of  $l_p$  close to  $\lfloor p^a \rfloor$  lead to an almost tri-diagonal matrix. However, only one of the two choices is suitable. We discriminate between these two kinds of values for  $l_p$  using this remark: if  $l_p^*$  is close to zero and  $l_p$  is chosen close to  $\lfloor p^a \rfloor$ , the last element of the lower “non-empty” diagonal is equal to zero, whereas, if  $l_p^*$  is close to  $\lfloor p^a \rfloor$  and  $l_p$  is chosen close to  $\lfloor p^a \rfloor$ , the last element of the lower diagonal is positive. This leads to the following more sophisticated choice procedure.

First define

$$N^{(l,p)}(s) = \frac{1}{p} \sum_{1 \leq i \leq p, p+1 \leq j \leq 2p-1} T_{i,j}^{(l,p)}(s) \quad \text{and} \quad \underline{N}^{(p)}(s) = \min_{l=0, \dots, \lfloor p^a \rfloor - 1} N^{(l,p)}(s)$$

and set

$$\mathcal{I}^{(p)}(s) = \{l = 0, \dots, \lfloor p^a \rfloor - 1 : N^{(l,p)}(s) = \underline{N}^{(p)}(s) \text{ or } N^{(l,p)}(s) = \underline{N}^{(p)}(s) + 1\}.$$

Then introduce the following quantities

$$l_p = \min_{l \in \mathcal{I}^{(p)}(s)} l, \quad \bar{l}_p = \max_{l \in \mathcal{I}^{(p)}(s)} l \quad \text{and} \quad i_p = \max\{i : T_{i, 2p-1}^{(\bar{l}_p, p)}(s) = 1\}.$$

Finally define  $l_p^e$  the following way: if  $i_p$  is well defined and  $T_{i_p+1, 2p}^{(\bar{l}_p, p)}(s) = 0$ , then  $l_p^e = l_p$  else  $l_p^e = \bar{l}_p$ . We have the following proposition.

**Proposition 3.4.** Assume  $\theta$  is a rational number. Let  $s_p = p^{-d}$  where  $0 < d < \min(a, 1/2)$ . Let  $(l_p^e)_{p \geq 1}$  be a sequence such that  $l_p^e \in \mathcal{I}^{(p)}(s_p)$ . Then

$$\frac{|l_p^e - l_p^*|}{m} = o_{a.s.}(1/p).$$

Consequently, Theorem 3.3 holds with  $l_p^o = l_p^e$ .

#### 4. Proofs

In the following,  $c$  denotes a positive constant that may vary from line to line.

##### 4.1. Proof of Proposition 2.1

We split the proof into three steps.

(i) *First step: an auxiliary sequence of matrices.* In this step we “replace”  $\theta$  by a value  $\theta_m$  which belongs to the grid with mesh  $1/m$ . So we define the sequence  $(\theta_m)_{m \geq 1}$  by

$$m(\theta_m - \lfloor \theta \rfloor_p) = \lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor.$$

Moreover, we take  $l_p = \bar{l}_p^* = m(\theta_m - \lfloor \theta \rfloor_p)$  so that the covariance matrices  $(\Sigma_p^*(m))_{m \geq 1}$  become tri-diagonal. Thus, we have

$$\begin{cases} 1 \leq i \leq p, 1 \leq j \leq p, i = j & (\Sigma_p^*(m))_{i,j} = p v_i^X(m) \\ p+1 \leq i \leq 2p, p+1 \leq j \leq 2p, i = j & (\Sigma_p^*(m))_{i,j} = p v_{j-p}^Y(m) \\ 1 \leq i \leq p, p+1 \leq j \leq 2p, j-p = i + p \lfloor \theta \rfloor_p & (\Sigma_p^*(m))_{i,j} = p v_i^{XY}(m) \\ p+1 \leq i \leq 2p, 1 \leq j \leq p, i-p = j + p \lfloor \theta \rfloor_p & (\Sigma_p^*(m))_{i,j} = p v_j^{XY}(m) \end{cases}$$

where

$$v_i^X(m) = \int_{(i-1)/p}^{i/p} K_{s+\theta_m}^2 ds, \quad v_i^Y(m) = \int_{(i-1)/p + (\theta_m - \lfloor \theta \rfloor_p)}^{i/p + (\theta_m - \lfloor \theta \rfloor_p)} (\rho^2 K_s^2 + L_s^2) ds, \quad v_i^{XY}(m) = \rho \int_{(i-1)/p}^{i/p} K_{s+\theta_m}^2 ds.$$

All the other terms are equal to zero.

**Lemma 4.1.** Let  $D_p = \Sigma_p^*(m) - \Sigma_{(l_p^o, p)}$  and  $\lambda_i(D_p)$  be the  $i$ -th eigenvalue of  $D_p$ . We have

$$\sup_{i=1, \dots, 2p} |\lambda_i(D_p)| = o(1).$$

**Proof.** Assume without loss of generality that  $l_p^0 \leq \lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor$ . We have

- For  $(i, j)$  such that  $i = 1, \dots, p$  and  $i = j$

$$(D_p)_{i,j} = p \left( \int_{(i-1)/p}^{i/p} K_{s+\theta_m}^2 ds - \int_{(i-1)/p}^{i/p} K_{s+\theta}^2 ds \right)$$

and so

$$|(D_p)_{i,j}| \leq c/m.$$

- For  $(i, j)$  such that  $i = p+1, \dots, 2p-1$  and  $i = j$

$$(D_p)_{i,j} = p \left( \int_{(i-1)/p+\tilde{l}_p^*/m}^{i/p+\tilde{l}_p^0/m} (\rho^2 K_s^2 + L_s^2) ds - \int_{(i-1)/p+l_p^0/m}^{i/p+l_p^0/m} (\rho^2 K_s^2 + L_s^2) ds \right)$$

and so

$$|(D_p)_{i,j}| \leq c \frac{p|l_p^0 - \tilde{l}_p^*|}{m}.$$

- For  $(i, j)$  such that  $i = j = 2p$

$$(D_p)_{i,j} = p \left( \int_{1-1/p+\tilde{l}_p^*/m}^1 (\rho^2 K_s^2 + L_s^2) ds - \int_{1-1/p+l_p^0/m}^1 (\rho^2 K_s^2 + L_s^2) ds \right) \\ + p \left( \int_0^{\tilde{l}_p^*/m} (\rho^2 K_s^2 + L_s^2) ds - \int_0^{l_p^0/m} (\rho^2 K_s^2 + L_s^2) ds \right)$$

and so

$$|(D_p)_{i,j}| \leq c \frac{p|l_p^0 - \tilde{l}_p^*|}{m}.$$

- For  $(i, j)$  such that  $1 \leq i \leq p$ ,  $p+1 \leq j \leq 2p-1$  and  $j-p = i + p\lfloor \theta \rfloor_p$

$$(D_p)_{i,j} = p \left( \rho \int_{(i-1)/p}^{i/p} K_{s+\theta_m}^2 ds - \rho \int_{(i-1)/p}^{i/p-(\theta-\theta_m)+(l_p^0-\tilde{l}_p^*)/m} K_{s+\theta}^2 ds \right)$$

and so

$$|(D_p)_{i,j}| \leq c \left( \frac{p}{m} + \frac{p|l_p^0 - \tilde{l}_p^*|}{m} \right).$$

- For  $(i, j)$  such that  $i = p(1 - \lfloor \theta \rfloor_p)$ ,  $j = 2p$

$$(D_p)_{i,j} = p\rho \left( \int_{1-\lfloor \theta \rfloor_p-1/p}^{1-\theta_m} K_{s+\theta_m}^2 ds - \int_{1-\lfloor \theta \rfloor_p-1/p}^{1-\theta} K_{s+\theta}^2 ds \right),$$

and so

$$|(D_p)_{i,j}| \leq c \frac{p}{m}.$$

- For  $(i, j)$  such that  $1 \leq i \leq p$ ,  $p+1 \leq j \leq 2p$  and  $j-p = i + p\lfloor \theta \rfloor_p + 1$

$$(D_p)_{i,j} = -p\rho \int_{i/p-(\theta-\theta_m)+(l_p^0-\tilde{l}_p^*)/m}^{i/p} K_{s+\theta}^2 ds$$

and so

$$|(D_p)_{i,j}| \leq c \left( \frac{p}{m} + \frac{p|l_p^0 - \tilde{l}_p^*|}{m} \right).$$

By Gerschgorin–Hadamard theorem, we deduce that

$$\sup_{i=1, \dots, 2p} |\lambda_i(D_p)| \leq c \left( \frac{p}{m} + \frac{p|l_p^0 - \tilde{l}_p^*|}{m} \right). \quad \square$$

(ii) *Second step: the limiting spectral distribution of  $\Sigma_p^*(m)$ .* We first establish the convergence of the moments of the population spectral distribution of  $\Sigma_p^*(m)$ . We also show that the sequence given by the limiting moments satisfies Carleman's condition. This condition implies that this sequence solves Hamburger moment problem. So there exists a distribution  $H_\infty$



which is characterised by the sequence of the limiting moments. Finally, by a classical result on the method of moments (see for example Theorem 4.5.5 in [3]), we deduce that the spectral distribution of  $\Sigma_p^*(m)$  weakly converges to  $H_\infty$ .

Recall that the matrix  $\Sigma_p^*(m)$  is tri-diagonal. It is important to note that the powers of  $\Sigma_p^*(m)$  have the same tri-diagonal structure. We now give notation for the coefficients of these matrices that will be convenient for deriving recurrence relations between the coefficients. We define

$$\begin{aligned} a_i^{(k)} &= (\Sigma_p^*(m)^k)_{i,j}, \quad \text{for } 1 \leq i \leq p, i = j \\ b_{i-p\lfloor\theta\rfloor_p}^{(k)} &= (\Sigma_p^*(m)^k)_{i,j}, \quad \text{for } p+1 \leq i \leq p+p\lfloor\theta\rfloor_p, i = j \\ b_{i-p-p\lfloor\theta\rfloor_p}^{(k)} &= (\Sigma_p^*(m)^k)_{i,j}, \quad \text{for } p+p\lfloor\theta\rfloor_p+1 \leq i \leq 2p, i = j \\ c_i^{(k)} &= (\Sigma_p^*(m)^k)_{i,j}, \quad \text{for } 1 \leq i \leq p, p+1 \leq j \leq 2p, j-p = i+p\lfloor\theta\rfloor_p \\ c_j^{(k)} &= (\Sigma_p^*(m)^k)_{i,j}, \quad \text{for } p+1 \leq i \leq 2p, 1 \leq j \leq p, i-p = j+p\lfloor\theta\rfloor_p. \end{aligned}$$

We also define  $c_j^{(k)} = 0$  for  $j = p-p\lfloor\theta\rfloor_p+1, \dots, p$ . The following recurrence relations between the coefficients hold: for  $j = 1, \dots, p$

$$\begin{aligned} a_j^{(k+1)} &= a_j^{(1)} a_j^{(k)} + c_j^{(1)} c_j^{(k)} \\ b_j^{(k+1)} &= b_j^{(1)} b_j^{(k)} + c_j^{(1)} c_j^{(k)} \\ c_j^{(k+1)} &= c_j^{(1)} a_j^{(k)} + c_j^{(k)} b_j^{(1)} = c_j^{(k)} a_j^{(1)} + c_j^{(1)} b_j^{(k)}. \end{aligned}$$

Let us define

$$P_p^{(k)}(t) = \sum_{i=1}^p a_i^{(k)} 1_{\{t \in [(i-1)/p, i/p]\}}, \quad Q_p^{(k)}(t) = \sum_{i=1}^p b_i^{(k)} 1_{\{t \in [(i-1)/p, i/p]\}}, \quad R_p^{(k)}(t) = \sum_{i=1}^p c_i^{(k)} 1_{\{t \in [(i-1)/p, i/p]\}}.$$

Let  $P^{(1)}(t) = K_{t+\theta}^2$ ,  $Q^{(1)}(t) = (\rho^2 K_{t+\theta}^2 + L_{t+\theta}^2) 1_{\{0 \leq t \leq 1-\theta\}} + (\rho^2 K_{t-(1-\theta)}^2 + L_{t-(1-\theta)}^2) 1_{\{1-\theta \leq t \leq 1\}}$  and  $R^{(1)}(t) = \rho K_{t+\theta}^2 1_{\{t \leq 1-\theta\}}$ .

We also define recursively  $P^{(k)}$ ,  $Q^{(k)}$  and  $R^{(k)}$  by

$$\begin{aligned} P^{(k)}(t) &= P^{(1)}(t) P^{(k-1)}(t) + R^{(1)}(t) R^{(k-1)}(t) 1_{\{t \leq 1-\theta\}}, \\ Q^{(k)}(t) &= Q^{(1)}(t) Q^{(k-1)}(t) + R^{(1)}(t) R^{(k-1)}(t) 1_{\{t \leq 1-\theta\}}, \\ R^{(k)}(t) &= R^{(1)}(t) P^{(k-1)}(t) + R^{(k-1)}(t) Q^{(1)}(t) 1_{\{t \leq 1-\theta\}}. \end{aligned}$$

We define for  $k \geq 0$  the recurrence hypothesis  $\mathcal{H}_k$  the following way:

$$\mathcal{H}_k: \text{ for all } p, \quad \sup_{j=1, \dots, p} |e_j^k| \leq d_k, \quad \sup_{j=1, \dots, p} \sup_{u \in [\frac{j-1}{p}, \frac{j}{p}]} |e_j^k - E^{(k)}(u)| \leq \frac{d_k}{p}$$

where  $(e, E)$  stands for  $(a, P)$ ,  $(b, Q)$  or  $(c, R)$  and the  $d_k$  are constants depending on  $k$ . Note that  $\mathcal{H}_k$  implies the uniform convergence of  $E_p^{(k)}$  to  $E^{(k)}$ . It is clear that  $\mathcal{H}_1$  holds. Using that

$$e_j^{(1)} f_j^{(k)} - E^{(1)}(u) F^{(k)}(u) = f_j^{(k)} (e_j^{(1)} - E^{(1)}(u)) + E^{(1)}(u) (f_j^{(k)} - F^{(k)}(u)),$$

with  $(e, E)$  and  $(f, F)$  standing for  $(a, P)$ ,  $(b, Q)$  or  $(c, R)$ , we easily obtain that for  $k \geq 1$ ,  $\mathcal{H}_k$  implies  $\mathcal{H}_{k+1}$ . It follows that, as  $p \rightarrow \infty$ ,

$$\frac{1}{2p} \text{Tr} (\Sigma_p^*(m)^k) = \frac{1}{2} \int_0^1 P_p^{(k)}(t) dt + \frac{1}{2} \int_0^1 Q_p^{(k)}(t) dt \rightarrow \mu^{(k)} = \frac{1}{2} \int_0^1 P^{(k)}(t) dt + \frac{1}{2} \int_0^1 Q^{(k)}(t) dt.$$

Let

$$\alpha = \max \left\{ \sup_{t \in [0, 1]} P_t^{(1)}, \sup_{t \in [0, 1]} Q_t^{(1)}, \sup_{t \in [0, 1]} R_t^{(1)} \right\}.$$

An easy recursion gives that for  $k \geq 1$ ,

$$\frac{1}{2p} \text{Tr} (\Sigma_p^*(m)^k) \leq (2\alpha)^k.$$

It follows that Carleman's condition (see [3]) holds for  $H_\infty$ .

(iii) *Third step: the limiting spectral distribution of  $\Sigma_{(p,p)}^k$ .* It is not difficult to show by Gershgorin–Hadamard theorem that for any  $k \geq 1$

$$\limsup_{p \rightarrow \infty} \frac{1}{2p} |\text{Tr} (\Sigma_{(p,p)}^k)| < \infty.$$

For two matrices  $A$  and  $B$ , we have

$$\text{Tr}(A^k - B^k) = \text{Tr}((A - B)M)$$

where

$$M = (A^{k-1} + A^{k-2}B + A^{k-3}B^2 + \dots + B^{k-1}).$$

If  $A$  and  $B$  are also symmetric, we have

$$|\text{Tr}(A^k - B^k)| \leq \{\text{Tr}((A - B)^2)\}^{1/2} \{\text{Tr}(M^T M)\}^{1/2}.$$

We now take  $A = \Sigma_p^*(m)$  and  $B = \Sigma_{(l_p^0, p)}$ . Since

$$\limsup_{p \rightarrow \infty} \frac{1}{2p} \text{Tr}(M^T M) < \infty.$$

Using Lemma 4.1, it follows that for  $k \geq 1$ , as  $p$  tends to infinity,

$$\frac{1}{2p} \left| \text{Tr}(\Sigma_p^*(m)^k) - \text{Tr}(\Sigma_{(l_p^0, p)}^k) \right| \rightarrow 0.$$

Consequently, both distributions have the same asymptotic moments.

#### 4.2. Proof of Proposition 3.1

We begin by the following remark: if  $f$  is a twice continuously differentiable, positive function, then for any  $0 < a < b < 1$ ,

$$\left( \int_a^b f(s) ds \right)^2 = (b-a) \int_a^b f^2(s) ds + r_{a,b} \quad (4.1)$$

where  $|r_{a,b}| \leq c(b-a)^4$ .

We only consider the case when  $l_p^0 \leq \lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor$ . The other case is treated the same way. We have

$$\begin{aligned} \mu_{p,p}^{(2)} &= \frac{p}{2} \sum_{i=1}^p \left( \int_{(i-1)/p}^{i/p} K_{s+\theta}^2 ds \right)^2 + \frac{p}{2} \sum_{i=1}^{p-1} \left( \int_{(i-1)/p+l_p^0/m}^{i/p+l_p^0/m} (\rho^2 K_s^2 + L_s^2) ds \right)^2 \\ &\quad + \frac{p}{2} \left( \int_{1-1/p+l_p^0/m}^1 (\rho^2 K_s^2 + L_s^2) ds + \int_0^{l_p^0/m} (\rho^2 K_s^2 + L_s^2) ds \right)^2 \\ &\quad + p \sum_{i=1}^{p(1-\lfloor \theta \rfloor_p)-1} \left( \rho \int_{(i-1)/p}^{i/p-(\theta-\lfloor \theta \rfloor_p)+l_p^0/m} K_{s+\theta}^2 ds \right)^2 + p \sum_{i=1}^{p(1-\lfloor \theta \rfloor_p)-1} \left( \rho \int_{i/p-(\theta-\lfloor \theta \rfloor_p)+l_p^0/m}^{i/p} K_{s+\theta}^2 ds \right)^2 \\ &\quad + p \left( \rho \int_{1-\lfloor \theta \rfloor_p-1/p}^{1-\theta} K_{s+\theta}^2 ds \right)^2 =: \text{I} + \text{II} + \text{III} + \text{IV} + \text{V} + \text{VI}. \end{aligned}$$

We now study each term.

– Term I: Using the initial remark, we get

$$\text{I} = \frac{1}{2} \int_0^1 K_{s+\theta}^4 ds + O(1/p^2).$$

– Term II and Term III: Still using the initial remark, we have that II + III is equal to

$$\begin{aligned} &\frac{1}{2} \int_0^1 (\rho^2 K_s^2 + L_s^2)^2 ds - \frac{p l_p^0}{2m} \int_{1-1/p+l_p^0/m}^1 (\rho^2 K_s^2 + L_s^2)^2 ds + \left( \frac{p l_p^0}{2m} - \frac{1}{2} \right) \int_0^{l_p^0/m} (\rho^2 K_s^2 + L_s^2)^2 ds \\ &\quad + p \int_{1-1/p+l_p^0/m}^1 (\rho^2 K_s^2 + L_s^2) ds \int_0^{l_p^0/m} (\rho^2 K_s^2 + L_s^2) ds + O(1/p^2). \end{aligned}$$

We obtain

$$\begin{aligned} \text{II} + \text{III} &= \frac{1}{2} \int_0^1 (\rho^2 K_s^2 + L_s^2)^2 ds + O(1/p^2) - \frac{1}{2} p \frac{l_p^0}{m} \left( \frac{1}{p} - \frac{l_p^0}{m} \right) ((\rho^2 K_0^2 + L_0^2) - (\rho^2 K_1^2 + L_1^2))^2 \\ &= \frac{1}{2} \int_0^1 (\rho^2 K_s^2 + L_s^2)^2 ds + o(1/p) - \frac{1}{2} (1 - p(\theta - \lfloor \theta \rfloor_p)) (\theta - \lfloor \theta \rfloor_p) ((\rho^2 K_0^2 + L_0^2) - (\rho^2 K_1^2 + L_1^2))^2. \end{aligned}$$

– Term IV: Note that here  $\varepsilon_p = (\theta - \lfloor \theta \rfloor_p) - l_p^0/m$ . We have

$$\begin{aligned} \text{IV} &= p \sum_{i=1}^{p(1-\lfloor \theta \rfloor_p)-1} \rho^2(1/p - \varepsilon_p) \int_{(i-1)/p}^{i/p - \varepsilon_p} K_{s+\theta}^4 ds + O(1/p^2) \\ &= p \sum_{i=1}^{p(1-\lfloor \theta \rfloor_p)-1} \rho^2(1/p - \varepsilon_p)^2 K_{(i-1)/p+1/2(1/p-\varepsilon_p)+\theta}^4 + O(1/p^2) \\ &= p \sum_{i=1}^{p(1-\lfloor \theta \rfloor_p)-1} \rho^2(1/p - \varepsilon_p)^2 K_{(i-1)/p+1/(2p)+\theta}^4 + o(1/p) \\ &= (1 - p\varepsilon_p)^2 \rho^2 \int_0^{(1-\lfloor \theta \rfloor_p)-1/p} K_{s+\theta}^4 ds + o(1/p) \\ &= \rho^2 \int_0^{(1-\lfloor \theta \rfloor_p)-1/p} K_{s+\theta}^4 ds + \rho^2(p^2\varepsilon_p^2 - 2p\varepsilon_p) \int_0^{1-\theta} K_{s+\theta}^4 ds + o(1/p). \end{aligned}$$

– Term V: We have

$$\begin{aligned} \text{V} &= p \sum_{i=1}^{p(1-\lfloor \theta \rfloor_p)-1} \rho^2 \varepsilon_p \int_{i/p - \varepsilon_p}^{i/p} K_{s+\theta}^4 ds + O(1/p^2) \\ &= p \sum_{i=1}^{p(1-\lfloor \theta \rfloor_p)-1} \rho^2 \varepsilon_p^2 K_{(i-1)/p+1/(2p)+\theta}^4 + o(1/p) \\ &= p^2 \varepsilon_p^2 \rho^2 \int_0^{(1-\lfloor \theta \rfloor_p)-1/p} K_{s+\theta}^4 ds + o(1/p) \\ &= p^2 \varepsilon_p^2 \rho^2 \int_0^{1-\theta} K_{s+\theta}^4 ds + o(1/p). \end{aligned}$$

– Term VI: We easily obtain

$$\begin{aligned} \text{VI} &= \rho^2 (p(\lfloor \theta \rfloor_p - \theta) + 1) \int_{(1-\lfloor \theta \rfloor_p)-1/p}^{1-\theta} K_{s+\theta}^4 ds + O(1/p^2) \\ &= \rho^2 \int_{(1-\lfloor \theta \rfloor_p)-1/p}^{1-\theta} K_{s+\theta}^4 ds + \rho^2 p(\lfloor \theta \rfloor_p - \theta + 1/p)(\lfloor \theta \rfloor_p - \theta) K_1^2 + O(1/p^2). \end{aligned}$$

#### 4.3. Proof of Proposition 3.2

We first prove the following general lemma.

**Lemma 4.2.** On a filtered probability space  $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$  and a  $n$ -dimensional  $\tilde{\mathcal{F}}$ -Brownian motion  $(\tilde{W}_t^{(1)}, \dots, \tilde{W}_t^{(n)})_{t \geq 0}$  on it. For  $d = 1, \dots, n$ , we define

$$\tilde{Z}_t^{(d)} = \tilde{z}_0 + \int_0^t H_s d\tilde{W}_s^{(d)}, \quad t \in [0, 1],$$

and for  $k = 2$  or  $k = 4$ ,

$$\tilde{V}_{t,k}^{(d)} = \frac{m^{k/2-1}}{k-1} \sum_{1 \leq j \leq \lfloor mt \rfloor} (\tilde{Z}_{j/m}^{(d)} - \tilde{Z}_{(j-1)/m}^{(d)})^k \quad \text{and} \quad \tilde{V}_{t,k} = \frac{1}{n} \sum_{d=1}^n \tilde{V}_{t,k}^{(d)}.$$

Let  $\varepsilon > 0$ . For  $k = 2$  or  $k = 4$ , we have

$$((mn) \wedge m^2)^{1/2-\varepsilon} \left( \tilde{V}_{t,k} - \int_0^t H_s^k ds \right) \xrightarrow{ucp} 0$$

where  $ucp$  denotes uniform convergence in probability over compact sets included in  $[0, 1]$ .

**Proof.** We first compute the bias and variance of  $\tilde{V}_{t,k}^{(1)}$ ,  $k = 2, 4$ . First, remark that for  $1 \leq j \leq \lfloor mt \rfloor$ , the  $(\tilde{Z}_{j/m}^{(d)} - \tilde{Z}_{(j-1)/m}^{(d)})$  are independent, centered, Gaussian random variables with variance  $\int_{(j-1)/m}^{j/m} H_s^2 ds$ . It follows that

$$E[\tilde{V}_{t,2}^{(1)}] = \int_0^{\lfloor mt \rfloor / m} H_s^2 ds, \quad E[\tilde{V}_{t,4}^{(1)}] = m \sum_{1 \leq j \leq \lfloor mt \rfloor} \left( \int_{(j-1)/m}^{j/m} H_s^2 ds \right)^2$$

and

$$\text{Var}[\tilde{V}_{t,2}^{(1)}] + \text{Var}[\tilde{V}_{t,4}^{(1)}] \leq c/m.$$

Now, we can write

$$\tilde{V}_{t,k} - \int_0^t H_s^k ds = M_{t,k} + E[\tilde{V}_{t,k}^{(1)}] - \int_0^t H_s^k ds$$

where

$$M_{t,k} = \tilde{V}_{t,k} - E[\tilde{V}_{t,k}^{(1)}] = \sum_{1 \leq j \leq \lfloor mt \rfloor} \xi_{j,k}$$

with

$$\xi_{j,k} = \frac{m^{k/2-1}}{n(k-1)} \sum_{d=1}^n \left( (\tilde{Z}_{j/m}^{(d)} - \tilde{Z}_{(j-1)/m}^{(d)})^k - E[(\tilde{Z}_{j/m}^{(d)} - \tilde{Z}_{(j-1)/m}^{(d)})^k] \right).$$

The process  $M_{t,k}$  is a centered martingale with respect to the filtration  $\tilde{\mathcal{G}}_t = \tilde{\mathcal{F}}_{\lfloor mt \rfloor / m}$  such that

$$E[M_{1,k}^2] = \text{Var}[\tilde{V}_{1,k}] \leq c/(nm).$$

By Doob's inequality, we obtain

$$E \left[ \sup_{t \in [0,1]} |M_{t,k}| \right] \leq c/(nm)^{1/2}.$$

In an obvious way for the case  $k = 2$  and using (4.1) for the case  $k = 4$ , we get

$$\left| E[\tilde{V}_{t,k}^{(1)}] - \int_0^t H_s^k ds \right| \leq c/m.$$

Thus,

$$E \left[ \sup_{t \in [0,1]} \left| \tilde{V}_{t,k} - \int_0^t H_s^k ds \right| \right] \leq c/m + c/(nm)^{1/2}.$$

The result follows.  $\square$

**Corollary 4.3.** We have

$$p(\hat{V}_t^X - V_t^X) \xrightarrow{ucp} 0, \quad p(\hat{V}_t^Y - V_t^Y) \xrightarrow{ucp} 0, \quad p(\hat{Q}_t^X - Q_t^X) \xrightarrow{ucp} 0, \quad p(\hat{Q}_t^Y - Q_t^Y) \xrightarrow{ucp} 0.$$

**Proposition 3.2** is an obvious consequence of the previous corollary and of the following lemma.

**Lemma 4.4.** Let  $\eta_m = m^{\frac{4-b}{4+2b}}$ , for some  $b > 4 + 6/a$ . As  $p \rightarrow +\infty$ , we have

$$p(\hat{\rho} - \rho) \xrightarrow{\mathbb{P}} 0.$$

**Proof.** Let

$$\zeta^{(d)} = \sum_{0 \leq k \leq m\bar{\theta}} \Gamma_k^{(d)} 1_{\{| \Gamma_k^{(d)} | \geq \eta_m\}} - \rho \int_0^{1-\bar{\theta}} H_s^2 ds.$$

· First step: we first prove that  $E[(\zeta^{(d)})^2] \leq c\eta_m^2$ . We have

$$\sum_{0 \leq k \leq m\bar{\theta}} \Gamma_k^{(d)} 1_{\{| \Gamma_k^{(d)} | \geq \eta_m\}}$$

is equal to

$$\Gamma_{m\lfloor \theta \rfloor m}^{(d)} + \Gamma_{m\lfloor \theta \rfloor m+1}^{(d)} - \Gamma_{m\lfloor \theta \rfloor m}^{(d)} 1_{\{| \Gamma_{m\lfloor \theta \rfloor m}^{(d)} | < \eta_m\}} - \Gamma_{m\lfloor \theta \rfloor m+1}^{(d)} 1_{\{| \Gamma_{m\lfloor \theta \rfloor m+1}^{(d)} | < \eta_m\}} + \sum_{\substack{0 \leq k \leq m\bar{\theta} \\ k \neq \lfloor m\theta \rfloor, k \neq m\lfloor \theta \rfloor m+1}} \Gamma_k^{(d)} 1_{\{| \Gamma_k^{(d)} | \geq \eta_m\}}.$$

Let  $\beta \geq 2$  and  $k < m\lfloor\theta\rfloor_m$ . Let  $\tilde{\mathcal{F}}_j^{(d)} = \sigma(X_{i/m}^{(d)}, Y_{i/m}^{(d)} : i \leq j)$ . By Rosenthal inequality for martingales, we get that  $E[|\Gamma_k^{(d)}|^\beta]$  is smaller than

$$c_\beta E \left[ \sum_{1 \leq j \leq m(1-\bar{\theta})} E[(\Delta^{(0,m)} X_j^{(d)} \Delta^{(0,m)} Y_{j+k}^{(d)})^2 | \tilde{\mathcal{F}}_{j+k-1}^{(d)}] \right]^{\beta/2} + c_\beta \sum_{1 \leq j \leq m(1-\bar{\theta})} E[|\Delta^{(0,m)} X_j^{(d)} \Delta^{(0,m)} Y_{j+k}^{(d)}|^\beta].$$

Hence

$$E[|\Gamma_k^{(d)}|^\beta] \leq cm^{-\beta/2} E \left[ \sum_{1 \leq j \leq m(1-\bar{\theta})} (\Delta^{(0,m)} X_j^{(d)})^2 \right]^{\beta/2} + cm^{1-\beta}.$$

Now using the fact that  $E \left[ \sum_{1 \leq j \leq m(1-\bar{\theta})} (\Delta^{(0,m)} X_j^{(d)})^2 \right]^{\beta/2}$  is smaller than

$$c E \left[ \sum_{1 \leq j \leq m(1-\bar{\theta})} \left( (\Delta^{(0,m)} X_j^{(d)})^2 - \int_{(j-1)/m}^{j/m} K_{s+\theta}^2 ds \right) \right]^{\beta/2} + c \left( \int_0^{1-\bar{\theta}} K_{s+\theta}^2 ds \right)^{\beta/2}$$

together with Rosenthal inequality, we deduce that

$$E \left[ \sum_{1 \leq j \leq m(1-\bar{\theta})} (\Delta^{(0,m)} X_j^{(d)})^2 \right]^{\beta/2} \leq c.$$

We eventually obtain for  $\beta \geq 2$

$$E[|\Gamma_k^{(d)}|^\beta] \leq cm^{-\beta/2}.$$

The case  $k > m\lfloor\theta\rfloor_m + 1$  is treated the same way. Now, for  $k \neq k'$ ,  $k \neq m\lfloor\theta\rfloor_m$ ,  $k \neq m\lfloor\theta\rfloor_m + 1$ ,  $k' \neq m\lfloor\theta\rfloor_m$ ,  $k' \neq m\lfloor\theta\rfloor_m + 1$ , applying Cauchy–Schwarz and Markov inequalities, we obtain for  $\beta \geq 2$

$$|E[\Gamma_k^{(d)} 1_{\{|\Gamma_k^{(d)}| \geq \eta_m\}} \Gamma_{k'}^{(d)} 1_{\{|\Gamma_{k'}^{(d)}| \geq \eta_m\}}]| \leq cm^{-1-\beta/4} \eta_m^{-\beta/2}.$$

Then,

$$E \left[ \left( \sum_{\substack{0 \leq k \leq m\bar{\theta} \\ k \neq \lfloor m\theta \rfloor, k \neq m\lfloor\theta\rfloor_m + 1}} \Gamma_k^{(d)} 1_{\{|\Gamma_k^{(d)}| \geq \eta_m\}} \right)^2 \right] \leq cm^{1-\beta/4} \eta_m^{-\beta/2}.$$

We now turn to  $\Gamma_{m\lfloor\theta\rfloor_m}^{(d)}$  and  $\Gamma_{m\lfloor\theta\rfloor_m+1}^{(d)}$ . We have

$$E[(\Gamma_{m\lfloor\theta\rfloor_m}^{(d)} 1_{\{\Gamma_{m\lfloor\theta\rfloor_m}^{(d)} < \eta_m\}})^2 + (\Gamma_{m\lfloor\theta\rfloor_m+1}^{(d)} 1_{\{\Gamma_{m\lfloor\theta\rfloor_m+1}^{(d)} < \eta_m\}})^2] \leq c\eta_m^2.$$

Moreover,  $\Gamma_{m\lfloor\theta\rfloor_m}^{(d)} + \Gamma_{m\lfloor\theta\rfloor_m+1}^{(d)}$  can be written

$$\begin{aligned} & \rho \sum_{1 \leq j \leq m(1-\bar{\theta})} (\Delta^{(0,m)} X_j^{(d)})^2 + \rho \sum_{1 \leq j \leq m(1-\bar{\theta})} \Delta^{(0,m)} X_j^{(d)} (X_{\lfloor\theta\rfloor_m - \theta + (j+1)/m}^{(d)} - X_{j/m}^{(d)}) \\ & + \rho \sum_{1 \leq j \leq m(1-\bar{\theta})} \Delta^{(0,m)} X_j^{(d)} (X_{(j-1)/m}^{(d)} - X_{(\lfloor\theta\rfloor_m - \theta + (j-1)/m) \vee 0}^{(d)}) + \sum_{1 \leq j \leq m(1-\bar{\theta})} \Delta^{(0,m)} X_j^{(d)} (\Delta^{(0,m)} Z_{m\lfloor\theta\rfloor_m}^{(d)} + \Delta^{(0,m)} Z_{m\lfloor\theta\rfloor_m+1}^{(d)}), \end{aligned}$$

with  $Z_t^{(d)} = \int_0^t L_s dW_s^{(d)}$ . We easily get

$$E \left[ \left( \sum_{1 \leq j \leq m(1-\bar{\theta})} \Delta^{(0,m)} X_j^{(d)} (\Delta^{(0,m)} Z_{m\lfloor\theta\rfloor_m}^{(d)} + \Delta^{(0,m)} Z_{m\lfloor\theta\rfloor_m+1}^{(d)}) \right)^2 \right] \leq c/m,$$

and we obtain

$$E \left[ \left( \sum_{1 \leq j \leq m(1-\bar{\theta})} \Delta^{(0,m)} X_j^{(d)} (X_{\lfloor\theta\rfloor_m - \theta + (j+1)/m} - X_{j/m}) \right)^2 \right] \leq c/m$$

and

$$\mathbb{E} \left[ \left( \sum_{1 \leq j \leq m(1-\bar{\theta})} \Delta^{(0,m)} X_j^{(d)} (X_{(j-1)/m} - X_{(\lfloor \theta \rfloor_m - \theta + (j-1)/m) \vee 0}) \right)^2 \right] \leq c/m.$$

Moreover we have

$$\mathbb{E} \left[ \left( \sum_{1 \leq j \leq m(1-\bar{\theta})} (\Delta^{(0,m)} X_j^{(d)})^2 - \int_0^{1-\bar{\theta}} H_s^2 ds \right)^2 \right] \leq c/m.$$

Hence, we eventually have that

$$\mathbb{E} \left[ \left( \sum_{0 \leq k \leq m\bar{\theta}} \Gamma_k^{(d)} 1_{\{|\Gamma_k^{(d)}| \geq \eta_m\}} - \rho \int_0^{1-\bar{\theta}} H_s^2 ds \right)^2 \right]$$

is smaller than

$$c(m^{1-\beta/4} \eta_m^{-\beta/2} + \eta_m^2 + m^{-1}).$$

We now take  $\beta = b$  and  $\eta_m = m^{\frac{4-b}{4+2b}}$ , we get

$$\mathbb{E} \left[ \left( \sum_{0 \leq k \leq m\bar{\theta}} \Gamma_k^{(d)} 1_{\{|\Gamma_k^{(d)}| \geq \eta_m\}} - \rho \int_0^{1-\bar{\theta}} H_s^2 ds \right)^2 \right] \leq c\eta_m^2.$$

· Second step: We now prove that  $|\mathbb{E}[\zeta^{(d)}]| \leq c\eta_m^2$ . Remark that

$$\mathbb{E} \left[ \sum_{0 \leq k \leq m\bar{\theta}} \Gamma_k^{(d)} 1_{\{|\Gamma_k^{(d)}| \geq \eta_m\}} \right] = -\mathbb{E}[\Gamma_{m\lfloor \theta \rfloor_m}^{(d)} 1_{\{|\Gamma_{m\lfloor \theta \rfloor_m}^{(d)}| < \eta_m\}}] - \mathbb{E}[\Gamma_{m\lfloor \theta \rfloor_m+1}^{(d)} 1_{\{|\Gamma_{m\lfloor \theta \rfloor_m+1}^{(d)}| < \eta_m\}}] + \rho \int_0^{1-\bar{\theta}} H_s^2 ds.$$

We have

$$|\mathbb{E}[\Gamma_{m\lfloor \theta \rfloor_m}^{(d)} 1_{\{|\Gamma_{m\lfloor \theta \rfloor_m}^{(d)}| < \eta_m\}}] + \mathbb{E}[\Gamma_{m\lfloor \theta \rfloor_m+1}^{(d)} 1_{\{|\Gamma_{m\lfloor \theta \rfloor_m+1}^{(d)}| < \eta_m\}}]| \leq \eta_m \mathbb{P}[|\Gamma_{m\lfloor \theta \rfloor_m}^{(d)} + \Gamma_{m\lfloor \theta \rfloor_m+1}^{(d)}| \leq 2\eta_m].$$

Using the preceding computations, we get

$$\mathbb{E} \left[ \left| \Gamma_{m\lfloor \theta \rfloor_m}^{(d)} + \Gamma_{m\lfloor \theta \rfloor_m+1}^{(d)} \right| \right] = \rho^2 \int_0^{1-\bar{\theta}} H_s^2 ds + R_1$$

with  $R_1^2 \leq c/m$ . Furthermore,

$$\mathbb{E} \left[ \left| \Gamma_{m\lfloor \theta \rfloor_m}^{(d)} + \Gamma_{m\lfloor \theta \rfloor_m+1}^{(d)} \right|^2 \right] = \mathbb{E} \left[ \left( \rho \sum_{1 \leq j \leq m(1-\bar{\theta})} (\Delta^{(0,m)} X_j^{(d)})^2 \right)^2 \right] + R_2,$$

with  $R_2 \leq c/m$ . So this is equal to

$$\rho^2 \left( \sum_{1 \leq j \leq m(1-\bar{\theta})} \mathbb{E}[(\Delta^{(0,m)} X_j^{(d)})^2] \right)^2 - \rho^2 \sum_{1 \leq j \leq m(1-\bar{\theta})} \left( \mathbb{E}[(\Delta^{(0,m)} X_j^{(d)})^2] \right)^2 + \rho^2 \sum_{1 \leq j \leq m(1-\bar{\theta})} \mathbb{E}[(\Delta^{(0,m)} X_j^{(d)})^4] + R_2,$$

which is equal to  $\rho^2 \left( \int_0^{1-\bar{\theta}} H_s^2 ds \right)^2 + R_3$ , with  $R_3 \leq c/m$ . Using Paley–Zygmund inequality, we get

$$\mathbb{P}[|\Gamma_{m\lfloor \theta \rfloor_m}^{(d)} + \Gamma_{m\lfloor \theta \rfloor_m+1}^{(d)}| \leq 2\eta_m] \leq c\eta_m.$$

· Third step: Since  $p\eta_m^2 \rightarrow 0$  and  $p^2\eta_m^2 n^{-1} \rightarrow 0$ , then

$$\mathbb{E}[p\zeta^{(d)}] \rightarrow 0, \quad \frac{\mathbb{E}[(p\zeta^{(d)})^2]}{n} \rightarrow 0,$$

and we obtain

$$\frac{1}{n} \sum_{d=1}^n p \zeta^{(d)} \xrightarrow{\mathbb{P}} 0.$$

Using [Corollary 4.3](#), the result follows.  $\square$

#### 4.4. Proof of [Theorem 3.3](#)

For  $0 \leq t \leq 1$ , let  $f(t) = \rho^2 Q_{1-t}^X$  and  $\hat{f}_p(t) = \hat{\rho}^2 \hat{Q}_{1-t}^X$ . The quantity  $p[f(\theta_p) - \hat{f}_p(\theta_p)]$  is equal to

$$2 \left( Q_{1-\theta_p}^X p(\rho^2 - \hat{\rho}^2) + \hat{\rho}^2 p(Q_{1-\theta_p}^X - \hat{Q}_{1-\theta_p}^X) \right)$$

and consequently, from [Corollary 4.3](#) and [Lemma 4.4](#)

$$p[f(\theta_p) - \hat{f}_p(\theta_p)] \xrightarrow{\mathbb{P}} 0.$$

In the same way, we obtain

$$p[(Q_1^X + Q_1^Y) - (\hat{Q}_1^X + \hat{Q}_1^Y) + \gamma_p(V_1^X + V_1^Y)^2 - \gamma_p(\hat{V}_1^X + \hat{V}_1^Y)^2] \xrightarrow{\mathbb{P}} 0.$$

By [Theorem 1.1](#) in [\[1\]](#)

$$2p[\hat{m}_{p,p}^{(2)} - m_{p,p}^{(2)}] = 2p \left[ \hat{m}_{p,p}^{(2)} - \frac{\gamma_p}{4}(V_1^X + V_1^Y)^2 - \frac{1}{2}(Q_1^X + Q_1^Y) - f(\theta_p) \right] + o(1)$$

converges weakly to a Gaussian random variable with mean  $\nu$  and variance  $\sigma^2$  respectively given taking  $z \mapsto z^2$  in [Eqs. \(1.6\)](#) and [\(1.7\)](#) in [\[1\]](#). Let

$$\hat{B}_p = \hat{m}_{p,p}^{(2)} - \frac{\gamma_p}{4}(\hat{V}_1^X + \hat{V}_1^Y)^2 - \frac{1}{2}(\hat{Q}_1^X + \hat{Q}_1^Y).$$

It follows that

$$2p[\hat{B}_p - \hat{f}_p(\theta_p)]$$

also converges weakly to a Gaussian random variable. Since  $|\hat{B}_p - \hat{f}_p(\hat{\theta}_p)| \leq c/m$ , the preceding result remains true when replacing  $\hat{B}_p$  by  $\hat{f}_p(\hat{\theta}_p)$ . Now,  $p[\hat{f}_p(\hat{\theta}_p) - \hat{f}_p(\theta_p)]$  can be written  $T_1 + T_2$  with

$$T_1 = p[f(\hat{\theta}_p) - f(\theta_p)]$$

$$T_2 = p[\hat{f}_p(\hat{\theta}_p) - f(\hat{\theta}_p) - \hat{f}_p(\theta_p) + f(\theta_p)].$$

The fact that  $T_2$  tends to zero in probability is a consequence of the ucp convergence of

$$p(\hat{\rho}^2 \hat{Q}_{1-t}^X - \rho^2 Q_{1-t}^X)$$

to 0 (see [Corollary 4.3](#)). We conclude using the Delta method.

#### 4.5. Proof of [Proposition 3.4](#)

Let  $s_p = p^{-d}$  where  $0 < d < \min(a, 1/2)$  and define

$$\mathcal{D}_p = \{(i, j), 1 \leq i \leq p, p+1 \leq j \leq 2p-1\}$$

$$\mathcal{E}_p = \{(i, j), 1 \leq i \leq p, p+1 \leq j \leq 2p-1, j-p = i + p\lfloor \theta \rfloor_p - 1\}$$

$$\mathcal{F}_p = \{(i, j), 1 \leq i \leq p, p+1 \leq j \leq 2p-1, j-p = i + p\lfloor \theta \rfloor_p\}$$

$$\mathcal{G}_p = \{(i, j), 1 \leq i \leq p, p+1 \leq j \leq 2p-1, j-p = i + p\lfloor \theta \rfloor_p + 1\}.$$

First note that, by [Lemma A.3](#) in [\[2\]](#), we have for any  $(i, j) \in \mathcal{D}_p \setminus (\mathcal{E}_p \cup \mathcal{F}_p \cup \mathcal{G}_p)$  and any  $0 \leq l \leq \lfloor p^a \rfloor - 1$

$$\mathbb{P} \left[ |s_{i,j}^{(l,p)}| > s_p \right] \leq C_1 \exp(-C_2 n s_p^2 / \log n).$$

It follows by Borel–Cantelli lemma that

$$\mathbb{P} \left[ \sup_{(i,j) \in \mathcal{D}_p \setminus (\mathcal{E}_p \cup \mathcal{F}_p \cup \mathcal{G}_p)} |s_{i,j}^{(l,p)}| > s_p \text{ i.o.} \right] = 0.$$

Let  $\underline{K} = \min_{s \in [0,1]} K_{s+\theta}^2$  and  $0 < \varepsilon < \underline{K}$ . Recall that since  $\theta$  is a rational number, the set

$$\{p(\theta - \lfloor \theta \rfloor_p), p \geq 1\}$$

contains zero and is finite. For large enough  $p$ , we only have to consider the two following cases:

- (1)  $p(\theta - \lfloor \theta \rfloor_p) \geq (\underline{K} - \varepsilon)^{-1} p^{-d}$  and  $1 - p(\theta - \lfloor \theta \rfloor_p) \geq (\underline{K} - \varepsilon)^{-1} p^{-d}$ .
- (2)  $p(\theta - \lfloor \theta \rfloor_p) = 0$ .

We now prove that for each sub sequence of integers  $p^{(i)}$  corresponding to the case  $i$ , the associated sequence  $\varepsilon_p^{(i)}$  is such that  $\varepsilon_p^{(i)} = o_{a.s.}(1/p^{(i)})$ . For simplicity, we keep the notation  $p$  instead of  $p^{(i)}$  for the two different cases, even though it only corresponds to a sub sequence of integers.

– Case (1):

· First step: We have for any  $(i, j) \in \mathcal{F}_p$  and  $0 \leq l \leq \lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor$

$$pv_{i,l,1}^{XY} \geq \frac{\underline{K}}{(\underline{K} - \varepsilon)} p^{-d}$$

and we have for any  $(i, j) \in \mathcal{F}_p$  and  $\lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor + 1 \leq l \leq \lfloor p^d \rfloor - 1$

$$pv_{i,l,2}^{XY} \geq \frac{\underline{K}}{(\underline{K} - \varepsilon)} p^{-d}.$$

Then we deduce that for any  $(i, j) \in \mathcal{F}_p$

$$\mathbb{P} \left[ |s_{i,j}^{(l,p)}| \leq s_p \right] \leq \mathbb{P} \left[ |s_{i,j}^{(l,p)} - pv_{i,l,k}^{XY}| > pv_{i,l,k}^{XY} - s_p \right] \leq C_1 \exp(-cC_2 np^{-2d} / \log n)$$

where  $k = 1$  if  $0 \leq l \leq \lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor$  and  $k = 2$  if  $\lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor + 1 \leq l \leq \lfloor p^d \rfloor - 1$ . It follows that by Borel–Cantelli lemma that for any  $0 \leq l \leq \lfloor p^d \rfloor - 1$

$$\mathbb{P} \left[ \inf_{(i,j) \in \mathcal{F}_p} |s_{i,j}^{(l,p)}| \leq s_p \text{ i.o.} \right] = 0,$$

and so

$$\liminf_{p \rightarrow \infty} \underline{N}^{(p)}(s_p) \geq_{a.s.} (1 - \theta).$$

· Second step: Consider a sequence  $l_p$  such that

$$\limsup_{p \rightarrow \infty} |l_p - l_p^*| \frac{p^{1+\bar{\gamma}}}{m} < +\infty$$

with  $d < \bar{\gamma} < a$ . For any  $(i, j) \in (\mathcal{E}_p \cup \mathcal{G}_p)$  we have

$$\limsup_{p \rightarrow \infty} p^{1+\bar{\gamma}} v_{i,l,k}^{XY} < +\infty$$

where  $k = 2$  if  $0 \leq l_p \leq \lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor$  and  $k = 1$  if  $\lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor + 1 \leq l_p \leq \lfloor p^d \rfloor - 1$ . By using the same type of arguments as previously, we have for any  $(i, j) \in (\mathcal{E}_p \cup \mathcal{G}_p)$

$$\mathbb{P} \left[ |s_{i,j}^{(l_p,p)}| > s_p \right] \leq \mathbb{P} \left[ |s_{i,j}^{(l_p,p)} - pv_{i,l,k}^{XY}| > s_p - pv_{i,l,k}^{XY} \right] \leq C_1 \exp(-cC_2 np^{-2d} / \log n).$$

It follows by Borel–Cantelli lemma that

$$\mathbb{P} \left[ \sup_{(i,j) \in (\mathcal{E}_p \cup \mathcal{G}_p)} |s_{i,j}^{(l_p,p)}| > s_p \text{ i.o.} \right] = 0.$$

· Third step: Consider a sequence  $l_p$  such that

$$\liminf_{p \rightarrow \infty} |l_p - l_p^*| \frac{p^{1+\underline{\gamma}}}{m} > 0$$

with  $0 < d < \underline{\gamma}$ . For the sub sequence of values of  $l_p$  such that  $0 \leq l_p \leq \lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor$ , for any  $(i, j) \in \mathcal{G}_p$ , we have

$$\liminf_{p \rightarrow \infty} p^{1+\underline{\gamma}} v_{i,l,2}^{XY} > 0$$

and consequently,

$$\mathbb{P} \left[ \inf_{(i,j) \in \mathcal{G}_p} |s_{i,j}^{(l_p,p)}| < s_p \text{ i.o.} \right] = 0.$$

For the sub sequence of values of  $l_p$  such that  $\lfloor m(\theta - \lfloor \theta \rfloor_p) \rfloor + 1 \leq l_p \leq \lfloor p^d \rfloor - 1$ , for any  $(i, j) \in \mathcal{E}_p$ , we have

$$\liminf_{p \rightarrow \infty} p^{1+\underline{\gamma}} v_{i,l,1}^{XY} > 0$$



and consequently,

$$\mathbb{P} \left[ \inf_{(i,j) \in \mathcal{E}_p} |s_{i,j}^{(l_p,p)}| < s_p \text{ i.o.} \right] = 0.$$

Thus,

$$\lim_{p \rightarrow \infty} \underline{N}^{(p)}(s_p) =_{\text{a.s.}} (1 - \theta)$$

and

$$\limsup_{p \rightarrow \infty} \frac{p^{1+\underline{\gamma}}}{m} (\bar{l}_p - l_p^*) = 0 \quad \text{and} \quad \limsup_{p \rightarrow \infty} \frac{p^{1+\underline{\gamma}}}{m} (l_p^* - \underline{l}_p) = 0.$$

Then it is easy to see that

$$\mathbb{P} [i_p = p(1 - \lfloor \theta \rfloor_p) - 1 \text{ i.o.}] = 1$$

and therefore  $l_p^e = \bar{l}_p$  a.s. since

$$\lim_{p \rightarrow \infty} T_{i_p+1,2p}^{(\bar{l}_p,p)}(s) =_{\text{a.s.}} 1.$$

So we conclude that

$$\frac{|l_p^e - l_p^*|}{m} = o_{\text{a.s.}}(1/p).$$

– Case (2):

· First step: By using the same type of arguments as previously, we have for any  $0 \leq l \leq \lfloor p^a \rfloor - 1$

$$\mathbb{P} \left[ \sup_{(i,j) \in \mathcal{G}_p} |s_{i,j}^{(l,p)}| > s_p \text{ i.o.} \right] = 0.$$

· Second step: Consider a sequence  $l_p$  such that

$$\limsup_{p \rightarrow \infty} l_p \frac{p^{1+\bar{\gamma}}}{m} < +\infty.$$

Then

$$\mathbb{P} \left[ \inf_{(i,j) \in \mathcal{F}_p} |s_{i,j}^{(l_p,p)}| \leq s_p \text{ i.o.} \right] = 0 \quad \text{and} \quad \mathbb{P} \left[ \sup_{(i,j) \in \mathcal{E}_p} |s_{i,j}^{(l_p,p)}| > s_p \text{ i.o.} \right] = 0.$$

Consider a sequence  $l_p$  such that

$$\liminf_{p \rightarrow \infty} l_p \frac{p^{1+\underline{\gamma}}}{m} > 0.$$

Then

$$\mathbb{P} \left[ \sup_{(i,j) \in \mathcal{F}_p} |s_{i,j}^{(l_p,p)}| > s_p \text{ i.o.} \right] = 0.$$

· Third step: Consider a sequence  $l_p$  such that

$$\limsup_{p \rightarrow \infty} |\lfloor p^a \rfloor - l_p| \frac{p^{1+\bar{\gamma}}}{m} < +\infty.$$

Then

$$\mathbb{P} \left[ \inf_{(i,j) \in \mathcal{E}_p} |s_{i,j}^{(l_p,p)}| < s_p \text{ i.o.} \right] = 0 \quad \text{and} \quad \mathbb{P} \left[ \sup_{(i,j) \in \mathcal{F}_p} |s_{i,j}^{(l_p,p)}| > s_p \text{ i.o.} \right] = 0.$$

Consider a sequence  $l_p$  such that

$$\liminf_{p \rightarrow \infty} |\lfloor p^a \rfloor - l_p| \frac{p^{1+\underline{\gamma}}}{m} > 0.$$

Then

$$\mathbb{P} \left[ \sup_{(i,j) \in \mathcal{E}_p} |s_{ij}^{(l_p, p)}| > s_p \text{ i.o.} \right] = 0.$$

Finally, it is easy to see that

$$\mathbb{P} [i_p = p(1 - \lfloor \theta \rfloor_p) \text{ i.o.}] = 1$$

and therefore  $l_p^e = l_p$  a.s. since

$$\lim_{p \rightarrow \infty} T_{i_p+1, 2p}^{(\bar{l}_p, p)}(s) =_{\text{a.s.}} 0.$$

So we conclude that

$$\frac{|l_p^e - l_p^*|}{m} = o_{\text{a.s.}}(1/p).$$

## Acknowledgment

We are grateful to Jack Silverstein for pointing out a helpful result in [8].

## References

- [1] Z.D. Bai, J.W. Silverstein, CLT of linear spectral statistics of large dimensional sample covariance matrices, *Annals of Probability* 32 (2004) 553–605.
- [2] P. Bickel, E. Levina, Regularized estimation of large covariance matrices, *Annals of Statistics* 36 (2008) 199–227.
- [3] K.L. Chung, *A Course in Probability Theory*, 3rd ed., Academic Press, 2001.
- [4] F. de Jong, T. Nijman, High frequency analysis of lead-lag relationships between financial markets, *Journal of Empirical Finance* 4 (1997) 259–277.
- [5] M. Hoffmann, M. Rosenbaum, N. Yoshida, Estimation of the Lead-Lag parameter from non-synchronous data, Working Paper, 2009.
- [6] I.G. Kowaller, P.D. Koch, T.W. Koch, The temporal price relationship between S&P 500 futures and S&P 500 index, *The Journal of Finance* 42 (1987) 1309–1329.
- [7] V.A. Marcenko, L.A. Pastur, Distribution of eigenvalues in certain sets of random matrices, *Matematicheskii Sbornik. New Series* 72 (1967) 507–536.
- [8] J.W. Silverstein, The limiting eigenvalue distribution of a multivariate  $F$  matrix, *SIAM Journal on Mathematical Analysis* 16 (1985) 641–646.
- [9] J.W. Silverstein, Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices, *Journal of Multivariate Analysis* 55 (1995) 331–339.
- [10] H. Weyl, Über die Gleichverteilung von Zahlen mod Eins, *Mathematische Annalen* 77 (1916) 313–352.
- [11] Y.Q. Yin, P.R. Krishnaiah, A limit theorem for the eigenvalues of product of two random matrices, *Journal of Multivariate Analysis* 13 (1983) 489–507.