Estimation of the lead-lag parameter from non-synchronous data

M. HOFFMANN[1], M. ROSENBAUM[2] and N. YOSHIDA[3]

[1] *ENSAE – CREST and CNRS UMR 8050. E-mail:* marc.hoffmann@ensae.fr

[2] *École polytechnique Paris and CREST. E-mail:* mathieu.rosenbaum@polytechnique.edu

[3] *University of Tokyo and Japan Science and Technology Agency. E-mail:* nakahiro@ms.u-tokyo.ac.jp

REVISED VERSION

We propose a simple continuous time model for modeling the lead-lag effect between two financial assets. A two dimensional process (X_t, Y_t) reproduces a lead-lag effect if for some time shift $\vartheta \in \mathbb{R}$, the process $(X_t, Y_{t+\vartheta})$ is a semi-martingale with respect to a certain filtration. The value of the time shift ϑ is the lead-lag parameter. Depending on the underlying filtration, the standard no-arbitrage case is obtained for $\vartheta = 0$. We study the problem of estimating the unknown parameter $\vartheta \in \mathbb{R}$, given randomly sampled non-synchronous data from (X_t) and (Y_t). By applying a certain contrast optimization based on a modified version of the Hayashi-Yoshida covariation estimator, we obtain a consistent estimator of the lead-lag parameter, together with an explicit rate of convergence governed by the sparsity of the sampling design.

Keywords: Hayashi-Yoshida covariation estimator, Lead-lag effect, Discretely observed continuous time processes, Contrast estimation.

1. Introduction

Market participants usually agree that certain pairs of assets (X, Y) share a "lead-lag effect", in the sense that the lagger (or follower) price process Y tends to partially reproduce the oscillations of the leader (or driver) price process X, with some temporal delay, or vice-versa. This property is usually referred to as the "lead-lag effect". The lead-lag effect may have some importance in practice, when assessing the quality of risk management indicators for instance, or more generally when considering statistical arbitrage strategies. It can be measured at various temporal scales (daily, hourly, or even at the level of seconds for flow products traded on electronic markets).

The lead-lag effect is a concept of common practice that has some history in financial econometrics. In time series for instance, this notion can be linked to the concept of Granger causality, and we refer to Comte and Renault [6] for a general approach. From a phenomenological perspective, the lead-lag effect is supported by empirical evidence reported in [4], [5] and [20], together with [22] and the references therein. To our knowledge

however, only few mathematical results are available from the point of view of statistical estimation from discretely observed continuous time processes. The purpose of this paper is to – partly – fill in this gap. (Also, recently, Robert and Rosenbaum study in [27] the lead-lag effect by means of random matrices, in a mixed asymptotic framework, a setting which is relatively different than in the present paper.)

1.1. Motivation

1) Our primary goal is to provide a simple – yet relatively general – model for capturing the lead-lag effect in continuous time, readily compatible with stochastic calculus in financial modeling. Informally, if $\tau_{-\vartheta}(Y)_t := Y_{t+\vartheta}$, with $\vartheta \in \mathbb{R}$, is the time shift operator, we say that the pair (X, Y) will produce a lead-lag effect as soon as $\big(X, \tau_{-\vartheta}(Y)\big)$ is a (regular) semi-martingale with respect to an appropriate filtration, for some ϑ, called the lead-lag parameter. The usual no-arbitrage case is embedded into this framework for $\vartheta = 0$. More in Section 2 below.

2) At a same level of importance, we aim at constructing a simple and efficient procedure for estimating the lead-lag parameter ϑ based on historical data. The underlying statistical model is generated by a – possibly random – sampling of both X and Y. The sampling typically happens at irregularly and non-synchronous times for X and Y. We construct in the paper an estimator of ϑ based on a modification of the Hayashi-Yoshida covariation estimator, see [13] and [15]. Our result is that the lead-lag parameter can be consistently estimated against a fairly general class of sampling schemes. Moreover, we explicit the rate of convergence of our procedure.

3) From a financial point of view, unless appropriate time shifts are operated, our model incapacitates our primary assets X and Y to be a semi-martingale with respect to the same filtration. This is consistent as far as modeling is concerned, but allows in principle for market imperfections such as statistical arbitrage if the lead-lag parameter ϑ is different from zero. More in Section 3.4 below. Addressing such a possibility is indeed the issue of the lead-lag effect, but we will content ourselves with detecting whether the lead-lag effect is present or not. The quantization of statistical arbitrage in terms of ϑ (and other parameters such as trading frequency, market friction, volatility and so on) lies beyond the scope of this paper.

4) From a statistical inference point of view, the statistician and the data provider are not necessarily the same agents, and this leads to technical difficulties linked to the sampling strategy. The data provider may choose the opening/closing for X and Y, possibly traded on different markets, possibly on different time clocks. He or she may also sample points at certain trading times or events which are randomly chosen in a particular time window. This typically happens if daily data are considered. At a completely different level, if high-frequency data are concerned, trading times are genuinely random and non-synchronous. Our approach will simultaneously incorporate these different points of view.

1.2. Organisation of the paper

In Section 2, we present our stochastic model for describing the lead-lag effect. We start with the simplest Bachelier model with no drift in Section 2.1. The issue boils down to defining properly the lead-lag effect between two correlated Brownian motions. In Section 2.2, a general lead-lag model is presented for two dimensional process, for which the marginal processes are semi-martingales with locally bounded drift and continuous local martingale part, with properly defined diffusion coefficients.

We present our main result in Section 3. Section 3.1 gives a precise construction of the underlying statistical experiment with the corresponding assumptions on the observation sampling schemes. The estimation procedure is constructed in Section 3.2. Our estimator is robust to non-synchronous data and does not require any pre-processing of the data such as previous tick algorithm or so. In Section 3.3, Theorem 1 shows that the lead-lag parameter parameter between X and Y can be consistently estimated with n non-synchronous historical data over a fixed time horizon $[0,T]$. The convergence rate is governed by Δ_n, the maximal distance between two data points. The rate of convergence of our estimator is essentially Δ_n^{-1} and *not* $\Delta_n^{-1/2}$, as one would expect from a regular estimation problem in diffusion processes, see *e.g.* [8]. This comes from the underlying structure of the statistical model which is not regular and which shares some analogy with change point problems. However, the problem of the optimality of our procedure is left open, and might prove difficult in the context of nonsynchronous data.

Theorem 1 is good news as far as practical implementation is concerned and is further addressed in the discussion Section 3.4, appended with numerical illustrations on simulated data in Section 5 and on real data in Section 6. The proofs are delayed until Section 4 and an appendix (Section 7) contains auxiliary technical results.

2. The lead-lag model

2.1. The Bachelier model

In its simplest form, a lead-lag Bachelier model with no drift between two Brownian motion components can be described as follows.

On a filtered space $\big(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\big)$, we consider a two-dimensional $\mathbb{F}$-Brownian motion $B = (B^{(1)}, B^{(2)})$ such that $\langle B^{(1)}, B^{(2)}\rangle_t = \rho t$ for every $t \geq 0$ and for some $\rho \in [-1,1]$.

Let $T > 0$ be some terminal time, fixed throughout the paper. For $t \in [0,T]$, set

$$\left\{ \begin{array}{rcl} X_t & := & x_0 + \sigma_1 B_t^{(1)} \\ \widetilde{Y}_t & := & y_0 + \sigma_2 B_t^{(2)}, \end{array} \right.$$

where $x_0, y_0 \in \mathbb{R}$ and $\sigma_1 > 0$, $\sigma_2 > 0$ are given constants. The corresponding Black-Scholes version of this model is readily obtained by exponentiating X and $\widetilde{Y}$.

The simplest way for introducing a lead-lag effect between X and $\widetilde{Y}$ is to operate a time shift: let $\vartheta \in \mathbb{R}$ represent the lead or lag time between X and $\widetilde{Y}$. Let us suppose for simplicity that $\vartheta \geq 0$. Put

$$\tau_\vartheta(\widetilde{Y})_t := \widetilde{Y}_{t-\vartheta}, \quad t \in [\vartheta, T]. \tag{1}$$

Our lead-lag model is the two-dimensional process

$$\big(X, \tau_\vartheta(\widetilde{Y})\big) = \big(X_t, \tau_\vartheta(\widetilde{Y})_t\big)_{t\in[\vartheta,T]}.$$

Since we have $B_t^{(2)} = \rho B_t^{(1)} + (1-\rho^2)^{1/2} W_t$ with $W = (W_t)_{t\in[0,T]}$ a Brownian motion independent of $B^{(1)}$, we obtain the simple and explicit representation

$$\left\{ \begin{array}{lcl} X_t & = & x_0 + \sigma_1 B_t^{(1)}, \\ \tau_\vartheta(\widetilde{Y})_t & = & y_0 + \rho\,\sigma_2 B_{t-\vartheta}^{(1)} + \sigma_2(1-\rho^2)^{1/2}\, W_{t-\vartheta} \end{array} \right. \tag{2}$$

for $t \in [\vartheta, T]$. In this representation, the interpretation of the lead-lag parameter ϑ is transparent. Alternatively, if we start with a process (X, Y) having representation

$$(X, Y) = \big(X, \tau_\vartheta(\widetilde{Y})\big) \tag{3}$$

as in (2), the lead-lag interpretation between X and Y readily follows. Since $\vartheta \geq 0$, the sample path of X anticipates on the path of Y by a time shift ϑ and to an amount – measured in normalized variance – proportional to $\rho\,\sigma_2/\sigma_1$. In that case, we say that X is the leader and Y is the lagger.

For the case $\vartheta < 0$, we intertwine the roles of X and Y in the terminology.

Remark 1. *Note that, except in the case $\vartheta = 0$, the process $\big(X_t, Y_t\big)_{t\in[\vartheta,T]}$ is not an $\mathbb{F}$-martingale. However, each component is a martingale with respect to a different filtration: X is an $\mathbb{F}$-martingale and $Y = \tau_\vartheta(\widetilde{Y})$ is an $\mathbb{F}^\vartheta$-martingale, with $\mathbb{F}^\vartheta = (\mathcal{F}_t^\vartheta)_{t\geq\vartheta}$, and $\mathcal{F}_t^\vartheta = \mathcal{F}_{t-\vartheta}$. In the representation* (3)*, the process $\tau_\vartheta(\widetilde{Y})$ is not "observed" at time $t-\vartheta$ from the "point of view" of $\mathbb{F}$.*

2.2. Lead-lag between two semi-martingales

We generalize the lead-lag model (3) to semi-martingales with local-martingales components that can be represented as Itô local-martingales.

We need some notation. Let $T > 0$ be some terminal time and let $\delta > 0$ represent the maximum temporal lead-lag allowed for the model, fixed throughout the paper. On a probability space $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$, let $\mathbb{F} = \big(\mathcal{F}_t\big)_{t\in[-\delta,T+\delta]}$ be a filtration satisfying the usual conditions. We denote by $\mathbb{F}_{[a,b]} = \big(\mathcal{F}_t\big)_{t\in[a,b]}$ the restriction of $\mathbb{F}$ to the time interval $[a, b]$.

Definition 1. *The two-dimensional process $(X,Y)_{t\in[0,T+\delta]}$ is a regular semi-martingale with lead-lag parameter $\vartheta \in [0,\delta)$ if the following decomposition holds:*

$$X = X^c + A, \;\; Y = Y^c + B,$$

with the following properties:

- *The process $(X^c_t)_{t\in[0,T+\delta]}$ is a continuous $\mathbb{F}_{[0,T+\delta]}$-local martingale, and the process $(Y^c_t)_{t\in[0,T+\delta]}$ is a continuous $\mathbb{F}^{\vartheta}_{[0,T+\delta]}$-local martingale.*
- *The quadratic variations $\langle X^c\rangle_{t\in[0,T+\delta]}$ and $\langle Y^c\rangle_{t\in[0,T+\delta]}$ are absolutely continuous w.r.t. the Lebesgue measure and their Radon-Nikodym derivatives admit a locally bounded version.*
- *The drifts A and B have finite variation over $[0,T+\delta]$.*

Definition 2. *The two-dimensional process $(X,Y)_{t\in[0,T+\delta]}$ is a regular semi-martingale with lead-lag parameter $\vartheta \in (-\delta,0]$ if the same properties as in Definition 1 hold, with X and Y intertwined and ϑ replaced by $-\vartheta$.*

Remark 2. *If $(X,Y)_{t\in[0,T+\delta]}$ is a regular semi-martingale with lead-lag parameter $\vartheta \in [0,\delta)$, then the process $(\tau_{-\vartheta}(Y^c))_{t\in[-\vartheta,T]}$ is a continuous $\mathbb{F}_{[-\vartheta,T]}$-local martingale, with $\tau_{-\vartheta}(Y)_t = Y_{t+\vartheta}$ the (inverse of the) shift operator defined in (1).*

Remark 3. *If $(X,Y)_{t\in[0,T+\delta]}$ is a regular semi-martingale with lead-lag parameter $\vartheta \in [0,\delta)$, then the process $(Y,X)_{t\in[0,T+\delta]}$ is a regular semi-martingale with lead-lag parameter $-\vartheta$.*

3. Main result

3.1. The statistical model

We observe a two-dimensional price process (X,Y) at discrete times. The components X and Y are observed over the time horizon $[0,T+\delta]$. The following assumption is in force throughout:

Assumption A. *The process $(X,Y) = (X_t,Y_t)_{t\in[0,T+\delta]}$ is a regular semi-martingale with lead-lag parameter $\vartheta \in \Theta = (-\delta,\delta)$.*

The – possibly random – observation times are given by the following subdivisions of $[0,T+\delta]$:

$$\mathcal{T}^X := \left\{ s_{1,n_1} < s_{2,n_1} < \cdots < s_{n_1,n_1} \right\} \tag{4}$$

for X and

$$\mathcal{T}^Y := \left\{ t_{1,n_2} < t_{2,n_2} < \cdots < t_{n_2,n_2} \right\} \tag{5}$$

for Y, with $n_1 = n_2$ or not. For simplicity, we assume $s_{1,n_1} = t_{1,n_2} = 0$ and $s_{n_1,n_1} = t_{n_2,n_2} = T + \delta$. The sample points are either chosen by the statistician or dictated for practical convenience by the data provider. They are usually neither equispaced in time nor synchronous, and may depend on the values of X and Y.

For some unknown $\vartheta \in \Theta := (-\delta, \delta)$, the process (X, Y) is a regular semi-martingale with lead-lag parameter ϑ, and we want to estimate ϑ based on the set of historical data

$$\big\{X_s,\ s \in \mathcal{T}^X\big\} \cup \big\{Y_t,\ t \in \mathcal{T}^Y\big\}. \tag{6}$$

In order to describe precisely the property of the sampling scheme $\mathcal{T}^X \cup \mathcal{T}^Y$, we need some notation, that we borrow from Hayashi and Yoshida [13]. The subdivision $\mathcal{T}^X$ introduced in (4) is mapped into a family of intervals

$$\mathcal{I} = \big\{I = (\underline{I}, \overline{I}] = (s_{i,n_1}, s_{i+1,n_1}],\ \ i = 1, \dots, n_1 - 1\big\}. \tag{7}$$

Likewise, the subdivision $\mathcal{T}^Y$ defined in (5) is mapped into

$$\mathcal{J} = \big\{J = (\underline{J}, \overline{J}] = (t_{j,n_2}, s_{j+1,n_2}],\ \ j = 1, \dots, n_2 - 1\big\}.$$

We will systematically employ the notation I (resp. J) for an element of $\mathcal{I}$ (resp. $\mathcal{J}$). We set

$$\Delta_n := \max\big\{\sup\big\{|I|, I \in \mathcal{I}\big\}, \sup\{|J|, J \in \mathcal{J}\}\big\},$$

where $|I|$ (resp. $|J|$) denotes the length of the interval I (resp. J) and n is a parameter tending to infinity.

Remark 4. *One may think of n being the number of data points extracted from the sampling, i.e. $n = \sharp\mathcal{I} + \sharp\mathcal{J}$. However, as we will see, only the (random) quantity Δ_n will prove relevant for measuring the accuracy of estimation of the lead-lag parameter.*

The assumptions on the sampling scheme is the following.

Assumption B.

[B1] *There exists a deterministic sequence of positive numbers v_n such that $v_n < \delta$ and $v_n \to 0$ as $n \to \infty$. Moreover*

$$v_n^{-1}\, \Delta_n \to 0$$

in probability as $n \to \infty$.

[B2] *For all $I \in \mathcal{I}$, the random times $\underline{I}$ and $\overline{I}$ are $\mathbb{F}^{v_n}$-stopping times if $\vartheta \geq 0$ (respectively $\mathbb{F}^{-\vartheta+v_n}$-stopping times if $\vartheta < 0$). For all $J \in \mathcal{J}$, the random times $\underline{J}$ and $\overline{J}$ are $\mathbb{F}^{\vartheta+v_n}$-stopping times if $\vartheta \geq 0$ (respectively $\mathbb{F}^{v_n}$-stopping times if $\vartheta < 0$).*

[B3] *There exists a finite grid $\mathcal{G}^n \subset \Theta$ such that $0 \in \mathcal{G}^n$ and*

- *For some $\gamma > 0$, we have $\sharp\,\mathcal{G}^n = O(v_n^{-\gamma})$.*

– *For some deterministic sequence $\rho_n > 0$, we have*

$$\bigcup_{\widetilde{\vartheta}\in\mathcal{G}^n} [\widetilde{\vartheta} - \rho_n, \widetilde{\vartheta} + \rho_n] \supset \Theta$$

and

$$\lim_{n\to\infty} \rho_n \min\left\{\mathbb{E}\big[\sharp\mathcal{I}\big], \mathbb{E}\big[\sharp\mathcal{J}\big]\right\} \to 0.$$

Remark 5. *Since both $\mathbb{E}\big[\sharp\mathcal{I}\big]$ and $\mathbb{E}\big[\sharp\mathcal{J}\big]$ diverge at rate no less than v_n^{-1}, Assumption B3 implies that $\rho_n = o(v_n)$. With no loss of generality, we thus may (and will) assume that $\rho_n \leq v_n$ for all n.*

3.2. The estimation procedure

Preliminaries

Assume first that the data arrive at regular and synchronous time stamps, *i.e.* we have data

$$(X_0, Y_0), (X_{\Delta_n}, Y_{\Delta_n}), (X_{2\Delta_n}, Y_{2\Delta_n}), \dots$$

over the time interval $[0, T]$. For every integer $k \in \mathbb{Z}$, we form the shifted time series

$$Y_{(k+i)\Delta_n}, \quad i = 1, 2, \dots$$

for every i such that $(k+i)\Delta_n$ is an admissible time stamp[1]. We can then construct the empirical covariation estimator

$$\mathcal{C}_n(k) := \sum_i \big(X_{i\Delta_n} - X_{(i-1)\Delta_n}\big)\big(Y_{(i+k)\Delta_n} - Y_{(i+k-1)\Delta_n}\big).$$

where the sum in i expands over all relevant data points. Over the time interval $[0, T]$, the number of elements used for the computation of $\mathcal{C}_n(k)$ should be of order Δ_n^{-1} as $n \to \infty$.

Assume further for simplicity that the process (X, Y) is a lead-lag Bachelier model in the sense of Section 2.1, with lead-lag parameter $\vartheta = k_0\Delta_n$. Heuristically, we have

$$\mathcal{C}_n(k) \approx \Delta_n^{-1}\mathbb{E}\big[(X_\bullet - X_{\bullet-\Delta_n})(Y_{\bullet+k\Delta_n} - Y_{\bullet+(k-1)\Delta_n})\big] + \Delta_n^{1/2}\xi^n,$$

as $\Delta_n \to 0$, where the noise remainder term ξ^n is a tight sequence by the central limit theorem. Now, thanks to the representation (2) we readily derive

$$\Delta_n^{-1}\mathbb{E}\big[(X_\bullet - X_{\bullet-\Delta_n})(Y_{\bullet+k\Delta_n} - Y_{\bullet+(k-1)\Delta_n})\big] = \begin{cases} 0 & \text{if} \quad k \neq k_0 \\ \rho\,\sigma_1\sigma_2 & \text{if} \quad k = k_0. \end{cases}$$

[1]Possibly, we end up with an empty data set.

Provided $\Delta_n^{1/2} \ll \rho\,\sigma_1\sigma_2$ as $n \to \infty$, we can detect asymptotically the value k_0 that defines ϑ in the very special case $\vartheta = k_0\Delta_n$ by maximizing in k the contrast sequence

$$k \rightsquigarrow \big|\mathcal{C}_n(k)\big|.$$

This is the essence of our method. For an arbitrary ϑ, we can anticipate that an approximation of ϑ taking the form $k_0\Delta_n$ would add an extra error term of the order of the approximation, *i.e.* Δ_n, which is a first guess for an achievable rate of convergence.

In a general context of regular semi-martingales with lead-lag effect, sampled at (random) non-synchronous data points, we consider the Hayashi-Yoshida (later abbreviated by HY) covariation estimator and modify it with an appropriate time shift on one component. We maximize the resulting empirical covariation estimator with respect to the time shift over an appropriate grid.

Construction of the estimator

We need some notation. If $H = (\underline{H}, \overline{H}]$ is an interval, for $\vartheta \in \Theta$, we define the shift interval $H_\vartheta := H + \vartheta = (\underline{H} + \vartheta, \overline{H} + \vartheta]$. We write

$$X(H)_t := \int_0^t 1_H(s)dX_s$$

for a (possibly random) interval, such that $s \rightsquigarrow 1_H(s)$ is an elementary predictable process. Also, for notational simplicity, we will often use the abbreviation

$$X(H) := X(H)_{T+\delta} = \int_0^{T+\delta} 1_H(s)dX_s.$$

For $\vartheta \in \Theta$, the shifted HY covariation contrast is defined as

$$\begin{aligned}\mathcal{U}^n(\vartheta) :=& 1_{\vartheta\geq 0} \sum_{I\in\mathcal{I},J\in\mathcal{J},\overline{I}\leq T} X(I)Y(J)1_{\{I\cap J_{-\vartheta}\neq\emptyset\}} \\ &+ 1_{\vartheta<0} \sum_{I\in\mathcal{I},J\in\mathcal{J},\overline{J}\leq T} X(I)Y(J)1_{\{J\cap I_{\vartheta}\neq\emptyset\}}.\end{aligned}$$

Our estimator $\widehat{\vartheta}_n$ is obtained by maximizing the contrast $\vartheta \rightsquigarrow \big|\mathcal{C}^n(\vartheta)\big|$ over the finite grid $\mathcal{G}^n$ constructed in Assumption [B3] in Section 3.1 above.

Eventually, $\widehat{\vartheta}_n$ is defined as a solution of

$$\big|\mathcal{U}^n(\widehat{\vartheta}_n)\big| = \max_{\vartheta\in\mathcal{G}^n} \big|\mathcal{U}^n(\vartheta)\big|. \tag{8}$$

3.3. Convergence result

Since $\tau_{-\vartheta}(Y^c)$ is a $\mathbb{F}$-local martingale, the quadratic variation process $\langle X^c, \tau_{-\vartheta}(Y^c)\rangle$ is well defined. We are now ready to assess our main result:

Theorem 1. *Work under Assumptions A and B. The estimator* $\widehat{\vartheta}_n$ *defined in* (8) *satisfies*

$$v_n^{-1}\big(\widehat{\vartheta}_n - \vartheta\big) \to 0$$

in probability, on the event $\big\{\langle X^c, \tau_{-\vartheta}(Y^c)\rangle_T \neq 0\big\}$, *as* $n \to \infty$.

3.4. Discussion

Covariation estimation of non-synchronous data

The estimation of the covariation between two semi-martingales from discrete data from non-synchronous observation times has some history. It was first introduced by Hayashi and Yoshida [13] and subsequently studies in various related contexts by several authors. A comprehensive list of references include: Malliavin and Mancino [21], Hayashi and Yoshida [13], [15], [12], [14], [16], Hayashi and Kusuoka [11], Ubukata and Oya [28], Hoshikawa *et al.* [17] and Dalalyan and Yoshida [7].

About the rate of convergence

The condition $\Delta_n = o(v_n)$ of Assumption [B1] is needed for technical reason, in order to manage the fact that Δ_n is random in general. In the case of regular sampling $\Delta_n = 1/n$, the obtained rate n^{-1} is substantially better than the usual $n^{-1/2}$-rate of regular parametric statistical model. This is due to the fact that the estimation of the lead-lag parameter is rather a change-point detection problem, see [18] for a general reference for the structure of parametric models. A more detailed analysis of the contrast function shows that its limit is not regular (not differentiable in the ϑ-variable) and this explains the presence of the rate n^{-1}. However, the optimality of our procedure is not granted and the rate Δ_n could presumably be improved in certain special situations.

At this stage, it would be interesting to investigate further the asymptotic distribution of the normalized error of the estimator $\widehat{\vartheta}$, for deterministic sampling in the simple Bachelier model. The limit, if it exists, is presumably non-Gaussian. We intend to describe this in a forthcoming work. In the general case of random sampling, the study of an asymptotic distribution of the normalized error of $\widehat{\vartheta}$ seems out of reach with the techniques used in the present paper.

Lead-lag effect and arbitrage

As stated, the lead-lag model for the two-dimensional process (X, Y) is not a semi-martingale, unless one component is appropriately shifted in time. This is *not compatible* in principle with the dominant theory of no-arbitrage models. This kind of modeling however seems to have some relevance in practice, and there is a natural way to reconcile both points of view.

We focus for example on the simplest Bachelier model of Section 2.1. We show in this paper that the lead-lag parameter ϑ can almost be identified in principle. Consequently, the knowledge of ϑ can then be incorporated into a trading strategy. If $\vartheta \neq 0$, we can

obtain in principle some *statistical arbitrage*, in the sense that we can find, in the Bachelier model without drift, a self financing portfolio of assets X and $\tau_{-\vartheta}(Y)$ with initial value zero and whose expectation at time T is positive. See also [9] for a general formulation of arbitrage in this context.

This statistical arbitrage can be erased by introducing further trading constraints such as a maximal trading frequency and transaction cost (slippage, execution risk and so on). In this setting, we can no longer guarantee a statistical arbitrage. Moreover, we may certainly incorporate risk constraints in order to define an admissible strategy.

This outlines that although we perturb the semi-martingale classical approach, our lead-lag model is compatible in principle with non statistical arbitrage constraints, under refined studies of risk profiles. We intend to set out in details this possibilities in a forthcoming work.

Microstructure noise

Our model does not incorporate microstructure noise. This is reasonable if Δ_n is thought of on a daily basis, say (if T is of the order of a year or more say), but is inconsistent in a high-frequency setting where T is of the order of one day. In that context, efficient semi-martingale prices of the assets are subject to the so-called microstructure noise, see among others Aït-Sahalia *et al.* [1], Bandi and Russell [2], Barndorff-Nielsen*et al.* [3], Hansen and Lunde [10], Jacod *et al.* [19], Rosenbaum [23], [24]. In [25] and [26], Robert and Rosenbaum introduce a model (model with uncertainty zones) where the efficient semi-martingale prices of the assets can be estimated at some random times from the observed prices. In particular, it is proved that the usual Hayashi-Yoshida estimator is consistent in this microstructure noise context as soon as it is computed using the estimated values of the efficient prices. Using the same approach, that is applying the lead-lag estimator to the estimated values of the efficient prices, one can presumably build an estimator which is robust to microstructure noise.

How to use high frequency data in practice

Nevertheless, when high frequency data are considered, we propose a simple pragmatic methodology that allows to implement our lead-lag estimation procedure without requiring the relatively involved data pre-processing suggested in the previous paragraph. A preliminary inspection of the signature plot in trading time – the realized volatility computed with different subsampling values for the trading times – enables to select a coarse subgrid among the trading times where microstructure noise effects can be neglected. Thanks to the non-synchronous character of high-frequency data, we can take advantage of this subsampling in trading time and obtain accurate estimation of the lead-lag parameter, at a scale that is significantly smaller than the average mesh size of the coarse grid itself. This would not be possible with a regular subsampling in calendar, time where the price at time t would be defined as the last traded price before t. This empirical approach is developed in the numerical illustration Section 6 on real data, in the particular case of measuring lead-lag between the future contract on Dax (FDAX) and the Euro-Bund future contract (FGBL) with same maturities.

Extension of the model

We consider this work as a first – and relatively simple – attempt for modeling the lead-lag effect in continuous time models. As a natural extension, it would presumably be more reasonable to consider more intricate correlations between assets in the model. For example, one could add a common factor in the two assets, without lead-lag effect, as suggested by the empirical study of Section 6. Through this, in addition to the "lead-lagged correlation", one would also obtain an instantaneous correlation between the assets. In order to estimate the lead-lag parameter in this context, one would presumably be required to consider local maxima of the contrast function we develop here. Such a development is again left out for future work.

4. Proof of Theorem 1

The proof of Theorem 1 is split in four parts. In the first three parts we work under supplementary assumptions on the processes and the parameter space (Assumption $\widetilde{A}$). We first show that if we compute the contrast function over points ϑ_n of the grid $\mathcal{G}^n$ such that the order of magnitude of $|\vartheta_n - \vartheta|$ is bigger than v_n, then the contrast function goes to zero (Proposition 1). Then we prove that, on the contrary, if the order of magnitude of $|\vartheta_n - \vartheta|$ is essentially smaller than v_n, then the contrast function goes to the covariation between X and $\tau_{-\vartheta}(Y)$ (Proposition 2). We put these two results together in the third part which ends the proof of Theorem 1 under the supplementary assumptions. The proof under the initial assumptions is given in the last part.

4.1. Preliminaries

Supplementary assumptions. For technical convenience, we will first prove Theorem 1 when the sign of ϑ is known and when the components X and Y are local martingales. Moreover, we introduce a localization tool. The quadratic variation processes of X and Y admitting locally bounded derivatives, there exists a sequence of stopping times tending almost surely to $T+\delta$ such that the associated stopped processes are bounded by deterministic constants. Since Theorem 1 is a convergence in probability result, we can without loss of generality work under the supplementary assumption that the quadratic variation processes are bounded over $[0, T+\delta]$. Therefore, we add-up the following restrictions:

Assumption $\widetilde{\mathbf{A}}$. *We have Assumption A and*

- **[$\widetilde{\mathbf{A}}$1]** *There exists $L > 0$ such that $\langle X \rangle'_{T+\delta} \leq L$ and $\langle Y \rangle'_{T+\delta} \leq L$.*
- **[$\widetilde{\mathbf{A}}$2]** *The parameter set is restricted to $\Theta = [0, \delta)$. Consequently, by $\mathcal{G}^n$ we mean here $\mathcal{G}^n \cap [0, \delta)$.*
- **[$\widetilde{\mathbf{A}}$3]** *$X = X^c$ and $Y = Y^c$.*

Notation. We now introduce further notation. For $I \in \mathcal{I}$ and $J \in \mathcal{J}$, let

$$\underline{I^n} = \underline{I} \wedge \inf\big\{t,\ \max_{I'}\big\{\overline{I'} \wedge t - \underline{I'} \wedge t\big\} \geq v_n\big\} \wedge T.$$

and

$$\underline{J^n} = \underline{J} \wedge \inf\big\{t,\ \max_{J'}\big\{\overline{J'} \wedge t - \underline{J'} \wedge t\big\} \geq v_n\big\} \wedge (T+\delta).$$

We define $\overline{I^n}$ and $\overline{J^n}$ in the same way for $\overline{I}$ and $\overline{J}$ respectively. Let $I^n = (\underline{I^n}, \overline{I^n}]$ and $J^n = (\underline{J^n}, \overline{J^n}]$.

Remark 6. *We have the following interpretation of $\underline{I^n}$ and $\overline{I^n}$: let τ^n denote the first time for which we know that an interval I will have a width that is larger than v_n. Then we keep only the $\underline{I}$ and $\overline{I}$ that are smaller than τ^n. If $\tau^n \leq T$, we also consider τ^n among the observation times. Note that τ^n is not a true observation time in general. However, this will not be a problem since the set where Δ_n is bigger than v_n will be asymptotically negligible. Obviously $\underline{I^n}$ and $\overline{I^n}$ are $\mathbb{F}$-stopping times, and $\underline{J^n}$ and $\overline{J^n}$ are $\mathbb{F}^{\vartheta+v_n}$-stopping times.*

Finally, for two intervals $H = (\underline{H}, \overline{H}]$ and $H' = (\underline{H'}, \overline{H'}]$, we define

$$K\big(H, H'\big) := 1_{H \cap H' \neq \emptyset}.$$

4.2. The contrast function

We consider here the case where the order of magnitude of $|\vartheta_n - \vartheta|$ is bigger than v_n. We first need to give a preliminary lemma that will ensure that the quantities we will use in the following are well defined.

Lemma 1. *Work under Assumption [B2], under the slightly more general assumption that for all $I = (\underline{I}, \overline{I}] \in \mathcal{I}$, the random variables $\underline{I}$ and $\overline{I}$ are $\mathbb{F}$-stopping times. Suppose that $\widetilde{\vartheta} \geq \vartheta + \varepsilon_n$ and $2v_n \leq \varepsilon_n$. Then for any random variable X' measurable w.r.t. $\mathcal{F}_{\overline{I^n}}$, the random variable $X'K(I^n_{\widetilde{\vartheta}}, J^n)$ is $\mathcal{F}^{\vartheta}_{\underline{J^n}}$-measurable. In particular, $f(\overline{I^n})X(I^n)K(I^n_{\widetilde{\vartheta}}, J^n)$ is $\mathcal{F}^{\vartheta}_{\underline{J^n}}$-measurable for any measurable function f.*

The proof of Lemma 1 is given in appendix. It is important to note that Lemma 1 implies that for $\widetilde{\vartheta} \geq \vartheta + \varepsilon_n$ and $2v_n \leq \varepsilon_n$, the random variable

$$1_{\{\overline{I^n} \leq T\}} X(I^n) K(I^n_{\widetilde{\vartheta}}, J^n) 1_{J^n}(s)$$

is $\mathcal{F}^{\vartheta}_s$-measurable. Indeed, $1_{J^n}(s)$ is $\mathcal{F}^{\vartheta}_s$ and $1_{J^n}(s) = 1$ implies $s \geq \underline{J^n}$. We now introduce a functional version of $\mathcal{U}^n$ by considering the random process

$$\mathbb{U}^n(\widetilde{\vartheta})_t := \sum_{I \in \mathcal{I}, J \in \mathcal{J}, \overline{I^n} \leq T} X(I^n) Y(J^n)_t K(I^n_{\widetilde{\vartheta}}, J^n).$$

We are now able to give the main proposition for the vanishing of the contrast function.

Proposition 1. *Let* $\varepsilon_n = 2v_n$, $\mathcal{G}_+^n = \{\widetilde{\vartheta} \in \mathcal{G}^n, \widetilde{\vartheta} \geq \vartheta + \varepsilon_n\}$ *and* $\mathcal{G}_-^n = \{\widetilde{\vartheta} \in \mathcal{G}^n, \widetilde{\vartheta} \leq \vartheta - \varepsilon_n\}$. *We have*

$$\max_{\widetilde{\vartheta} \in \mathcal{G}_+^n \cup \mathcal{G}_-^n} |\mathbb{U}^n(\widetilde{\vartheta})_{T+\delta}| \to 0,$$

in probability.

Proof. Assume first $\widetilde{\vartheta} \geq \vartheta + \varepsilon_n$. Thanks to Lemma 1, we obtain a martingale representation of the process $\mathbb{U}^n(\widetilde{\vartheta})$ that takes the form

$$\mathbb{U}^n(\widetilde{\vartheta})_t = \sum_{I \in \mathcal{I}, J \in \mathcal{J}} \int_0^t 1_{\{\overline{I^n} \leq T\}} X(I^n) K(I^n_{\widetilde{\vartheta}}, J^n) 1_{J^n}(s) dY_s,$$

where the stochastic integral with respect to Y is taken for the filtration $\mathbb{F}^\vartheta$. As a result, the $\mathbb{F}^\vartheta$-quadratic variation of $\mathbb{U}^n$ is given by

$$\langle \mathbb{U}^n(\widetilde{\vartheta}) \rangle_t = \int_0^t \Big(\sum_{I \in \mathcal{I}, J \in \mathcal{J}} 1_{\{\overline{I^n} \leq T\}} X(I^n) K(I^n_{\widetilde{\vartheta}}, J^n) 1_{J^n}(s) \Big)^2 d\langle Y \rangle_s.$$

Using that the intervals J^n are disjoint, we obtain

$$\langle \mathbb{U}^n(\widetilde{\vartheta}) \rangle_t = \int_0^t \sum_{J \in \mathcal{J}} \Big(\sum_{I \in \mathcal{I}} 1_{\{\overline{I^n} \leq T\}} X(I^n) K(I^n_{\widetilde{\vartheta}}, J^n) \Big)^2 1_{J^n}(s) d\langle Y \rangle_s.$$

For a given interval J^n, the union of the intervals I^n that have a non empty intersection with J^n is an interval of width smaller than $3v_n$. Indeed, the maximum width of J^n is v_n and add to this (if it exists) the width of the interval I^n such that $\underline{I^n} \leq \overline{J^n}$, $\overline{I^n} \geq \overline{J^n}$ and the width of the interval I^n such that $\underline{I^n} \leq \underline{J^n}$, $\overline{I^n} \geq \underline{J^n}$. Thus,

$$\begin{aligned} \sum_{I \in \mathcal{I}} 1_{\{\overline{I^n} \leq T\}} X(I^n) K(I^n_{\widetilde{\vartheta}}, J^n) &\leq \sup_{s \leq T} \sup_{0 \leq u \leq 3v_n} |X_{(s+u) \wedge T} - X_u| \\ &\leq 2 \max_{1 \leq k \leq \lfloor (3v_n)^{-1} T \rfloor} \sup_{t \in [3v_n(k-1), 3v_n k]} |X_{t \wedge T} - X_{3v_n(k-1)}|. \end{aligned}$$

Consequently, we obtain for every $t \in [0, T+\delta]$ and $\widetilde{\vartheta} \in [\vartheta + \varepsilon_n, \delta]$,

$$\langle \mathbb{U}^n(\widetilde{\vartheta}) \rangle_t \leq 4L(T+\delta) \max_{1 \leq k \leq \lfloor (3v_n)^{-1} T \rfloor} \sup_{t \in [3v_n(k-1), 3v_n k]} |X_{t \wedge T} - X_{3(k-1)v_n}|^2.$$

For every $p > 1$, it follows from the Bürkholder-Davis-Gundy inequality that

$$\mathbb{E}\big[|\mathbb{U}^n(\widetilde{\vartheta})_{T+\delta}|^{2p}\big] \lesssim \sum_{k=1}^{\lfloor (3v_n)^{-1} T \rfloor} \mathbb{E}\big[\sup_{t \in [3v_n(k-1), 3v_n k]} |X_{t \wedge T} - X_{3(k-1)v_n}|^{2p} \big] \lesssim v_n^{p-1},$$

where the symbol $\lesssim$ means inequality in order, up to constant that does not depend on n. Pick $\varepsilon > 0$. We derive

$$\begin{aligned}\mathbb{P}\big[\max_{\widetilde{\vartheta}\in\mathcal{G}_+^n}|\mathbb{U}^n(\widetilde{\vartheta})_{T+\delta}| > \varepsilon\big] &\leq \varepsilon^{-2p}\sum_{\widetilde{\vartheta}\in\mathcal{G}_+^n}\mathbb{E}\big[|\mathbb{U}^n(\widetilde{\vartheta})_{T+\delta}|^{2p}\big]\\ &\lesssim v_n^{p-1}\sharp\,\mathcal{G}_+^n \to 0\end{aligned}$$

as $n\to\infty$, provided $p > \gamma + 1$ where γ is defined in Assumption [B3]. A choice that is obviously possible.

The same argument holds for the case $\widetilde{\vartheta} \leq \vartheta - \varepsilon_n$, but with an X-integral representation in that latter case. The result follows.

□

4.3. Stability of the HY estimator

We consider now the case where the order of magnitude of $|\vartheta_n - \vartheta|$ is essentially smaller than v_n. We have the following proposition

Proposition 2. *Work under Assumptions $\widetilde{A}$ and B. For any sequence ϑ_n in $[0,\delta)$ such that $\vartheta_n \leq \vartheta$ and $|\vartheta_n - \vartheta| \leq \rho_n$ (remember that ρ_n is defined in Assumption B3), we have*

$$\mathcal{U}^n(\vartheta_n) \to \langle X, \tau_{-\vartheta}(Y)\rangle_{[0,T]}, \tag{9}$$

in probability as $n\to\infty$.

Proof. The proof goes into several steps.

Step 1. In this step, we show that our contrast function can be regarded as the Hayashi-Yoshida estimator applied to X and to the properly shifted values of Y plus a remainder term. If $\vartheta = 0$, then $\vartheta_n = 0$ and Proposition 2 asserts nothing but the consistency of the standard HY-estimator, see Hayashi and Yoshida [13] and Hayashi and Kusuoka [11]. Thus we may assume $\vartheta > 0$.

By symmetry, we only need to consider the case where

$$\mathbb{E}\big[\sharp\mathcal{I}\big] \geq \mathbb{E}\big[\sharp\mathcal{J}\big].$$

Set $\delta_n = \vartheta_n - \vartheta$, $\widetilde{Y}_t = \tau_{-\vartheta}(Y)_t$ and $\widetilde{J}^n = J^n_{-\vartheta}$ and

$$\mathbb{U}^n(\vartheta_n) = \sum_{I\in\mathcal{I},J\in\mathcal{J},\overline{I}\leq T} X(I^n)Y(J^n)1_{\{I^n\cap J^n_{-\vartheta_n}\neq\emptyset\}}.$$

We then have

$$\mathbb{U}^n(\vartheta_n) = \sum_{I\in\mathcal{I},J\in\mathcal{J},\overline{I}\leq T} X(I^n)\widetilde{Y}(\widetilde{J}^n)1_{\{I^n\cap \widetilde{J}^n_{-\delta_n}\neq\emptyset\}}.$$

This can be written $\mathcal{V}^n + \mathcal{R}^n$ with

$$\mathcal{V}^n = \sum_{I\in\mathcal{I},J\in\mathcal{J},\overline{I}\le T} X(I^n)\widetilde{Y}(\widetilde{J}^n_{-\delta_n})1_{\{I^n\cap\widetilde{J}^n_{-\delta_n}\neq\emptyset\}}$$
$$\mathcal{R}^n = \sum_{I\in\mathcal{I},J\in\mathcal{J},\overline{I}\le T} X(I^n)\big\{\widetilde{Y}(\widetilde{J}^n) - \widetilde{Y}(\widetilde{J}^n_{-\delta_n})\big\}1_{\{I^n\cap\widetilde{J}^n_{-\delta_n}\neq\emptyset\}}.$$

Remark that $\widetilde{Y}(\widetilde{J}^n)$ and $\widetilde{Y}(\widetilde{J}^n_{-\delta_n})$ are well defined since $\widetilde{Y}$ is defined on $[-\vartheta, T]$ and $\vartheta_n \le \vartheta$. For every $J \in \mathcal{J}$, $\overline{J^n}$ is a $\mathbb{F}^{\vartheta+v_n}$-stopping time, therefore $\overline{\widetilde{J}^n_{-\delta_n}} = \overline{J^n_{-\vartheta_n}}$ is a $\mathbb{F}^{v_n-\delta_n}$-stopping time, and a $\mathbb{F}$-stopping time as well. Thus $\mathcal{V}^n$ is a variant of the HY-estimator: more precisely,

$$\widetilde{\mathcal{V}}^n := \sum_{I,J:\overline{I}\le T} X(I^n)\widetilde{Y}(\widetilde{J}^n_{-\delta_n}\cap\mathbb{R}_+)1_{\{I^n\cap(\widetilde{J}^n_{-\delta_n}\cap\mathbb{R}_+)\neq\emptyset\}}$$

is the original HY-estimator and we have $\mathcal{V}^n - \widetilde{\mathcal{V}}^n \to 0$ in probability as $n\to\infty$. It follows that $\widetilde{\mathcal{V}}^n \to \langle X, \widetilde{Y}\rangle_T$ in probability as $n\to\infty$, see [13] and [11].

Step 2. Before turning to the term $\mathcal{R}^n$, we give a technical lemma and explain a simplifying procedure. For an interval $I = [\underline{I}, \overline{I}) \in \mathcal{I}$, set

$$M^I := \sup\Big\{\overline{\widetilde{J}^n_{-\delta_n}},\ J\in\mathcal{J},\ \overline{\widetilde{J}^n_{-\delta_n}} \le \overline{I^n}\Big\}.$$

Note that if we consider the interval J at the extreme left end of the family $\mathcal{J}$, we have, for large enough n

$$\overline{\widetilde{J}^n_{-\delta_n}} \le v_n - \vartheta - \delta_n < -\frac{\vartheta}{2}$$

say, so we may assume that the set over which we take the supremum is non empty.

Lemma 2. *Work under Assumption [B2]. The random variables M^I are $\mathbb{F}$-stopping times.*

The proof of this lemma is given in appendix. We now use a simplifying operation. For each I^n, we merge all the J^n such that $\widetilde{J}^n_{-\delta_n} \subset I^n$. We call this procedure Π-reduction. The Π-reduction produces a new sequence of increasing random intervals extracted from the original sequence $(\widetilde{J}^n_{-\delta_n})$, which are $\mathbb{F}$-predictable by Lemma 2. More precisely, the end points are $\mathbb{F}$-stopping times. It is important to remark that the Π-reduction implies that there are at most two points of type J between any $\underline{I^n}$ and $\overline{I^n}$. Moreover, since $\mathcal{R}^n$ is a bilinear form of the increments of X and $\widetilde{Y}$, it is invariant under Π-reduction. Likewise for the maximum length Δ_n. Thus, without loss of generality, we may assume that the $\widetilde{J}^n_{-\delta_n}$ are Π-reduced.

Step 3. We now turn to $\mathcal{R}^n$. We write

$$I^n(\widetilde{J}^n_{-\delta_n}) = \bigcup_{I\in\mathcal{I},\overline{I}\le T,\{I^n\cap\widetilde{J}^n_{-\delta_n}\neq\emptyset\}} I^n.$$

We have

$$|\mathcal{R}^n| \le \sum_{J\in\mathcal{J}} |\widetilde{Y}(\widetilde{J}^n) - \widetilde{Y}(\widetilde{J}^n_{-\delta_n})||X\big(I^n(\widetilde{J}^n_{-\delta_n})\big)|.$$

We now index the intervals $\widetilde{J}^n$ by j and set $\widetilde{J}^n_j = \{0\}$ if $j > \sharp\{J\}$. Thus, the preceding line can be written

$$|\mathcal{R}^n| \le \sum_{j} |\widetilde{Y}(\widetilde{J}^n_j) - \widetilde{Y}(\widetilde{J}^n_{-\delta_n,j})||X\big(I^n(\widetilde{J}^n_{-\delta_n,j})\big)|.$$

Then, Cauchy-Schwarz inequality gives that $(\mathbb{E}[|\mathcal{R}^n|])^2$ is smaller than

$$\sum_j \mathbb{E}[|\widetilde{Y}(\widetilde{J}^n_j) - \widetilde{Y}(\widetilde{J}^n_{-\delta_n,j})|^2] \sum_j \mathbb{E}[|X\big(I^n(\widetilde{J}^n_{-\delta_n,j})\big)|^2].$$

We easily get that

$$\sum_j \mathbb{E}[|\widetilde{Y}(\widetilde{J}^n_j) - \widetilde{Y}(\widetilde{J}^n_{-\delta_n,j})|^2] \lesssim \delta_n \sharp\mathcal{J}$$

and we claim that (see next step)

$$\sum_j \mathbb{E}[|X\big(I^n(\widetilde{J}^n_{-\delta_n,j})\big)|^2] \lesssim 1. \tag{10}$$

Since $\delta_n \le \rho_n$, Proposition 2 readily follows.

Step 4. It remains to prove (10). Here we extend $(X_t)_{t\in\mathbb{R}_+}$ as $X_s = 0$ for $s < 0$, and denote the extended one by the same "X". This extension is just for notational convenience, and causes no problem because in what follows, we use the martingale property of X only over the time interval $\mathbb{R}_+$. For ease of notation, we also stop writing the index j for the intervals. We begin with the following remark. Take an interval $\widetilde{J}^n_{-\delta_n}$, say $(J_1, J_2]$ and $I^n(\widetilde{J}^n_{-\delta_n})$ associated, say $(I_1, I_2]$. Call J_0 the last observation point of type J occurring before J_1 and J_{-1} the last observation point of type J occurring before J_0. Two situations are possible:
- If there is no observation point of type I between I_1 and I_2, then, if it exists, J_0 is necessarily before I_1. If it does not exist, we have $J_1 \le v_n$.
- If there are some observation points of type I between I_1 and I_2, then J_0 might also be between I_1 and I_2. However, thanks to the Π-reduction, we know that J_{-1} is necessarily smaller than I_1. Consequently, we have that $|X\big(I^n(\widetilde{J}^n_{-\delta_n})\big)|$ is smaller than

$$\sup_{t\in[\underline{\widetilde{J}^{n,-1}_{-\delta_n}},\overline{I^{n,-1}}]} |X_t - X_{\underline{\widetilde{J}^{n,-1}_{-\delta_n}}}| + \sup_{t\in[\underline{\widetilde{J}^{n,-2}_{-\delta_n}},\overline{I^{n,-2}}]} |X_t - X_{\underline{\widetilde{J}^{n,-2}_{-\delta_n}}}| + \sup_{t\in[\underline{\widetilde{J}^{n}_{-\delta_n}},\overline{I^{n}_{+}}]} |X_t - X_{\underline{\widetilde{J}^{n}_{-\delta_n}}}|,$$

where we used the following notation:

- I^n_+ is the first interval I^n such that $\overline{I^n}$ exits to the right of $\widetilde{J}^n_{-\delta_n}$.
- $\widetilde{J}^{n,-1}_{-\delta_n}$ denotes the interval of the form $\widetilde{J}^n_{-\delta_n}$ which is the nearest neighbor to $\widetilde{J}^n_{-\delta_n}$ on the left.
- $\widetilde{J}^{n,-2}_{-\delta_n}$ denotes the interval of the form $\widetilde{J}^n_{-\delta_n}$ which is the nearest neighbor to $\widetilde{J}^{n,-1}_{-\delta_n}$ on the left.
- $I^{n,-1}$ is the first exit time to the right of $\widetilde{J}^{n,-1}_{-\delta_n}$ among the I^n.
- $I^{n,-2}$ is the first exit time to the right of $\widetilde{J}^{n,-2}_{-\delta_n}$ among the I^n.
- For $k=1,2$, if $\widetilde{J}^{n,-k}_{-\delta_n}$ is not defined, $\sup_{t\in[\underline{\widetilde{J}^{n,-k}_{-\delta_n}},\overline{I^{n,-k}}]}|X_t - X_{\underline{\widetilde{J}^{n,-k}_{-\delta_n}}}| = 0$.

Hence we obtain

$$\sum_j \mathbb{E}[|X\big(I^n(\widetilde{J}^n_{-\delta_n,j})\big)|^2] \lesssim \sum_j \mathbb{E}\big[\sup_{t\in[\underline{\widetilde{J}^n_{-\delta_n}},\overline{I^n_+}]}|X_t - X_{\underline{\widetilde{J}^n_{-\delta_n}}}|^2\big]$$

and so $\sum_j \mathbb{E}[|X\big(I^n(\widetilde{J}^n_{-\delta_n,j})\big)|^2]$ can be bound in order by

$$\sum_j \mathbb{E}\big[\big(\sup_{t\in[\underline{\widetilde{J}^n_{-\delta_n}},\overline{\widetilde{J}^n_{-\delta_n}}]}|X_t - X_{\underline{\widetilde{J}^n_{-\delta_n}}}|\big)^2\big] + \mathbb{E}\big[\big(\sum_j \sup_{t\in[\underline{I^n_+},\overline{I^n_+}]}|X_t - X_{\underline{I^n_+}}|\big)^2\big].$$

Thanks to the Π-reduction, we know that a given interval of the form $(\underline{I^n_+},\overline{I^n_+}]$ can be associated to at most two values of type J. Thus, the second term of the preceding quantity is smaller than

$$2\mathbb{E}\big[\big(\sum_i \sup_{t\in[\underline{I^n},\overline{I^n})}|X_t - X_{\underline{I^n}}|1_{i\leq \sharp\mathcal{I}}\big)^2\big],$$

where i is an indexing of the intervals $[\underline{I^n},\overline{I^n})$. Note that each $\underline{I^n_+}$ is an $\mathbb{F}$-stopping time as it is the maximum among all

$$\overline{I^n} \leq \overline{\widetilde{J}^n_{-\delta_n}}$$

together with a strong predictability property, see Lemma 2 for a similar statement. So, using Bürkholder-Davis-Gundy inequality, (10) is proved and Proposition 2 follows. □

4.4. Completion of proof of Theorem 1 under Assumption $\widetilde{A}$

Write $\mathcal{A} = \big\{\langle X, \tau_{-\vartheta}(Y)\rangle_T \neq 0\big\}$. By Assumption [B3], we have

$$\bigcup_{\widetilde{\vartheta}\in\mathcal{G}^n}[\widetilde{\vartheta}-\rho_n, \widetilde{\vartheta}+\rho_n) \supset \Theta.$$

Therefore, there exists a sequence ϑ_n in $\mathcal{G}^n$ such that $\vartheta_n \leq \vartheta$ and $|\vartheta_n - \vartheta| \leq 2\rho_n$. For sufficiently large n we have $\rho_n \leq \varepsilon_n = 2v_n$. Moreover, on the event $\mathcal{A}$

$$\mathcal{U}^n(\hat{\vartheta}_n) > \sup_{\widetilde{\vartheta}\in\mathcal{G}^n_+\cup\mathcal{G}^n_-}\big|\mathcal{U}^n(\widetilde{\vartheta})|$$

implies $|\hat{\vartheta}_n - \vartheta| < \varepsilon_n$. It follows that

$$\mathbb{P}\big[\big\{|\hat{\vartheta}_n - \vartheta| \geq \varepsilon_n\big\} \cap \mathcal{A}\big] \leq \mathbb{P}\big[\big\{ \sup_{\widetilde{\vartheta} \in \mathcal{G}_+^n \cup \mathcal{G}_-^n} \big|\mathcal{U}^n(\widetilde{\vartheta})\big| \geq \big|\mathcal{U}^n(\vartheta_n)\big|\big\} \cap \mathcal{A}\big].$$

Let $\varepsilon > 0$. For large enough n, the probability to have Δ_n smaller than v_n is larger than $1 - \varepsilon$ and consequently,

$$\mathbb{P}\Big[\big\{|\hat{\vartheta}_n - \vartheta| \geq \varepsilon_n\big\} \cap \mathcal{A}\Big] \leq \mathbb{P}\Big[\big\{ \sup_{\widetilde{\vartheta} \in \mathcal{G}_+^n \cup \mathcal{G}_-^n} \big|\mathbb{U}^n(\widetilde{\vartheta})_{T+\delta}\big| \geq \big|\mathcal{U}^n(\vartheta_n)\big|\big\} \cap \mathcal{A}\Big] + \varepsilon.$$

This can be bounded in order by

$$\mathbb{P}\Big[|\mathcal{U}^n(\vartheta_n)| < \tfrac{1}{2}|\langle X, \tau_{-\vartheta}(Y)\rangle_T|\Big] + \mathbb{P}\Big[\big\{ \sup_{\widetilde{\vartheta} \in \mathcal{G}_+^n \cup \mathcal{G}_-^n} \big|\mathbb{U}^n(\widetilde{\vartheta})_{T+\delta}\big| > \tfrac{1}{2}|\langle X, \tau_{-\vartheta}(Y)\rangle_T|\big\} \cap \mathcal{A}\Big] + \varepsilon$$

and this last quantity converges to ε as $n \to \infty$ by applying Proposition 1 and Proposition 2.

4.5. The case with drifts

We now give the proof of Theorem 1 under Assumptions $\widetilde{A}1$, $\widetilde{A}2$ and B. The contrast $\mathcal{U}^n(\widetilde{\vartheta})$ admits the decomposition $\mathcal{U}^n(\widetilde{\vartheta}) = \widetilde{\mathcal{U}}^n(\widetilde{\vartheta}) + \mathcal{R}^n(\widetilde{\vartheta})$ with

$$\widetilde{\mathcal{U}}^n(\widetilde{\vartheta}) = \sum_{I \in \mathcal{I}, J \in \mathcal{J}, \overline{I} \leq T} X^c(I) Y^c(J) 1_{\{I \cap J_{-\widetilde{\vartheta}} \neq \emptyset\}}$$

and

$$\mathcal{R}^n(\widetilde{\vartheta}) = \sum_{I \in \mathcal{I}, J \in \mathcal{J}} \big(X(I)B(J) + A(I)Y^c(J)\big) 1_{\{I \cap J_{-\widetilde{\vartheta}} \neq \emptyset\}}$$

For a function $t \to Z_t$ defined on the interval H, introduce the modulus of continuity

$$w_Z(a, H) = \sup\big\{|Z_t - Z_s|,\ s, t \in H, |s - t| < a\big\},\ \ a > 0.$$

We have

$$\sup_{\widetilde{\vartheta} \in [0,\delta)} \big|\mathcal{R}^n(\widetilde{\vartheta})\big| \leq\ w_X(3\Delta_n,\ [0, T]) \sup_{t \in [0, T+\delta]} |B_t| + w_{Y^c}(3\Delta_n,\ [0, T + \delta]) \sup_{t \in [0,T]} |A_t|$$

and this term goes to 0 in probability as $n \to \infty$.

Finally, the result is obtained in a similar way as in the no drift case, using (X^c, Y^c) in place of (X, Y).

4.6. The case where $\vartheta \in (-\delta, \delta)$

We now give the proof of Theorem 1 under Assumptions $\widetilde{A}1$ and B. Even in the case where ϑ is negative, Proposition 1 is still in force and we obtain

$$\sup_{\widetilde{\vartheta} \in \mathcal{G}^n \cap [0,\delta)} |\mathcal{U}^n(\widetilde{\vartheta})| \to 0$$

in probability as $n \to \infty$. The result follows from Remark 3.

5. A numerical illustration on simulated data

5.1. Synchronous data: methodology

We first superficially analyze the performances of $\widehat{\vartheta}_n$ on a simulated lead-lag Bachelier model without drift. More specifically, we take a random process $\big(X, \tau_{-\vartheta}(Y)\big)$ following the representation given in (2) in Section 2.1, having

$$T = 1, \;\; \delta = 1, \;\; \vartheta = 0.1, \; x_0 = \widetilde{y}_0 = 0, \sigma_1 = \sigma_2 = 1.$$

In this simple model, we consider again synchronous equispaced data with period Δ_n and correlation parameters ρ. In that very simple model, we construct $\widehat{\vartheta}_n$ with a grid $\mathcal{G}^n$ with equidistant points with mesh $h_n = \Delta_n$. We consider the following variations:

1. Mesh size: $h_n \in \{10^{-3}, 3.\,10^{-3}, 6.\,10^{-3}\}$.
2. Correlation value: $\rho \in \{0.25, 0.5, 0.75\}$.

5.2. Synchronous data: estimation results and their analysis

We repeat 300 simulations of the experiment and compute the value of $\widehat{\vartheta}_n$ each time, the true value being $\vartheta = 0.1$, letting ρ vary in $\{0.25, 0.5, 0.75\}$. We adopt the following terminology:

1. The fine grid estimation (abbreviated FG) with $h_n = 10^{-3}$.
2. The moderate grid estimation (abbreviated MG) with $h_n = 3.\,10^{-3}$.
3. The coarse grid estimation (abbreviated CG) with $h_n = 6.\,10^{-3}$.

The estimation results are displayed in Table 1 below. With no surprise, for a given mesh h_n, the difficulty of the estimation problem increases as ρ decreases.

In the fine grid approximation case (FG) with mesh $h_n = 10^{-3}$, the lead-lag parameter ϑ belongs to $\mathcal{G}^n$ exactly. Therefore, the contrast $\mathcal{U}^n(\widehat{\vartheta})$ is close to 0 for all values $\widetilde{\vartheta} \in \mathcal{G}^n$, except perhaps for the exact value $\widetilde{\vartheta} = \vartheta$. This is illustrated in Figure 1 and Figure 2 below, where we display the values or $\mathcal{U}^n(\widehat{\vartheta}_n)$. Note how more scattered are the values of $\mathcal{U}^n(\widehat{\vartheta}_n)$ for $\rho = 0.25$ compared to $\rho = 0.75$. This is of course no surprise.

$\widehat{\vartheta}_n$	0.096	0.099	0.1	0.102	Other
FG, $\rho = 0.75$	0	0	300	0	0
MG, $\rho = 0.75$	0	300	0	0	0
CG, $\rho = 0.75$	1	0	0	299	0
FG, $\rho = 0.50$	0	0	300	0	0
MG, $\rho = 0.50$	0	299	0	1	0
CG, $\rho = 0.50$	13	0	0	280	7
FG, $\rho = 0.25$	0	0	300	0	0
MG, $\rho = 0.25$	0	152	0	11	137
CG, $\rho = 0.25$	10	0	0	66	124

Table 1: *Estimation of* $\vartheta = 0.1$ *on 300 simulated samples for* $\rho \in \{0.25, 0.5, 0.75\}$.

For the moderate grid (MG) and the coarse grid (CH) cases, the lead-lag parameter $\vartheta \notin \mathcal{G}^n$ Hence, $\mathcal{U}^n(\widetilde{\vartheta})$ is close to 0 for almost all values of $\mathcal{G}^n$ except but two. When ρ is small, the statistical error in the estimation of ρ is such that $\left|\max_{\widetilde{\vartheta}\in\mathcal{G}^n} \mathcal{U}^n(\widetilde{\vartheta})\right|$ is not well located anymore. The error in the estimation can then be substantial, but is nevertheless consistent with our convergence result. This is illustrated in Figures 3 to 8 below.

When ρ decreases or when the mesh h_n of the grid increases, the performance of $\widehat{\vartheta}_n$ deteriorates, as shown in Figures 7 and 8 below.

5.3. Non-synchronous data

We randomly pick 300 sampling times for X over $[0, 1]$ uniformly over a grid of mesh size 10^{-3}. We randomly pick 300 sampling times for Y likewise, and independently of the sampling for X. The data generating process is the same as in Section 5.1. In Table 2, we display the estimation results for 300 simulations, in the fine gird case (FG) with $\vartheta = 0.1$ and $\rho = 0.75$.

The histograms for the case $\rho = 0.5$ and $\rho = 0.25$ are displayed in Figures 9 and 10.

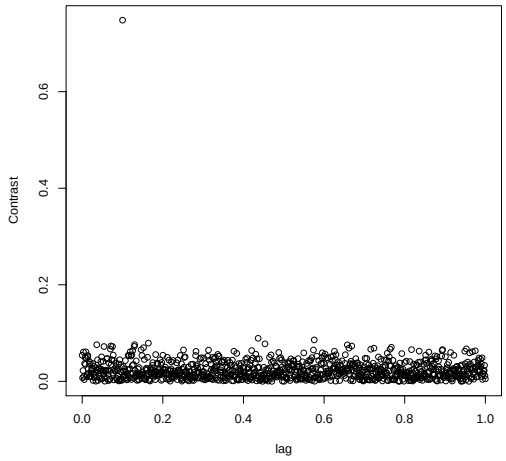

Figure 1. *Fine grid case (FG). Over one simulation: displayed values of* $\left|\mathcal{U}^n(\widetilde{\vartheta})\right|$ *for* $\widetilde{\vartheta} \in \mathcal{G}^n$ *with mesh* $h_n = 10^{-3}$ *and* $\rho = 0.75$. *The value* $\max_{\widetilde{\vartheta}\in\mathcal{G}^n} \left|\mathcal{U}^n(\widetilde{\vartheta})\right|$ *is well located.*

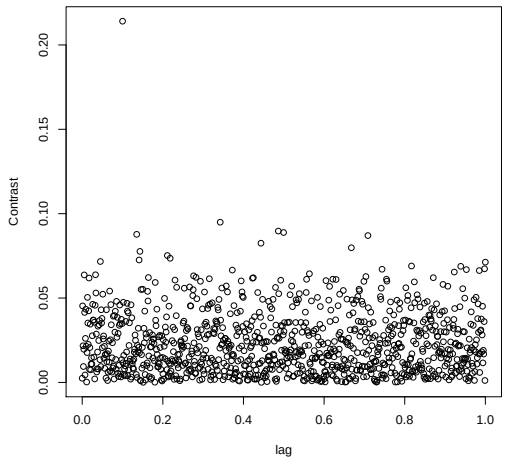

Figure 2. *Same setting as in Figure 1 for* $\rho = 0.25$. *The value* $\max_{\widetilde{\vartheta}\in\mathcal{G}^n} \left|\mathcal{U}^n(\widetilde{\vartheta})\right|$ *is still correctly located.*

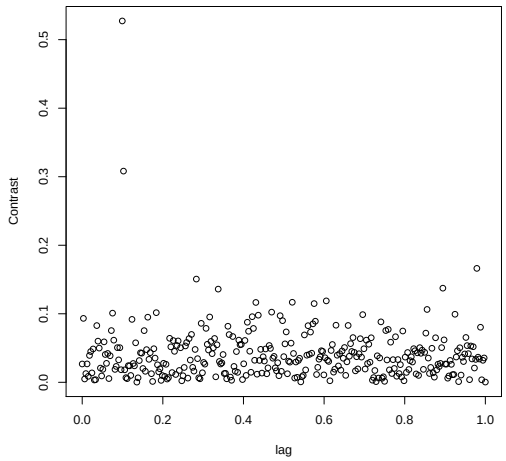

Figure 3. *Moderate grid case (MG). Over one simulation: displayed values of* $\left|\mathcal{U}^n(\widetilde{\vartheta})\right|$ *for* $\widetilde{\vartheta} \in \mathcal{G}^n$ *with mesh* $h_n = 10^{-3}$ *and* $\rho = 0.75$. *The value* $\max_{\widetilde{\vartheta}\in\mathcal{G}^n} \left|\mathcal{U}^n(\widetilde{\vartheta})\right|$ *is still well located. We begin to see the effect of the maximization over a grid* $\mathcal{G}^n$ *which does not match exactly with the true value* ϑ *with the appearance of a second maximum.*

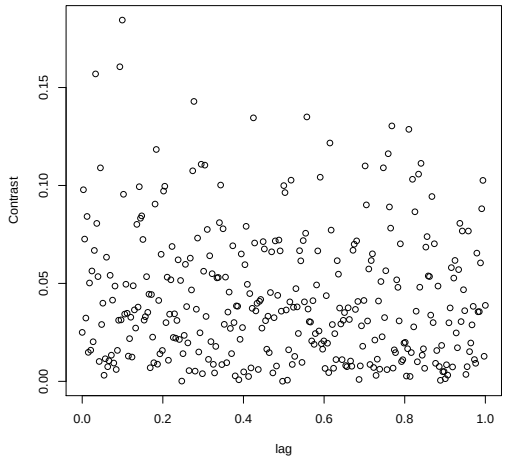

Figure 4. *Same setting as in Figure 3 for* $\rho = 0.25$. *The value* $\max_{\widetilde{\vartheta}\in\mathcal{G}^n} \left|\mathcal{U}^n(\widetilde{\vartheta})\right|$ *is still correctly located, but the overall shape of* $\left|\mathcal{U}^n(\widetilde{\vartheta})\right|$ *deteriorates.*

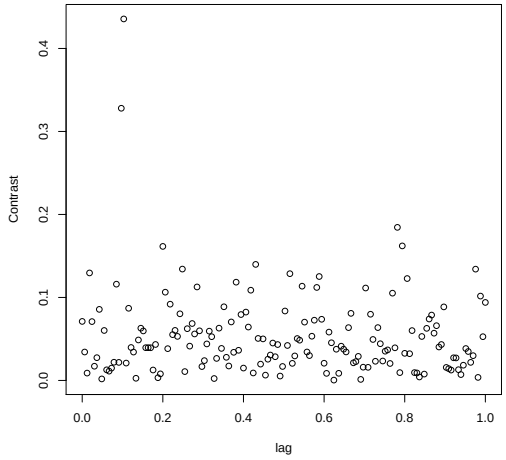

Figure 5. *Coarse grid case (CG). Over one simulation: displayed values of* $\left|\mathcal{U}^n(\widetilde{\vartheta})\right|$ *for* $\widetilde{\vartheta} \in \mathcal{G}^n$ *with mesh* $h_n = 10^{-3}$ *and* $\rho = 0.75$. *The value* $\max_{\widetilde{\vartheta}\in\mathcal{G}^n} \left|\mathcal{U}^n(\widetilde{\vartheta})\right|$ *is still well located. The fact that* $\mathcal{G}^n$ *does not match* ϑ *appears more clearly than in Figure 3.*

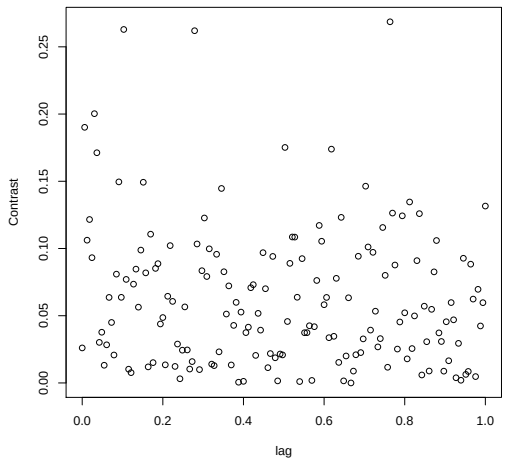

Figure 6. *Same setting as in Figure 5 for* $\rho = 0.25$. *The value* $\max_{\widetilde{\vartheta}\in\mathcal{G}^n} \left|\mathcal{U}^n(\widetilde{\vartheta})\right|$ *is no more correctly located.*

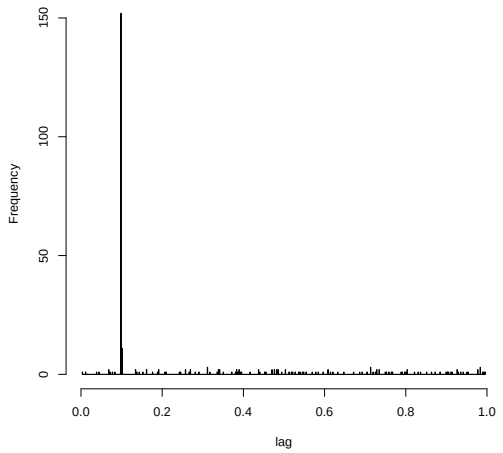

Figure 7. *Moderate grid case (MG). Histogram of the values of* $\widehat{\vartheta}_n$ *with true value* $\vartheta = 0.1$ *over* 300 *simulations for* $\rho = 0.25$.

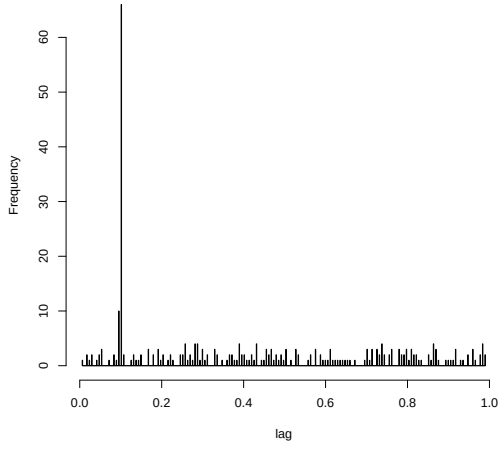

Figure 8. *Coarse grid case (CG). Histogram of the values of* $\widehat{\vartheta}_n$ *with true value* $\vartheta = 0.1$ *over* 300 *simulations for* $\rho = 0.25$.

$\widehat{\vartheta}$	0.099	0.1	0.101	0.102	0.103	0.104	0.105
FG, $\rho = 0.75$	16	106	107	46	19	4	2

Table 2: *Estimation of* $\vartheta = 0.1$ *on 300 simulated samples for* $\rho = 0.75$ *and non-synchronous data.*

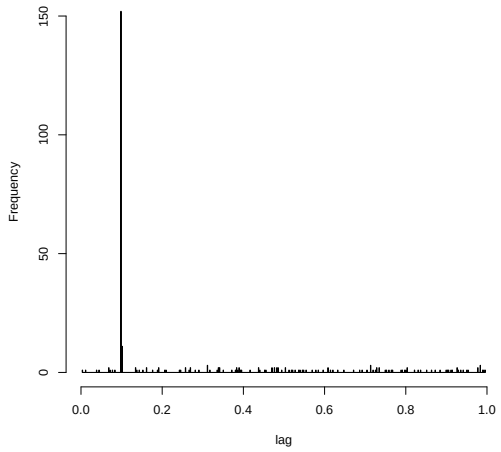

Figure 9. *Fine grid case (FG), non-synchronous data. Histogram of the values of* $\widehat{\vartheta}_n$ *with true value* $\vartheta = 0.1$ *over* 300 *simulations for* $\rho = 0.5$.

6. A numerical illustration on real data

6.1. The dataset

We study here the lead-lag relationship between the following two financial assets:

- The future contract on the DAX index (FDAX for short), with maturity December 2010.

- The Euro-Bund future contract (Bund for short), with maturity December 2010, which is an interest rate product based on a notional long-term debt instruments issued by the Federal Republic of Germany.

These two assets are electronically traded on the EUREX market, and are known to be highly liquid. Our data set has been provided by the company QuantHouse EUROPE/ASIA[2]. It consists in all the trades for 20 days of October 2010. Each trading day starts at 8.00 am CET and finishes at 22.00 CET, and the accuracy in the timestamp values is one millisecond.

[2]http://www.quanthouse.com.

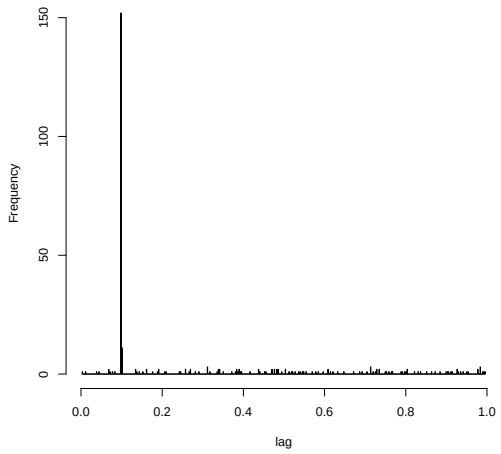

Figure 10. *Fine grid case (FG), non-synchronous data. Histogram of the values of* $\widehat{\vartheta}_n$ *with true value* $\vartheta = 0.1$ *over* 300 *simulations for* $\rho = 0.25$. *The performances of* $\widehat{\vartheta}_n$ *clearly deteriorates as compared to Figure 9.*

6.2. Methodology: a one day analysis

In order to explain our methodology, we take the example of a representative day: 2010, October 13.

Microstructure noise

Since high frequency data are concerned, we need to incorporate microstructure noise effects, at least at an empirical level. A classical way to study the intensity of the microstructure noise is to draw the signature plot (here in trading time). The signature plot is a function from $\mathbb{N}$ to $\mathbb{R}^+$. To a given integer k, it associates the sum of the squared increments of the traded price (the realized volatility) when only 1 trade out of k is considered for computing the traded price. If the price were coming from a continuous-time semi-martingale, the signature plot should be approximately flat. In practice, it is decreasing, as shown by Figure 11.

According to Figure 11, for all our considered day, we subsample our data so that we keep one trade out of 20. On 2010, October 13, after subsampling, it remains 2018 trades for the Bund and 3037 trades for the FDAX.

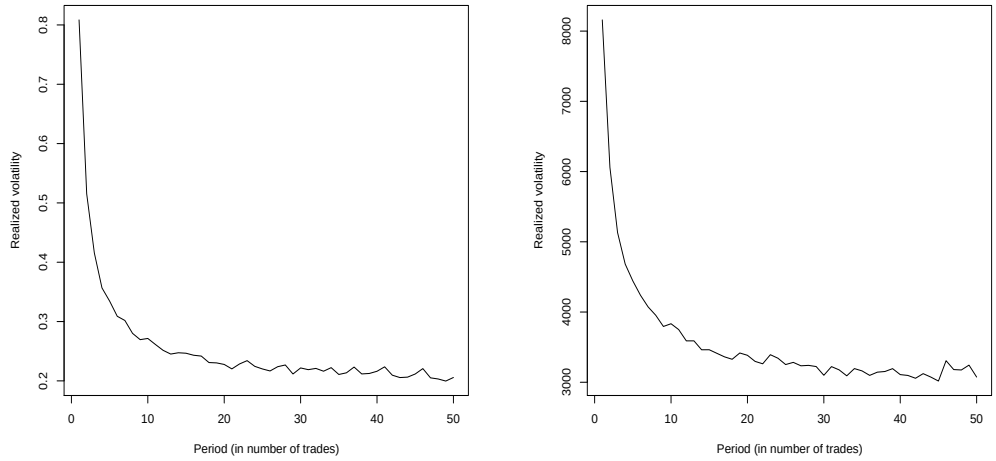

Figure 11. *Signature plot for the Bund (left) and the FDAX (right) for 2010, October 13.*

Construction of the contrast function

The second step is to compute our contrast function. Here the Bund plays the role of X and the FDAX the role of Y. Therefore, if the estimated value is positive, it means that the Bund is the leader asset and the FDAX the lagger asset, and conversely. To have a first idea of the lead-lag value, we consider our contrast function for a time shift between -10 minutes and 10 minutes, on a grid with mesh 30 seconds. The result of this computation for October 2010, 13 is given in Figure 12.

From Figure 12, we see that the lead-lag value is close to zero. Thus, we then compute the contrast function for a time shift between -5 seconds and 5 seconds, on a grid with mesh 0.1 second. The result of this computation for 2010, October 13 is given in Figure 13.

From Figure 13, we can conclude that on 2010, October 13, the FDAX seems to lead the Bund, with a small lead lag value of -0.8 second.

6.3. Systematic results over a one-month period

We now give in Figure 14 the results for all the days of October 2010.

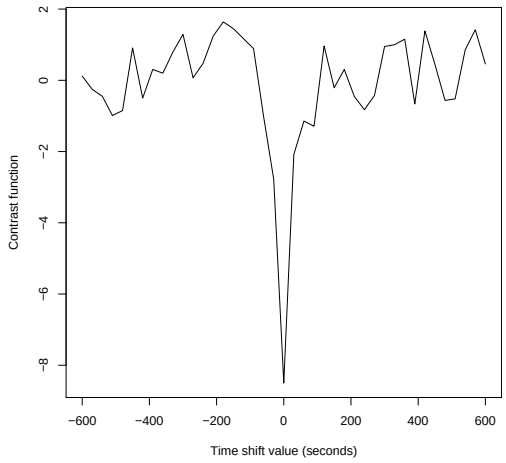

Figure 12. *The function $\mathcal{U}^n$ for 2010, October 13, time shift values between -10 minutes and 10 minutes, on a grid with mesh 30 seconds. The contrast is obtained by taking the absolute value of $\mathcal{U}^n$.*

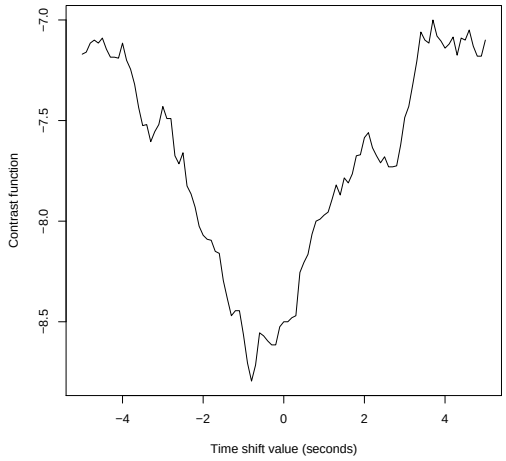

Figure 13. *The function $\mathcal{U}^n$ for 2010, October 13, time shift values between -5 seconds and 5 seconds, on a grid with mesh 0.1 second. The contrast is obtained by taking the absolute value of $\mathcal{U}^n$.*

Day	Number of Trades for the Bund (after subsampling)	Number of Trades for the FDAX (after subsampling)	Lead-Lag (seconds)
1 October 2010	2847	4215	-0.2
5 October 2010	2213	3302	-1.1
6 October 2010	2244	2678	-0.1
7 October 2010	1897	3121	-0.5
8 October 2010	2545	2852	-0.6
11 October 2010	1050	1497	-1.4
12 October 2010	2265	3018	-0.8
13 October 2010	2018	3037	-0.8
14 October 2010	2057	2625	-0.0
15 October 2010	2571	3269	-0.7
18 October 2010	1727	2326	-2.1
19 October 2010	2527	3162	-1.6
20 October 2010	2328	2554	-0.5
21 October 2010	2263	3128	-0.1
22 October 2010	1894	1784	-1.2
25 October 2010	1501	2065	-0.4
26 October 2010	2049	2462	-0.1
27 October 2010	2606	2864	-0.6
28 October 2010	1980	2632	-1.3
29 October 2010	2262	2346	-1.6

Figure 14. *Estimated lead-lag values for October 2010.*

The results of Figure 14 seem to indicate that on average, the FDAX tends to lead the Bund. Indeed, the estimated lead-lag values are systematically negative. Of course these results have to be taken with care since the estimated values are relatively small (the order of one second); however, dealing with highly traded assets on electronic markets, the order of magnitude of the lead-lag values that we find are no surprise and are consistent with common knowledge. A possible interpretation – yet speculative at the exploratory level intended in this paper – for the presence of such lead-lag effects is the existence of very quick portfolio adjustments between risky and non risky assets. The negative value could mean that market participants adjust first the risky part and then the non risky part.

7. Appendix

7.1. Proof of Lemma 1

Preliminary results We first prove the following results.

Lemma 3. *Work under Assumption [B2], under the slightly more general assumption that for all* $I = [\underline{I}, \overline{I}) \in \mathcal{I}$*, the random variables* $\underline{I}$ *and* $\overline{I}$ *are* $\mathbb{F}$*-stopping times.*

(a) *If* $\widetilde{\vartheta} \geq \vartheta + v_n$*, then for any* $\mathbb{F}$*-stopping time* σ *and* $t \in \mathbb{R}_+$*,* $\sigma + \widetilde{\vartheta}$ *is an* $\mathbb{F}^{\vartheta+v_n}$*-stopping time. In particular, the random variables* $\underline{I^n_{\widetilde{\vartheta}}}$ *and* $\overline{I^n_{\widetilde{\vartheta}}}$ *are* $\mathbb{F}^{\vartheta+v_n}$*-stopping times.*

(b) *For each $J \in \mathcal{J}$, we have*

$$\mathcal{F}^{\vartheta+v_n}_{\overline{J^n}} \subset \mathcal{F}^{\vartheta+v_n}_{\underline{J^n}+v_n} = \mathcal{F}^{\vartheta}_{\underline{J^n}}.$$

and for each $I \in \mathcal{I}$,

$$\mathcal{F}_{\overline{I^n}} = \mathcal{F}^{\vartheta+v_n}_{\overline{I^n_{(\vartheta+v_n)}}}.$$

(c) *Suppose that $\widetilde{\vartheta} \geq \vartheta + \varepsilon_n$ and $2v_n \leq \varepsilon_n$. Then for any random variable X' measurable w.r.t. $\mathcal{F}_{\overline{I^n}}$, the random variables $X'1_{\{\underline{I^n_{\widetilde{\vartheta}}} \leq \overline{J^n}\}}$ and $X'1_{\{\underline{I^n_{\widetilde{\vartheta}}} < \overline{J^n}\}}$ are $\mathcal{F}^{\vartheta}_{\underline{J^n}}$-measurable.*

Proof. **Proof of (a)**. For any $\mathbb{F}$-stopping time σ and $t \in \mathbb{R}_+$,

$$\begin{aligned}\big\{\sigma + \widetilde{\vartheta} \leq t\big\} = \big\{\sigma \leq t - \widetilde{\vartheta}\big\} &= \big\{\sigma \leq (t - (\widetilde{\vartheta} - \vartheta - v_n)) - \vartheta - v_n\big\} \\ &\in \mathcal{F}^{\vartheta+v_n}_{t-(\widetilde{\vartheta}-\vartheta-v_n)} \subset \mathcal{F}^{\vartheta+v_n}_t.\end{aligned}$$

Proof of (b). Note first that under Assumption [B2], the $\mathbb{F}^{\vartheta+v_n}$-stopping time $\underline{J^n}$ is in particular an $\mathbb{F}^{\vartheta}$-stopping time, thus $\mathcal{F}^{\vartheta}_{\underline{J^n}}$ is a σ-field. Moreover, since $\overline{J^n}$ and $\underline{J^n} + v_n$ are $\mathbb{F}^{\vartheta+v_n}$-stopping times by definition, both $\mathcal{F}^{\vartheta+v_n}_{\overline{J^n}}$ and $\mathcal{F}^{\vartheta+v_n}_{\underline{J^n}+v_n}$ are σ-fields, and also the inclusion is trivial from $\overline{J^n} \leq \underline{J^n} + v_n$. To obtain the equality, it suffices to observe that each of the conditions "$\mathcal{A} \in \mathcal{F}^{\vartheta+v_n}_{\underline{J^n}+v_n}$" and "$\mathcal{A} \in \mathcal{F}^{\vartheta}_{\underline{J^n}}$" is equivalent to the condition

$$\mathcal{A} \cap \{\underline{J^n} \leq t - v_n\} \in \mathcal{F}^{\vartheta}_{t-v_n}$$

for all $t \in \mathbb{R}_+$. The second equality is proved in the same way.
Proof of (c). Since $\overline{J^n}$ and $\underline{I^n_{\widetilde{\vartheta}}}$ are $\mathbb{F}^{\vartheta+v_n}$-stopping times by assumption, we have

$$\{\underline{I^n_{\widetilde{\vartheta}}} \leq \overline{J^n}\} \in \mathcal{F}^{\vartheta+v_n}_{\overline{J^n}} \subset \mathcal{F}^{\vartheta}_{\underline{J^n}},$$

the last inclusion following from **(b)**. If $\underline{I^n_{\widetilde{\vartheta}}} \leq \overline{J^n}$, then

$$\overline{I^n} \leq \underline{I^n} + v_n \leq \overline{J^n} - \widetilde{\vartheta} + v_n \leq \overline{J^n} - \vartheta - v_n, .$$

which implies $\overline{I^n_{\vartheta+v_n}} \leq \overline{J^n}$. Thus

$$X'1_{\{\underline{I^n_{\widetilde{\vartheta}}} \leq \overline{J^n}\}} = X'1_{\{\overline{I^n_{(\vartheta+v_n)}} \leq \overline{J^n}\}} \times 1_{\{\underline{I^n_{\widetilde{\vartheta}}} \leq \overline{J^n}\}}.$$

We have that X' is measurable with respect to $\mathcal{F}_{\overline{I^n}} = \mathcal{F}^{\vartheta+v_n}_{\overline{I^n_{(\vartheta+v_n)}}}$. Also $\overline{I^n_{(\vartheta+v_n)}}$ is a stopping time with respect to $\mathbb{F}^{\vartheta+v_n}$ by **(a)**. Consequently, $X'1_{\{\overline{I^n_{(\vartheta+v_n)}} \leq \overline{J^n}\}}$ is $\mathcal{F}^{\vartheta+v_n}_{\overline{J^n}}$-measurable hence $\mathcal{F}^{\vartheta}_{\underline{J^n}}$-measurable. Eventually, $X'1_{\{\underline{I^n_{\widetilde{\vartheta}}} \leq \overline{J^n}\}}$ is $\mathcal{F}^{\vartheta}_{\underline{J^n}}$-measurable. The other statement is proved the same way. □

Proof of Lemma 1 We have

$$X'K(I^n_{\underline{\vartheta}}, J^n) = X'1_{\{\underline{I^n_{\vartheta}} \le \underline{J^n}\}} 1_{\{\underline{J^n} < \overline{I^n_{\vartheta}}\}} + X'1_{\{\underline{I^n_{\vartheta}} > \underline{J^n}\}} 1_{\{\overline{J^n} > \underline{I^n_{\vartheta}}\}}.$$

Since $\widetilde{\vartheta} \ge \vartheta + \varepsilon_n \ge \vartheta + v_n$, both $\underline{I^n_{\widetilde{\vartheta}}}$ and $\overline{I^n_{\widetilde{\vartheta}}}$ are $\mathbb{F}^{\vartheta+v_n}$-stopping times. Therefore, the second term on the right-hand side of the above equality is $\mathcal{F}^{\vartheta}_{\underline{J^n}}$-measurable by **(c)** of Lemma 3.

Now we notice that if $\underline{I^n_{\widetilde{\vartheta}}} \le \underline{J^n}$, then $\underline{I^n_{\widetilde{\vartheta}}} \le \overline{J^n}$, therefore

$$X'1_{\{\underline{I^n_{\widetilde{\vartheta}}} \le \underline{J^n}\}} 1_{\{\underline{J^n} < \overline{I^n_{\widetilde{\vartheta}}}\}} = \big(X'1_{\{\underline{I^n_{\widetilde{\vartheta}}} \le \overline{J^n}\}}\big) \times \big(1_{\{\underline{I^n_{\widetilde{\vartheta}}} \le \underline{J^n}\}} 1_{\{\underline{J^n} < \overline{I^n_{\widetilde{\vartheta}}}\}}\big).$$

The first factor on the right-hand side of the above equality is $\mathcal{F}^{\vartheta}_{\underline{J^n}}$-measurable by **(c)** of Lemma 3, and the second factor is obviously $\mathcal{F}^{\vartheta}_{\underline{J^n}}$-measurable. This completes the proof.

7.2. Proof of Lemma 2

Let us fix $I \in \mathcal{I}$. Let

$$T_J = \begin{cases} \overline{I^n} - v_n & \text{on} \quad \left\{\overline{\widetilde{J^n}_{-\delta_n}} > \overline{I^n}\right\} \\ \overline{\widetilde{J^n}}_{-\delta_n} & \text{on} \quad \left\{\overline{\widetilde{J^n}_{-\delta_n}} \le \overline{I^n}\right\}. \end{cases}$$

We know that $\overline{I^n} - v_n$ is an $\mathbb{F}$-stopping time by Assumption [B2], and also that $\overline{\widetilde{J^n}_{-\delta_n}} - v_n$ is an $\mathbb{F}$-stopping time due to $\delta_n \le 0$. Let us show first that the T_J's are $\mathbb{F}$-stopping times. Let $t \in [-\delta, T+\delta]$. Let

$$\mathcal{A}_1 = \left\{\overline{I^n} - v_n \le t,\ \overline{\widetilde{J^n}_{-\delta_n}} - v_n > \overline{I^n} - v_n\right\}$$

and

$$\mathcal{A}_2 = \left\{\overline{\widetilde{J^n}}_{-\delta_n} \le t,\ \overline{\widetilde{J^n}_{-\delta_n}} - v_n \le \overline{I^n} - v_n\right\}.$$

It is obvious that $\mathcal{A}_1 \in \mathcal{F}_t$ since $\overline{I^n} - v_n$ is an $\mathbb{F}$-stopping time and also

$$\left\{\overline{\widetilde{J^n}_{-\delta_n}} - v_n \ge \overline{I^n} - v_n\right\} \in \mathcal{F}_{\overline{I^n} - v_n}.$$

For the term $\mathcal{A}_2$, if $t \in [-\delta, -\delta + v_n]$, then $\mathcal{A}_2 = \emptyset \in \mathcal{F}_{-\delta} \subset \mathcal{F}_t$. Otherwise, if $t \in (-\delta + v_n, T+\delta]$, then

$$\mathcal{A}_2 = \left\{\overline{\widetilde{J^n}_{-\delta_n}} - v_n \le t - v_n,\ \overline{\widetilde{J^n}_{-\delta_n}} - v_n \le \overline{I^n} - v_n\right\} \in \mathcal{F}_{t-v_n} \subset \mathcal{F}_t.$$

Eventually, we have $\{T_J \le t\} \in \mathcal{F}_t$, hence T_J is an $\mathbb{F}$-stopping time.

In conclusion, there exists at least one $\overline{\widetilde{J}^n_{-\delta_n}}$ in $[\overline{I^n} - v_n, \overline{I^n}]$. Therefore, we have $M^I = \sup_J T_J$ and this implies that M^I is also an $\mathbb{F}$-stopping time.

Acknowledgements. This work was originated by discussions between M. Hoffmann and M. Rosenbaum with S. Pastukhov from the Electronic Trading Group research team of P. Guével at BNP-Paribas. We are grateful to M. Musiela, head of the Fixed Income research at BNP-Paribas, for his constant support and encouragements. We also thank E. Bacry and K. Al Dayri for inspiring discussions. Japan Science and Technology supported the theoretical studies in this work. N. Yoshida's research was also supported by Grants-in-Aid for Scientific Research No. 19340021, the global COE program, "The research and training center for new development in mathematics" of Graduate School of Mathematical Sciences, University of Tokyo, JST Basic Research Programs PRESTO, and by Cooperative Reserch Program of the Institue of Statistical Mathematics.

References

[1] Aït-Sahalia, Y., Mykland, P. and Zhang, L. (2005). A tale of Two Time Scales: Determining integrated volatiity with noisy high-frequency data. *Journal Am. Statist. Assoc.*, **100**,1394–1411.

[2] Bandi, F. M. and Russell, J. R. (2008) Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies*, **75**, 339–369.

[3] Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2008). Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise. *Econometrica*, **76**, 1481–1536.

[4] de Jong, F. and Nijman, T. (1997). High frequency analysis of Lead-Lag relationships between financial markets *Journal of empirical finance*, **4**, 259–277.

[5] Chiao, C., Hung, K. and Lee, C. F. (2004). The price adjustment and lead-lag relations between stock returns: microstructure evidence from the Taiwan stock market. *Empirical Finance*, **11**, 709–731.

[6] Comte F., and Renaut, E. (1996). Non-Causality in Continuous Time Models. *Econometric Theory*, **12**, 215–256.

[7] Dalalyan, A. and Yoshida, N. (2008). Second-order asymptotic expansion for the covariance estimator of two asynchronously observed diffusion processes. Arxiv preprint arXiv:0804.0676, 2008-arxiv.org

[8] Genon-Catalot V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Ann. Inst. H. Poincar. Probab. Statist* **29**, 119–151.

[9] Gnedenko, B. (2007). Quantitative Trading Strategies and Multiscale Structure of Price Trends. General Analysis Framework. Working paper.

[10] Hansen, P.R. and Lunde, A. (2006). Realized Variance and Market Microstructure Noise. *Journal of Business and Economics Statistics*, **24**, 127-161.

[11] Hayashi, T. and Kusuoka, S. (2008). Consistent Estimation of Covariation under Nonsynchronicity. *Statist. Infer. Stoch. Process.* **11**, 93–106.

[12] Hayashi, T. and Yoshida, N. (2005a). Estimating correlations with missing observations in continuous diffusion models. Preprint.
[13] Hayashi, T. and Yoshida, N. (2005b). On the covariance estimation of nonsynchronously observed diffusion processes. *Bernoulli*, **11**, 359–379.
[14] Hayashi, T. and Yoshida, N. (2006). Nonsynchronous Covariance Estimator and Limit Theorem. Research memorandum No. 1020, Institute of Statistical Mathematics.
[15] Hayashi, T. and Yoshida, N. (2008). Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *AISM*, **60**, 367–406.
[16] Hayashi, T. and Yoshida, N. (2008). Nonsynchronous Covariance Estimator and Limit Theorem II. Research memorandum No. 1067, Institute of Statistical Mathematics.
[17] Hoshikawa, T. Kanatani, T., Nagai, K. and Nishiyama Y. (2008). Nonparametric Estimation Methods of Integrated Multivariate Volatilities. *Econometric reviews*, **27**, 112–138.
[18] Ibragimov I.A. and Hasminskii R.Z. (1981). *Statistical Estimation, asymptotic theory*. New York, Springer-Verlag.
[19] Jacod, J., Li Y., Mykland, P.A., Podolskij M, and Vetter, M. (2009). Microstructure Noise in the Continuous Case: The Pre-Averaging Approach. *Stochastic Processes and Their Applications*, **119**, 2249–2276.
[20] Kang, J., Lee, C. and Lee, S. (2006). Empirical Investigation of the Lead-Lag Relations of Returns and Volatilities among the KOSPI200 Spot, Futures and Options Markets and their Explanations. *Journal of Emerging Market Finance*, **5**, 235–261.
[21] Malliavin, P. and Mancino, M.E. (2002). Fourier Series Method for Measurement of Multivariate Volatilities, *Finance Stoch.*, **6**, 49–61.
[22] O'Connor, M. (1999). The Cross-Sectional Relationship Between Trading Costs and Lead/lag Effects in Stock & Option Markets *The Financial Review*, **34**.
[23] Rosenbaum, M. (2007). A new microstructure noise index. To appear in *Quantitative Finance*.
[24] Rosenbaum, M. (2009). Integrated volatility and round-off error. *Bernoulli*, **15**, 687–720.
[25] Robert, C.Y. and Rosenbaum, M. (2009a). A new approach for the dynamics of ultra high frequency data: the model with uncertainty zones. To appear in *Journal of Financial Econometrics*.
[26] Robert, C.Y. and Rosenbaum, M. (2009b). Volatility and covariation estimation when microstructure noise and trading times are endogenous. To appear in *Mathematical Finance*.
[27] Robert, C.Y. and Rosenbaum, M. (2010). On the limiting spectral distribution of the covariance matrices of time-lagged processes, *Journal of Multivariate Analysis*, **101**, 2434–2451.
[28] Ubukata, M. and Oya, K. (2008). A Test for Dependence and Covariance Estimator of Market Microstructure Noise. Discussion paper in *Economics and Business*, 07-03-Rev.2, 112–138.