VOLATILITY ESTIMATION UNDER ENDOGENOUS MICROSTRUCTURE NOISE

Christian Y. Robert
Claude Bernard University Lyon 1

VOLATILITY ESTIMATION UNDER ENDOGENOUS MICROSTRUCTURE NOISE

By Christian Y. Robert and Mathieu Rosenbaum

CREST-ENSAE Paris Tech,
CMAP-École Polytechnique Paris

This paper considers practically appealing procedures for estimating intraday volatility measures of financial assets. The underlying microstructure model accommodates the inherent properties of ultra high frequency data with the assumption of continuous efficient price processes. In this model, the microstructure noise is endogenous but does not only depend on the prices. Using the (observed) last traded prices of the assets, we develop a new approach that enables to approximate the values of the efficient prices at some random times. Based on these approximated values, we build an estimator of the integrated volatility and give its asymptotic theory. We also give a consistent estimator of the integrated co-volatility when two assets (asynchronous by construction of the model) are observed.

1. Introduction. In the recent years, a large number of papers has been devoted to the problem of estimating the integrated volatility of a financial asset from high frequency data. Since the dynamics of these data largely differ from the semi-martingale type behavior of low frequency data, their naïve use leads to a biased behavior of quadratic variation-type estimators. Consequently, it is now usual to view intraday market prices as noisy observations of the efficient price and to build estimators in this so called microstructure noise context.

Different types of microstructure noise models are encountered in the financial econometrics literature. The more usual type is the case of an additive noise that is independent of the efficient price ([3], [4], [5], [15], [16], [34], [36]). An additive endogenous component is also considered in [6] through a linear combination of past returns and in [23] through the Brownian motion driving the price. A major drawback of these approaches is that they do not allow for prices discreteness. More convincing types of microstructure noise are given by the cases of a contamination of the efficient price through a Markov kernel ([22], [26]) and a rounding error ([9], [25], [31]). The microstructure noise is respectively partially and fully endogenous in these models.

When estimating the integrated volatility, all the preceding models, even when allowing for prices discreteness, are not completely satisfying in practice since they are treated through a deterministic exogenous sampling (or some stochastic sampling through a regular time-change, which in fact can be boiled down in the deterministic case, see [6]) and so do not take into account the information contained in the intertrade durations. Hence, their use in practice lead to the following question: what is the right sampling frequency to use : 1 second, 1 minute, 5 minutes ? Moreover, one may ask about the right price to use : bid price, mid-quote price, last traded price ? Different answers to these questions often lead to significantly

AMS 2000 subject classifications: 60F05, 60G44, 62P05, 62F12

Keywords and phrases: Microstructure noise, Ultra high frequency data, Volatility, Co-volatility, Durations, Asynchronous data, Stopping times, Martingales

January 20, 2009

different results in volatility measures, see [19].

The problem of estimating the integrated co-volatility under microstructure noise has been much less studied. Indeed, in that case, beyond microstructure noise, one also faces the issue of the asynchronicity of the data. Both lead to a highly biased behavior of naïve high frequency co-volatility measures, see [20], [32], [35]. The asynchronicity has been first treated in [20]. The additional presence of microstructure noise is considered in [7], [32] and [35], in additive settings where the two preceding questions are still in force.

In this paper, we introduce a model with endogenous, structural, microstructure noise that enables to get rid of the problem of the choice of a price and a sampling frequency. This model for the last traded price, called model with uncertainty zones, allows for transaction price increments of one or several ticks, the size of the price jumps being determined by explanatory variables involving for example the order book. Moreover, as shown in details in [29], it enables to reproduce the main stylized facts of returns, durations and microstructure noise. The main idea behind this model is that, if a transaction occurs at some value on the tick grid and leads to a change in the transaction price, then the efficient price has been quite close to this value shortly before the transaction. So, we call uncertainty zones the bands around the mid-tick grid where the efficient price is too far from the tick grid to trigger a price change. In our setting, the width of these uncertainty zones quantifies the aversion to price changes of the market participants. Finally, note that this model can be easily interpreted by practitioners and its results on a large amount of real data are quite promising, see [29].

Our estimation procedure for the integrated volatility is based on a tick time sampling. It consists in deriving estimated values of the efficient price at some random times and then computing the realized volatility over these values. Our estimator is consistent and we provide its asymptotic theory as the tick size goes to zero. We also give a consistent estimator of the integrated co-volatility in the case where two assets are observed. One of the technical difficulties of our approach is that we deal with endogenous times and so usual limit theorems do not apply.

The paper is organized as follows. We describe and discuss the model in Section 2. The estimators and associated theorems are given in Section 3. Section 4 contains the proofs and the results of a simulation study are given in Section 5. We conclude in Section 6.

2. Model with uncertainty zones.

2.1. *Description of the model.* We build in this section a model on the last traded price. In an idealistic framework, where the efficient price would be observed, market participants would trade when the efficient price crosses the tick grid. In practice, there is some uncertainty about the efficient price value so that market participants are reluctant to price changes. Hence, there is a modification of the transaction price only if some buyers and sellers are truly convinced that the efficient price is sufficiently far from the last traded price. We introduce a parameter η that quantifies the aversion to price changes (with respect to the tick size) of the market participants and propose a model that takes into account this aversion.

Let $(X_t)_{t\geq 0}$ denote the efficient price of the asset. On a rich enough filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, we assume that the logarithm of the efficient price $(Y_t)_{t\geq 0}$ is a $\mathcal{F}_t$-adapted continuous semi-martingale of the form

$$Y_t = \log X_t = \log X_0 + \int_0^t a_u \mathrm{d}u + \int_0^t \sigma_{u-} \mathrm{d}W_u,$$

where $(W_t)_{t\geq 0}$ is a standard $\mathcal{F}$-Brownian motion, $(a_t)_{t\geq 0}$ is a progressively measurable process with locally bounded sample paths and $(\sigma_t)_{t\geq 0}$ is a positive $\mathcal{F}_t$-adapted process with càdlàg sample paths. Note that by taking the predictable projection of a, which is still locally bounded, one can also consider the usual assumption of a predictable drift. However, this is not useful in our continuous efficient price setting.

The tick grid where transaction prices are bound to lie on is defined as $\{k\alpha; k \in \mathbb{N}\}$, with α the tick size. For $k \in \mathbb{N}$ and $0 < \eta < 1$, we define the zone U_k by $U_k = [0, \infty) \times (d_k, u_k)$ with

$$d_k = (k + 1/2 - \eta)\alpha \text{ and } u_k = (k + 1/2 + \eta)\alpha.$$

Thus, U_k is a band around the mid-tick grid value $(k + 1/2)\alpha$, see Figure 1. Note that when η is smaller than $1/2$, there is no overlap between the zones.

Let $t_0 = 0$ and P_0 be the opening price. For $i \geq 1$, denote by t_i the i-th time where a change in the transaction price of the asset is observed, by P_{t_i} its associated transaction price, and define the last traded price $(P_t)_{t\geq 0}$ as the càdlàg piecewise constant process built from the $(t_i, P_{t_i})_{i\geq 0}$.

We assume that the transaction price may jump from price $k'\alpha$ to price $k\alpha$ with $k' \neq k$ only once the efficient price exited down the zone U_k or exited up the zone U_{k-1} and provided that market conditions are favorable for a transaction to occur. In a way, the transaction price only changes when the efficient price is close from a new multiple value of α and market participants want to trade. The zones $(U_k)_{k\in\mathbb{N}}$ represent bands inside of which the efficient price can not trigger a change of the transaction price. Consequently, they will be referred to as the uncertainty zones.

More specifically, let us precise the construction of the sequence $(\tau_i)_{i\geq 0}$ of the exit times from the uncertainty zones which will lead to a change in the transaction price. Let $\tau_0 = 0$ and assume without loss of generality that τ_1 is the exit time of $(X_t)_{t\geq 0}$ from the set (d_{k_0-1}, u_{k_0}) where $k_0 = X_0^{(\alpha)}$, with $X_0^{(\alpha)}$ the value of X_0 rounded to the nearest multiple of α. We introduce a sequence $(L_i)_{i\geq 1}$ of $\mathcal{F}_{\tau_i}$-measurable discrete random variables which represent the absolute value in number of ticks of the price jump between the i-th and the $(i+1)$-th transaction leading to a price change. As explained later, the distribution of this variable will depend on the value of some market quantities at time τ_i. Then define recursively τ_{i+1} as the exit time of $(X_t)_{t>\tau_i}$ from the set $(d_{k_i - L_i}, u_{k_i + L_i - 1})$, where $k_i = X_{\tau_i}^{(\alpha)}$, that is

$$\tau_{i+1} = \inf\Big\{t : t > \tau_i, X_t = X_{\tau_i}^{(\alpha)} - \alpha(L_i - \frac{1}{2} + \eta) \text{ or } X_t = X_{\tau_i}^{(\alpha)} + \alpha(L_i - \frac{1}{2} + \eta)\Big\}. \tag{1}$$

In particular, if $X_{\tau_i} = d_j$ for some $j \in \mathbb{N}$, τ_{i+1} is the exit time of $(X_t)_{t>\tau_i}$ from the set (d_{j-L_i}, u_{j+L_i-1}), and if $X_{\tau_i} = u_j$ for some $j \in \mathbb{N}$, τ_{i+1} is the exit time of $(X_t)_{t>\tau_i}$ from the set (d_{j-L_i+1}, u_{j+L_i}), see Figure 1.

Finally, let $t_0 = 0$ and $P_0 = X_0^{(\alpha)}$. We assume that the couples (t_i, P_{t_i}) satisfy for $i \geq 1$

$$\tau_i \leq t_i < \tau_{i+1} \text{ and } P_{t_i} = X_{\tau_i}^{(\alpha)}.$$

It means that between τ_i and τ_{i+1}, at least one transaction has occurred at price P_{t_i} and t_i is the time of the first of these transactions. The difference $t_i - \tau_i$ can be viewed as the delay caused by the reaction times of the market participants and/or by the trading process.

Note also that αL_i is the absolute value of the price jump between the i-th and the $(i+1)$-th transactions with price change and that

$$P_{t_i} = X_{\tau_i} + sign(X_{\tau_i} - X_{\tau_{i-1}})(1/2-\eta)\alpha = X_{\tau_i} + sign(P_{t_i} - P_{t_{i-1}})(1/2-\eta)\alpha. \tag{2}$$

Hence, if one knows (estimates) η, one can recover (estimate) X_{τ_i} from $P_{t_{i-1}}$ and P_{t_i}. Figure 1 displays an example of the different trajectories in a case where $0<\eta<1/2$.

The measurement error at transaction time t_i is given by

$$P_{t_i} - X_{t_i} = -(X_{t_i} - X_{\tau_i}) + sign(X_{\tau_i} - X_{\tau_{i-1}})(1/2-\eta)\alpha. \tag{3}$$

This error has to be compared with the simple linear models of endogeneity for exogenous sampling introduced in [23] or in Section 5.5 of [6]. Note that it depends on the price and the τ_i in an intricate way since the τ_i are stopping times with respect to a bigger filtration than those generated by the price.

We eventually precise the conditional distribution of the jump sizes in ticks between consecutive transaction prices. We assume that the jump sizes are bounded (what is empirically not restrictive) and denote by m their maximal value. For $k=1,...,m$ and $t>0$, let

$$N^{(a)}_{\alpha,t,k} = \sum_{\tau_i \le t} \mathbb{I}_{\{|X_{\tau_i} - X_{\tau_{i-1}}| = \alpha(k-1+2\eta)\}} \text{ and } N^{(c)}_{\alpha,t,k} = \sum_{\tau_i \le t} \mathbb{I}_{\{|X_{\tau_i} - X_{\tau_{i-1}}| = \alpha k\}}$$

be respectively the number of alternations and continuations of k ticks. An alternation (continuation) of k ticks is a jump of k ticks whose direction is opposite to (the same as) that of the preceding jump, see Figure 1. Remark that for small (large) values of η, one will mainly observe alternations (continuations). Let $(\chi_t)_{t\ge 0}$ be a continuous M-dimensional $\mathcal{F}_t$-adapted process. We define the filtration $\mathcal{E}$ as the complete right-continuous filtration generated by $(X_t, \chi_t, N^{(a)}_{\alpha,t,k}, N^{(c)}_{\alpha,t,k},\ k=1,\dots,m)$. We assume that conditional on $\mathcal{E}_{\tau_i}$, L_i is a discrete random variable on $[\![1,m]\!]$ satisfying

$$\mathbb{P}_{\mathcal{E}_{\tau_i}}[L_i = k] = p_k(\chi_{\tau_i}), \quad 1 \le k \le m, \tag{4}$$

for some unknown positive differentiable with bounded derivative functions p_k. In practice, χ_t may represent quantities related for example to the traded volume, the bid-ask spread, or the bid and ask depths. For the applications, specific form for the p_k are given in [29].

2.2. *Discussion.*

• The model with uncertainty zones accommodates the inherent properties of prices, durations and microstructure noise (see [10], [17], [19] for the different features of these quantities) together with a semi-martingale efficient price. In particular, this model allows for discrete prices, a bid-ask bounce and an inverse relation between durations and volatility. Moreover the usual behaviors of the autocorrelograms and cross correlograms of returns and microstructure noise, both in calendar and tick time, are reproduced. Eventually, it leads to jumps in the price of several ticks, the size of the jumps being determined by explanatory variables involving for example the order book. Mostly, the model with uncertainty zones is clearly validated on real data. These results are studied in details in [29].

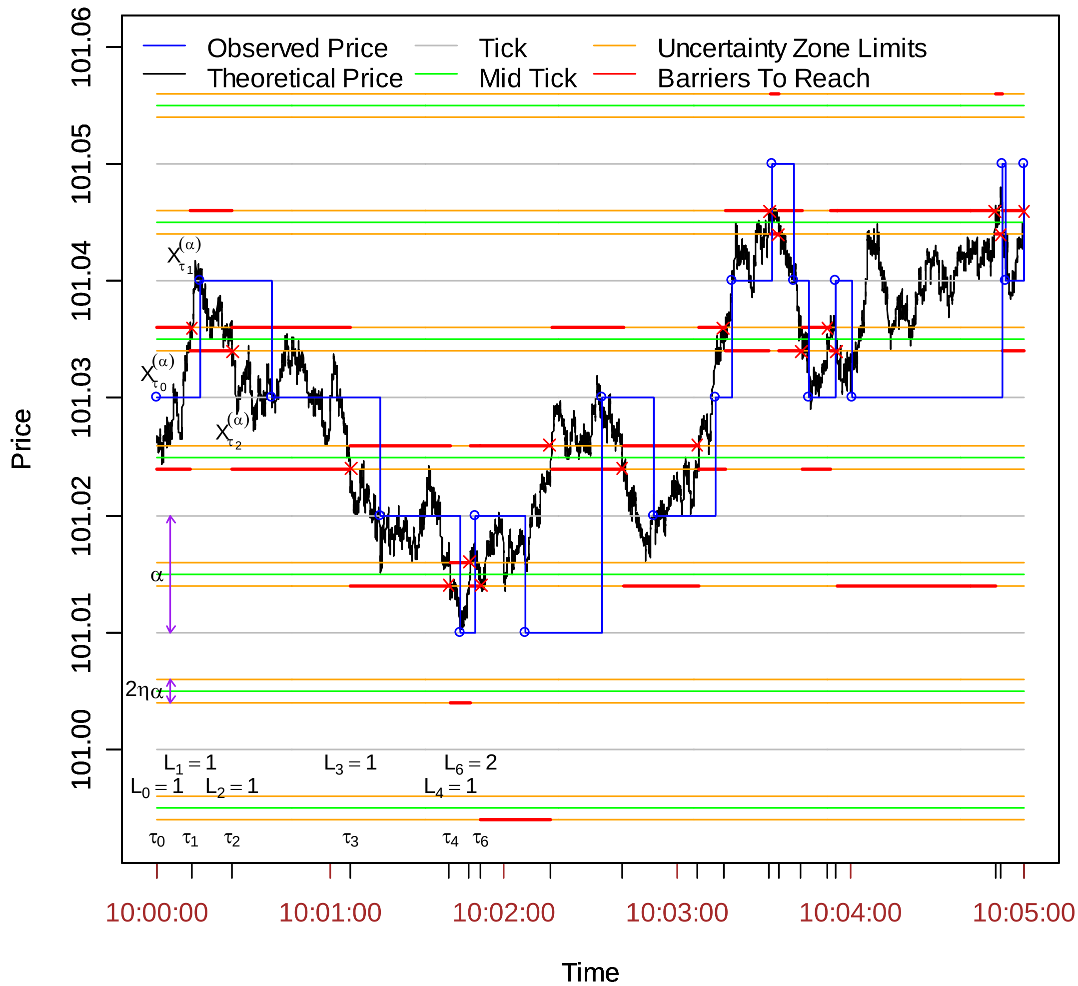

FIG 1. *Example of trajectories of the latent price and of the observed price. The red crosses denote the exit points associated to the* τ_i.

• Some restrictive cases of our model are mentioned in the literature. The case $\eta = 0$, $L_i = 1$, $t_i = \tau_i$ for all i corresponds to the pure rounding case, which is not realistic because of the infinite number of oscillations of the price. The case $\eta = 1/2$, $L_i = 1$, $t_i = \tau_i$ corresponds to the model studied in [11]. Note that in general, this specification of η and of the L_i does not seem to be convenient for real data, see [29]. The case $\eta < 1/2$, $L_i = 1$, $t_i = \tau_i$ is mentioned in [24]. Note also that a discrete version of this model was introduced in [30].

• For expository purpose, we impose the not really restrictive assumption that the upward and downward barriers to reach are only defined through the L_i, see Figure 1. However, a model with some variables L_i^+ and L_i^- for the upward and downward barriers could be considered.

• As explained in the previous section, η quantifies the aversion to price changes (with respect to the tick size) of the market participants. Indeed, η controls the width of the uncertainty zones. In tick unit, the larger η, the farther from the last traded price the efficient price has to be so that a price change occurs. In some sense, a small η ($< 1/2$) means that the tick size appears too large to the market participants and a large η means that the tick size appears too small.

• There are several other ways to interpret the parameter η, notably from a practitioner's perspective. For example, one can think that in the very high frequencies, the order book can not "follow" the efficient price and is reluctant to price changes. This reluctancy could be characterized by η. Another possibility is to view η as a measure of the usual prices depth explored by the transaction volumes.

• It is shown in [29] that the parameter η remains remarkably stable in time for a large number of assets. Moreover, all types of configuration between the tick size and the orders of magnitude of the volatility and η can be found on the market.

• Although the majority of the transactions does not lead to a price change, see for example [18], [27], we only model transactions with price change and consider a tick time sampling scheme $(t_i, P_{t_i})_{i\geq 0}$. In [18], an empirical and theoretical study shows that sampling in tick time is generally preferable to sampling in transaction time or to the common practice of sampling in calendar time when estimating the integrated volatility.

• We consider a structural model for the microstructure noise and so, from the efficient price process, we directly model transaction prices. Moreover, if one can estimate the parameter η, up to an estimation error, one can retrieve the true value of the latent price at time τ_i, see Equation (2). This is very convenient in the purpose of building statistical procedures, see Section 3.

3. Estimation procedures.

For a fixed objective time T, our goal is to estimate the integrated volatility of the asset and the integrated co-volatility between two assets over $[0, t]$, $t \leq T$. In our model, we work with random observation times, whose structure depends on the efficient price. This context differs from those using deterministic sampling schemes and lead to new technical issues, see in particular [1], [2] and [14]. Our asymptotics is to consider that the tick size is going to zero. Even if the tick size is fixed on the markets, it is just a reasonable way to make the number of observations go to infinity. This kind of asymptotics is also used in [9], [11] or [31].

3.1. *Estimation of the integrated volatility.*

The integrated volatility of (X_t) on $[0, t]$, $t \leq T$, is defined by

$$IV_t = \int_0^t \sigma_s^2 \mathrm{d}s.$$

Note that, contrary to some authors, we do not focus on the quadratic variation of $(X_t)_{t\geq 0}$ but of $(\log X_t)_{t\geq 0}$. Although the two quantities are close (up to a price scale factor), according to the mathematical finance theory, the second one is probably more relevant. Moreover, its inference is more intricate since price discreteness happens on the original scale and not on the log scale.

In our framework, a natural idea for estimating this quantity is to consider the following realized volatility

$$RV_{\alpha,t} = \sum_{\tau_i \leq t} \Big(\frac{X_{\tau_i} - X_{\tau_{i-1}}}{X_{\tau_{i-1}}}\Big)^2.$$

We use this form of the realized volatility for technical convenience. It is as natural as the logarithmic form since our model is not based on the log price. The X_{τ_i} are not observed but can be obtained through Equation (2) up to the knowledge of η. So, the first step consists in estimating the parameter η and the second step in replacing η by its estimate to approach $RV_{\alpha,t}$.

We define the estimator of η by

$$\hat{\eta}_{\alpha,t} = \Big(0 \vee \sum_{k=1}^{m} \lambda_{\alpha,t,k} u_{\alpha,t,k}\Big) \wedge 1,$$

with

$$\lambda_{\alpha,t,k} = \frac{N^{(a)}_{\alpha,t,k} + N^{(c)}_{\alpha,t,k}}{\sum_{j=1}^{m}\big[N^{(a)}_{\alpha,t,j} + N^{(c)}_{\alpha,t,j}\big]} \text{ and } u_{\alpha,t,k} = \frac{1}{2}\big(k\big(\frac{N^{(c)}_{\alpha,t,k}}{N^{(a)}_{\alpha,t,k}} - 1\big) + 1\big).$$

The idea behind this estimator is that the $u_{\alpha,t,k}$ are consistent estimators of η for each k. The $\lambda_{\alpha,t,k}$ are then natural weighting factors. Note in particular that $N^{(c)}_{\alpha,t,1}/N^{(a)}_{\alpha,t,1}$ is an estimator of 2η. Consequently, if η is smaller than $1/2$, we may expect more alternations than continuations in the last traded price and conversely.

Next we define our estimator of the integrated volatility by

$$\widehat{RV}_{\alpha,t} = \sum_{t_i \le t} \big(\frac{\hat{X}^t_{\tau_i} - \hat{X}^t_{\tau_{i-1}}}{\hat{X}^t_{\tau_{i-1}}}\big)^2,$$

where for $t_i < t$,

$$\hat{X}^t_{\tau_i} = P_{t_i} - \alpha\big(\frac{1}{2} - \hat{\eta}_{\alpha,t}\big) sign(P_{t_i} - P_{t_{i-1}}).$$

The key tool to establish the consistency of our estimator of η is the Dambis-Schwarz theorem. Indeed, it allows us to transform our process into a Brownian motion in a modified time in which the distributions of the exit times from the uncertainty zones are explicit. Then we use classical results on the quadratic variation over stopping times, see for example Theorem I.4.47 in [21]. In the following, $\overset{u.c.p.}{\to}$ denotes uniform convergence in probability over compact sets included in $[0,T]$. Abusing notation slightly, we say that the family of processes Z^α converges uniformly in probability towards Z as α tends to zero if for any sequence α_n tending to zero, $Z^{\alpha_n} \overset{u.c.p.}{\to} Z$. We have the following result.

Theorem 1. *As α tends to 0,*

$$\hat{\eta}_{\alpha,t} \overset{u.c.p.}{\to} \eta \text{ and } \widehat{RV}_{\alpha,t} \overset{u.c.p.}{\to} \int_0^t \sigma_s^2 ds.$$

For our next theorem, we work under the following assumption.

Assumption 1. *The process χ is a $\mathcal{F}_t$-adapted continuous Ito semi-martingale with progressively measurable with locally bounded sample paths and positive $\mathcal{F}_t$-adapted volatility matrix whose elements have càdlàg sample paths.*

Let us now give the definition of stable convergence in law. Let Z^α be a family of random variables (taking their values in the space of càdlàg functions endowed with the Skorokhod topology J_1). Let α_n be a deterministic sequence tending to zero as n tends to infinity and $\mathcal{I}$ be a sub-σ-field of $\mathcal{F}$. We say that Z^{α_n} converges $\mathcal{I}$-stably to Z as α_n tends to zero ($Z^{\alpha_n} \overset{\mathcal{I}-\mathcal{L}s}{\to} Z$) if for every $\mathcal{I}$-measurable bounded real random variable V, (V, Z^{α_n}) converges in law to (V, Z) as n tends to infinity. This is a slightly stronger mode of convergence than the weak convergence, see [21] for details and equivalent definitions. Finally, we say that Z^α converges $\mathcal{I}$-stably to Z as α tends to zero if for any sequence α_n tending to zero, $Z^{\alpha_n} \overset{\mathcal{I}-\mathcal{L}s}{\to} Z$.

Next we introduce the following notation: ∇_1 and ∇_2 are two $(2m+1)$ valued vectors defined by $\nabla_{1,1} = 1$, $\nabla_{1,i} = 0$ for $i = 2, \ldots, 2m+1$, $\nabla_{2,1} = 0$ and for $i = 1, \ldots, m$,

$$\nabla_{2,2i} = i + \eta - 1/2, \ \nabla_{2,2i+1} = -i^{-1}(i + \eta - 1/2)(i + 2\eta - 1).$$

The processes $(f_t)_{t\geq 0}$ and $(\mu_t)_{t\geq 0}$ are defined by

$$f_t = \int_0^t \varphi(\chi_u)\sigma_u^2 X_u^2 \mathrm{d}u, \quad \mu_t = \int_0^t \sum_{k=1}^m \frac{2k(k-1+2\eta)}{2k-1+2\eta} p_k(\chi_u)\varphi(\chi_u)\sigma_u^2 \mathrm{d}u,$$

with

$$\varphi(\chi_u) = (\sum_{j=1}^m p_j(\chi_u) j(j-1+2\eta))^{-1}.$$

We are now able to state our limit theorem. Note that the observation times are random, endogenous. So, usual theorems for deterministic or exogenous sampling can not be applied. The key idea for the proof is to work in a modified time in which the observation times are equidistant and to use stability properties of the convergence in law in the Skorohod space.

Let $\mathbb{D}[0,T]$ denote the space of càdlàg functions on $[0,T]$ and $^\top$ the transpose operator. We have the following result.

THEOREM 2. *Let $\mathcal{I}$ be the filtration generated by the processes X and χ. Under Assumption 1, as α tends to 0, we have*

$$\alpha^{-1}(\widehat{RV}_{\alpha,t} - \int_0^t \sigma_s^2 ds) \overset{\mathcal{I}-\mathcal{L}s}{\to} (\nabla_1^\top + \frac{\mu_t}{f_t}\nabla_2^\top) \int_0^t b_{f_s} d\mathcal{W}_{f_s},$$

in $\mathbb{D}[0,T]$, where $\mathcal{W}$ is a $(2m+1)$ Brownian motion which is defined on an extension of the filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and is independent of all the preceding quantities and b_s is a $(2m+1)\times(2m+1)$ matrix defined in Lemma 14.

Note that the assumptions on the efficient price are very weak. In particular, σ_u is not necessarily an Ito semi-martingale as in [6] or [22].

3.2. *Estimation of the integrated co-volatility.* We now turn to the problem of estimating the integrated co-volatility when two assets are observed. In our context, one can not use the classical realized covariation estimator for two reasons: the asynchronicity of the data and the presence of microstructure noise. The problem of the asynchronicity of the data has to be taken with great care since intuitive ideas such as the previous tick interpolation may lead to a systematic bias called Epps effect, see [20] and [35] for details. This issue has been treated in [20] in the case when the correlation coefficient and the volatility functions are deterministic and when the transaction times are independent of the price. Nevertheless, the proposed estimator seems in general not robust to microstructure noise, see [32]. We show in this paper that a modified version of the Hayashi-Yoshida estimator, which uses the estimated values of the efficient price given by Equation (2), is consistent.

More precisely we consider a $\mathcal{F}_t$-adapted bidimensional continuous Ito semi-martingale $(X_t^{(1)}, X_t^{(2)})$ such that for $j = 1, 2$

$$Y_t^{(j)} = \log X_t^{(j)} = \log X_0^{(j)} + \int_0^t a_u^{(j)} \mathrm{d}u + \int_0^t \sigma_{u-}^{(j)} \mathrm{d}W_u^{(j)}$$

and

$$\langle W^{(1)}, W^{(2)} \rangle_t = \int_0^t \rho_s \mathrm{d}s,$$

where ρ_s is an adapted process with càdlàg sample paths such that for all s, $-1 < \rho_s < 1$. We impose the same assumptions on $a^{(j)}$ and $\sigma^{(j)}$ as in Section 2.1. The quantities $\alpha^{(j)}$, $\eta^{(j)}$, $L_i^{(j)}$, $\tau_i^{(j)}$, $t_i^{(j)}$ and $P_{t_i^{(j)}}^{(j)}$ are also defined in the same way as in Section 2.1.

The usual Hayashi-Yoshida covariation estimator is given by

$$(5)\quad HY_t = \sum_{t_{i_1}^{(1)} \leq t} \sum_{t_{i_2}^{(2)} \leq t} \big(\log(P_{t_{i_1}^{(1)}}^{(1)}) - \log(P_{t_{i_1-1}^{(1)}}^{(1)}) \big) \big(\log(P_{t_{i_2}^{(2)}}^{(2)}) - \log(P_{t_{i_2-1}^{(2)}}^{(2)}) \big) \mathbb{I}_{[t_{i_1-1}^{(1)}, t_{i_1}^{(1)}] \cap [t_{i_2-1}^{(2)}, t_{i_2}^{(2)}] \neq \emptyset}.$$

We build our modified estimator the following way. First we consider new sequences of stopping times $(\lambda_i^{(1)})_{i \geq 0}$ and $(\lambda_i^{(2)})_{i \geq 0}$ which limit the issue of asynchronicity. Then we use the estimated values of the efficient price instead of the observed price. We define our new sequences by $\lambda_0^{(1)} = 0$, $\lambda_0^{(2)} = 0$ and for $i \geq 0$, if $\tau_1^{(1)} > \tau_1^{(2)}$,

$$\lambda_{i+1}^{(1)} = \min \{ \tau_j^{(1)} : \tau_j^{(1)} > \lambda_{i+1}^{(2)} \}, \quad \lambda_{i+1}^{(2)} = \min \{ \tau_j^{(2)} : \tau_j^{(2)} \geq \lambda_i^{(1)} \}$$

and if $\tau_1^{(1)} \leq \tau_1^{(2)}$,

$$\lambda_{i+1}^{(1)} = \min \{ \tau_j^{(1)} : \tau_j^{(1)} \geq \lambda_i^{(2)} \}, \quad \lambda_{i+1}^{(2)} = \min \{ \tau_j^{(2)} : \tau_j^{(2)} > \lambda_{i+1}^{(1)} \}.$$

Note that if $\tau_1^{(1)} > \tau_1^{(2)}$, then $\lambda_1^{(2)} < \lambda_1^{(1)} \leq \lambda_2^{(2)} < \lambda_2^{(1)} \ldots \leq \lambda_j^{(2)} < \lambda_j^{(1)}$. We denote by $l_i^{(j)}$ the value of the time $t_{\cdot}^{(j)}$ associated to $\lambda_i^{(j)}$. Finally, our estimator is defined by

$$\widehat{RCV}_t = \sum_{l_{i_1}^{(1)} \leq t} \sum_{l_{i_2}^{(2)} \leq t} \big(\log(\hat{X}_{\lambda_{i_1}^{(1)}}^{(1)}) - \log(\hat{X}_{\lambda_{i_1-1}^{(1)}}^{(1)}) \big) \big(\log(\hat{X}_{\lambda_{i_2}^{(2)}}^{(2)}) - \log(\hat{X}_{\lambda_{i_2-1}^{(2)}}^{(2)}) \big) \mathbb{I}_{[l_{i_1-1}^{(1)}, l_{i_1}^{(1)}] \cap [l_{i_2-1}^{(2)}, l_{i_2}^{(2)}] \neq \emptyset}.$$

To establish our result, we will require the following assumption.

Assumption 2. *For all k and m, the rank of $t_k^{(m)}$ among the set of all the $t_i^{(1)}$ and $t_j^{(2)}$ is the same as the rank of $\tau_k^{(m)}$ among the set of all the $\tau_i^{(1)}$ and $\tau_j^{(2)}$.*

Remark that under Assumption 2, $\mathbb{I}_{[l_{i_1-1}^{(1)}, l_{i_1}^{(1)}] \cap [l_{i_2-1}^{(2)}, l_{i_2}^{(2)}] \neq \emptyset}$ is equal to $\mathbb{I}_{[\lambda_{i_1-1}^{(1)}, \lambda_{i_1}^{(1)}] \cap [\lambda_{i_2-1}^{(2)}, \lambda_{i_2}^{(2)}] \neq \emptyset}$. We have the following theorem.

THEOREM 3. *Assume that $\alpha^{(2)} = c\alpha^{(1)}$ with $c > 0$. Under Assumption 2, as $\alpha^{(1)}$ tends to zero, we have*

$$\widehat{RCV}_t \overset{u.c.p.}{\to} \int_0^t \rho_s \sigma_s^{(1)} \sigma_s^{(2)} ds.$$

Thus, the problem of estimating the integrated co-volatility of two assets is another example which shows that our method consisting in estimating the values of the efficient price is very convenient to adapt classical statistical procedures to the microstructure noise context. Remark that a limit theorem is probably hard to obtain through our techniques because it would require a time change adapted to both assets.

4. Proofs. In all the proofs, c denotes a positive constant that may vary from line to line and $(\alpha_n)_{n\geq 0}$ is a sequence tending to zero. So, we write $\tau_{i,n}$ for τ_i, $t_{i,n}$ for t_i and $L_{i,n}$ for L_i. We define $\mathcal{E}^n$ as the complete right-continuous filtration generated by $(X_t, \chi_t, N^{(a)}_{\alpha_n,t,k}, N^{(c)}_{\alpha_n,t,k},\ k = 1,\dots,m)$. Moreover, without loss of generality, we consider that the semi-martingale χ is a one-dimensional process of the form

$$\chi_t = \chi_0 + \int_0^t a_u^\chi \mathrm{d}u + \int_0^t \sigma_{u-}^\chi \mathrm{d}\hat{W}_u, \text{ for } t \leq T,$$

with $\hat{W}$ a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ and we set $\chi_t = \chi_T$ for $t > T$.

4.1. *Preliminary remarks.* We introduce in this section some tools we use throughout the proofs.

A convenient construction of the $L_{i,n}$. We write $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$. The processes X and χ are measurable with respect to $\mathcal{F}_1$ and $\mathcal{F}_2$ is the filtration generated by a Brownian motion W', independent of $\mathcal{F}_1$. Let Φ denote the cumulative distribution function of a standard Gaussian random variable. We define

$$g_{t,n} = \sup\{\tau_{j,n} : \tau_{j,n} < t\},$$

$$L'_t = \sum_{k=1}^m k\mathbb{I}\{\Phi(\frac{W'_t - W'_{g_{t,n}}}{\sqrt{t - g_{t,n}}}) \in [\sum_{j=1}^{k-1} p_j(\chi_t), \sum_{j=1}^{k} p_j(\chi_t))\}$$

and $L_{i,n} = L'_{\tau_{i,n}}$, with the usual convention $\sum_{j=1}^{0} p_j(\chi_t) = 0$. So defined, for $k = 1,\dots,m$, we have $\mathbb{P}_{\mathcal{E}^n_{\tau_{i,n}}}[L_{\tau_{i,n}} = k] = p_k(\chi_{\tau_{i,n}})$. This construction is particularly convenient for the localization procedure and change of probability that we now explain.

Localization and change of probability. From the assumptions of Section 2.1, there exist an increasing sequence of stopping times T_q and a real sequence $K_q > 0$ such that $T_q \overset{a.s.}{\to} T$ as q tends to infinity and for $0 \leq t \leq T_q$,

$$|Y_t| + |\chi_t| + |\sigma_{t-}| + |a_t| + |\sigma_{t-}^{-1}| + |\sigma_{t-}^\chi| + |a_t^\chi| \leq K_q.$$

For $q > 0$, let $(Y_t^{(q)})_{t\geq 0}$ and $(\chi_t^{(q)})_{t\geq 0}$ be defined by $Y_0^{(q)} = y_0$, $\chi_0^{(q)} = \chi_0$ and

$$\mathrm{d}Y_t^{(q)} = a_t^{(q)} \mathrm{d}t + \sigma_{t-}^{(q)} \mathrm{d}W_t,$$

$$\mathrm{d}\chi_t^{(q)} = (a_t^\chi)^{(q)}\mathrm{d}t + (\sigma_{t-}^\chi)^{(q)}\mathrm{d}\hat{W}_t,$$

with $a_t^{(q)} = a_{t\wedge T_q}$, $(a_t^\chi)^{(q)} = a_{t\wedge T_q}^\chi$, $\sigma_t^{(q)} = \sigma_t$, $(\sigma_t^\chi)^{(q)} = \sigma_t^\chi$ if $t < T_q$ and $\sigma_t^{(q)} = \sigma_{T_q-}$, $(\sigma_t^\chi)^{(q)} = \sigma_{T_q-}^\chi$ if $t \geq T_q$. We have that $Y_t^{(q)}$ (resp. $\chi_t^{(q)}$) coincides with Y_t (resp. χ_t) on $[0, T_q]$.

Now consider a twice differentiable real function Φ_q on $\mathbb{R}$ such that for $x \in [-K_q, K_q]$, $\Phi_q(x) = x$ and for all $|x| \geq K_q$, $|\Phi_q(x)| \leq c_1$, $|\Phi_q^{'}(x)| + |\Phi_q^{''}(x)| \leq c_2$ and $|\Phi_q^{''}(x)/\Phi_q^{'}(x)| \leq c_3$, where c_1, c_2 and c_3 are positive constants. Let $\tilde{Y}_t^{(q)} = \Phi_q(Y_t^{(q)})$. As soon as $y_0 \in [-K_q, K_q]$, $\tilde{Y}_0^{(q)} = y_0$ and by Ito's formula

$$\mathrm{d}\tilde{Y}_t^{(q)} = \tilde{a}_t^{(q)}\mathrm{d}t + \tilde{\sigma}_{t-}^{(q)}\mathrm{d}W_t,$$

where

$$\tilde{a}_t^{(q)} = \Phi_q^{'}(Y_t^{(q)})a_t^{(q)} + \Phi_q^{''}(Y_t^{(q)})(\sigma_{t-}^{(q)})^2/2 \text{ and } \tilde{\sigma}_t^{(q)} = \Phi_q^{'}(Y_t^{(q)})\sigma_t^{(q)}.$$

Again, we have that $\tilde{Y}_t^{(q)}$ coincides with Y_t on $[0, T_q]$. Define

$$W_t^{(q)} = W_t + \int_0^t \frac{(\tilde{a}_s^{(q)} + (\tilde{\sigma}_{s-}^{(q)})^2/2)}{\tilde{\sigma}_{s-}^{(q)}}\mathrm{d}s.$$

In this setting, we easily see that Novikov's criterion holds and so, applying Girsanov's theorem, $(W_t^{(q)})_{t\in[0,T]}$ is a Brownian motion under a measure $\mathbb{P}^{(q)}$ absolutely continuous with respect to $\mathbb{P}$ on $[0, T]$. Note that conditional on $\mathcal{E}_{\tau_{i,n}}^n$, the distribution of $L_{i,n}$ is not modified when considering $\mathbb{P}^{(q)}$ instead of $\mathbb{P}$. Let $\tilde{X}_t^{(q)} = \exp(\tilde{Y}_t^{(q)})$. We finally have

$$\mathrm{d}\tilde{X}_t^{(q)} = \tilde{X}_t^{(q)}\tilde{\sigma}_{t-}^{(q)}\mathrm{d}W_t^{(q)}.$$

Using the preceding construction and the fact that the convergence in probability and the stable convergence in law are preserved by absolutely continuous change of probability, it is no restriction for proving our theorems to consider from now that the following additional assumption is granted.

Assumption 3. *For all $t \in [0, T]$, $a_t = -\sigma_{t-}^2/2$ and there exists a constant $K > 0$ such that,*

$$|Y_t| + |\chi_t| + |\sigma_t| + |\sigma_t^\chi| + |a_t^\chi| \leq K.$$

Time change. In the following, it is sometimes useful to view our price process as a time-changed Brownian motion. For that purpose, we introduce the process $(Z_t)_{t\geq 0}$ with infinite bracket at infinity defined by

$$Z_t = X_{t\wedge T} + \int_0^t \mathbb{I}_{s>T}\mathrm{d}W_s.$$

Let

$$\mathcal{T}(s) = \inf\{t \geq 0 : \langle Z\rangle_t > s\}.$$

By Dubins-Schwarz Theorem for continuous local martingales (see for example [28], Theorem V.1.6), there exists a $\mathcal{F}_{\mathcal{T}(s)}$-adapted Brownian motion $(B_s)_{s\geq 0}$ such that for $t \geq 0$,

$$B_{\langle Z\rangle_t} + x_0 = Z_t.$$

Hence, for $t \in [0, T]$, $B_{\langle X\rangle_t} + x_0 = X_t$. From now, we redefine $\tau_{i,n}$ and $L_{i,n}$ in the same way as in Equation (1) and Equation (4) replacing X_t by Z_t. For simplicity, we keep from now the notation X for Z. We set $\nu_{i,n} = \langle X\rangle_{\tau_{i,n}}$ and $\Delta\nu_{i,n} = \nu_{i+1,n} - \nu_{i,n}$.

4.2. *Technical lemmas.* We state in this section some lemmas that we widely use throughout the proofs of our three theorems. The first technical lemma is an analogous of Lemma 9 in [13] in the case of random sampling.

LEMMA 1. *Let $(\mathcal{G}_t)_{t\geq 0}$ be a filtration, $(\nu_{i,n})$ be an array of increasing sequences of stopping times and $(\xi_{\nu_{i,n}})$ an array of $\mathcal{G}_{\nu_{i,n}}$-mesurable random variables. Let ν be another stopping time such that $\nu \leq c_\nu$ with c_ν a positive constant. Let $N_{n,t} = \sup\{i : \nu_{i,n} \leq t\}$. Suppose that N_{n,c_ν} is finite almost surely and that there exists a positive deterministic sequence v_n such that the sequence $(v_n N_{n,c_\nu})_{n\geq 0}$ is tight. Moreover, assume that*

$$\sum_{\nu_{i,n}\leq\nu} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi^2_{\nu_{i,n}}] + \sup_{i\geq 1} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi^2_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}\leq\nu}] \xrightarrow{\mathbb{P}} 0$$

and

$$\mathbb{E}[\sup_{i\leq N_{n,c_\nu}+1} \xi^2_{\nu_{i,n}}] \to 0.$$

Then, $\displaystyle\sum_{\nu_{i,n}\leq\nu} \xi_{\nu_{i,n}} \xrightarrow{\mathbb{P}} U$ *is equivalent to* $\displaystyle\sum_{\nu_{i,n}\leq\nu} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}}] \xrightarrow{\mathbb{P}} U.$

PROOF. Let us consider the implication $(\rightarrow)$. Define

$$\xi_{\nu_{i,n}}(t) = \xi_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}\leq t\leq\nu} - \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}\leq t\leq\nu}].$$

Since $\displaystyle\sum_{i\geq 1} \xi_{\nu_{i,n}}(\nu)$ is equal to

$$\sum_{\nu_{i,n}\leq\nu} \xi_{\nu_{i,n}} - \sum_{\nu_{i,n}\leq\nu} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}}] + \sum_{\nu_{i,n}\leq\nu} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}}] - \sum_{i\geq 1} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}\leq\nu}],$$

it is enough to show that

$$\sum_{\nu_{i,n}\leq\nu} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}}] - \sum_{i\geq 1} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}\leq\nu}] \xrightarrow{\mathbb{P}} 0$$

and

$$\sum_{i\geq 1} \xi_{\nu_{i,n}}(\nu) \xrightarrow{\mathbb{P}} 0.$$

The first term is equal to $e_{1,n} - e_{2,n}$ with

$$e_{1,n} = \sum_{\nu_{i,n}\leq\nu} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}>\nu}] \text{ and } e_{2,n} = \sum_{\nu_{i,n}>\nu} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}\leq\nu}].$$

Using that

$$\sum_{i\geq 1} \mathbb{I}_{\nu_{i-1,n}>\nu} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}\leq\nu}] = 0,$$

we easily obtain

$$e_{2,n} = \sum_{i\geq 1} (\mathbb{I}_{\nu_{i,n}>\nu} - \mathbb{I}_{\nu_{i-1,n}>\nu}) \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}\leq\nu}].$$

Hence

$$|e_{2,n}| \leq \sup_{i\geq 1} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[|\xi_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}\leq \nu}|] \leq \sqrt{\sup_{i\geq 1} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi^2_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}\leq \nu}|]}$$

and so $|e_{2,n}| \xrightarrow{\mathbb{P}} 0$. Since

$$\sum_{i\geq 1} \mathbb{I}_{\nu_{i,n}\leq \nu} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}} \mathbb{I}_{\nu_{i-1,n}>\nu}] = 0,$$

we get in the same way

$$e_{1,n} = \sum_{i\geq 1} \mathbb{I}_{\nu_{i,n}\leq \nu} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}}(\mathbb{I}_{\nu_{i,n}>\nu} - \mathbb{I}_{\nu_{i-1,n}>\nu})].$$

Remarking that

$$\mathbb{E}[|e_{1,n}|] \leq \mathbb{E}\Big[\sum_{i\geq 1} |\xi_{\nu_{i,n}}| \mathbb{I}_{\nu_{i-1,n}\leq \nu \leq \nu_{i,n}}\Big] \leq \mathbb{E}[\sup_{i\leq N_{n,c_\nu}+1} |\xi_{\nu_{i,n}}|] \leq \sqrt{\mathbb{E}[\sup_{i\leq N_{n,c_\nu}+1} \xi^2_{\nu_{i,n}}]},$$

we obtain $|e_{1,n}| \xrightarrow{\mathbb{P}} 0$. Let

$$e_{3,n}(t) = \sum_{i\geq 1} \xi_{\nu_{i,n}}(t).$$

We now prove that $e_{3,n}(\nu) \xrightarrow{\mathbb{P}} 0$. Let $\varepsilon > 0$ and $M_\varepsilon > 0$ be such that

$$\sup_n \mathbb{P}[v_n N_{n,c_\nu} > M_\varepsilon] \leq \varepsilon/3.$$

Let $A > 0$, we have

$$\mathbb{P}[e_{3,n}(\nu)^2 \geq A] \leq \mathbb{P}[e_{3,n}(\nu)^2 \mathbb{I}_{\{v_n N_{n,c_\nu} \leq M_\varepsilon\}} \geq A] + \varepsilon/3.$$

Let $M_{\varepsilon,n} = \lfloor M_\varepsilon / v_n \rfloor + 1$ and

$$e_{4,\varepsilon,n}(t) = \sum_{i=1}^{M_{\varepsilon,n}} \xi_{\nu_{i,n}}(t).$$

For a given stopping time $\rho \leq \nu$, $\xi_{\nu_{i,n}}(\rho)$ is a martingale increment with respect to $(\mathcal{G}_{\nu_{i,n}})$. So, we have

$$\mathbb{E}[e_{4,\varepsilon,n}(\rho)^2] = \mathbb{E}\Big[\sum_{i=1}^{M_{\varepsilon,n}} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi_{\nu_{i,n}}(\rho)^2]\Big] \leq \mathbb{E}\Big[\sum_{i=1}^{M_{\varepsilon,n}} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi^2_{\nu_{i,n}} \mathbb{I}_{\nu_{i,n}\leq \rho \leq c_\nu}]\Big].$$

This is also less than

$$\mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi^2_{\nu_{i,n}} \mathbb{I}_{\nu_{i-1,n}<\rho\leq c_\nu}]\Big] + \mathbb{E}[\sup_{i\leq N_{n,c_\nu}+1} \xi^2_{\nu_{i,n}}].$$

It follows that the process $t \to e_{4,\varepsilon,n}(t)$ is L-dominated by the predictable process

$$t \to \sum_{i=1}^{M_{\varepsilon,n}} \mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi^2_{\nu_{i,n}} \mathbb{I}_{\nu_{i-1,n}<t\leq \nu}] + \mathbb{E}[\sup_{i\leq N_{n,c_\nu}+1} \xi^2_{\nu_{i,n}}]$$

on $[0,\nu]$, see for details I.3.29 in [21]. Now, since the preceding process is an increasing process on $[0,\nu]$, we can apply Lenglart's inequality and we obtain that for all $\eta>0$ and $A>0$,

$$\mathbb{P}[e_{4,\varepsilon,n}(\nu)^2\geq A]\leq \mathbb{P}[\sup_{t\in[0,\nu]}\{e_{4,\varepsilon,n}(t)^2\}\geq A]\leq \frac{\eta}{A}+\mathbb{P}\Big[\mathbb{E}[\sup_{i\leq N_{n,c\nu}+1}\xi^2_{\nu_{i,n}}]+\sum_{i=1}^{M_{\varepsilon,n}}\mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi^2_{\nu_{i,n}}\mathbb{I}_{\nu_{i,n}\leq\nu}]\geq\eta\Big].$$

Eventually, we get

$$\mathbb{P}[e_{4,\varepsilon,n}(\nu)^2\geq A]\leq\frac{\eta}{A}+\mathbb{P}\Big[\mathbb{E}[\sup_{i\leq N_{n,c\nu}+1}\xi^2_{\nu_{i,n}}]+\sum_{i\geq1}\mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi^2_{\nu_{i,n}}\mathbb{I}_{\nu_{i,n}\leq\nu}]\geq\eta\Big].$$

By the same arguments as for the first term, we have

$$\sum_{i\geq1}\mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi^2_{\nu_{i,n}}\mathbb{I}_{\nu_{i,n}\leq\nu}]-\sum_{\nu_{i,n}\leq\nu}\mathbb{E}_{\mathcal{G}_{\nu_{i-1,n}}}[\xi^2_{\nu_{i,n}}]\xrightarrow{\mathbb{P}}0.$$

Let us take $\eta=A\varepsilon/3$. For big enough n, the second term of the inequality is less than $\varepsilon/3$ and so the result follows. The other implication is proved exactly the same way. □

We now give results on the first exit time of a Brownian motion $(B_s)_{s\geq0}$. Let α be a positive constant, k a positive integer and $0<\eta<1$. We write

$$\nu_{(k)}=\min\{s: B_s\notin(-\alpha(k-1+2\eta),\alpha k)\}.$$

The following formulas can be found in [8].

Lemma 2. *For $u>0$,*

$$\begin{aligned}\mathbb{E}[e^{-u\nu_{(k)}}\mathbb{I}_{\{B_{\nu_{(k)}}=\alpha k\}}] &= \frac{sh(\alpha(k-1+2\eta)\sqrt{2u})}{sh(\alpha(2k-1+2\eta)\sqrt{2u})},\\ \mathbb{E}[e^{-u\nu_{(k)}}\mathbb{I}_{\{B_{\nu_{(k)}}=-\alpha(k-1+2\eta)\}}] &= \frac{sh(\alpha k\sqrt{2u})}{sh(\alpha(2k-1+2\eta)\sqrt{2u})}.\end{aligned}$$

Using the preceding lemma, tedious but straightforward computations lead to the following corollary.

Corollary 1.

$$P[B_{\nu_{(k)}}=\alpha k]=\frac{k-1+2\eta}{2k-1+2\eta},\quad P[B_{\nu_{(k)}}=-\alpha(k-1+2\eta)]=\frac{k}{2k-1+2\eta}$$

and

$$\mathbb{E}[B_{\nu_{(k)}}]=0,\quad \mathbb{E}[|B_{\nu_{(k)}}|]=\alpha\frac{2k(k-1+2\eta)}{2k-1+2\eta},\quad \mathbb{E}[B^2_{\nu_{(k)}}]=\alpha^2k(k-1+2\eta)=\mathbb{E}[\nu_{(k)}],$$

$$\begin{aligned}
\mathbb{E}[B^4_{\nu_{(k)}}] &= \alpha^4 k(k-1+2\eta)(k^2-k+2\eta k+4\eta^2-4\eta+1),\\
\mathbb{E}[\nu_{(k)}\mathbb{I}_{\{B_{\nu_{(k)}}=\alpha k\}}] &= \alpha^2\frac{k(k-1+2\eta)(3k-2+4\eta)}{3(2k-1+2\eta)},\\
\mathbb{E}[\nu^2_{(k)}\mathbb{I}_{\{B_{\nu_{(k)}}=\alpha k\}}] &= \alpha^4\frac{k(k-1+2\eta)(3k-2+4\eta)(25k^2-22k+4+16\eta^2+44k\eta-16\eta)}{90(2k-1+2\eta)},\\
\mathbb{E}[\nu_{(k)}\mathbb{I}_{\{B_{\nu_{(k)}}=-\alpha(k-1+2\eta)\}}] &= \alpha^2\frac{k(k-1+2\eta)(3k-1+2\eta)}{3(2k-1+2\eta)},\\
\mathbb{E}[\nu^2_{(k)}\mathbb{I}_{\{B_{\nu_{(k)}}=-\alpha(k-1+2\eta)\}}] &= \alpha^4\frac{k(k-1+2\eta)(3k-1+2\eta)(25k^2-28k+7+28\eta^2+56k\eta-28\eta)}{90(2k-1+2\eta)},\\
\mathbb{E}[(\nu_{(k)})^2] &= \alpha^4\frac{k(k-1+2\eta)(5k^2-5k+1+4\eta^2+10k\eta-4\eta)}{6}.
\end{aligned}$$

4.3. *Proof of Theorem 1.* We begin this section with the following result.

LEMMA 3. *We have*

$$\sup_{\{i:[\tau_{i,n},\tau_{i+1,n}]\subset[0,T]\}}(\tau_{i+1,n}-\tau_{i,n})\to 0,\ \ a.s.$$

PROOF. For $\omega\in\Omega$, let $S(\omega)$ be defined by

$$S(\omega)=\limsup\sup_{\{i:[\tau_{i,n}(\omega),\tau_{i+1,n}(\omega)]\subset[0,T]\}}\big(\tau_{i+1,n}(\omega)-\tau_{i,n}(\omega)\big).$$

Assume that S is not almost surely equal to zero. Let Ω' be such that $\mathbb{P}[\Omega']>0$ and such that for all $\omega\in\Omega'$, $t\to X_t(\omega)$ is $(1/2-\gamma)$ Hölder continuous for some γ, $0<\gamma<1/2$ and $S(\omega)>0$. For given $\omega\in\Omega'$, we define $\tau_n^1(\omega)$ and $\tau_n^2(\omega)$ the left and right bound of the first interval $[\tau_{i,n}(\omega),\tau_{i+1,n}(\omega)]$ such that

$$i=\operatorname*{argmax}_{\{j:[\tau_{j,n}(\omega),\tau_{j+1,n}(\omega)]\subset[0,T]\}}\big(\tau_{j+1,n}(\omega)-\tau_{j,n}(\omega)\big).$$

Thus

$$S(\omega)=\limsup\big(\tau_n^2(\omega)-\tau_n^1(\omega)\big).$$

Let

$$M_n(\omega)=\sup_{\{t\in[\tau_n^1(\omega),\tau_n^2(\omega)]\}}\big(X_t(\omega)-X_{\tau_n^1(\omega)}(\omega)\big).$$

We have $M_n(\omega)\le c\alpha_n$ and so $M_n(\omega)\to 0$. There exists an increasing function ψ and two constants $0\le c_1<c_2$ such that

$$\lim\tau^1_{\psi(n)}(\omega)=c_1 \text{ and } \lim\tau^2_{\psi(n)}(\omega)=c_2.$$

Let $0<\varepsilon<(c_2-c_1)/3$ and n_0 be such that for all $m\ge n_0$, $|\tau^1_{\psi(m)}(\omega)-c_1|<\varepsilon$ and $|\tau^2_{\psi(m)}(\omega)-c_2|<\varepsilon$. Let $t_0\in[c_1+(c_2-c_1)/3,c_2-(c_2-c_1)/3]$. For all $m\ge n_0$, $t_0\in[\tau^1_{\psi(m)}(\omega),\tau^2_{\psi(m)}(\omega)]$ and so

$$|X_{t_0}(\omega)-X_{c_1}(\omega)|\le M_{\psi(m)}(\omega)+c\varepsilon^{1/2-\gamma},$$

with $0<\gamma<1/2$. So we obtain that $t\to X_t(\omega)$ is constant on $[c_1+(c_2-c_1)/3, c_2-(c_2-c_1)/3]$. Finally, on $[0,T]$, with positive probability, $t\to X_t$ is constant on some interval with non empty interior which is absurd since σ_t is positive. We deduce that S is equal to zero almost surely, which concludes. $\square$

Let

$$N_{\alpha_n,t}=\sup\{i:\tau_{i,n}\in[0,t]\}=\sum_{k=1}^m(N^{(a)}_{\alpha_n,t,k}+N^{(c)}_{\alpha_n,t,k})$$

denote the number of transactions with price change on $[0,t]$. We have the following lemma.

Lemma 4. *For any $0<t\leq T$, the sequence $(\alpha_n^2 N_{\alpha_n,t})_{n\geq 0}$ is tight.*

Proof. There exists a positive constant c such that

$$c\alpha_n^2 N_{\alpha_n,t}\leq \sum_{i=1}^{N_{\alpha_n,t}}(X_{\tau_{i,n}}-X_{\tau_{i-1,n}})^2.$$

Moreover,

$$\sum_{i=1}^{N_{\alpha_n,t}}(X_{\tau_{i,n}}-X_{\tau_{i-1,n}})^2=-(X_t-X_{N_{\alpha_n,t}})^2+\sum_{i\geq 1}(X_{\tau_{i,n}\wedge t}-X_{\tau_{i-1,n}\wedge t})^2.$$

Using Lemma 3 together with Theorem I.4.47 in [21], we easily obtain that

$$\sum_{i=1}^{N_{\alpha_n,t}}(X_{\tau_{i,n}}-X_{\tau_{i-1,n}})^2\overset{\mathbb{P}}{\to}\int_0^t X_s^2\sigma_s^2\mathrm{d}s,$$

which concludes. $\square$

We deduce the following result.

Lemma 5. *We have*

$$RV_{\alpha_n,t}=\sum_{i=1}^{N_{\alpha_n,t}}\Big(\frac{X_{\tau_{i,n}}-X_{\tau_{i-1,n}}}{X_{\tau_{i-1,n}}}\Big)^2\overset{u.c.p.}{\to}\int_0^t\sigma_s^2\,ds.$$

Proof. Using the same method as in the preceding lemma, we get

$$\sum_{i=1}^{N_{\alpha_n,t}}\big(\log(X_{\tau_{i,n}})-\log(X_{\tau_{i-1,n}})\big)^2\overset{u.c.p.}{\to}\int_0^t\sigma_s^2\mathrm{d}s.$$

Then, using the fact that $|X_{\tau_{i,n}}-X_{\tau_{i-1,n}}|\leq c\alpha_n$ together with Assumption 3, it is easily shown that

$$\sum_{i=1}^{N_{\alpha_n,t}}\Big(\big(\log(X_{\tau_{i,n}})-\log(X_{\tau_{i-1,n}})\big)^2-\Big(\frac{X_{\tau_{i,n}}-X_{\tau_{i-1,n}}}{X_{\tau_{i-1,n}}}\Big)^2\Big)\overset{u.c.p.}{\to}0.$$

$\square$

We now state a result on the number of continuations of the process.

LEMMA 6. *For* $1 \leq k \leq m$,

$$\alpha_n^2 N_{\alpha,t,k}^{(c)} \overset{u.c.p.}{\to} \frac{k-1+2\eta}{2k-1+2\eta} \int_0^t p_k(\chi_u)\varphi(\chi_u)\sigma_u^2 X_u^2 du.$$

PROOF. We have

$$\alpha_n^2 N_{\alpha_n,t,k}^{(c)} = \alpha_n^2 \sum_{i=1}^{N_{\alpha_n,t}} \mathbb{I}_{\{|X_{\tau_{i,n}} - X_{\tau_{i-1,n}}| = \alpha k\}} = \alpha_n^2 \sum_{\nu_{i,n} \leq \langle X \rangle_t} \mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n k\}}.$$

Note that $N_{\alpha_n,t} = N'_{n,\langle X \rangle_t}$ with $N'_{n,u} = \sup\{i : \nu_{i,n} \leq u\}$. By Assumption 3, $\langle X \rangle_t \leq d$, where d is a positive constant. In the same way as in Lemma 4, we easily show that the sequence $(\alpha_n^2 N'_{n,d})_{n \geq 0}$ is tight. So, it is clear that

$$\alpha_n^4 \sum_{\nu_{i,n} \leq \langle X \rangle_t} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[\mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n k\}}] \overset{\mathbb{P}}{\to} 0$$

and

$$\alpha_n^4 \sup_{i \geq 1} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[\mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n k\}} \mathbb{I}_{\nu_{i,n} \leq \langle X \rangle_t}] \overset{\mathbb{P}}{\to} 0.$$

Lemma 1 gives that $\alpha_n^2 N_{\alpha_n,t,k}^{(c)}$ and

$$\alpha_n^2 \sum_{\nu_{i,n} \leq \langle X \rangle_t} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[\mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n k\}}]$$

have the same limit in probability. Using Corollary 1, we get that the preceding quantity is equal to

$$\begin{aligned}
&\alpha_n^2 \sum_{\nu_{i,n} \leq \langle X \rangle_t} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[\mathbb{E}_{\mathcal{F}_{\mathcal{T}(\nu_{i-1,n})}}[\mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n k\}}]] \\
&= \alpha_n^2 \sum_{\nu_{i,n} \leq \langle X \rangle_t} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[\mathbb{I}_{\{L_{i-1,n} = k\}} \frac{k-1+2\eta}{2k-1+2\eta}] \\
&= \alpha_n^2 \frac{k-1+2\eta}{2k-1+2\eta} \sum_{\nu_{i,n} \leq \langle X \rangle_t} p_k(\chi_{\mathcal{T}(\nu_{i-1,n})}).
\end{aligned}$$

By Corollary 1,

$$\mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[\Delta\nu_{i-1,n}] = \alpha_n^2 \sum_{k=1}^m p_k(\chi_{\mathcal{T}(\nu_{i-1,n})}) k(k-1+2\eta) \tag{6}$$

and

$$\alpha_n^2 \sum_{\nu_i \leq \langle X \rangle_t} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[p_k(\chi_{\mathcal{T}(\nu_{i-1,n})})] = \sum_{\nu_{i,n} \leq \langle X \rangle_t} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[p_k(\chi_{\mathcal{T}(\nu_{i-1,n})})\varphi(\chi_{\mathcal{T}(\nu_{i-1,n})})\Delta\nu_{i,n}].$$

Using that $\mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[(\Delta\nu_{i,n})^2] \leq c\alpha_n^4$, Lemma 1 gives that

$$\alpha_n^2 \sum_{\nu_{i,n}\leq\langle X\rangle_t} p_k(\chi_{\mathcal{T}(\nu_{i-1,n})})$$

and

$$\sum_{\nu_{i,n}\leq\langle X\rangle_t} p_k(\chi_{\mathcal{T}(\nu_{i-1,n})})\varphi(\chi_{\mathcal{T}(\nu_{i-1,n})})\Delta\nu_{i,n}$$

have the same limit in probability. The function

$$s \to p_k(\chi_{\mathcal{T}(s)})\varphi(\chi_{\mathcal{T}(s)})$$

being almost surely continuous, using Lemma 3, we get

$$\sum_{\nu_{i,n}\leq\langle X\rangle_t} p_k(\chi_{\mathcal{T}(\nu_{i-1,n})})\varphi(\chi_{\mathcal{T}(\nu_{i-1,n})})\Delta\nu_{i,n} \overset{a.s.}{\to} \int_0^{\langle X\rangle_t} p_k(\chi_{\mathcal{T}(s)})\varphi(\chi_{\mathcal{T}(s)})\mathrm{d}s$$

and the result follows. □

The same method as in the preceding lemma gives the following result for the number of alternations.

Lemma 7. *For $1 \leq k \leq m$,*

$$\alpha_n^2 N^{(a)}_{\alpha_n,t,k} \overset{u.c.p.}{\to} \frac{k}{2k-1+2\eta}\int_0^t p_k(\chi_u)\varphi(\chi_u)\sigma_u^2 X_u^2\,du.$$

We now give the proof of Theorem 1.

Proof. Let

$$\widehat{RV}_{\alpha_n,t} = \sum_{i=1}^{N_{\alpha_n,t}} \Big(\frac{\hat{X}^t_{\tau_{i,n}} - \hat{X}^t_{\tau_{i-1,n}}}{\hat{X}^t_{\tau_{i-1,n}}}\Big)^2$$

where

$$\hat{X}^t_{\tau_{i,n}} = X_{\tau_{i,n}} m(\hat{\eta}_{\alpha_n,t}, P_{t_{i,n}})$$

with

$$m(\hat{\eta}_{\alpha_n,t}, P_{t_{i,n}}) = 1 - \frac{\alpha_n(\eta-\hat{\eta}_{\alpha_n,t})sign(P_{t_{i,n}} - P_{t_{i-1,n}})}{P_{t_{i,n}} - \alpha_n(\frac{1}{2}-\eta)sign(P_{t_{i,n}} - P_{t_{i-1,n}})}.$$

First note that by Lemma 6 and 7, $\hat{\eta}_{\alpha_n,t} \overset{u.c.p.}{\to} \eta$. Then it is easily shown that

$$\sum_{i=1}^{N_{\alpha_n,t}} \Big(\big(\log(\hat{X}^t_{\tau_{i,n}}) - \log(\hat{X}^t_{\tau_{i-1,n}})\big)^2 - \big(\frac{\hat{X}^t_{\tau_{i,n}} - \hat{X}^t_{\tau_{i-1,n}}}{\hat{X}^t_{\tau_{i-1,n}}}\big)^2\Big) \overset{u.c.p.}{\to} 0.$$

We have

$$\widehat{RV}_{\alpha_n,t} = \widetilde{RV}_{\alpha_n,t} + \sum_{i=1}^{N_{\alpha_n,t}} \big(\log(\frac{m(\hat{\eta}_{\alpha_n,t}, P_{t_{i,n}})}{m(\hat{\eta}_{\alpha_n,t}, P_{t_{i-1,n}})})\big)^2 + 2\sum_{i=1}^{N_{\alpha_n,t}} \big(\log(X_{\tau_{i,n}}) - \log(X_{\tau_{i-1,n}})\big)\big(\log(\frac{m(\hat{\eta}_{\alpha_n,t}, P_{t_{i,n}})}{m(\hat{\eta}_{\alpha_n,t}, P_{t_{i-1,n}})})\big),$$

where, using Lemma 5, $\widetilde{RV}_{\alpha_n,t} \overset{u.c.p.}{\to} \int_0^t \sigma_s^2 \mathrm{d}s$. We now prove that the other terms tend to 0. We have

$$|\log\big(\frac{m(\hat{\eta}_{\alpha_n,t}, P_{t_{i,n}})}{m(\hat{\eta}_{\alpha_n,t}, P_{t_{i-1,n}})}\big)| \leq c\alpha_n|\eta - \hat{\eta}_{\alpha_n,t}|$$

and

$$\sum_{i=1}^{N_{\alpha_n,t}}\big(\log(\frac{m(\hat{\eta}_{\alpha_n,t}, P_{t_{i,n}})}{m(\hat{\eta}_{\alpha_n,t}, P_{t_{i-1,n}})})\big)^2 \leq c\alpha_n^2 N_{\alpha_n,t}(\eta - \hat{\eta}_{\alpha_n,t})^2.$$

Moreover

$$|\log(X_{\tau_{i,n}}) - \log(X_{\tau_{i-1,n}})| \leq c\alpha_n$$

and

$$|\sum_{i=1}^{N_{\alpha_n,t}}\big(\log(X_{\tau_{i,n}}) - \log(X_{\tau_{i-1,n}})\big)\big(\log(\frac{m(\hat{\eta}_{\alpha_n,t}, P_{t_{i,n}})}{m(\hat{\eta}_{\alpha_n,t}, P_{t_{i-1,n}})})\big)| \leq c\alpha_n^2 N_{\alpha_n,t}|\eta - \hat{\eta}_{\alpha_n,t}|.$$

By Lemma 4, the result follows. □

4.4. *Proof of Theorem 2.* Let θ denote the reciprocal of the process f introduced before Theorem 2. Consider the time-changed process $(Z_u)_{u\geq 0}$ defined by

$$Z_u = X_{\theta_u}.$$

This process is adapted to the filtration $(\mathcal{H}_u)_{u\geq 0}$ with $\mathcal{H}_u = \mathcal{F}_{\theta_u}$. Moreover it is a $\mathcal{H}_u-$local martingale such that $\langle Z\rangle_u = \langle X\rangle_{\theta_u}$, see for example [28] p.181.

For each n, we define the filtration $\mathcal{H}^n$ by

$$\mathcal{H}_u^n = \mathcal{E}^n_{\tau_{\lfloor \alpha_n^{-2}u\rfloor,n}}.$$

Moreover, we define the process $Z(n)$ by

$$Z(n)_u = Z_{f_{\tau_{\lfloor \alpha_n^{-2}u\rfloor,n}}}.$$

LEMMA 8. *The process $Z(n)$ is a $\mathcal{H}^n$ martingale such that*

$$\langle Z(n)\rangle_u \overset{\mathbb{P}}{\to} \langle Z\rangle_u.$$

PROOF. We have

$$Z(n)_u = \sum_{i=1}^{\lfloor \alpha_n^{-2}u\rfloor}(Z(n)_{\alpha_n^2 i} - Z(n)_{\alpha_n^2(i-1)}) + x_0.$$

Since by Corollary 1

$$\mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[Z(n)_{\alpha_n^2 i} - Z(n)_{\alpha_n^2(i-1)}] = \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[B_{\nu_{i,n}} - B_{\nu_{i-1,n}}] = 0,$$

we easily obtain that $Z(n)$ is a $\mathcal{H}^n$ martingale such that

$$\langle Z(n)\rangle_u = \sum_{i=1}^{\lfloor \alpha_n^{-2}u\rfloor} \mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[(Z(n)_{\alpha_n^2 i} - Z(n)_{\alpha_n^2(i-1)})^2] = \sum_{i=1}^{\lfloor \alpha_n^{-2}u\rfloor} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[(B_{\nu_{i,n}} - B_{\nu_{i-1,n}})^2].$$

By Corollary 1, we get

$$\langle Z(n)\rangle_u = \sum_{\nu_{i,n}\leq \langle X\rangle_{\tau_{\lfloor \alpha_n^{-2}u\rfloor}}} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[\Delta \nu_{i,n}].$$

This can be written $T_1 + T_2 + T_3$ with

$$\begin{aligned} T_1 &= \sum_{\nu_{i,n}\leq \langle X\rangle_{\theta_u}} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[\Delta \nu_{i,n}] \\ T_2 &= \sum_{\langle X\rangle_{\tau_{\lfloor \alpha_n^{-2}u\rfloor}}\leq \nu_{i,n} < \langle X\rangle_{\theta_u}} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[\Delta \nu_{i,n}] \\ T_3 &= \sum_{\langle X\rangle_{\theta_u} < \nu_{i,n}\leq \langle X\rangle_{\tau_{\lfloor \alpha_n^{-2}u\rfloor}}} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[\Delta \nu_{i,n}]. \end{aligned}$$

By Lemma 1, T_1 tends to $\langle X\rangle_{\theta_u} = \langle Z\rangle_u$ in probability. By Corollary 1, we easily have

$$T_2 \leq c\alpha_n^2 |N_{\alpha_n,\theta_u} - \lfloor \alpha_n^{-2}u\rfloor|.$$

By Lemma 6 and Lemma 7, we have $\alpha_n^2 N_{\alpha_n,t} \overset{u.c.p}{\to} f_t$. Consequently, T_2 tends to 0 in probability. The same result holds for T_3. □

Let $\mathcal{M}_b$ denote the set of all bounded martingales on $(\Omega, \mathcal{H}, \mathbb{P})$. Let $N \in \mathcal{M}_b$ and $N(n)$ be defined by

$$N(n)_u = N_{f_{\tau_{\lfloor \alpha_n^{-2}u\rfloor,n}}}.$$

Using the sampling theorem, we get that $N(n)$ is a $\mathcal{H}^n$ martingale. Let $(N^1,\dots,N^m)$ be a finite family with elements in $\mathcal{M}_b$. Using that $f_{\tau_{\lfloor \alpha_n^{-2}u\rfloor,n}} \overset{u.c.p}{\to} u$ together with Theorem VI.6.37 b) in [21], we obtain the following convergence for the Skorohod topology on $\mathbb{D}^{m+1}[0,\infty)$

$$(Z(n), N^1(n),\dots,N^m(n)) \xrightarrow{\mathbb{P}} (Z, N^1,\dots,N^m).$$

Hence we have Property IX.7.1 in [21].

Let us now introduce the $(2m+1)$-dimensional process $K(n)$ defined by

$$K(n)_u = \sum_{i=1}^{\lfloor \alpha_n^{-2}u\rfloor} K_i(n)$$

where $K_i(n) = \big(K_{i,1}(n),\dots,K_{i,2m+1}(n)\big)$ and for $1 \leq k \leq m$,

$$\begin{aligned} K_{i,1}(n) &= \alpha_n^{-1}\Big(\frac{(B_{\nu_{i,n}} - B_{\nu_{i-1,n}})^2}{B^2_{\nu_{i-1,n}\wedge\langle X\rangle_T}} - \varpi_{i,n}\mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}\big[\frac{(B_{\nu_{i,n}} - B_{\nu_{i-1,n}})^2}{B^2_{\nu_{i-1,n}\wedge\langle X\rangle_T}}\big]\Big) \\ K_{i,2k}(n) &= \alpha_n^{-1}\Big(\alpha_n^2 \mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n k\}} - \varpi_{i,n}\alpha_n^2\mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}\big[\mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n k\}}\big]\Big) \\ K_{i,2k+1}(n) &= \alpha_n^{-1}\Big(\alpha_n^2 \mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n(2\eta+k-1)\}} - \varpi_{i,n}\alpha_n^2\mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}\big[\mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n(2\eta+k-1)\}}\big]\Big), \end{aligned}$$

with

$$\varpi_{i,n} = \frac{\Delta\nu_{i-1,n}}{\mathbb{E}_{\mathcal{E}^n_{T(\nu_{i-1,n})}}[\Delta\nu_{i-1,n}]}.$$

The random variables $K_{i,j}(n)$ are $\mathcal{H}^n_{\alpha_n^2 i}$ measurable and so $K(n)$ is a $\mathcal{H}^n$ martingale. To establish the stable convergence of this process, we will apply Theorem IX.7.3 in [21].

Let $N \in \mathcal{M}_b$ and orthogonal to Z and $N(n)$ being such that $N(n)_u = N_{f_{\tau_{\lfloor \alpha_n^{-2}u\rfloor,n}}}$.

LEMMA 9. *We have*

$$\langle K(n), N(n)\rangle_u \xrightarrow{\mathbb{P}} 0.$$

PROOF. Since $N(n)$ is a $\mathcal{H}^n$ martingale, we have that $\langle K_1(n), N(n)\rangle_u$ is equal to

$$\alpha_n^{-1}\sum_{i=1}^{\lfloor \alpha_n^{-2}u\rfloor} \frac{1}{Z(n)^2_{\alpha_n^2(i-1)\wedge\alpha_n^2 N_{\alpha_n,T}}}\mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[(Z(n)_{\alpha_n^2 i} - Z(n)_{\alpha_n^2(i-1)})^2(N(n)_{\alpha_n^2 i} - N(n)_{\alpha_n^2(i-1)})]$$
$$-\alpha_n^{-1}\sum_{i=1}^{\lfloor \alpha_n^{-2}u\rfloor} \mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[\frac{(Z(n)_{\alpha_n^2 i} - Z(n)_{\alpha_n^2(i-1)})^2}{Z(n)^2_{\alpha_n^2(i-1)\wedge\alpha_n^2 N_{\alpha_n,T}}}]\mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[\varpi_{i,n}(N(n)_{\alpha_n^2 i} - N(n)_{\alpha_n^2(i-1)})].$$

Let $1 \le k \le m$. We define $\tau_{i,n}^{(k)}$ the same way as $\tau_{i,n}$ by Equation (1) except that $L_{i,n}$ is replaced by k. Let $\nu_{i,n}^{(k)} = \langle X\rangle_{\tau_{i,n}^{(k)}}$, $\mathcal{B}_u = B_{u+\nu_{i-1,n}} - B_{\nu_{i-1,n}}$, $\mathcal{F}^{\mathcal{B}}$ be the filtration generated by $\mathcal{B}$ and $\gamma_u = \mathbb{E}_{\mathcal{F}^{\mathcal{B}}_u}[(B_{\nu_{i,n}^{(k)}} - B_{\nu_{i-1,n}})^2]$. Since γ is a martingale with respect to $\mathcal{F}^{\mathcal{B}}$, by the martingale representation theorem, we have for $u \ge 0$

$$\gamma_u = \gamma_0 + \int_0^u v_s \mathrm{d}\mathcal{B}_s = \gamma_0 + \int_{f_{\tau_{i-1,n}}}^{f_{\langle X\rangle^{-1}_{u+\nu_{i-1,n}}}} v_{\langle X\rangle_{\theta_s}-\nu_{i-1,n}}\mathrm{d}Z_s,$$

for some predictable process v. We set $v_u = 0$ for $u < 0$ and

$$\tilde\gamma_u = \gamma_0 + \int_0^u v_{\langle X\rangle_{\theta_s}-\nu_{i-1,n}}\mathrm{d}Z_s.$$

Hence, for given n, since Z is a $\mathcal{H}$ martingale, $\tilde\gamma_u$ is a bounded $\mathcal{H}$ martingale, orthogonal to N. Remarking that $\tilde\gamma_{f_{\tau_{i,n}^{(k)}}} = (B_{\nu_{i,n}^{(k)}} - B_{\nu_{i-1,n}})^2$ and using the fact that $\tilde\gamma N$ is a martingale together with the sampling theorem, we get

$$\mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[(Z(n)_{\alpha_n^2 i} - Z(n)_{\alpha_n^2(i-1)})^2\mathbb{I}_{\{L_{i-1,n}=k\}}(N(n)_{\alpha_n^2 i} - N(n)_{\alpha_n^2(i-1)})]$$
$$= \mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[\mathbb{I}_{\{L_{i-1,n}=k\}}\mathbb{E}_{\mathcal{F}_{\tau_{i-1,n}}}[\tilde\gamma_{f_{\tau_{i,n}^{(k)}}}(N_{f_{\tau_{i,n}^{(k)}}} - N_{f_{\tau_{i-1,n}}})]] = 0.$$

Using the same kind of computations for the second term, we get $\langle K_1(n), N(n)\rangle_u = 0$. In the same way, we obtain that for $1 \le k \le m$, $\langle K_{2k}(n), N(n)\rangle_u = \langle K_{2k+1}(n), N(n)\rangle_u = 0$. □

LEMMA 10. *For $\varepsilon > 0$ and $1 \le j \le 2m+1$, we have*

$$\sum_{i=1}^{\lfloor \alpha_n^{-2}u\rfloor} \mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[K_{i,j}(n)^2\mathbb{I}_{\{|K_{i,j}(n)|>\varepsilon\}}] \xrightarrow{\mathbb{P}} 0.$$

PROOF. It is clear that

$$|K_{i,j}(n)| \leq c\alpha_n(1+\varpi_{i,n}).$$

Using the fact that for $p \in \mathbb{N}$, $\mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[\varpi^p_{i,n}] \leq c_p$ together with Cauchy-Schwarz and Markov inequalities, we get

$$\mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[K_{i,j}(n)^2 \mathbb{I}_{\{|K_{i,j}(n)>\varepsilon\}}] \leq c\varepsilon^{-1}\alpha_n^{5/2}.$$

□

An obvious application of the Burkholder-Davis-Gundy inequality leads to the following lemma.

LEMMA 11. *Let h be a differentiable real function with bounded derivative. For $p > 0$, we have*

$$\mathbb{E}[|h(\chi_{\mathcal{T}(\nu_{i,n})}) - h(\chi_{\mathcal{T}(\nu_{i-1,n})})|^p] \leq c_p\alpha_n^p.$$

LEMMA 12. *For $1 \leq k \leq m$, we have*

$$\langle K_1(n), Z(n)\rangle_u \xrightarrow{\mathbb{P}} 0,\ \langle K_{2k}(n), Z(n)\rangle_u \xrightarrow{\mathbb{P}} 0,\ \langle K_{2k+1}(n), Z(n)\rangle_u \xrightarrow{\mathbb{P}} 0.$$

PROOF. i) $\langle K_1(n), Z(n)\rangle_u$ is equal to $A_n^1 + A_n^2$ with

$$A_n^1 = \alpha_n^{-1}\sum_{i=1}^{\lfloor \alpha_n^{-2}u\rfloor} \frac{1}{B^2_{\nu_{i-1,n}\wedge\langle X\rangle_T}} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}\big[(B_{\nu_{i,n}} - B_{\nu_{i-1,n}})^3\big]$$

and

$$A_n^2 = -\alpha_n^{-1}\sum_{i=1}^{\lfloor \alpha_n^{-2}u\rfloor} \frac{\mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}\big[(B_{\nu_{i,n}} - B_{\nu_{i-1,n}})^2\big]}{B^2_{\nu_{i-1,n}\wedge\langle X\rangle_T}} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}\big[\varpi_{i,n}(B_{\nu_{i,n}} - B_{\nu_{i-1,n}})\big].$$

The term A_n^1 can be written

$$\alpha_n^2\sum_{i=1}^{\lfloor \alpha_n^{-2}u\rfloor} \frac{\mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{U}} - \mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{D}}}{B^2_{\nu_{i-1,n}\wedge\langle X\rangle_T}} \sum_{k=1}^m a_k p_k(\chi_{\mathcal{T}(\nu_{i-1,n})}),$$

with $\mathcal{D} = \cup d_k$, $\mathcal{U} = \cup u_k$ and $a_k = -k(1-2\eta)(k-1+2\eta)$.

Using Lemma 11, we get that it is also equal to $R_n^1 + R_n^2$ with

$$R_n^1 = \alpha_n^2\sum_{i=1}^{\lfloor \alpha_n^{-2}u\rfloor} \frac{\mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{U}} - \mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{D}}}{B^2_{\nu_{i-2,n}\wedge\langle X\rangle_T}} \sum_{k=1}^m a_k p_k(\chi_{\mathcal{T}(\nu_{i-2,n})})$$

and $\mathbb{E}[|R_n^2|] \leq c\alpha_n$. Let

$$R^3_{n,i-1} = \frac{\mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{U}} - \mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{D}}}{B^2_{\nu_{i-2,n}\wedge\langle X\rangle_T}} \frac{\sum_{k=1}^m a_k p_k(\chi_{\mathcal{T}(\nu_{i-2,n})})}{2\sum_{k=1}^m b_k p_k(\chi_{\mathcal{T}(\nu_{i-2,n})})},$$

with $b_k = k(2k-1+2\eta)^{-1}$ and

$$R_n^3 = \alpha_n^2 \sum_{i=1}^{\lfloor \alpha_n^{-2} u \rfloor} \big(R_{n,i-1}^3 - \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-2,n})}}[R_{n,i-1}^3]\big).$$

From Lemma 9 in [13], the last term tends to zero in probability. We have

$$\mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-2,n})}}[R_{n,i-1}^3] = \frac{\big(1-2\sum_{k=1}^m b_k p_k(\chi_{\mathcal{T}(\nu_{i-2,n})})\big)\big(\mathbb{I}_{B_{\nu_{i-2,n}}\in\mathcal{U}} - \mathbb{I}_{B_{\nu_{i-2,n}}\in\mathcal{D}}\big)}{B^2_{\nu_{i-2,n}\wedge\langle X\rangle_T}} \frac{\sum_{k=1}^m a_k p_k(\chi_{\mathcal{T}(\nu_{i-2,n})})}{2\sum_{k=1}^m b_k p_k(\chi_{\mathcal{T}(\nu_{i-2,n})})}.$$

This is also equal to $S^1_{n,i-1} + S^2_{n,i-1}$ with

$$S^1_{n,i-1} = \frac{\big(1-2\sum_{k=1}^m b_k p_k(\chi_{\mathcal{T}(\nu_{i-3,n})})\big)\big(\mathbb{I}_{B_{\nu_{i-2,n}}\in\mathcal{U}} - \mathbb{I}_{B_{\nu_{i-2,n}}\in\mathcal{D}}\big)}{B^2_{\nu_{i-3,n}\wedge\langle X\rangle_T}} \frac{\sum_{k=1}^m a_k p_k(\chi_{\mathcal{T}(\nu_{i-3,n})})}{2\sum_{k=1}^m b_k p_k(\chi_{\mathcal{T}(\nu_{i-3,n})})}$$

and $\mathbb{E}[|S^2_{n,i-1}|] \leq c\alpha_n$. Finally, $R_n^3 = R_n^1 + R_n^4$ with $R_n^4 \leq c\alpha_n$ and so R_n^1 tends to zero in probability. We now turn to A_n^2. By Corollary 1, $\mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[\Delta\nu_{i-1,n}(B_{\nu_{i,n}} - B_{\nu_{i-1,n}})]$ is equal to

$$(\mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{U}} - \mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{D}})\Big(\sum_{k=1}^m p_k(\chi_{\mathcal{T}(\nu_{i-1,n})})\alpha_n^3 \frac{k(k-1+2\eta)}{3(2k-1+2\eta)}\big(k(3k-2+4\eta)-(k-1+2\eta)(3k-1+2\eta)\big)\Big).$$

By the same arguments as for A_n^1, A_n^2 tends to zero in probability.

ii) We have $\langle K_{2k}(n), Z(n)\rangle_u = B_n^{2k,1} + B_n^{2k,2}$, with

$$B_n^{2k,1} = \alpha_n^{-1} \sum_{i=1}^{\lfloor \alpha_n^{-2} u \rfloor} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}\big[\alpha_n^2 \mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n k\}}(B_{\nu_{i,n}} - B_{\nu_{i-1,n}})\big]$$

and

$$B_n^{2k,2} = -\alpha_n^{-1} \sum_{i=1}^{\lfloor \alpha_n^{-2} u \rfloor} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[\alpha_n^2 \mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n k\}}]\mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}\big[\varpi_{i,n}(B_{\nu_{i,n}} - B_{\nu_{i-1,n}})\big].$$

Moreover,

$$\begin{aligned} B_{\nu_{i,n}} - B_{\nu_{i-1,n}} &= (\mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{U}} - \mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{D}}) \\ &\quad \sum_{j=1}^m \mathbb{I}_{\{L_{i-1,n}=j\}}\big(\alpha_n j \mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n j\}} - \alpha_n(2\eta + j - 1)\mathbb{I}_{\{|B_{\nu_{i,n}} - B_{\nu_{i-1,n}}| = \alpha_n(2\eta+j-1)\}}\big). \end{aligned}$$

Hence we get

$$B_n^{2k} = \alpha_n^2 k \frac{k-1+2\eta}{2k-1+2\eta} \sum_{i=1}^{\lfloor \alpha_n^{-2} u \rfloor} (\mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{U}} - \mathbb{I}_{B_{\nu_{i-1,n}}\in\mathcal{D}}) p_k(\chi_{\mathcal{T}(\nu_{i-1,n})}).$$

Using the same trick as in i) we obtain $\langle K_{2k}(n), Z(n)\rangle_u \xrightarrow{\mathbb{P}} 0$. In the same way, we get the last result. □

Let

$$\psi_k^{(c)}(\chi_{\mathcal{T}(v)}) = \frac{(k-1+2\eta)p_k(\chi_{\mathcal{T}(v)})}{(2k-1+2\eta)j(j-1+2\eta)}, \quad \psi_k^{(a)}(\chi_{\mathcal{T}(v)}) = \frac{kp_k(\chi_{\mathcal{T}(v)})}{(2k-1+2\eta)j(j-1+2\eta)}$$

and

$$\begin{aligned} m_{k,1} &= k-1+2\eta,\ m_{k,2} = 2k-1+2\eta,\ m_{k,3} = m_{k,1}/m_{k,2}, \\ m_{k,4} &= k(3k-2+4\eta),\ m_{k,5} = k(3k-1+2\eta), \\ m_{k,6} &= (k^2(3k-2+4\eta) + (k-1+2\eta)^2(3k-1+2\eta)), \\ v(\chi_u) &= \frac{1}{6}\sum_{j=1}^{m} j(j-1+2\eta)p_j(\chi_u)(5j^2-5j+1+4\eta^2+10j\eta-4\eta). \end{aligned}$$

Lemma 13. *For $1 \le l, l' \le 2m+1$,*

$$\sum_{i=1}^{\lfloor \alpha_n^{-2}u \rfloor} \mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[K_{i,l}(n)K_{i,l'}(n)] \xrightarrow{\mathbb{P}} c_{l,l',u},$$

with

$$c_{1,1,u} = \frac{2}{3}\int_0^{\langle X\rangle_{\theta_u}} \sum_{k=1}^{m} km_{k,1}(k-2\eta+(k-1+2\eta)^2)p_k(\chi_{\mathcal{T}(v)})X^{-4}_{\mathcal{T}(v\wedge\langle X\rangle_T)}\varphi(\chi_{\mathcal{T}(v)})dv,$$

for $1 \le k \le m$,

$$c_{1,2k,u} = k^2\int_0^{\langle X\rangle_{\theta_u}} p_k(\chi_{\mathcal{T}(v)})X^{-2}_{\mathcal{T}(v\wedge\langle X\rangle_T)}\varphi(\chi_{\mathcal{T}(v)})dv - \frac{k}{3}m_{k,3}m_{k,4}\int_0^{\langle X\rangle_{\theta_u}} p_k(\chi_{\mathcal{T}(v)})X^{-2}_{\mathcal{T}(v\wedge\langle X\rangle_T)}\varphi(\chi_{\mathcal{T}(v)})dv$$
$$-\frac{k}{3}m_{k,3}m_{k,6}\int_0^{\langle X\rangle_{\theta_u}} p_k(\chi_{\mathcal{T}(v)})\psi_k^{(c)}(\chi_{\mathcal{T}(v)})X^{-2}_{\mathcal{T}(v\wedge\langle X\rangle_T)}\varphi(\chi_{\mathcal{T}(v)})dv + \int_0^{\langle X\rangle_{\theta_u}} \psi_k^{(c)}(\chi_{\mathcal{T}(v)})X^{-2}_{\mathcal{T}(v\wedge\langle X\rangle_T)}v(\chi_{\mathcal{T}(v)})\varphi(\chi_{\mathcal{T}(v)})dv,$$

for $1 \le k \le m$,

$$c_{1,2k+1,u} = m_{k,1}^2\int_0^{\langle X\rangle_{\theta_u}} p_k(\chi_{\mathcal{T}(v)})X^{-2}_{\mathcal{T}(v\wedge\langle X\rangle_T)}\varphi(\chi_{\mathcal{T}(v)})dv - \frac{k}{3}m_{k,3}m_{k,5}\int_0^{\langle X\rangle_{\theta_u}} p_k(\chi_{\mathcal{T}(v)})X^{-2}_{\mathcal{T}(v\wedge\langle X\rangle_T)}\varphi(\chi_{\mathcal{T}(v)})dv$$
$$-\frac{k}{3}m_{k,3}m_{k,6}\int_0^{\langle X\rangle_{\theta_u}} p_k(\chi_{\mathcal{T}(v)})\psi_k^{(a)}(\chi_{\mathcal{T}(v)})\tilde{X}^{-2}_{\mathcal{T}(v)}\varphi(\chi_{\mathcal{T}(v)})dv + \int_0^{\langle X\rangle_{\theta_u}} \psi_k^{(a)}(\chi_{\mathcal{T}(v)})X^{-2}_{\mathcal{T}(v\wedge\langle X\rangle_T)}v(\chi_{\mathcal{T}(v)})\varphi(\chi_{\mathcal{T}(v)})dv,$$

for $1 \le k, k' \le m$,

$$c_{2k,2k',u} = m_{k,3}\mathbb{I}_{k=k'}\int_0^{\langle X\rangle_{\theta_u}} p_k(\chi_{\mathcal{T}(v)})\varphi(\chi_{\mathcal{T}(v)})dv - \frac{k}{3}m_{k,3}m_{k,4}\int_0^{\langle X\rangle_{\theta_u}} \psi_{k'}^{(c)}(\chi_{\mathcal{T}(v)})p_k(\chi_{\mathcal{T}(v)})\varphi(\chi_{\mathcal{T}(v)})dv$$
$$-\frac{k'}{3}m_{k',3}m_{k',4}\int_0^{\langle X\rangle_{\theta_u}} \psi_k^{(c)}(\chi_{\mathcal{T}(v)})p_{k'}(\chi_{\mathcal{T}(v)})\varphi(\chi_{\mathcal{T}(v)})dv + \int_0^{\langle X\rangle_{\theta_u}} \psi_k^{(c)}(\chi_{\mathcal{T}(v)})\psi_{k'}^{(c)}(\chi_{\mathcal{T}(v)})v(\chi_{\mathcal{T}(v)})\varphi(\chi_{\mathcal{T}(v)})dv,$$

for $1 \le k, k' \le m$,

$$c_{2k+1,2k'+1,u} = km_{k,2}^{-1}\mathbb{I}_{k=k'}\int_0^{\langle X\rangle_{\theta_u}} p_k(\chi_{\mathcal{T}(v)})\varphi(\chi_{\mathcal{T}(v)})dv - \frac{k}{3}m_{k,3}m_{k,5}\int_0^{\langle X\rangle_{\theta_u}} \psi_{k'}^{(a)}(\chi_{\mathcal{T}(v)})p_k(\chi_{\mathcal{T}(v)})\varphi(\chi_{\mathcal{T}(v)})dv$$
$$-\frac{k'}{3}m_{k',3}m_{k',5}\int_0^{\langle X\rangle_{\theta_u}} \psi_k^{(a)}(\chi_{\mathcal{T}(v)})p_{k'}(\chi_{\mathcal{T}(v)})\varphi(\chi_{\mathcal{T}(v)})dv + \int_0^{\langle X\rangle_{\theta_u}} \psi_k^{(a)}(\chi_{\mathcal{T}(v)})\psi_{k'}^{(a)}(\chi_{\mathcal{T}(v)})v(\chi_{\mathcal{T}(v)})\varphi(\chi_{\mathcal{T}(v)})dv,$$

for $1 \leq k, k' \leq m$,

$$c_{2k,2k'+1,u} = -\frac{k}{3} m_{k,3} m_{k,4} \int_0^{\langle X\rangle_{\theta_u}} \psi_{k'}^{(a)}(\chi_{\mathcal{T}(v)}) p_k(\chi_{\mathcal{T}(v)}) \varphi(\chi_{\mathcal{T}(v)})\, dv$$
$$-\frac{k'}{3} m_{k',3} m_{k',5} \int_0^{\langle X\rangle_{\theta_u}} \psi_k^{(c)}(\chi_{\mathcal{T}(v)}) p_{k'}(\chi_{\mathcal{T}(v)}) \varphi(\chi_{\mathcal{T}(v)})\, dv + \int_0^{\langle X\rangle_{\theta_u}} \psi_k^{(c)}(\chi_{\mathcal{T}(v)}) \psi_{k'}^{(a)}(\chi_{\mathcal{T}(v)}) v(\chi_{\mathcal{T}(v)}) \varphi(\chi_{\mathcal{T}(v)})\, dv.$$

Proof. For simplicity we just prove the convergence to $c_{1,1,u}$. The other results are proved the same way. Using Corollary 1, we have that

$$\sum_{i=1}^{\lfloor \alpha_n^{-2} u \rfloor} \mathbb{E}_{\mathcal{H}^n_{\alpha_n^2(i-1)}}[K_{i,1}(n)^2]$$

is equal to

$$\frac{1}{B^4_{\nu_{i-1,n}\wedge\langle X\rangle_T}} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[((B_{\nu_{i,n}} - B_{\nu_{i-1,n}})^2 - \Delta\nu_{i-1,n})^2] = \frac{2}{3B^4_{\nu_{i-1,n}\wedge\langle X\rangle_T}} \mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}[(B_{\nu_{i,n}} - B_{\nu_{i-1,n}})^4].$$

Using the same arguments as in the end of the proof of Lemma 12, the result follows. □

Lemma 14. *Let* $c_u = (c_{i,j,u})_{i\in[1:2m+1],\ j\in[1:2m+1]}$. *We have* $K(n)_u \overset{\mathcal{H}-\mathcal{L}s}{\to} K_u$ *in* $\mathbb{D}^{2m+1}[0,\infty)$ *with*

$$K_u = \int_0^u b_s\, d\mathcal{W}_s,$$

where b_s *is a matrix with dimension* $(2m+1)\times(2m+1)$ *such that*

$$\int_0^u b_s b_s^\top\, ds = c_u.$$

Proof. The result follows from Lemma 8 to Lemma 13 together with Theorem IX.7.3 in [21]. □

Lemma 15. *Let* $\tilde{K}_t = K_{f_t}$ *and* $\tilde{K}(n)_t = K(n)_{\alpha_n^2 N_{\alpha_n,t}}$. *We have* $\tilde{K}(n)_t \overset{\mathcal{I}-\mathcal{L}s}{\to} \tilde{K}_t$ *in* $\mathbb{D}^{2m+1}[0,\infty)$.

Proof. We first use the obvious fact that the process $(K(n)_u, Z_u, \chi_{\theta_u})$ converges $\mathcal{H}$ stably towards the process $(K_u, Z_u, \chi_{\theta_u})$ in $\mathbb{D}^{2m+3}[0,\infty)$. Since $\alpha_n^2 N_{\alpha_n,t} \overset{u.c.p}{\to} f_t$, the fact that the stable convergence in law implies the convergence in law together with the composition mapping theorem (see Theorem 13.2.2 in [33]) and the Skorohod representation theorem, the preceding result gives

$$(\tilde{K}(n)_t, Z_{\alpha_n^2 N_{\alpha_n,t}}, \chi_{\theta_{\alpha_n^2 N_{\alpha_n,t}}}) \to (\tilde{K}_t, X_t, \chi_t)$$

and finally

$$(\tilde{K}(n)_t, X_t, \chi_t) \to (\tilde{K}_t, X_t, \chi_t),$$

for the weak convergence in $\mathbb{D}^{2m+3}[0,\infty)$. Then the result follows from the definition of the stable convergence in law.

□

LEMMA 16. *For $0 \leq t \leq T$, we have*

$$\alpha_n^{-1}\Big(\sum_{\nu_{i,n}\leq \langle X\rangle_t} \varpi_{i,n}\mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}\big[\frac{(B_{\nu_{i,n}}-B_{\nu_{i-1,n}})^2}{B^2_{\nu_{i-1,n}\wedge\langle X\rangle_T}}\big] - \int_0^t \sigma_u^2 du\Big) \overset{u.c.p.}{\to} 0,$$

$$\alpha_n^{-1}\Big(\sum_{\nu_{i,n}\leq \langle X\rangle_t} \varpi_{i,n}\alpha_n^2\mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}\big[\mathbb{I}_{\{|B_{\nu_{i,n}}-B_{\nu_{i-1,n}}|=\alpha_n k\}}\big] - \frac{k-1+2\eta}{2k-1+2\eta}\int_0^t p_k(\chi_u)\varphi(\chi_u)\sigma_u^2 X_u^2 du\Big) \overset{u.c.p.}{\to} 0,$$

$$\alpha_n^{-1}\Big(\sum_{\nu_{i,n}\leq \langle X\rangle_t} \varpi_{i,n}\alpha_n^2\mathbb{E}_{\mathcal{E}^n_{\mathcal{T}(\nu_{i-1,n})}}\big[\mathbb{I}_{\{|B_{\nu_{i,n}}-B_{\nu_{i-1,n}}|=\alpha_n(2\eta+k-1)\}}\big] - \frac{k}{2k-1+2\eta}\int_0^t p_k(\chi_u)\varphi(\chi_u)\sigma_u^2 X_u^2 du\Big) \overset{u.c.p.}{\to} 0.$$

PROOF. The first quantity is equal to

$$\begin{aligned}
&\alpha_n^{-1}\big(\sum_{\nu_{i,n}\leq \langle X\rangle_t}\frac{\Delta\nu_{i-1,n}}{B^2_{\nu_{i-1,n}}} - \int_0^{\langle X\rangle_t} B_s^{-2}\mathrm{d}s\big)\\
=&\alpha_n^{-1}\big(\sum_{\nu_{i,n}\leq \langle X\rangle_t}\big\{\frac{\Delta\nu_{i-1,n}}{B^2_{\nu_{i-1,n}}} - \int_{\nu_{i-1,n}}^{\nu_{i,n}} B_s^{-2}\mathrm{d}s\big\}\big) + \alpha_n^{-1}\int_{\nu_{N_{\alpha_n,t},n}}^{\langle X\rangle_t} B_s^{-2}\mathrm{d}s.
\end{aligned}$$

The result easily follows. The second quantity is equal to

$$\alpha_n^{-1}\frac{k-1+2\eta}{2k-1+2\eta}\Big(\sum_{\nu_{i,n}\leq\langle X\rangle_t}\alpha_n^2\Delta\nu_{i-1,n}p_k(\chi_{\mathcal{T}(\nu_{i-1,n})})\varphi(\chi_{\mathcal{T}(\nu_{i-1,n})}) - \int_0^{\langle X\rangle_t} p_k(\chi_{\mathcal{T}(v)})\varphi(\chi_{\mathcal{T}(v)})\mathrm{d}v\Big).$$

The result is obtained using Lemma 11. We get the result for the last term the same way. □

We have the following lemma.

LEMMA 17. *Let g_t be the $2\times(2m+1)$ matrix defined by $g_t = (\nabla_1, (f_t)^{-1}\nabla_2)^\top$ and*

$$V_t^n = (RV_{\alpha_n,t} - \langle \log X\rangle_t, \hat{\eta}_{\alpha_n,t} - \eta).$$

We have

$$\alpha_n^{-1}V_t^n \overset{\mathcal{I}-\mathcal{L}s}{\to} g_t\int_0^t b_{f_s}\, d\mathcal{W}_{f_s},\ \text{ in } \mathbb{D}^2[0,\infty).$$

PROOF. The result follows from Lemma 15 and Lemma 16 together with the Delta method and Proposition V.1.5 in [28]. □

We now give the proof of Theorem 2.

PROOF. Using the same method as in the proof of Lemma 5, we can show that

$$\alpha_n^{-1}\sum_{i=1}^{N_{\alpha_n,t}}\big(\frac{\hat{X}^t_{\tau_{i,n}} - \hat{X}^t_{\tau_{i-1,n}}}{\hat{X}^t_{\tau_{i-1,n}}}\big)^2$$

has the same limit as

$$\alpha_n^{-1}\sum_{i=1}^{N_{\alpha_n,t}}\big(\log(\hat{X}^t_{\tau_{i,n}}) - \log(\hat{X}^t_{\tau_{i-1,n}})\big)^2.$$

This quantity is equal to

$$\alpha_n^{-1}\sum_{i=1}^{N_{\alpha_n,t}}\big(\log(X_{\tau_{i,n}})-\log(X_{\tau_{i-1,n}})\big)^2$$
$$+\alpha_n^{-1}\sum_{i=1}^{N_{\alpha_n,t}}\big(\log(1-\frac{\alpha_n(\eta-\hat{\eta}_{\alpha_n,t})sign(P_{t_i}-P_{t_{i-1}})}{X_{\tau_{i,n}}})-\log(1-\frac{\alpha_n(\eta-\hat{\eta}_{\alpha_n,t})sign(P_{t_{i-1,n}}-P_{t_{i-2,n}})}{X_{\tau_{i-1,n}}})\big)^2$$
$$+2\alpha_n^{-1}\sum_{i=1}^{N_{\alpha_n,t}}\big(\log(X_{\tau_{i,n}})-\log(X_{\tau_{i-1,n}})\big)\big(\log(1-\frac{\alpha_n(\eta-\hat{\eta}_{\alpha_n,t})sign(P_{t_{i,n}}-P_{t_{i-1,n}})}{X_{\tau_{i,n}}})\big)$$
$$-2\alpha_n^{-1}\sum_{i=1}^{N_{\alpha_n,t}}\big(\log(X_{\tau_{i,n}})-\log(X_{\tau_{i-1,n}})\big)\big(\log(1-\frac{\alpha_n(\eta-\hat{\eta}_{\alpha_n,t})sign(P_{t_{i-1,n}}-P_{t_{i-2,n}})}{X_{\tau_{i-1,n}}})\big).$$

Obvious computations give that the second term tends to zero. For the first term, from Lemma 17, we get

$$\alpha_n^{-1}\Big(\sum_{i=1}^{N_{\alpha_n,t}}\big(\log(X_{\tau_{i,n}})-\log(X_{\tau_{i-1,n}})\big)^2-\langle\log X\rangle_t\Big)\overset{\mathcal{I}-\mathcal{L}s}{\longrightarrow}\nabla_1^\top\int_0^t b_{f_s}\mathrm{d}\mathcal{W}_{f_s}.$$

The sum of the last two terms is equal to

$$(\alpha_n^{-1}(\hat{\eta}_{\alpha_n,t}-\eta))\alpha_n\sum_{i=1}^{N_{\alpha_n,t}}\frac{(\log(X_{\tau_{i,n}})-\log(X_{\tau_{i-1,n}}))}{X_{\tau_{i-1,n}}}(\frac{X_{\tau_{i-1,n}}}{X_{\tau_{i,n}}}sign(X_{\tau_{i,n}}-X_{\tau_{i-1,n}})-sign(X_{\tau_{i-1,n}}-X_{\tau_{i-2,n}}))+R_n,$$

with R_n tending to zero. The term

$$\alpha_n\sum_{i=1}^{N_{\alpha_n,t}}\frac{\big(\log(X_{\tau_{i,n}})-\log(X_{\tau_{i-1,n}})\big)}{X_{\tau_{i-1,n}}}(\frac{X_{\tau_{i-1,n}}}{X_{\tau_{i,n}}}sign(X_{\tau_{i,n}}-X_{\tau_{i-1,n}})-sign(X_{\tau_{i-1,n}}-X_{\tau_{i-2,n}}))$$

has the same limit as

$$\alpha_n\sum_{i=1}^{N_{\alpha_n,t}}\frac{|\log(X_{\tau_{i,n}})-\log(X_{\tau_{i-1,n}})|}{X_{\tau_{i-1,n}}}.$$

Using the same method as in Lemma 6, this tends to μ_t. Finally, the last term tends to

$$\mu_t(f_t)^{-1}\nabla_2^\top\int_0^t b_{f_s}\mathrm{d}\mathcal{W}_{f_s}.$$

□

4.5. *Proof of Theorem 3.*

PROOF. Using the stability of the convergence in probability by absolutely continuous change of probability, we can set $a_t^{(j)}=0$. We consider without loss of generality that for all n $\tau_{1,n}^{(1)}>\tau_{1,n}^{(2)}$. We define

$$N_{\alpha_n,t}^{\lambda}=\sup\{i:\lambda_{i,n}^{(1)}\in[0,t]\}$$

and

$$RCV_{\alpha_n,t} = \sum_{i=2}^{N^{\lambda}_{\alpha_n,t}} \big(\log(X^{(1)}_{\lambda^{(1)}_{i,n}}) - \log(X^{(1)}_{\lambda^{(1)}_{i-2,n}})\big)\big(\log(X^{(2)}_{\lambda^{(2)}_{i,n}}) - \log(X^{(2)}_{\lambda^{(2)}_{i-1,n}})\big).$$

We have

$$\begin{aligned} RCV_{\alpha_n,t} &= \sum_{i=2}^{N^{\lambda}_{\alpha_n,t}} \big(\log(X^{(1)}_{\lambda^{(2)}_{i,n}}) - \log(X^{(1)}_{\lambda^{(2)}_{i-1,n}})\big)\big(\log(X^{(2)}_{\lambda^{(2)}_{i,n}}) - \log(X^{(2)}_{\lambda^{(2)}_{i-1,n}})\big) \\ &\quad + \sum_{i=2}^{N^{\lambda}_{\alpha_n,t}} \big(\log(X^{(1)}_{\lambda^{(2)}_{i-1,n}}) - \log(X^{(1)}_{\lambda^{(1)}_{i-2,n}})\big)\big(\log(X^{(2)}_{\lambda^{(2)}_{i,n}}) - \log(X^{(2)}_{\lambda^{(2)}_{i-1,n}})\big) \\ &\quad + \sum_{i=2}^{N^{\lambda}_{\alpha_n,t}} \big(\log(X^{(1)}_{\lambda^{(1)}_{i,n}}) - \log(X^{(1)}_{\lambda^{(2)}_{i,n}})\big(\log(X^{(2)}_{\lambda^{(2)}_{i,n}}) - \log(X^{(2)}_{\lambda^{(2)}_{i-1,n}})\big). \end{aligned}$$

By Theorem I.4.47 in [21],

$$\sum_{i=2}^{N^{\lambda}_{\alpha_n,t}} \big(\log(X^{(1)}_{\lambda^{(2)}_{i,n}}) - \log(X^{(1)}_{\lambda^{(2)}_{i-1,n}})\big)\big(\log(X^{(2)}_{\lambda^{(2)}_{i,n}}) - \log(X^{(2)}_{\lambda^{(2)}_{i-1,n}})\big) \overset{u.c.p.}{\to} \int_0^t \rho_s \sigma^{(1)}_s \sigma^{(2)}_s \mathrm{d}s.$$

For the second term, we use Lemma 1 and we get that it has the same limit as

$$\sum_{i=2}^{N^{\lambda}_{\alpha_n,t}} \mathbb{E}_{\mathcal{F}_{\lambda^{(2)}_{i-1,n}\wedge t}}\big[\big(\log(X^{(1)}_{\lambda^{(2)}_{i-1,n}\wedge t}) - \log(X^{(1)}_{\lambda^{(1)}_{i-2,n}\wedge t})\big)\big(\log(X^{(2)}_{\lambda^{(2)}_{i,n}\wedge t}) - \log(X^{(2)}_{\lambda^{(2)}_{i-1,n}\wedge t})\big)\big] = 0.$$

We now treat the last term. Let $\varepsilon > 0$. This term is equal to

$$\begin{aligned} &\sum_{i=2}^{\lfloor \alpha_n^{-(2+\varepsilon)}\rfloor} \big(\log(X^{(1)}_{\lambda^{(1)}_{i,n}\wedge t}) - \log(X^{(1)}_{\lambda^{(2)}_{i,n}\wedge t})\big)\big(\log(X^{(2)}_{\lambda^{(2)}_{i,n}\wedge t}) - \log(X^{(2)}_{\lambda^{(2)}_{i-1,n}\wedge t})\big) \\ &+ \sum_{\{i:\lfloor \alpha_n^{-(2+\varepsilon)}\rfloor \le i \le N^{\lambda}_{\alpha_n,t}\}} \big(\log(X^{(1)}_{\lambda^{(1)}_{i,n}}) - \log(X^{(1)}_{\lambda^{(2)}_{i,n}})\big)\big(\log(X^{(2)}_{\lambda^{(2)}_{i,n}}) - \log(X^{(2)}_{\lambda^{(2)}_{i-1,n}})\big). \end{aligned}$$

Since $\mathbb{E}_{\mathcal{F}_{\lambda^{(2)}_{i,n}\wedge t}}\big[\log(X^{(1)}_{\lambda^{(1)}_{i,n}\wedge t}) - \log(X^{(1)}_{\lambda^{(2)}_{i,n}\wedge t})\big] = 0$, we have

$$\begin{aligned} &\mathbb{E}\Big[\Big(\sum_{i=2}^{\lfloor \alpha_n^{-(2+\varepsilon)}\rfloor} \big(\log(X^{(1)}_{\lambda^{(1)}_{i,n}\wedge t}) - \log(X^{(1)}_{\lambda^{(2)}_{i,n}\wedge t})\big)\big(\log(X^{(2)}_{\lambda^{(2)}_{i,n}\wedge t}) - \log(X^{(2)}_{\lambda^{(2)}_{i-1,n}\wedge t})\big)\Big)^2\Big] \\ &= \sum_{i=2}^{\lfloor \alpha_n^{-(2+\varepsilon)}\rfloor} \mathbb{E}\big[\big(\log(X^{(1)}_{\lambda^{(1)}_{i,n}\wedge t}) - \log(X^{(1)}_{\lambda^{(2)}_{i,n}\wedge t})\big)^2\big(\log(X^{(2)}_{\lambda^{(2)}_{i,n}\wedge t}) - \log(X^{(2)}_{\lambda^{(2)}_{i-1,n}\wedge t})\big)^2\big] \\ &\le c\alpha_n^2 \sum_{i=2}^{\lfloor \alpha_n^{-(2+\varepsilon)}\rfloor} \mathbb{E}\big[\big(\log(X^{(2)}_{\lambda^{(2)}_{i,n}\wedge t}) - \log(X^{(2)}_{\lambda^{(2)}_{i-1,n}\wedge t})\big)^2\big] \\ &\le c\alpha_n^2 \mathbb{E}\big[\sum_{i=2}^{\lfloor \alpha_n^{-(2+\varepsilon)}\rfloor} \int_{\lambda^{(2)}_{i-1,n}\wedge t}^{\lambda^{(2)}_{i,n}\wedge t} (\sigma^{(2)}_s)^2 \mathrm{d}s\big]. \end{aligned}$$

Hence the first part of the last term tends to zero in probability. Moreover, the second part is smaller than $c(N^{(1)}_{\alpha_n,t} - \alpha_n^{-(2+\varepsilon)})^+$. Using the fact that $N^{\lambda}_{\alpha_n,t} \leq N^{(1)}_{\alpha_n,t} + N^{(2)}_{\alpha_n,t}$ together with the tightness of the sequences $\alpha_n^2 N^{(1)}_{\alpha_n,t}$ and $\alpha_n^2 N^{(2)}_{\alpha_n,t}$, we finally obtain that this second part tends to zero and so

$$RCV_{\alpha_n,t} \overset{u.c.p.}{\rightarrow} \int_0^t \rho_s \sigma_s^{(1)} \sigma_s^{(2)} \mathrm{d}s.$$

Moreover, using the same method as in the proof of Theorem 1, one can show that the difference $\widehat{RCV}_{\alpha_n,t} - RCV_{\alpha_n,t}$ tends to zero. The result follows.

□

5. Simulation study. In this section, we compare our estimators to other estimation procedures through simulations of the price process in the model with uncertainty zones. We consider the following model

$$\mathrm{d}X_t = \sigma_t X_t \mathrm{d}W_t,\ x_0 = 100,\ t \in [0,1],$$

where $\sigma_t = 0.01 \times \big(1 + 0.5 \times \sin(2\pi t + \pi/4)\big)$, which gives a classical U-shape intraday volatility curve. We fix $\alpha = 0.05$, $\eta = 0.05$ and, for simplicity, we assume that, for $i \geq 1$, $L_i = 1$ and $t_i = \tau_i$.

Our simulation accuracy is equivalent to 0.1 second. More precisely, the interval [0,1] corresponds to one trading day of eight hours and the discretization mesh is $(3600 \times 8 \times 10)^{-1}$ on $[0,1]$.

We compute the following estimators of the integrated volatility on $[0,1]$ over 1000 simulations:

- the realized volatility estimator,
- the kernel estimator of [6] with a Tukey-Hanning kernel,
- the pure rounding estimator presented in [31],
- the two scales estimator from [36] (ZMA for short),
- the Garman-Klass estimator, see [12] for details,
- our new integrated volatility estimator (RR for short).

The first three estimators are computed for different dyadic subsampling frequencies (from 0 to 10, that is from 1 second to 1024 seconds) whereas the ZMA estimator is computed for subsampling frequencies of 1, 2, 4 and 8 seconds (the results for other frequencies being not relevant). The results are given in Figure 2. We provide the averages of the estimated values together with 90% confidence intervals. The true value of the integrated volatility on $[0,1]$ (1.125×10^{-4}) is given by the grey line.

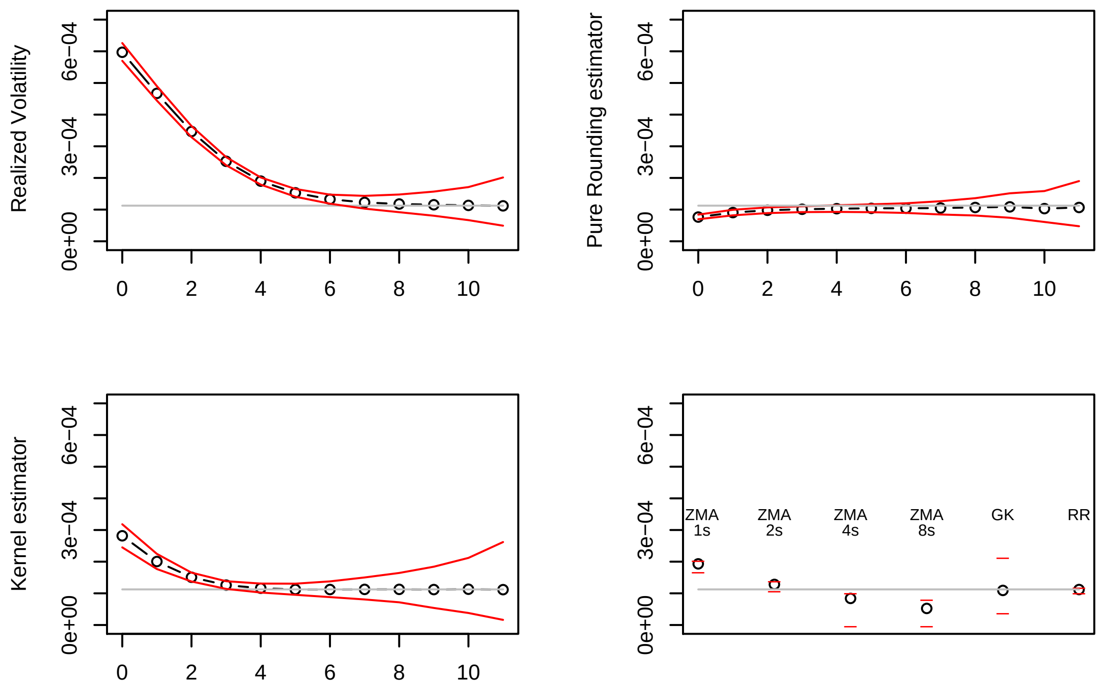

Fig 2. *Estimators in the case of a deterministic integrated volatility. The circles represent the average of the estimators and the red lines give the confidence intervals.*

The realized volatility becomes reasonable from a sampling frequency of two minutes whereas the kernel estimator is quite sharp when considering a sampling frequency of about sixteen seconds. This agrees with the simulation results about the noise that can be found in [29]. Indeed, at a first glance, in this specific setting of parameters, the assumptions required in [6] seem relatively fulfilled for a sampling frequency of sixteen seconds. The pure rounding estimator is almost unbiased for all frequencies and its confidence intervals are tight. The two scales estimator seems convenient only for a two seconds sampling whereas the Garman-Klass estimator is unbiased but leads to a large confidence interval. Finally, our estimator, which does not necessitate the choice of a sampling frequency, appears unbiased and very sharp.

We now consider the same model for two assets, where the two Brownian motions are correlated with a constant correlation coefficient ρ_t equal to 0.4. Thus the integrated co-volatility on $[0,1]$ is deterministic and equal to 4.5×10^{-5}. We compute the Hayashi-Yoshida estimator and our new estimator. The results over 1000 simulations are given in Table 1.

Statistics	HY	$\widehat{RCV}$
Mean	8.55×10^{-5}	4.48×10^{-5}
Standard Deviation	7.8×10^{-6}	5.7×10^{-6}

Table 1

Monte Carlo comparison between the Hayashi-Yoshida estimator and the new estimator.

The fact of correcting the last traded price values and considering the new stopping times considerably improves the Hayashi-Yoshida estimator. Indeed, whereas in average this estimator is twice larger than the true value, our estimator is almost unbiased and its variance is reasonable.

6. Conclusion. This paper studies volatility estimation issues in a model with endogenous microstructure noise.

- We work in the model with uncertainty zones. It is chosen for two reasons : first, the underlying efficient price is a semi-martingale and second, it reproduces the main stylized facts of transaction prices, durations and microstructure noise and its results on real data are very promising, see [29]. From a practical point of view, contrary to many other models, the computation of estimators does not require any choice of a sampling frequency and there is no ambiguity about the price to use.

- Our statistical procedures are based on the idea of approximating the value of the efficient price at some random times. Because of these random, endogenous times, usual semimartingale convergence theorems do not apply in our framework and so a new methodology is proposed.

- Our estimator of the integrated volatility is naturally given by a realized volatility computed on the approximated values of the efficient price. It is proved to satisfy a central limit theorem. The proof of the theorem uses a time change method together with stability properties of the weak convergence in the Skorokhod space. This idea is not specific to our model and can probably be used to treat other inference issues where times are endogenous.

- When two assets are observed, a slightly modified version of the Hayashi-Yoshida estimator computed on the approximated values of the efficient price gives us a consistent estimator of the integrated co-volatility. Thus, we have built an estimator for the co-volatility in the presence of asynchronicity of the data and microstructure noise.

- Simulations show that our results are very satisfying and that, in this quite realistic model, our method outperforms several other estimators.

References.

[1] Aït-Sahalia, Y. and Mykland, P.A. (2003). The Effects of Random and Discrete Sampling when Estimating Continuous-Time Diffusions. Econometrica **71** 483-549.

[2] Aït-Sahalia, Y. and Mykland, P.A. (2004). Estimators of Diffusions with Randomly Spaced Observations: A General Theory. The Annals of Statistics **32** 2186-2222.

[3] Aït-Sahalia, Y., Mykland, P.A. and Zhang, L. (2005). How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. Review of Financial Studies **18** 351-416.

[4] Aït-Sahalia, Y., Mykland, P.A. and Zhang, L. (2005). Ultra High Frequency Estimation with Dependent Microstructure Noise. Working paper.

[5] Bandi, F. and Russel, J. (2005). Separating Microstructure Noise from Volatility. Journal of Financial Economics **79** 655-692.

[6] Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2009). Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise. To appear in Econometrica.

[7] Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2009). Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading. Working paper.

[8] Borodin, A.N. and Salminen, P. (2002). Handbook of Brownian Motion-Facts and Formulae. Probability and Its Applications, Birkhäuser.

[9] Delattre, S. and Jacod, J. (1997). A Central Limit Theorem for Normalized Functions of the Increments of a Diffusion Process, in the Presence of Round-Off Errors. Bernoulli **3** 1-28.

[10] Engle, R. (2000). The Econometrics of Ultra High Frequency Data. Econometrica **68** 1-22.

[11] Fuksawa, M. (2007). Realized Volatility Based on Tick Time Sampling. To appear in Finance and Stochastics.

[12] Garman, M.B. and Klass, M.J. (1980). On the Estimation of Security Price Volatilities from Historical Data. The Journal of Business **53**, 67-78.

[13] Genon-Catalot, V. and Jacod, J. (1993). On Estimating the Diffusion Coefficient for Multidimensional Processes. Annales de l'IHP-Probabilité **29** 119-151.

[14] Genon-Catalot, V. and Jacod, J. (1994). Estimation of the Diffusion Coefficient for Difffusion Processes: Random Sampling. Scandinavian Journal of Statistics **21** 193-221.

[15] Gloter, A. and Jacod, J. (2001). Diffusions with Measurement Errors. I - Local Asymptotic Normality. ESAIM: Probability and Statistics **5** 225-242.

[16] Gloter, A. and Jacod, J. (2001). Diffusions with Measurement Errors. II - Optimal Estimators. ESAIM: Probability and Statistics **5** 243-260.

[17] Gourieroux, C. and Jasiak, J. (2001). Financial Econometrics: Problems, Models and Methods. Princeton University Press.

[18] Griffin, J.E. and Oomen, R.C.A. (2008). Sampling Returns for Realized Variance Calculations: Tick Time or Transaction Time? Econometric Reviews **27** 230-253.

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

[20] Hayashi, T. and Yoshida, N. (2005). On Covariance Estimation of Non-synchronously Observed Diffusion Processes. Bernoulli **11** 359-379.

[21] Jacod, J. and Shiryaev, A.N. (2002). Limit Theorems for Stochastic Processes. Second Edition, Springer-Verlag.

[22] Jacod, J., Li, Y., Mykland, P.A., Podolskij, M. and Vetter, M. (2009). Microstructure Noise in the Continuous Case: The Pre-Averaging Approach. To appear in Stochastic Processes and their Applications.

[23] Kalnina, I. and Linton, O.B. (2008). Estimating Quadratic Variation Consistently in the Presence of Correlated Measurement Error. Journal of Econometrics **147** 47-59.

[24] Large, J. (2007). Estimating Quadratic Variation when Quoted Prices Change by a Constant Increment. Working paper.

[25] Li, Y. and Mykland, P.A. (2008). Rounding Errors and Volatility Estimation. Working paper.

[26] Li, Y. and Mykland, P.A. (2007). Are Volatility Estimators Robust with Respect to Modeling Assumptions? Bernoulli **13** 601-622.

[27] Phillips, P. and Yu, J. (2006). Information Loss in Volatility Measurement with Flat Price Trading. Working paper.

[28] Revuz, D. and Yor, M. (1999). Continuous Martingales and Brownian Motion. Third Edition, Springer.

[29] Robert, C.Y. and Rosenbaum, M. (2009). A New Approach for the Dynamics of Ultra High Frequency Data: the Model with Uncertainty Zones. Working paper, available at http://www.crest.fr/pageperso/rosenbaum/rosenbaum.htm.

[30] Rosenbaum, M. (2007). A New Microstructure Noise Index. Working paper.

[31] Rosenbaum, M. (2009). Integrated Volatility and Round-Off Error. To appear in Bernoulli.

[32] Voev, V. and Lunde, A. (2007). Integrated Covariance Estimation using High-frequency Data in the Presence of Noise. Journal of Financial Econometrics **5** 68-104.

[33] Whitt, W. (2002). Stochastic-Process Limits, Springer.

[34] Zhang, L. (2006). Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach. Bernoulli **12** 1019-1043.

[35] Zhang, L. (2006). Estimating Covariation: Epps Effect and Microstructure Noise. Working Paper.

[36] Zhang, L., Mykland, P.A. and Aït-Sahalia, Y. (2005). A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. Journal of the American Statistical Association **100** 1394-1411.

Christian Yann Robert
CREST-ENSAE Paris Tech
Timbre J120, 3 Avenue Pierre Larousse,
92245 Malakoff Cedex, France.
E-mail: chrobert@ensae.fr

Mathieu Rosenbaum
CMAP-École Polytechnique Paris
UMR CNRS 7641, 91128 Palaiseau Cedex, France.
E-mail: mathieu.rosenbaum@polytechnique.edu