APPENDIX

7 Consistency in a simple model

The purpose of this section is to provide an outline of the LPM and the conditions of the CLT by investigating the simpler problem of consistency in the case of a simple model. Two toy examples, the estimation of volatility with regular not noisy observations, and the estimation of the rate of a Poisson process, are discussed extensively. Morever, techniques of proofs are also mentioned throughout the section. The obtained conditions are illustrative. Proofs of the conditions along with proofs that conditions hold in the two toy examples can be found in Section 8. Finally, some detailed mathematical definitions can be found in what follows too.

The simple model

We focus on a simple setting in this section. First, we work with one dimensional returns, i.e. $d := 1$. Also, we assume that the observations are regular, so that $\tau_{i,n} = \frac{i}{n}T$. The parametric model is assumed to be very simple, in particular there is no past dependence in the returns. It assumes that there exists a parameter $\theta^* \in K$ such that $R_{i,n}$ are independent and identically distributed (IID) random functions of θ^*. If we introduce $U_{i,n}$ an adequate IID sequence of random variables with distribution U which can depend on n, we can express the returns as

$$R_{i,n} := F_n\big(U_{i,n}, \theta^*\big), \tag{62}$$

where $F_n(x,y)$ is a non-random function. In (62), $U_{i,n}$ can be seen as the *random innovation*.

Since θ_t^* can in fact be time-varying, $R_{i,n}$ do not necessarily follow (62) in the time-varying parameter model. A formal time-varying generalization of (62) will be given in (65). In general, $R_{i,n}$ are neither identically distributed nor independent. $R_{i,n}$ are not even necessarily conditionally independent given the true parameter process θ_t^*, as we can see in the following two toy examples.

Example 1. (estimating volatility) Consider when $\theta_t^* := \sigma_t^2$ (the volatility is thus assumed to follow (21)), and $R_{i,n} := \int_{\tau_{i-1,n}}^{\tau_{i,n}} \sigma_s dW_s$, where W_t is a standard 1-dimensional

Brownian motion. In this case, the parameter space is $K := \mathbb{R}_*^+$. The parametric model assumes $\theta^* := \sigma^2$ and that the distribution of the returns is $R_{i,n} := \sigma \Delta W_{\tau_{i,n}}$, where $\Delta W_{\tau_{i,n}} := W_{\tau_{i,n}} - W_{\tau_{i-1,n}}$ is the increment of the Brownian motion between the $(i-1)$th observation time and the ith observation time and σ^2 is the fixed volatility. Under that assumption, the returns are IID. Under the time-varying parameter model, $R_{i,n}$ are clearly not necessarily IID, and they are also not necessarily conditionally independent given the whole volatility process σ_t^2 if there is a leverage effect.

Example 2. (estimating the rate of a Poisson process) Suppose the statistician observes data on the number of events (such as trades) in an arbitrary asset, and thinks the number of events happening between 0 and t, N_t, follows a homogeneous Poisson process with rate λ. The parameter rate $\theta_t^* := \lambda_t$ will be assumed to follow (21), with possibly a null-volatility $\sigma_t^\theta = 0$ if the homogeneity assumption turns out to be true. Because the econometrician does not have access to the raw data, she can't observe directly the exact time of each event. Instead, she only observes the number of events happening on a period (for instance a ten-minute block) $[\tau_{i-1,n}, \tau_{i,n})$, that is $R_{i,n} = N_{\tau_{i,n}}^- - N_{\tau_{i-1,n}}$. If the statistician's assumption of homogeneity is true, the returns are IID. In case of heterogeneity, N_t will be a inhomogeneous Poisson process, and the returns $R_{i,n}$ will most likely be neither identically distributed nor independent.

We need to introduce some notation and definitions. On a given block $i = 1, \cdots, B_n$ the observed returns will be called $R_{i,n}^1, \cdots, R_{i,n}^{h_n}$. Formally, it means that $R_{i,n}^j := R_{(i-1)h_n+j,n}$ for any $j = 1, \cdots, h_n$. In analogy with $R_{i,n}^j$, we introduce the approximated returns $\tilde{R}_{i,n}^1, \cdots, \tilde{R}_{i,n}^{h_n}$ on the ith block. We also introduce the corresponding observation times $\tau_{i,n}^j := \tau_{(i-1)h_n+j,n}$ for $j = 0, \cdots, h_n$. Note that $\tau_{i,n}^0 = \tau_{i-1,n}^{h_n}$. Finally, for $j = 1, \cdots, h_n$ we define the time increment between the $(j-1)$th return and the jth return of the ith block as $\Delta\tau_{i,n}^j := \tau_{i,n}^j - \tau_{i,n}^{j-1}$.

We provide a time-varying generalization of the parametric model (62) as well as a formal expression for the approximated returns. To deal with the former, we assume that in general

$$R_{i,n} := F_n\big(U_{i,n}, \{\theta_s^*\}_{\tau_{i-1,n} \leq s \leq \tau_{i,n}}\big). \tag{63}$$

The time-varying parameter model in (63) is a natural extension of the parametric model (62) because the returns $R_{i,n}$ can depend on the parameter process path from the previous sampling time $\tau_{i-1,n}$ to the current sampling time $\tau_{i,n}$. As $R_{i,n}$ depend on the parameter path, it seems natural to allow $U_{i,n}$ to be themselves process paths. For example, when the parameter is equal to the volatility process $\theta_t^* := \sigma_t^2$, we will assume that $U_{i,n}$ are equal to the underlying Brownian motion W_t path (see Example 3 for more details). Also, as $U_{i,n}$ are random innovation, they should be independent of the parameter process path past, but not on the current parameter path. In the case of volatility, it means that we allow for the leverage effect. A simple particular case of (63) is given by

$$R_{i,n} := F_n\big(U_{i,n}, \theta^*_{\tau_{i-1,n}}\big), \tag{64}$$

i.e. the returns depend on the parameter path only through its initial value. Finally, the approximated returns $\tilde{R}_{i,n}$ follow a mixture of the parametric model (62) with initial block parameter value. We are now providing a formal definition of our intuition. We assume that

$$R_{i,n}^j := F_n\big(U_{i,n}^j, \{\theta_s^*\}_{\tau_{i,n}^{j-1} \leq s \leq \tau_{i,n}^j}\big), \tag{65}$$

$$\tilde{R}_{i,n}^j := F_n\big(U_{i,n}^j, \tilde{\Theta}_{i,n}\big), \tag{66}$$

where the random innovation $U_{i,n}^j$ take values on a space $\mathcal{U}_n$ that can be functional[14] and that can depend on n, $U_{i,n}^j$ are IID for a fixed n but the distribution can depend on n, and $F_n(x,y)$ is a non-random function[15]. Note that (65) is a mere re-expression of (63) using a different notation. For any block $i = 1, \cdots, B_n$ and for any observation time $j = 0, \cdots, h_n$ of the *ith* block, we define $\mathcal{I}_{i,n}^j$[16] the filtration up to time $\tau_{i,n}^j$. The crucial

[14]$\mathcal{U}_n$ is a Borel space, for example the space $\mathcal{C}_1[0, \Delta\tau_n]$ of 1-dimensional continuous paths parametrized by time $t \in [0, \tau_n]$.

[15]Let $\mathcal{C}_p(\mathbb{R}^+)$ be the space of p-dimensional continuous paths parametrized by time $t \in \mathbb{R}^+$, which is a Borel space. Consequently, $\mathcal{U}_n \times \mathcal{C}_p(\mathbb{R}^+)$ is also a Borel space. We assume that $F_n(x,y)$ is a jointly measurable real-valued function on $\mathcal{U}_n \times \mathcal{C}_p(\mathbb{R}^+)$. Note that the advised reader will have seen that a priori $\{\theta_s^*\}_{\tau_{i,n}^{j-1} \leq s \leq \tau_{i,n}^j}$ is defined on $\mathcal{C}_p[0, \tau_n]$ (after translation of the domain by $-\tau_{i,n}^{j-1}$) in (65) and $\tilde{\Theta}_{i,n}$ is a vector in (66), whereas both should be defined on the space $\mathcal{C}_p(\mathbb{R}^+)$ according to the definition. We match the definitions by extending them as continuous paths on $\mathbb{R}^+$. Formally, if $\theta_t \in \mathcal{C}_p[0, \tau_n]$, we extend it as $\theta_t := \theta_{\tau_n}$ for all $t > \tau_n$. Similarly, if $\theta \in K$, we extend it as $\theta_t := \theta$ for all $t \geq 0$.

[16]Let $(\Omega, \mathcal{F}, P)$ be a probability space. Define the *sorted filtration* $\{\mathcal{I}_{k,n}\}_{k \geq 0}$ such that for any non-negative integer k that we can decompose as $k = (i-1)h_n + j$ where $i \in \{1, \cdots, B_n\}$ and $j \in \{0, \cdots, h_n\}$, $\mathcal{I}_{k,n} := \mathcal{I}_{i,n}^j$. We assume that $\mathcal{I}_{k,n}$ is a (discrete-time) filtration on $(\Omega, \mathcal{F}, P)$. In addition, we assume that $\{\theta_s^*\}_{0 \leq s \leq \tau_{i,n}^j}$ and $U_{i,n}^j$ are $\mathcal{I}_{i,n}^j$-measurable.

assumption is that $U_{i,n}^j$ has to be independent of the past filtration[17] (and in particular of $\tilde{\Theta}_{i,n}$). Note that we do not assume any independence between the random innovation $U_{i,n}^j$ and the parameter process $\{\theta_s^*\}_{\tau_{i,n}^{j-1}\leq s\leq \tau_{i,n}^j}$. We provide directly the definitions of F_n and $U_{i,n}^j$ in the two toy examples.

Example 3. (estimating volatility) In this case, $\mathcal{U}_n$ is defined as the space $\mathcal{C}_1[0,\Delta\tau_n]$ of continuous paths parametrized by time $t \in [0,\tau_n]$, $U_{i,n}^j := \{\Delta W_{[\tau_{i,n}^{j-1},s]}\}_{\tau_{i,n}^{j-1}\leq s\leq \tau_{i,n}^j}$ are the Brownian motion increment path processes between two consecutive observation times. We assume that (W_t^θ, W_t) is jointly a (possibly non-standard) 2-dimensional Brownian motion. Thus, the random innovation $U_{i,n}^j$ are indeed independent of the past in view of the Markov property of Brownian motions. We also define $F_n(u_t,\theta_t) := \int_0^{\tau_n} \theta_s^{\frac{1}{2}} du_s$. We thus obtain that the returns are defined as $R_{i,n}^j := \int_{\tau_{i,n}^{j-1}}^{\tau_{i,n}^j} \sigma_s dW_s$ and that the approximated returns $\tilde{R}_{i,n}^j := \sigma_{\tau_{i,n}^0} \Delta W_{[\tau_{i,n}^{j-1},\tau_{i,n}^j]}$ are the same quantity when holding the volatility constant on the block.

Example 4. (estimating the rate of a Poisson process) We assume that the rate of the (possibly inhomogeneous) Poisson process is $\alpha_n\lambda_t$, where α_n is a non time-varying and non-random quantity such that $\alpha_n\Delta\tau_n := 1$. In this case, we assume that $\mathcal{U}_n$ is the space of increasing paths on $\mathbb{R}^+$ starting from 0 which takes values in $\mathbb{N}$ and whose jumps are equal to 1. We also assume that for any path in $\mathcal{U}_n$, the number of jumps is finite on any compact of $\mathbb{R}^+$. $U_{i,n}^j$ can be defined as standard Poisson processes $\{N_t^{i,j,n}\}_{t\geq 0}$, independent of each other. We also have $F_n(u_t,\theta_t) := u_{\int_0^{\tau_n}\alpha_n\theta_s du_s}$. Thus, if we let $t_{i,n}^j := \int_{\tau_{i,n}^{j-1}}^{\tau_{i,n}^j} \alpha_n\lambda_s ds$, the returns are the time-changed Poisson processes

$$R_{i,n}^j = N_{t_{i,n}^j}^{i,j,n}, \tag{67}$$

$$\tilde{R}_{i,n}^j = N_{\alpha_n\Delta\tau_{i,n}^j\lambda_{\tau_{i,n}^0}}^{i,j,n}. \tag{68}$$

Consistency

In the following of this section, we will make the block size h_n go to infinity

$$h_n \to \infty. \tag{69}$$

[17]past filtration means up to time $\tau_{i,n}^{j-1}$

Furthermore, we will make the block length $\Delta \mathrm{T}_{i,n}$ vanish asymptotically. Because we assume observations are regular in this section, this can be expressed as

$$h_n n^{-1} \to 0. \tag{70}$$

We can rewrite the consistency of $\widehat{\Theta}_n$ as

$$\sum_{i=1}^{B_n} \big(\widehat{\Theta}_{i,n} - \Theta_{i,n}\big) \Delta \mathrm{T}_{i,n} \xrightarrow{\mathbb{P}} 0. \tag{71}$$

where the formal definition of $\widehat{\Theta}_{i,n}$ can be found in (75). In order to show (71), we can decompose the increments $(\widehat{\Theta}_{i,n} - \Theta_{i,n})$ into the part related to misspecified distribution error, the part on estimation of approximated returns error and the evolution in the spot parameter error

$$\begin{aligned} \widehat{\Theta}_{i,n} - \Theta_{i,n} = \big(\widehat{\Theta}_{i,n} - \widehat{\tilde{\Theta}}_{i,n}\big) + \big(\widehat{\tilde{\Theta}}_{i,n} - \theta^*_{\mathrm{T}i-1,n}\big) \\ + \big(\theta^*_{\mathrm{T}i-1,n} - \Theta_{i,n}\big), \end{aligned} \tag{72}$$

where $\widehat{\tilde{\Theta}}_{i,n}$, which is defined formally in (76), is the parametric estimator used on the underlying non-observed approximated returns. It is not a feasible estimator and appears in (72) only to shed light on the way we can obtain the consistency of the estimator in the proofs. We first deal with the last error term in (72), which is due to the non-constancy of the spot parameter θ^*_t. Note that

$$\sum_{i=1}^{B_n} \big(\tilde{\Theta}_{i,n} - \Theta_{i,n}\big) \Delta \mathrm{T}_{i,n} = \sum_{i=1}^{B_n} \big(\theta^*_{\mathrm{T}_{i-1,n}} \Delta \mathrm{T}_{i,n} - \int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}} \theta^*_s ds\big) \tag{73}$$

and thus we deduce from Riemann-approximation[18] that

$$\sum_{i=1}^{B_n} \big(\tilde{\Theta}_{i,n} - \Theta_{i,n}\big) \Delta \mathrm{T}_{i,n} \xrightarrow{\mathbb{P}} 0. \tag{74}$$

To deal with the other terms in (72), we assume that for any positive integer k, the practitioner has at hand an estimator $\hat{\theta}_{k,n} := \hat{\theta}_{k,n}(r_{1,n}; \cdots ; r_{k,n})$, which depends on the

[18] see i.e. Proposition 4.44 in p.51 of Jacod and Shiryaev (2003)

input of returns $\{r_{1,n}; \cdots ; r_{k,n}\}$. On each block $i = 1, \cdots , B_n$ we estimate the local parameter as

$$\widehat{\Theta}_{i,n} := \hat{\theta}_{h_n,n}\big(R^1_{i,n}; \cdots ; R^{h_n}_{i,n}\big). \tag{75}$$

The non-feasible estimator $\widehat{\tilde{\Theta}}_{i,n}$ is defined as the same parametric estimator with approximated returns as input instead of observed returns

$$\widehat{\tilde{\Theta}}_{i,n} := \hat{\theta}_{h_n,n}\big(\tilde{R}^1_{i,n}; \cdots ; \tilde{R}^{h_n}_{i,n}\big). \tag{76}$$

Note that (76) is infeasible because the approximated returns $\tilde{R}^j_{i,n}$ are non-observable quantities.

Example 5. (estimating volatility) The estimator is the scaled usual RV, i.e. $\hat{\theta}_{k,n}(r_{1,n}; \cdots ; r_{k,n}) := T^{-1}k^{-1}n \sum_{j=1}^k r^2_{j,n}$. Note that $\hat{\theta}_{k,n}$ can also be seen as the MLE (see the discussion pp. 112-115 in Mykland and Zhang (2012)).

Example 6. (estimating the rate of a Poisson process) The estimator to be used is the return mean $\hat{\theta}_{k,n}(r_{1,n}; \cdots ; r_{k,n}) := k^{-1} \sum_{j=1}^k r_{j,n}$.

In order to tackle the second term in (72), we make the assumption that the parametric estimator is $\mathbf{L}^1$-convergent, locally uniformly in the model parameter θ if we actually observe returns coming from the parametric model. This can be expressed in the following condition.

Condition (C). Let the innovation of a block $(V_{1,n}, \cdots , V_{h_n,n})$ be IID with distribution U_n. For any $M > 0$,

$$\sup_{\theta \in K_M} \mathbb{E}\left[\left|\hat{\theta}_{h_n,n}(F_n(V_{1,n}, \theta); \cdots ; F_n(V_{h_n,n}, \theta)) - \theta\right|\right] \to 0.$$

Remark 8. (practicability) Under Condition (C), results on regular conditional distributions[19] give us that the error made on the estimation of the underlying non-observed returns tends to 0, i.e.

$$\sum_{i=1}^{B_n} \big(\widehat{\tilde{\Theta}}_{i,n} - \tilde{\Theta}_{i,n}\big)\Delta \mathrm{T}_{i,n} \xrightarrow{\mathbb{P}} 0. \tag{77}$$

[19]see for instance Leo Breiman (1992), see Section 8 for more details.

This proof technique is the main idea of the paper. Regular conditional distributions are used to deduce results on the time-varying parameter model using uniform results in the parametric model.

Remark 9. (consistency) Note that $\mathbf{L}^1$-convergence is slightly stronger than the simple consistency of the parametric estimator. Nonetheless, in most applications, we will have both.

We can now summarize the consistency result in this very simple case where observations occur at equidistant time intervals and returns are IID under the parametric model. Under Condition (C) and assuming that

$$\sum_{i=1}^{B_n}(\widehat{\Theta}_{i,n} - \widehat{\widetilde{\Theta}}_{i,n})\Delta \mathrm{T}_{i,n} \xrightarrow{\mathbb{P}} 0, \tag{78}$$

we have the consistency of (3), i.e.

$$\widehat{\Theta}_n \xrightarrow{\mathbb{P}} \Theta. \tag{79}$$

We obtain the consistency in the couple of toy examples[20].

Remark 10. (LPE equal to the parametric estimator) The reader will have noticed that in the couple of examples, the LPE is equal to the parametric estimator. This is because in those very basic examples, the parametric estimator is linear, i.e. for any positive integer k and $l = 1, \cdots, k-1$

$$\hat{\theta}_{k,n}(r_{1,n}; \cdots ; r_{k,n}) = \frac{l}{k}\hat{\theta}_{l,n}(r_{1,n}; \cdots ; r_{l,n}) + \frac{k-l}{k}\hat{\theta}_{k-l,n}(r_{l+1,n}; \cdots ; r_{k,n})$$

In more general examples, this equation will break, and we will obtain two distinct estimators.

[20] see Section 8 for proofs

8 Proofs

8.1 Preliminaries

In view of our assumptions on θ_t^*, we can follow standard localisation arguments (see, e.g., pp. $160-161$ of Mykland and Zhang (2012)) and assume without loss of generality that K is a compact space. In case θ_t^* is an Itô semimartingale satisfying Condition (P1), we can also assume without loss of generality that there exists $0 \leq \sigma^+$ such that for any eigen value λ_t of σ_t^θ, we have $0 \leq \lambda_t \leq \sigma^+$ and that there exists $0 \leq a^+$ such that $\mid a_t^\theta \mid \leq a^+$.

Finally, we fix some notation. In the following of this paper, we will be using C for any constant $C > 0$, where the value can change from one line to the next.

We start with the proofs related to the consistency in the simple model introduced in Section 7. This provides an overview of the proof techniques, although the techniques will be more intricate when proving Theorem 2 (Central limit theorem), which includes non-regular observations.

8.2 Proof of Condition (C) $\Rightarrow$(77)

It is sufficient to show that Condition (C) implies that

$$\sup_{i \geq 0} \mathbb{E}\left[\left|\widehat{\widetilde{\Theta}}_{i,n} - \theta^*_{\mathrm{T}i-1,n}\right|\right] = o_p(1). \tag{80}$$

By (66) and (76), we can build g_n such that we can write

$$\left|\widehat{\widetilde{\Theta}}_{i,n} - \theta^*_{\mathrm{T}i-1,n}\right| = g_n(U^1_{i,n}, \cdots, U^{h_n}_{i,n}, \theta^*_{\mathrm{T}i-1,n}),$$

where g_n is a jointly measurable real-valued function such that

$$\mathbb{E}\left|g_n(U^1_{i,n}, \cdots, U^{h_n}_{i,n}, \theta^*_{\mathrm{T}i-1,n}\right| < \infty.$$

We have that

$$\begin{aligned}
\mathbb{E}\left[g_n(U^1_{i,n}, \cdots, U^{h_n}_{i,n}, \theta^*_{\mathrm{T}i-1,n})\right] &= \mathbb{E}\left[\mathbb{E}\left[g_n(U^1_{i,n}, \cdots, U^{h_n}_{i,n}, \theta^*_{\mathrm{T}i-1,n})\middle|\theta^*_{\mathrm{T}i-1,n}\right]\right] \\
&= \mathbb{E}\left[\int g_n(u, \theta^*_{\mathrm{T}i-1,n})\mu_\omega(du)\right]
\end{aligned}$$

where $\mu_\omega(du)$ is a regular conditional distribution for $(U^1_{i,n}, \cdots, U^{h_n}_{i,n})$ given $\tilde{\Theta}_{i,n}$ (see, e.g., Breiman (1992)). From Condition (C), we obtain (80).

8.3 Proof of the consistency in Example 1

Let's show Condition (C) first. For any $M > 0$, the quantity

$$\left|\hat{\theta}_{h_n,n}\big(F_n(V_{1,n},\theta);\cdots;F_n(V_{h_n,n},\theta)\big)-\theta\right|$$

can be shown to go to 0 in probability as a straightforward consequence of Theorem I.4.47 of p.52 in Jacod and Shiryaev (2003).

To show the condition (78), it is sufficient to show that the following quantity

$$nh_n^{-1}\mathbb{E}_{\mathrm{T}_{i-1,n}}\left[\left|\big(\theta^*_{\mathrm{T}_{i-1,n}}\Delta W_{[\mathrm{T}_{i-1,n};\mathrm{T}_{i,n}]}\big)^2-\Big(\int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}}\theta^*_s dW_s\Big)^2\right|\right] \tag{81}$$

goes to 0 uniformly in i. To prove this, we can use the formula $(a^2-b^2)=(a+b)(a-b)$, together with conditional Burkholder-Davis-Gundy inequality (BDG, see inequality (2.1.32) of p. 39 in Jacod and Protter (2011)).

8.4 Proof of Consistency in Example 2

Condition (C) can be shown easily. Similarly, the condition (78) is a direct consequence of the definition in (67), (68) together with (70).

8.5 Proof of Theorem 2 (Central limit theorem with non regular observation times)

We prove directly the central limit theorem in this general case. As a by-product, this implies the case with regular observations, i.e. Theorem 1. We can decompose $n^{\frac{1}{2}}\sum_{i=1}^{B_n}\big(\widehat{\Theta}_{i,n}-\Theta_{i,n}\big)\Delta\mathrm{T}_{i,n}$ as

$$I+II+III+IV, \tag{82}$$

with

$$I = n^{\frac{1}{2}} \sum_{i=1}^{B_n} \big(\widehat{\Theta}_{i,n} \Delta \mathrm{T}_{i,n} - \hat{\widehat{\Theta}}{}^{\mathbf{P}}_{i,n} \Delta \widetilde{\mathrm{T}}^{\mathbf{P}}_{i,n}\big),$$

$$II = n^{\frac{1}{2}} \sum_{i=1}^{B_n} \big(\hat{\widehat{\Theta}}{}^{\mathbf{P}}_{i,n} - \theta^*_{\mathrm{T}_{i-1,n}}\big) \Delta \widetilde{\mathrm{T}}^{\mathbf{P}}_{i,n},$$

$$III = n^{\frac{1}{2}} \sum_{i=1}^{B_n} \theta^*_{\mathrm{T}_{i-1,n}} \big(\Delta \widetilde{\mathrm{T}}^{\mathbf{P}}_{i,n} - \Delta \mathrm{T}_{i,n}\big),$$

$$IV = n^{\frac{1}{2}} \sum_{i=1}^{B_n} \big(\theta^*_{\mathrm{T}_{i-1,n}} - \Theta_{i,n}\big) \Delta \mathrm{T}_{i,n}.$$

It is clear that $I \xrightarrow{\mathbb{P}} 0$ by (41) and $III \xrightarrow{\mathbb{P}} 0$ by (42) along with Lemma 2.2.10 (p. 55) in Jacod and Protter (2011) and the fact that θ^*_t takes values in a compact set. We prove in what follows that $IV \xrightarrow{\mathbb{P}} 0$ and that $II \to \widetilde{Z}$, where $\widetilde{Z}$ follows the definition of Theorem 2.

We show $IV \xrightarrow{\mathbb{P}} 0$

We consider first the case where θ^*_t satisfies Condition (P2). We introduce

$$e_{i,n} := n^{\frac{1}{2}} \big(\theta^*_{\mathrm{T}_{i-1,n}} - \Theta_{i,n}\big) \Delta \mathrm{T}_{i,n}. \tag{83}$$

It is sufficient to show that $\sum_{i=1}^{B_n} \mid e_{i,n} \mid \xrightarrow{\mathbb{P}} 0$, and by virtue of Lemma 2.2.10 (p. 55) in Jacod and Protter (2011) that $\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[\mid e_{i,n} \mid \big] \xrightarrow{\mathbb{P}} 0$. We compute

$$\begin{aligned}
\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[\mid e_{i,n} \mid \big] &= n^{\frac{1}{2}} \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\Big[\Big| \int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}} (\theta^*_u - \theta^*_{\mathrm{T}_{i-1,n}}) du\Big|\Big] \\
&\leq C n^{\frac{1}{2}} \sum_{i=1}^{B_n} \underbrace{\Big(\mathbb{E}_{\mathrm{T}_{i-1,n}}\big[(\Delta \mathrm{T}_{i,n})^2\big]\Big)^{\frac{1}{2}}}_{O_p(h_n n^{-1})} \\
&\quad \underbrace{\Big(\mathbb{E}_{\mathrm{T}_{i-1,n}}\Big[\sup_{\mathrm{T}_{i-1,n} \leq s \leq \mathrm{T}_{i,n}} \big|\theta^*_s - \theta^*_{\mathrm{T}_{i-1,n}}\big|^2\Big]\Big)^{\frac{1}{2}}}_{o_p(n^{-\frac{1}{2}})} \\
&= o_p(1),
\end{aligned}$$

where we used onditional Cauchy-Schwarz to obtain the inequality, Condition (T) along with Condition (P2) to obtain the last equality. We deduce that $IV \xrightarrow{\mathbb{P}} 0$ in this case too.

We now consider the case where θ_t^* satisfies Condition (P1) and (26) holds. We start by decomposing $e_{i,n}$ into its bias and its martingale part. We have

$$e_{i,n} = \underbrace{n^{\frac{1}{2}} \int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}} \int_{\mathrm{T}_{i-1,n}}^{s} a_u^\theta du ds}_{e_{i,n}^{(b)}} + \underbrace{n^{\frac{1}{2}} \int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}} \int_{\mathrm{T}_{i-1,n}}^{s} \sigma_u^\theta dW_u ds}_{e_{i,n}^{(m)}}.$$

We will show in what follows that $\sum_{i=1}^{B_n} e_{i,n}^{(b)} = o_{\mathbb{P}}(1)$ and $\sum_{i=1}^{B_n} e_{i,n}^{(m)} = o_{\mathbb{P}}(1)$. We start with the first assertion. As for the previous case, it is sufficient to show that $\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[\mid e_{i,n}^{(b)} \mid\big] \xrightarrow{\mathbb{P}} 0$. As a_t^θ is bounded, we can bound the expression via

$$\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[\mid e_{i,n}^{(b)} \mid\big] \leq C n^{\frac{1}{2}} \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[(\Delta \mathrm{T}_{i,n})^2\big].$$

Then, using Condition (T) along with (26), we conclude that this is $o_{\mathbb{P}}(1)$.

We show now that $\sum_{i=1}^{B_n} e_{i,n}^{(m)} = o_{\mathbb{P}}(1)$. As it is a martingale, it is sufficient to show that $\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[\mid e_{i,n}^{(m)} \mid^2\big] \xrightarrow{\mathbb{P}} 0$. We compute

$$\begin{aligned}
\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[\mid e_{i,n}^{(m)} \mid^2\big] &= n \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\Big[\Big| \int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}} \int_{\mathrm{T}_{i-1,n}}^{s} \sigma_u^\theta dW_u ds\Big|^2\Big] \\
&\leq C n \sum_{i=1}^{B_n} \Big(\mathbb{E}_{\mathrm{T}_{i-1,n}}\big[(\Delta \mathrm{T}_{i,n})^3\big]\Big)^{\frac{2}{3}} \\
&\quad \Big(\mathbb{E}_{\mathrm{T}_{i-1,n}}\Big[\sup_{\mathrm{T}_{i-1,n} \leq s \leq \mathrm{T}_{i,n}} \Big| \int_{\mathrm{T}_{i-1,n}}^{s} \sigma_u^\theta dW_u\Big|^6\Big]\Big)^{\frac{1}{3}}, \\
&\leq C n \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[(\Delta \mathrm{T}_{i,n})^3\big] \\
&= o_p(1),
\end{aligned}$$

where we used conditional Hölder's inequality with $p = 3/2$ and $q = 3$ in the first inequality, BDG with $p = 3$ in the second inequality, Condition (T) along with (26) in the last equality.

We show $II \to \widetilde{Z}$

We aim to use Theorem 2-2 (p. 242) in Jacod (1997). Conditions are further specified in Theorem 3-2 (p. 244) in the case when observations are regular. Following the proof of Theorem 3-2, we can actually show that such conditions hold in the more general case when observations are not regular, choosing the filtration $\mathcal{J}_{\mathrm{T}_{i,n}}$. It is crucial to note that we are not working with the filtration $\mathcal{J}_{\tau_{i,n}}$.

Consequently, our goal is to show the conditions (3.10)-(3.14) from Theorem 3-2 (p. 244) in Jacod (1997). Note that (3.12) and (3.14) are respectively implied by (39) and (40). The bias condition (3.10) is satisfied as an application of (36) along with regular conditional distribution.

In this step, we prove that (3.11) is satisfied. We introduce $A_{i,n} := n^{\frac{1}{2}}\big(\hat{\widetilde{\Theta}}^{\mathbf{P}}_{i,n} - \theta^*_{\mathrm{T}_{i-1,n}}\big)\Delta\widetilde{\mathrm{T}}^{\mathbf{P}}_{i,n}$ and

$$C_{i,n} := \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[A_{i,n}A^T_{i,n}\big] - \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[A_{i,n}\big]\mathbb{E}_{\mathrm{T}_{i-1,n}}\big[A^T_{i,n}\big].$$

The condition (3.11) can be expressed as

$$\sum_{i=1}^{B_n} C_{i,n} \xrightarrow{\mathbb{P}} T\int_0^T V_{\theta^*_s} ds. \tag{84}$$

By regular conditional distribution, (36) and (37), we have that

$$\sum_{i=1}^{B_n} C_{i,n} = T\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[V_{\theta^*_{\mathrm{T}_{i-1,n}}}\Delta\widetilde{\mathrm{T}}^{\mathbf{P}}_{i,n}\big] + o_p(1).$$

In view of (42), the conditional Cauchy-Schwarz inequality and the boundedness of V_θ, we get

$$\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\Big[V_{\theta^*_{\mathrm{T}_{i-1,n}}}\Delta\widetilde{\mathrm{T}}^{\mathbf{P}}_{i,n}\Big] = \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\Big[V_{\theta^*_{\mathrm{T}_{i-1,n}}}\Delta\mathrm{T}_{i,n}\Big] + o_p(1).$$

Using Lemma 2.2.11 of Jacod and Protter (2011) together with conditional Cauchy-Schwarz inequality, (35) and the boundedness of V_θ, we obtain

$$T\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\Big[V_{\theta^*_{\mathrm{T}_{i-1,n}}}\Delta\mathrm{T}_{i,n}\Big] = T\sum_{i=1}^{B_n} V_{\theta^*_{\mathrm{T}_{i-1,n}}}\Delta\mathrm{T}_{i,n} + o_p(1).$$

We can apply now Proposition I.4.44 (p. 51) in Jacod and Shiryaev (2003) and we get

$$T\sum_{i=1}^{B_n} V_{\theta^*_{\mathrm{T}_{i-1,n}}} \Delta \mathrm{T}_{i,n} \xrightarrow{\mathbb{P}} T\int_0^T V_{\theta^*_s} ds.$$

In this final step, we prove that the Lindeberg condition (3.13) is satisfied. We will show in this step that for all $\epsilon > 0$,

$$\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[\mid A_{i,n} \mid^2 \mathbf{1}_{\{|A_{i,n}|>\epsilon\}} \right] \xrightarrow{\mathbb{P}} 0. \tag{85}$$

Actually, (85) can be shown using regular conditional distribution along with (38).

8.6 Proof of Theorem 3 (QMLE)

We want to show that the conditions of Theorem 1 are satisfied. We start with the case $\alpha > \frac{1}{2}$. The key result is Theorem 6 in Xiu (2010, p. 241). We choose $\mathbf{P} = (0,0)$.

We show first Condition (E). We can see easily from the key result that if we choose $V_{\theta^*_t} = 6\sigma^2_t$, then (37) is satisfied.

We can verify the Lindeberg condition (38) using conditional Cauchy-Schwarz inequality and the fact that the fourth moment of

$$h_n^{\frac{1}{2}} \left(\hat{\hat{\Theta}}^{\mathbf{P},\theta}_{i,n} - \theta \right)$$

is bounded.

As for the bias condition (36), we can see that as the noise shrinks faster than the order of the returns to 0, then the bias tends to the sum of the diagonal elements of W_1 defined in (23) in Xiu (2010, p. 241) minus unity. This equals 0 and thus (36) is satisfied.

The condition (29) is satisfied combining the fact that the noise is independent from X_t, the aforementioned theorem with the rationale in Section 8.3.

We now show that (27) and (28) are satisfied. Actually, we can show trivially that (27) holds for the reference continuous martingale $M_t = 0$. We recall that we are "only" showing stable convergence with respect to $\mathcal{F}^X_t$, and we show now the condition related to stable convergence (28). Actually, we can assume that $N_t = X_t$, this will imply that the result holds for any $N \in \mathcal{M}_b(M^\perp)$. From Theorem 6 in Xiu (2010), we have that

$$\hat{\tilde{\Theta}}^{\mathbf{P}}_{i,n} = \sum_{k=(i-1)h_n}^{ih_n-1} \sum_{l=k+1}^{ih_n-1} \omega_{k,l,n} (Z_{\tau_{k+1,n},n} - Z_{\tau_{k,n},n})(Z_{\tau_{l+1,n},n} - Z_{\tau_{l,n},n}).$$

We can develop $(Z_{\tau_{k+1,n},n} - Z_{\tau_{k,n},n})(Z_{\tau_{l+1,n},n} - Z_{\tau_{l,n},n}) = I + II + III + IV$, where $I = (X_{\tau_{k+1,n}} - X_{\tau_{k,n}})(X_{\tau_{l+1,n}} - X_{\tau_{l,n}})$, $II = (X_{\tau_{k+1,n}} - X_{\tau_{k,n}})(\epsilon_{l+1,n} - \epsilon_{l,n})$, $III = (\epsilon_{k+1,n} - \epsilon_{k,n})(X_{\tau_{l+1,n}} - X_{\tau_{l,n}})$ and $IV = (\epsilon_{k+1,n} - \epsilon_{k,n})(\epsilon_{l+1,n} - \epsilon_{l,n})$. Because the noise is independent from X_t, it is clear that $\mathbb{E}_{\mathrm{T}_{i-1,n}}\big[II * (X_{\mathrm{T}_{i,n}} - X_{\mathrm{T}_{i-1,n}})\big] = 0$, $\mathbb{E}_{\mathrm{T}_{i-1,n}}\big[III * (X_{\mathrm{T}_{i,n}} - X_{\mathrm{T}_{i-1,n}})\big] = 0$ and $\mathbb{E}_{\mathrm{T}_{i-1,n}}\big[IV * (X_{\mathrm{T}_{i,n}} - X_{\mathrm{T}_{i-1,n}})\big] = 0$. As for I, we can express

$$I * (X_{\mathrm{T}_{i,n}} - X_{\mathrm{T}_{i-1,n}}) = I * \sum_{k=(i-1)h_n}^{ih_n-1} (X_{\tau_{k+1,n}} - X_{\tau_{k,n}})$$

and from this expression straightforward computation leads to

$$\mathbb{E}_{\mathrm{T}_{i-1,n}}\Big[\big(\hat{\tilde{\Theta}}^{\mathbf{P}}_{i,n} - \tilde{\Theta}_{i,n}\big)\big(X_{\mathrm{T}_{i,n}} - X_{\mathrm{T}_{i-1,n}}\big)\Big] = 0.$$

We consider now the case $\alpha = 1/2$, i.e. when both the noise variance and the returns are of the same rate. In that case, we need to use the bias-corrected estimator $\widehat{\Theta}_n^{(BC)}$ so that we can verify the conditions of Theorem 1. The key result here is Proposition 1 (p. 369) along with its proof (p. 391-393) in Aït-Sahalia et al. (2005).

The bias condition (36) is satisfied on the account that we have reduced the bias of the estimator. Actually, the de-bias of the estimator doesn't affect the rest of the proof. Moreover, the increment condition (37) and the Lindeberg condition (38) are satisfied using similar techniques of proof. Finally, the conditions (27), (28) and (29) are satisfied using the same line of reasoning as in the previous case.

8.7 Proof of Theorem 4 (powers of volatility)

We aim to show that we can verify the conditions of Theorem 1. The idea is to use a Taylor expansion as in the delta method. Then the conditions will be satisfied partly following the proof of Theorem 3. More specifically, we will do the proof for (23) and

(24) of Condition (E), but will not explicit the proof of (25), (27), (28), (29) which can be proven using the same ideas. We use the following notation:

$$\hat{\hat{\Theta}}_{i,n}^{\mathbf{P},\theta} := g(\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma}) - B_{i,n}, \tag{86}$$

where $B_{i,n}$ can correspond to either one of the two bias-correction expressions found in (49) and (50). We have that

$$\hat{\hat{\Theta}}_{i,n}^{\mathbf{P},\theta} - \theta := g(\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma}) - B_{i,n} - g(\sigma^2), \tag{87}$$

for some σ^2. Using a Taylor expansion, we obtain that:

$$\begin{aligned} g(\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma}) - g(\sigma^2) = (\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma} - \sigma^2)g'(\sigma^2) + \frac{1}{2}(\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma} - \sigma^2)^2 g''(\sigma^2) \\ + \frac{1}{6}(\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma} - \sigma^2)^3 g^{(3)}(\eta), \end{aligned} \tag{88}$$

where η is between σ^2 and $\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma}$. Combining (87) and (88) and several assumptions (including the conditions on g), we obtain:

$$\operatorname{Var}\left[h_n^{\frac{1}{2}}\big(\hat{\hat{\Theta}}_{i,n}^{\mathbf{P},\theta} - \theta\big)\right] = (g'(\sigma^2))^2 \operatorname{Var}\left[h_n^{\frac{1}{2}}(\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma} - \sigma^2)\right] + o(1). \tag{89}$$

From here, one can conclude (24) using the proof of Theorem 3.

As for the bias condition (23), combining (87) and (88) and several assumptions we deduce that:

$$\begin{aligned} \mathbb{E}\big[\big(\hat{\hat{\Theta}}_{i,n}^{\mathbf{P},\theta} - \theta\big)\big] = \mathbb{E}\big[(\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma} - \sigma^2)g'(\sigma^2) + \frac{1}{2}(\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma} - \sigma^2)^2 g''(\sigma^2) - B_{i,n}\big] \\ + o(n^{-\frac{1}{2}}). \end{aligned} \tag{90}$$

We can show that

$$\mathbb{E}\big[(\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma} - \sigma^2)g'(\sigma^2)\big] = o(n^{-\frac{1}{2}}) \tag{91}$$

as in the proof of Theorem 3. We can also show that

$$\mathbb{E}\big[\frac{1}{2}(\hat{\hat{\sigma}}_{i,n}^{2,\mathbf{P},\sigma} - \sigma^2)^2 g''(\sigma^2) - B_{i,n}\big] = o(n^{-\frac{1}{2}}) \tag{92}$$

following the same line of reasoning as that of the case $v = 4$ in the proof of Lemma 4.4 in Jacod and Rosenbaum (2013, p. 1480). In view of (90), (91) and (92), we can show the bias condition (23).

8.8 Proof of Theorem 5 (E-(LPE of QMLE))

The strategy of the proof consists in showing that the estimation error in ν does not affect asymptotically the behavior of the QMLE so that we can apply directly Theorem 3. To do that, the key results will be Theorem 3(i) (p. 37) in Li et al. (2016) and Theorem 6 (p. 241) in Xiu (2010).

We recall that $\widehat{X}_{\tau_{i,n}} = Z_{\tau_{i,n},n} - g(I_{i,n}, \hat{\nu})$ and we define the n-dimensional vector $\widehat{Y}_n = (\widehat{X}_{\tau_{1,n}} - \widehat{X}_0, \cdots, \widehat{X}_T - \widehat{X}_{\tau_{n-1,n}})$. We also define $Y_n = ((X_{\tau_{1,n}} - X_0) + (\tilde{\epsilon}_{1,n} - \tilde{\epsilon}_{0,n}), \cdots, (X_T - X_{\tau_{n-1,n}}) + (\tilde{\epsilon}_{n,n} - \tilde{\epsilon}_{n-1,n}))$ and $\delta_n = (g(I_{1,n}, \hat{\nu}) - g(I_{0,n}, \hat{\nu})) - (g(I_{1,n}, \nu) - g(I_{0,n}, \nu)), \cdots, g(I_{n,n}, \hat{\nu}) - g(I_{n-1,n}, \hat{\nu})) - (g(I_{n,n}, \nu) - g(I_{n-1,n}, \nu)))$. It is clear that

$$\widehat{Y}_n = Y_n + \delta_n. \tag{93}$$

Finally, we recall that $\widehat{\Theta}_n$ is the LPE of QMLE on $\widehat{Y}_n$ and we define $\widetilde{\Theta}_n$ as the LPE of QMLE on Y_n.

Consider the case $\alpha < 1/2$ (the case $\alpha = 1/2$ is done following the same line of reasoning). The goal is to show that stably in distribution

$$n^{1/2}\Big(\widehat{\Theta}_n - \Theta\Big) \to \Big(6T^{-1}\int_0^T \sigma_s^4 ds\Big)^{\frac{1}{2}} \mathcal{N}(0,1). \tag{94}$$

We decompose the left hand-side term in (94) as

$$n^{1/2}(\widehat{\Theta}_n - \Theta) = \underbrace{n^{1/2}(\widetilde{\Theta}_n - \Theta)}_{A_n} + \underbrace{n^{1/2}(\widehat{\Theta}_n - \widetilde{\Theta}_n)}_{B_n}.$$

On the account of Theorem 3, we have that $A_n \to (6T^{-1}\int_0^T \sigma_s^4 ds)^{\frac{1}{2}}\mathcal{N}(0,1)$. Thus, if we can show that $B_n \xrightarrow{\mathbb{P}} 0$, then this implies (94).

We show now that $B_n \xrightarrow{\mathbb{P}} 0$. We define $\mathcal{M}_n$ the set of real $n \times n$ matrices. In view of Theorem 6 (p. 241) in Xiu (2010), there exists a function

$$\begin{aligned} M : K &\to \mathcal{M}_n \times \mathcal{M}_n \\ \theta &\mapsto (M^{(1)}(\theta), M^{(2)}(\theta)) \end{aligned}$$

such that $\widehat{\Theta}_{i,n} = \widehat{Y}_n' M(\widehat{\Theta}_{i,n})\widehat{Y}_n$ and $\tilde{\Theta}_{i,n} = Y_n' M(\tilde{\Theta}_{i,n})Y_n$, where we define for any $\theta \in K$ and any n dimensional vector Y:

$$Y'M(\theta)Y = (Y'M^{(1)}(\theta)Y, Y'M^{(2)}(\theta)Y).$$

We have that

$$\begin{aligned}
B_n &= n^{1/2}(\widehat{\Theta}_n - \tilde{\Theta}_{i,n}), \\
&= n^{1/2}(\widehat{Y}_n' M(\widehat{\Theta}_{i,n})\widehat{Y}_n - Y_n' M(\tilde{\Theta}_{i,n})Y_n), \\
&= n^{1/2}((Y_n + \delta_n)' M(\widehat{\Theta}_{i,n})(Y_n + \delta_n) - Y_n' M(\tilde{\Theta}_{i,n})Y_n), \\
&= n^{1/2}(Y_n'(M(\widehat{\Theta}_{i,n}) - M(\tilde{\Theta}_{i,n}))Y_n \\
&\quad + (\delta_n' M(\widehat{\Theta}_{i,n})Y_n + Y_n' M(\widehat{\Theta}_{i,n})\delta_n + \delta_n' M(\widehat{\Theta}_{i,n})\delta_n)), \\
&= n^{1/2} Y_n'(M(\widehat{\Theta}_{i,n}) - M(\tilde{\Theta}_{i,n}))Y_n + o_p(1), \\
&= o_p(1).
\end{aligned}$$

where we used (93) in the third equality, Assumption A along with Theorem 3(i) in Li et al. (2016) in the fifth equality, and Theorem 6 in Xiu (2010) along with Assumption A, Theorem 3(i) in Li et al. (2016) in the sixth equality.

8.9 Proof of Theorem 6 (powers of volatility)

The proof follows the proof of Theorem 5 along with the proof of Theorem 4.

8.10 Proof of Theorem 7 (Time-varying friction parameter model with uncertainty zones)

In order to prove the theorem, we will show that the conditions of Theorem 2 are satisfied. For this purpose we set $\mathbf{P} = (1, 1)$.

First, Condition (T) follows exactly from Corollary 4.4 (p. 14) in Robert and Rosenbaum (2012).

We aim to show now Condition (E*). We start with the bias condition (36). To avoid more involved notation, we keep the notation introduced in Section 4.4 to prove this part. We recall the definition of the estimator of η as

$$\widehat{\eta}_{t,n} := \sum_{k=1}^{m} \lambda_{t,k,n} u_{t,k,n},$$

with

$$\lambda_{t,k,n} := \frac{N_{t,k,n}^{(a)} + N_{t,k,n}^{(c)}}{\sum_{j=1}^{m} \left(N_{t,j,n}^{(a)} + N_{\alpha,t,j}^{(c)} \right)}, \tag{95}$$

$$u_{t,k,n} := \max\left\{ 0, \min\left\{ 1, \frac{1}{2}\left(k\left(\frac{N_{t,k,n}^{(c)}}{N_{t,k,n}^{(a)}} - 1 \right) + 1 \right) \right\} \right\}. \tag{96}$$

One can see easily from (96) that $u_{t,k,n}$ are consistent estimators of η with bias which satisfies the condition (36). Moreover, as $\widehat{\eta}_{t,n}$ is a linear combination of $u_{t,k,n}$, it also satisfies (36). It remains to show that the estimator of volatility which we recall to be defined as

$$\widehat{RV}_{t,n} = \sum_{i=1}^{N_n(t)} (\widehat{X}_{\tau_{i,n}} - \widehat{X}_{\tau_{i-1,n}})^2, \text{ where} \tag{97}$$

$$\widehat{X}_{\tau_{i,n}} = Z_{\tau_{i,n},n} - \alpha_n (1/2 - \widehat{\eta}_{t,n}) \text{sign}(R_{i,n}), \tag{98}$$

also satisfies the bias condition. In fact, combining (97) and (98) along with the key relation between $Z_{\tau_{i,n},n}$ and $X_{\tau_{i,n}}$ which can be found in (2.3) on p. 5 in Robert and Rosenbaum (2012), we can deduce that the bias of (97) is a function of the bias of $\widehat{\eta}_{t,n}$ which satisfies the condition (36).

We prove now the condition (37). We set an arbitrary $M > 0$. In view of the form of the sampling times (55), we have uniformly in $\theta \in K_M$ and in $i = 1, \cdots, B_n$ that

$$\begin{aligned}
&\text{Var}\left[h_n^{\frac{1}{2}} \big(\hat{\theta}_{h_n,n}(R_{i,n}^{1,\mathbf{P},\theta}; \cdots; R_{i,n}^{h_n,\mathbf{P},\theta}) - \theta \big) \Delta \mathrm{T}_{i,n}^{\mathbf{P},\theta} \right] \\
= &\text{Var}\left[h_n^{\frac{1}{2}} \big(\hat{\theta}_{h_n,n}(R_{i,n}^{1,\mathbf{P},\theta}; \cdots; R_{i,n}^{h_n,\mathbf{P},\theta}) - \theta \big) \right] \big(\mathbb{E}\big[\Delta \mathrm{T}_{i,n}^{\mathbf{P},\theta} \big] \big)^2 \\
&+ o_p(h_n^2 n^{-2}), \\
= &\text{Var}\left[h_n^{\frac{1}{2}} \big(\hat{\theta}_{h_n,n}(R_{i,n}^{1,\mathbf{P},\theta}; \cdots; R_{i,n}^{h_n,\mathbf{P},\theta}) - \theta \big) \right] \mathbb{E}\big[\Delta \mathrm{T}_{i,n}^{\mathbf{P},\theta} \big] \Delta \mathrm{T}_{i,n}^{\mathbf{P},\theta} \\
&+ o_p(h_n^2 n^{-2}), \\
= &S_{\theta,n}^{(1)} S_{\theta,n}^{(2)} \Delta \mathrm{T}_{i,n}^{\mathbf{P},\theta} T h_n n^{-1} + o_p(h_n^2 n^{-2}),
\end{aligned}$$

with

$$S_{\theta,n}^{(1)} := \text{Var}\left[h_n^{\frac{1}{2}} \big(\hat{\theta}_{h_n,n}(R_{i,n}^{1,\mathbf{P},\theta}; \cdots; R_{i,n}^{h_n,\mathbf{P},\theta}) - \theta \big) \right]$$

and $S_{\theta,n}^{(2)} := \mathbb{E}\big[\Delta \mathrm{T}_{i,n}^{\mathbf{P},\theta} \big] T^{-1} h_n^{-1} n$. By Lemma 4.19 in p. 26 of Robert and Rosenbaum (2012) in the special case where the volatility is constant, we obtain the existence and the value of $S_\theta^{(1)}$ such that $S_{\theta,n}^{(1)} \to S_\theta^{(1)}$. Also, by Corollary 4.4 in p. 14 of Robert and Rosenbaum (2012), there exists $S_\theta^{(2)}$ such that $S_{\theta,n}^{(2)} \to S_\theta^{(2)}$. If we define $V_\theta = S_\theta^{(1)} S_\theta^{(2)}$, (37) is satisfied.

The Lindeberg condition (38) can be obtained using conditional Cauchy-Schwarz inequality, together with the fact that the fourth moment of

$$h_n^{\frac{1}{2}}\big(\hat{\theta}_{h_n,n}(R_{i,n}^{1,\mathbf{P},\theta};\cdots;R_{i,n}^{h_n,\mathbf{P},\theta}) - \theta\big)$$

is bounded, and Condition (T).

We prove now the conditions (39) and (40). Here again we choose the reference martingale $M_t = 0$, and thus we obtain trivially (39). To show (40), if we decompose $\big(\hat{\tilde{\Theta}}_{i,n}^{\mathbf{P}} - \theta_{\tau_{i,n}}^*\big)$ following the definition of the estimator, $\Delta\widetilde{\mathrm{T}}_{i,n}^{\mathbf{P}}$ as

$$\sum_{j=1}^{h_n}(\widetilde{\tau}_{(i-1)h_n+j,n}^{\mathbf{P}} - \widetilde{\tau}_{(i-1)h_n+j-1,n}^{\mathbf{P}}),$$

and

$$N_{\mathrm{T}_{i,n}} - N_{\mathrm{T}_{i-1,n}} = \sum_{j=1}^{h_n}(N_{(i-1)h_n+j,n}^{\mathbf{P}} - N_{(i-1)h_n+j-1,n}^{\mathbf{P}}),$$

and develop the product of those three expressions, we can easily get rid of the cross terms, and the other terms can be shown going to 0 following the same line of reasoning as the proof of Lemma 4.11 (pp. 20-21) and Lemma 4.14 (pp. 22-23) in Robert and Rosenbaum (2012).

We turn now to (41) and (42). We start by showing the latter condition. We can decompose $\Delta\widetilde{\mathrm{T}}_{i,n}^{\mathbf{P}} - \Delta\mathrm{T}_{i,n}$ into

$$\big(\Delta\widetilde{\mathrm{T}}_{i,n}^{\mathbf{P}} - \Delta\breve{\mathrm{T}}_{i,n}^{\mathbf{P}}\big) + \big(\Delta\breve{\mathrm{T}}_{i,n}^{\mathbf{P}} - \Delta\mathrm{T}_{i,n}\big), \tag{99}$$

where $\Delta\breve{\mathrm{T}}_{i,n}^{\mathbf{P}}$ follows the same definition as $\widetilde{\mathrm{T}}_{i,n}^{\mathbf{P}}$ (i.e. we hold volatility constant on the block) except that the starting point of the past is not set to $\mathbf{P}$ but kept to the random past $P_{\mathrm{T}_{i-1,n},n}$. We deal with the first term in (99). We can see that under the parametric model the past $P_{\tau_{i,n}}$ follows a discrete Markov chain on the space $\{1,\cdots,m\}\times\{-1,1\}$. Following the same line of reasoning as in the proof of Lemma 14 in Potiron and Mykland (2017), we can easily show that

$$n^{\frac{1}{2}}\sum_{i=1}^{B_n}\mathbb{E}_{\mathrm{T}_{i-1,n}}\Big[\big|\Delta\widetilde{\mathrm{T}}_{i,n}^{\mathbf{P}} - \Delta\breve{\mathrm{T}}_{i,n}^{\mathbf{P}}\big|\Big] \xrightarrow{\mathbb{P}} 0$$

We turn now to the second term in (99). Using the same idea as in the proof of Lemma 11 in Potiron and Mykland (2017), we deduce

$$n^{\frac{1}{2}} \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[\left| \Delta \breve{\mathrm{T}}_{i,n}^{\mathbf{P}} - \Delta \mathrm{T}_{i,n} \right| \right] \xrightarrow{\mathbb{P}} 0.$$

We have thus shown that (42) holds. The same line of reasoning can lead us to (41).

8.11 Proof of Theorem 8 (Time-varying MA(1))

The key result for this proof is the connection between the MA(1) process and the observations in the model described in Section 4.1 in the case $\alpha = 1/2$. Such connection can be seen in view of the proof of Proposition I (pp. 391-393) in Aït-Sahalia et al. (2005). More specifically, we can use Taylor expansions to re-express this estimator as the estimator in Section 4.1 and then use Theorem 3 (ii) to conclude. Similar Taylor expansions were already obtained in the proof of Theorem 4, and we will not further explain the details in this specific case.

8.12 Estimation of the friction parameter bias and standard deviation in the model with uncertainty zones

In this section, we provide the formal definitions, along with some theoretical derivation, of the friction parameter bias and standard deviation used in our empirical illustration. The notation of Section 4.4 and Section 10 are in force.

We estimate the standard deviation as

$$\widehat{s}_n := \widehat{s}_n(\widehat{\eta}_{T,n}),$$

where a formal expression or an estimator of the variance of $V(\eta) := \widehat{s}_n(\eta)$ is provided in what follows, depending on the setting. We also derive an expression or estimator of the bias of $\widehat{\eta}_{T,n}$, that we call $B(\eta)$. They are both obtained assuming that the friction parameter is fixed to η. In our numerical study we find that this bias is very close to 0 so that it is relatively safe to assume that it equals 0 for the purpose of statistical inference.

We consider first the case where the absolute jump size is constant equal to the tick size, i.e. $L_{i,n} := 1$, and $N_n(t)$ is non-random. In view of (57), we have

$$\widehat{\eta}_{t,n} := \min\Big(1, \frac{N^{(c)}_{t,1,n}}{2N^{(a)}_{t,1,n}}\Big).$$

We also have by definition that the number of alternations is $N^{(a)}_{t,1,n} = N_n(t) - N^{(c)}_{t,1,n}$. Then

$$N^{(c)}_{t,1,n} \sim Bin(N_n(t), \frac{2\eta}{2\eta+1}), \tag{100}$$

where $Bin(n,p)$ is a binomial distribution with n observations and probability p. Let $B \sim Bin(N_n(t), \frac{2\eta}{2\eta+1})$. We can define the bias as

$$B(\eta) := \mathbb{E}\left[\min\Big(1, \frac{B}{2(N_n - B)}\Big)\right] - \eta$$

and the variance as

$$V(\eta) := \mathrm{Var}\left[\min\Big(1, \frac{B}{2(N_n - B)}\Big)\right].$$

In this case we have thus shown that B and V can be computed easily numerically.

We assume now that $N_n(t)$ can be random. We can work conditional on $N_n(t)$. As the sampling times are endogenous, (100) is not true in that case. Nonetheless, we can still approximate $N^{(c)}_{t,1,n}$ by $Bin(N_n(t), \frac{2\eta}{2\eta+1})$ if the number of observations is large enough.

We now turn out to the general case, i.e. when $L_{i,n}$ can be different from 1. For $k = 1, \cdots, m$ we define $\widetilde{p}_k := \frac{2\eta+k-1}{2\eta+k}$ and we let B_k be an independent sequence of distribution $Bin(N^{(c)}_{t,k,n} + N^{(a)}_{t,k,n}, \widetilde{p}_k)$, and

$$C_k := \max\Big(0, \min\Big(1, \frac{1}{2}\big(k\big(\frac{B_k}{N^{(a)}_{t,k,n} + N^{(c)}_{t,k,n} - B_k} - 1\big) + 1\big)\Big)\Big).$$

The distribution of $\widehat{\eta}_{t,n}$ can be approximated by the distribution of

$$\sum_{i=1}^{m} \lambda_{t,k,n} C_k,$$

and we can estimate the bias as $\widehat{B}(\eta) := \sum_{i=1}^m \lambda_{t,i,n}\mathbb{E}\big[C_k\big]$ and the variance as $\widehat{V}(\eta) := \sum_{i=1}^m \lambda^2_{\alpha,t,i}\,\mathrm{Var}\,\big[C_k\big]$.

9 Additional numerical study: the time-varying MA(1) case

9.1 Goal of the study

To investigate the finite sample performance of the LPE, we consider the time-varying MA(1) with null-mean introduced in Section 4.5, where the related local estimator is the MLE. The goal of the study is twofold. First, we want to investigate how the LPE performs compared to some naive concurrent approaches. Second, we want to discuss about the choice of the tuning parameter h_n in practice.

We consider the following simple concurrent approaches:

> *MLE*: the global MLE when considering that the parameters are not time-varying on $[0, T]$.
>
> *Fitting Recent Observations (FRO)*: This approach consists in fitting the MLE on a recent sub-block with less observations (e.g. on $[T_F, T]$ where $T_F > 0$) so that the parameter is roughly constant on that block.

To compute the bias-corrected estimator $\widehat{\Theta}_n^{(BC)} = \widehat{\Theta}_{i,n} - b(\widehat{\Theta}_{i,n}, h_n)$, we can either compute and implement the function $b(\theta, h)$ or carry out Monte-Carlo simulations to compute $b(\theta, h)$ for any (θ, h) prior to the numerical study. We choose the latter option as this allows to get also rid of bias terms which appears in the Taylor expansion in a higher order than $O(h^{-2})$. Indeed, although those terms vanish asymptotically, they can pop up in a finite sample context. To be more specific, we first compute the sample mean for a grid of parameter values and block length (θ, h) with, say, 100,000 Monte Carlo paths[21] of the parametric model. Then on each block, we estimate the bias by $b(\widehat{\Theta}_{i,n}, h_n)$.

We discuss here what we expect theoretically from the bias-correction. In view of the decomposition (9), we can disentangle the bias of $\widehat{\Theta}_{i,n}$ on first approximation as the sum of two terms, namely the bias of the parametric estimator and the bias due to the fact that the parameter is time-varying. The former can be corrected by the

[21] Actually, the number of Monte Carlo paths can be significantly lower when h increases

econometrician, and we define the bias-corrected local estimate $\widehat{\Theta}_{i,n}^{(BC)}$ accordingly. On the contrary, as the the parameter path is unknown, we cannot correct for the latter. This is one reason why we have to work with (up to constant terms) a $h_n < n^{\frac{1}{2}}$. The theory shows that the normalized latter bias will vanish asymptotically under that condition. Conversely, the econometrician who chooses to work locally with $h_n > n^{\frac{1}{2}}$ will most likely obtain a significant latter bias which she cannot identify, and correcting for the former bias might not improve the estimation in that case.

9.2 Model design

We recall that the time-varying parameter is $\theta_t^* = (\beta_t, \kappa_t)$. We set $T = 1$, which stands for one day (or one week, one month). We fix the number of observations $n = 10,000$. We consider one toy model where the parameters move around a target parameter deterministically. We assume that the noisy parameter follows a cos function $\theta_t^* = \nu + A\cos(\frac{2\pi t\delta}{T})$, where $\nu = (\beta, \kappa)$ is the parameter, $A = (A^{(\beta)}, A^{(\kappa)})$ corresponds to the amplitude, and $\delta = (\delta^{(\beta)}, \delta^{(\kappa)})$ stands for the number of oscillations on $[0, T]$. With this model, we set $\Theta = (\beta, \kappa)$. We fix the parameter $\nu = (0.5, 1)$ and the amplitude $A = (0.2, 0.4)$. We also choose one setting with a small number of oscillations $\delta = (4, 4)$ and one with a bigger number of oscillations $\delta = (10, 10)$. We simulate $M = 1,000$ Monte-carlo repetitions.

In view of Theorem 8 and Condition (P2), the tuning parameter h_n should (up to constant terms) satisfy $n^{1/4} < h_n < n^{1/2}$. In our case, we have that $n^{1/4} = 10$ and $n^{1/2} = 100$. Accordingly we set $h_n = 25$, 100, 500, 1000, 2000, 5000. For the FRO approach, we set[22] $T_F = 0.95$, which means that we consider the last 500 observations to fit the MLE.

9.3 Results

The results are reported in Table 2 when $\delta = (4, 4)$ and Table 3 when $\delta = (10, 10)$. First, note that the results are similar for both values. Second, as expected from the theory, the LPE performs at its best with the choice $h_n = n^{\frac{1}{2}} = 100$, and the bias-corrected version is much better. Moreover, it outperforms the two concurrent approaches.

[22]This choice is arbitrary, but different values would yield to similar results.

The case $h_n = 25$ allows us to check what can happen when we have blocks with very few observations. The bias-corrected estimator performs well to estimate κ, but somehow the bias-correction to estimate β does not provide better estimates. This is most likely due to the fact that we have not enough observations on each block.

The estimation made with $h_n = 500$ is very decent in the case with small number of oscillations. The bias-corrected estimator is actually not as good. This corroborates the theory that when $h_n >> 100$ the main source of bias is due to the parameter which is time-varying rather than the parametric estimator bias itself. If we have a bigger number of oscillations, the estimates are not as accurate. When using bigger h_n, we see the same pattern, and the accuracy of the estimation decreases as h_n increases.

The global MLE performs relatively well to estimate β, but have a strong bias in κ. This indicates that even in a simple deterministic model which oscillates around the target value, the MLE cannot be trusted. Finally, the FRO is far off and the standard deviation is bigger.

Remark 11. (block size) The conditions of our paper provides the asymptotic order to use for the tuning parameter h_n. Thus, it gives a rule of thumb to use on finite sample, but it is left to the practitioner to ultimately choose h_n. If the parametric estimator is badly biased, the practitioner should increase the value of h_n. Also, if the parameter seems roughly constant, h_n can be chosen to be bigger. In our simulation study, this rule of thumb could be trusted. In our empirical illustration, we can see that the estimated volatility is robust to the value of h_n if we choose $h_n \approx N_n^{1/2}$. As n can be chosen such that $n = N_n$, this indicates that the rule of thumb seems to be robust to the actual choice of h_n in our empirical study too.

10 Empirical illustration in the model with uncertainty zones

In this section, we implement the LPE in the model with uncertainty zones introduced in Section 4.4. We recall that the parameter of interest is defined as $\xi_t^* = (\sigma_t^2, \eta_t)$. We are looking at Orange (ORA.PA) stock price traded actively on the CAC 40 on one random day, Monday March 4th, 2013. To prevent from opening and closing effect, we

		β	β	κ	κ
estimator	block size	sample bias	s.d.	sample bias	s.d.
MLE		-0.0052	0.0085	0.1041	0.0148
FRO		0.2913	0.0355	0.1666	0.0355
LPE	25	-0.0168	0.0131	-0.0985	0.0096
BC LPE	25	-0.0172	0.0132	-0.0062	0.0097
LPE	100	0.0035	0.0083	-0.0256	0.0096
BC LPE	100	-0.0010	0.0082	-0.0065	0.0096
LPE	500	-0.0021	0.0094	0.0073	0.0101
BC LPE	500	-0.0049	0.0095	0.0098	0.0104
LPE	1000	-0.0030	0.0099	0.0425	0.0125
BC LPE	1000	-0.0056	0.0100	0.0438	0.0126
LPE	2000	-0.0032	0.0102	0.1029	0.0143
BC LPE	2000	-0.0055	0.0101	0.1035	0.0143
LPE	5000	-0.0052	0.0087	0.1037	0.0148
BC LPE	5000	-0.0060	0.0087	0.1044	0.0147

Table 2: In this table, we report the sample bias and the standard deviation for the different estimators in the case of a small number of oscillations $\delta = (4, 4)$. The parameter $(\beta, \kappa) = (0.5, 1)$. The number of Monte-carlo simulations is 1,000. Note that BC stands for "bias-corrected".

		β	β	κ	κ
estimator	block size	sample bias	s.d.	sample bias	s.d.
MLE		-0.0069	0.0105	0.1094	0.0222
FRO		0.0065	0.0391	0.0882	0.0678
LPE	25	-0.0148	0.0183	-0.0876	0.0144
BC LPE	25	-0.0155	0.0184	0.0046	0.0143
LPE	100	0.0017	0.0092	-0.0164	0.0183
BC LPE	100	0.0012	0.0092	0.0039	0.0183
LPE	500	-0.0053	0.0094	0.1046	0.0219
BC LPE	500	-0.0085	0.0094	0.1086	0.0221
LPE	1000	-0.0071	0.0102	0.1078	0.0216
BC LPE	1000	-0.0115	0.0102	0.1098	0.0217
LPE	2000	-0.0071	0.0106	0.1087	0.0220
BC LPE	2000	-0.0108	0.0106	0.1096	0.0221
LPE	5000	-0.0071	0.0106	0.1087	0.0220
BC LPE	5000	-0.0093	0.0106	0.1090	0.0219

Table 3: In this table, we report the sample bias and the standard deviation for the different estimators in the case of a bigger number of oscillations $\delta = (10, 10)$. The parameter $(\beta, \kappa) = (0.5, 1)$. The number of Monte-carlo simulations is 1,000. Note that BC stands for "bias-corrected".

assume that restrict to data obtained from 9am to 4pm. The number of transactions inducing to a price change during this time period is equal to $N_n = 3306$, the tick size $\alpha_n = 0.001$ euro, and the price is equal to 8 euros on average.

We report the global estimate $\widehat{\eta}_{T,n} = 0.155$, and the standard deviation[23] $\widehat{s}_n = 0.008$. Moreover, Figure 2 documents the friction parameter estimates over time for several values of block size. Based on those estimates and the local standard deviation estimate $\hat{s}_{i,n}$, we compute the associated chi-square statistic[24]

$$\chi_n^2 := \sum_{i=1}^{B_n-1} \Big(\frac{\hat{\eta}_{i,n} - \widehat{\eta}_{T,n}}{\hat{s}_{i,n}}\Big)^2.$$

Under the null hypothesis which states that η_t is constant, χ_n^2 follows approximately a chi-square distribution with $B_n - 1$ degrees of freedom. We report χ_n^2 for different values of h_n in Table 4. The obtained values indicate that we have strong evidence against the null hypothesis, revealing that the friction parameter is time-varying.[25]

We report in Figure 3 the estimated volatility for different values of h_n. Because $N_n^{1/2} \approx 57.5$ we set $h_n = 43, \cdots, 63$. We also report the estimates with RV and the global model with uncertainty zones. The estimates of the latter seems to slightly underestimate the integrated volatility. In addition, the former, which is positively biased under the presence of microstructure noise, is far off, most likely overestimating the integrated volatility. Finally, the estimates are very similar for different values of h_n, which seems to indicate that the method is robust to small block size variation.

[23]related definitions and derivation of $\widehat{s}_n$ and $\hat{s}_{i,n}$ can be found in Section 8.12

[24]Note that since the number of observations of the last block is arbitrary, the last block estimate is not used to compute the chi-square statistic.

[25]This analysis has been carried out on other days and other stocks. We consistently conclude that the friction parameter is time-varying.

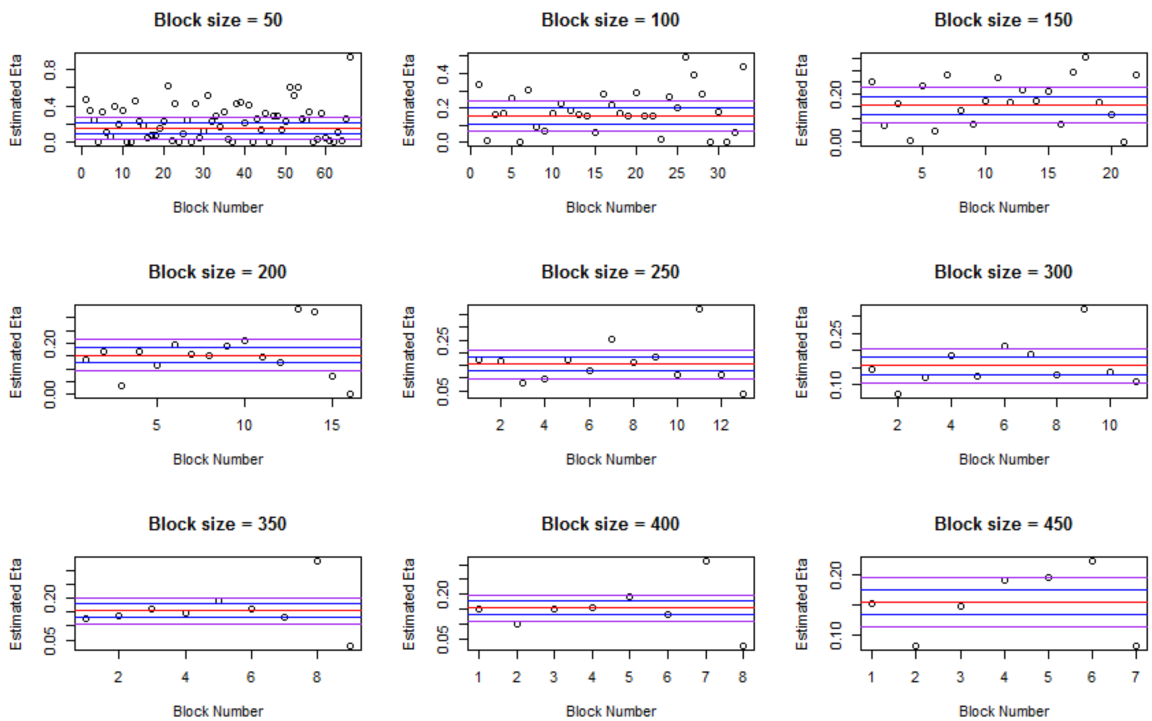

Figure 2: Estimated friction parameter over time for different values of h_n. The red line corresponds to the global estimate. The blue lines are one (local) standard deviation away from the global estimator. The purple lines are two (local) standard deviations away.

h_n	B_n	Chi Sq. Stat	Dg. Fr.	p-value
50	67	719	66	≈ 0
100	34	268	33	≈ 0
150	23	155	22	≈ 0
200	17	116	16	≈ 0
250	14	109	13	≈ 0
300	12	68.5	11	≈ 0
350	10	90.6	9	≈ 0
400	9	91.5	8	≈ 0
450	8	42.6	7	≈ 0

Table 4: Summary chi-square statistics χ_n^2 based on the block size h_n.

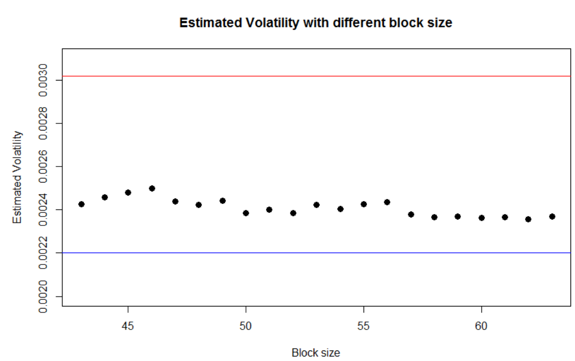

Figure 3: Estimated volatility with the LPE for different values of h_n. The red line corresponds to the RV estimator. The blue line stands for the global model with uncertainty zones volatility estimator.