THE UNIVERSITY OF CHICAGO

# ESTIMATING THE INTEGRATED PARAMETER OF THE LOCALLY PARAMETRIC MODEL IN HIGH-FREQUENCY DATA

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS

BY
YOANN POTIRON

CHICAGO, ILLINOIS
MARCH 2016

# CONTENTS

# ABSTRACT

Three chapters are included in this work. The first chapter introduces the remaining chapters. The description of the research chapters (Chapters 2 and 3) can be found in the following paragraphs.

In Chapter 2, we consider the problem of estimating high-frequency covariance (quadratic covariation) of two arbitrary assets observed asynchronously. Simple assumptions, such as independence, are usually imposed on the relationship between the prices process and the observation times. In Chapter 2, we introduce a general endogenous two-dimensional nonparametric model. Because an observation is generated whenever an auxiliary process called *observation time process* hits one of the two boundary processes, it is called the *hitting boundary process with time process* (HBT) model. We establish a central limit theorem for the Hayashi-Yoshida estimator under HBT in the case where the price process and the observation price process follow a continuous Itô process. We obtain an asymptotic bias. We provide an estimator of the latter as well as a bias-corrected estimator of the high-frequency covariance. In addition, we give a consistent estimator of the associated standard error.

In Chapter 3, we show that the techniques used to solve the high-frequency covariance problem can actually be applied to other problems in the high-frequency literature. We give a general time-varying parameter model, where the multidimensional parameter follows a continuous local martingale. As such, we call it the *locally parametric model* (LPM). The quantity of interest is defined as the *integrated value over time* of the parameter process $\Theta := T^{-1} \int_0^T \theta_t^* dt$. We provide an estimator of $\Theta$ based on a parametric estimator in the original (non time-varying) parametric model and conditions under which we can show consistency and the corresponding central limit theorem.

Since the estimator is obtained by chopping the data into small blocks, then estimating the parameter on each block while pretending it is constant locally and finally taking a block length weighted mean of the estimates on each block, we call it the *local parametric estimator* (LPE). The class of estimators is very broad, and can contain estimators that are (not too) biased, such as the *bias-corrected MLE*. We show that the LPM class contains some models that come from popular problems in the high-frequency financial econometrics literature (estimating volatility, high-frequency covariance, integrated betas, leverage effect, volatility of volatility), as well as a new general asset-price diffusion model which allows for endogenous observations and time-varying noise which can be auto-correlated and correlated with the efficient price. Finally, as another example of how to apply the limit theory provided in Chapter 3, we build a time-varying friction parameter extension of the (semi-parametric) model with uncertainty zones (Robert and Rosenbaum (2012)) and we show that we can easily verify the conditions for the estimation of integrated volatility.

# ACKNOWLEDGEMENTS

I would like to thank first my advisor Per Mykland who helped me a lot. Also, I would like to thank Dacheng Xiu, Dan Nicolae, Simon Clinet, Mathieu Rosenbaum, Steven Lalley, Takaki Hayashi, Nakahiro Yoshida and Greg Lawler for helpful discussions and advice. Finally, I am very thankful to my classmates Eileen Li and Yongrui Chen for proofreading the manuscript.

# LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1
# INTRODUCTION

High-frequency data analysis has been one of the main challenging fields in statistics, both because of the quantity of data available and the highly technical theory required to tackle problems. Ever since the seminal paper of Black and Scholes (1973), researchers have been interested in the estimation of integrated volatility. One typical reason why practitioners care very much about integrated volatility is that they can hedge knowing its value. Similar problems found in the high-frequency literature include but are not limited to high-frequency covariance, microstructure noise, integrated betas, leverage effect, volatility of volatility, etc.

Because the theory involved in estimating high-frequency quantities is very demanding, researchers have to make assumptions, usually stronger than what empirical studies reveal, in order to show consistency and the central limit theorem of a given estimator, as is the case with the Hayashi-Yoshida estimator (HY) of high-frequency covariance introduced in Hayashi and Yoshida (2005). It is usually assumed that the order arrival times are independent of the asset price. In Chapter 2, we investigate what happens to the HY when sampling times are correlated with the stochastic price. We obtain an asymptotic bias and give a bias-corrected estimator of high-frequency covariance.

Indeed, covariation between two assets is a crucial quantity in finance. Fundamental examples include optimal asset allocation and risk management. In the past few years, using the increasing amount of high-frequency data available, many papers have been published about estimating this covariance. Suppose that the latent log-price of two arbitrary assets $X_t = (X_t^{(1)}, X_t^{(2)})$ follows a continuous Itô process

$$dX_t^{(1)} := \mu_t^{(1)} dt + \sigma_t^{(1)} dW_t^{(1)}, \tag{1.1}$$

$$dX_t^{(2)} := \mu_t^{(2)} dt + \sigma_t^{(2)} dW_t^{(2)}, \tag{1.2}$$

where $\mu_t^{(1)}, \mu_t^{(2)}, \sigma_t^{(1)}, \sigma_t^{(2)}$ are random processes, and $W_t^{(1)}$ and $W_t^{(2)}$ are standard Brownian motions, with (random) high-frequency correlation $d\langle W^{(1)}, W^{(2)}\rangle_t = \rho_t dt$. Econometrics usually seeks to infer the *integrated covariation*

$$\langle X^{(1)}, X^{(2)}\rangle_t := \int_0^t \rho_u \sigma_u^{(1)} \sigma_u^{(2)} du. \tag{1.3}$$

Earlier results were focused on estimating the integrated variance of a single asset, starting from the probabilistic point of view (Genon-Catalot and Jacod (1993), Jacod (1994)). Barndorff-Nielsen and Shephard (2001, 2002) introduced the problem in econometrics. Adapted to two dimensions, if each process is observed simultaneously at (possibly random) times $\tau_{0,n} := 0$, $\tau_{1,n}$ , ..., $\tau_{N_n,n}$ the *realized covariation* $\left[X^{(1)}, X^{(2)}\right]_t$ is defined as the sum of cross log returns

$$\left[X^{(1)}, X^{(2)}\right]_t = \sum_{\tau_{i,n} \leq t} \Delta X_{\tau_{i,n}}^{(1)} \Delta X_{\tau_{i,n}}^{(2)}, \tag{1.4}$$

where for any positive integer $i$, $\Delta X_{\tau_{i,n}}^{(k)} = X_{\tau_{i,n}}^{(k)} - X_{\tau_{i-1,n}}^{(k)}$ corresponds to the increment of the $k$th process between the last two sampling times. As the observation intervals $\Delta\tau_{i,n}$ get closer (and the number of observations $N_n$ goes to infinity), $\left[X^{(1)}, X^{(2)}\right]_t \xrightarrow{\mathbb{P}} \langle X^{(1)}, X^{(2)}\rangle_t$ (see e.g. Theorem I.4.47 in Jacod and Shiryaev (2003)). Furthermore, when the observation times $\tau_{i,n}$ are independent of the prices process $X_t$, its estimation error follows a mixed normal distribution (Jacod and Protter (1998), Zhang (2001), Mykland and Zhang (2006)). This gives us insight on how to estimate the integrated covariation. However, in practice, these two assumptions are usually not satisfied. The observation times of the two assets are rarely *synchronous* and there is *endogeneity* in the price sampling times.

The first issue has been studied for a long time. The lack of synchronicity often creates undesirable effects in inference. If we sample at very high frequencies, we observe the Epps effect (Epps (1979)), i.e. the correlation estimates are drastically decreased compared to an estimate with sparse observations. We can observe the same effect for exchange rates (Guillaume et al. (1997), Muthuswamy et al. (2001)). Additionnally, asynchronicity can cause difficulties in daily data (Scholes and Williams (1977)). At first, the literature on estimating the covariance mostly relied on a forced synchronization of the data (see, e.g., Lundin et al. (2001), Brandt and Diebold (2003)), for instance choosing beforehand a window of size h, and interpolating the values of the two assets at times $\{h, 2h, \ldots, Mh\}$. Hayashi and Yoshida (2005) introduced the so-called *Hayashi-Yoshida estimator* (HY)

$$\widehat{\langle X^{(1)}, X^{(2)} \rangle}_t^{HY} = \sum_{\tau_{i,n}^{(1)}, \tau_{j,n}^{(2)} < t} \Delta X_{\tau_{i,n}^{(1)}}^{(1)} \Delta X_{\tau_{j,n}^{(2)}}^{(2)} \mathbf{1}_{\left\{[\tau_{i-1,n}^{(1)}, \tau_{i,n}^{(1)}) \cap [\tau_{j-1,n}^{(2)}, \tau_{j,n}^{(2)}) \neq \emptyset\right\}}, \tag{1.5}$$

where $\tau_{i,n}^{(k)}$ are the observation times of the $k$th asset. Note that if the observations of both processes occur simultaneously, (1.4) and (1.5) are equal. One of main advantages of HY is that it is nonparametric in nature.

We insist on the fact that the primary goal of Chapter 2 is to provide a better estimator of the high-frequency covariance than the usual Hayashi-Yoshida estimator. To obtain this new estimator, we will estimate the second-order bias, and remove it from the Hayashi-Yoshida estimator. Note that estimating this bias is much more challenging than in the volatility case because observations are asynchronous. In particular, the estimator will involve a quantity that can be considered as the *tricity* of Li et al. (2014), but with a more intricate definition because of the asynchronicity in sampling times. This new definition can be seen as an analogy with the generalization of the RV estimator (1.4) by the HY estimator (1.5).

Another very important issue to address is the estimation of the asymptotic standard deviation. First, because the model is more general than in the no-endogeneity work, the theoretical asymptotic variance will be different. Consequently, a new variance estimator, which takes into proper account the endogeneity, will be given. This estimator is built mostly on the observed price of the assets, whereas the variance estimate of the Hayashi-Yoshida machinery (see e.g. Hayashi and Yoshida (2011) and Koike (2013)) relied heavily on the actual observation times. In practice, there is a lag between the time when a stock market's agent is giving a bid-ask order and the recorded (observed) time. As such, we cannot trust completely the observation times. Consequently, even in the no-endogeneity case, the variance estimate provided in Chapter 2 has a big advantage compared to the previous variance estimator.

The proof of the high-frequency covariance problem goes as follows. We first show that we can bound uniformly the variance of HY when assuming a parametric problem, i.e. fixing the volatility and high-frequency correlation between the two assets. Then, we chop the data into non-overlapping blocks of observations and show that the normalized discrepancy between the nonparametric error of HY and the parametric approximation of the error uniformly vanishes asymptotically.

These techniques are not problem-specific and they can very much be applied to other problems in the high-frequency literature. This is the basic idea behind Chapter 3. We assume in Chapter 3 that the stochastic multidimensional parameter $\theta_t^*$ is driving the structure of the returns. This leads us to the problem of estimating the integrated parameter

$$\Theta := \frac{1}{T}\int_0^T \theta_s^* ds. \tag{1.6}$$

The parameter process can be for example equal to the volatility, the covariation between several assets (1.3), the time-varying variance of the microstructure noise, the friction parameter of the model with uncertainty zones (see Section 3.7.1 for more details), the betas, the volatility of volatility, the leverage effect process, the event arrival rate or any other parameter driving the observation times (see Section 3.7.1), the parameters of a self-exciting process (Hawkes (1971)) or in general of any time-varying parameter extension of a parametric model, etc. As an example of application of (1.6), Chen and Mykland (2015) measured the liquidity risk when the parameter is equal to the time-varying variance of the microstructure noise.

We assume that the econometrician has a time-varying parameter model (and in particular a parametric model) at hand, that can be written as the *Locally Parametric Model* (LPM) defined in Section 3.5. The LPM covers a large class of models, as we will see in the examples of Section 3.7. In particular, in the case where we model the price of an asset as a continuous *efficient* stochastic process, it allows for endogeneity (the sampling times can be correlated with the efficient price and the microstructure noise), auto-correlated time-varying noise and correlation between the efficient return and the noise, as well as multidimensional asynchronous observations.

Indeed, in high-frequency data, when estimating a quantity such as volatility, one has to first build and rely on a model of the observations. When observations are close, empirical studies strongly suggest that *the market microstructure* generates a divergence between the observed price process and efficient price process. This divergence could be induced among other things by transaction price changes occurring on the tick grid (price discreteness) or by the existence of a waiting-list for sellers and buyers at each level of price (bid-ask spreads). Accordingly, the market microstructure is very often assumed to be stationary (i.e. with non time-varying variance) and not autocorrelated, and independent of the true price process. Nonetheless, Hansen and Lunde (2006) have shown empirically that the microstructure noise is time-dependent and correlated with the efficient price itself. Correlation between the efficient price and the noise can be explained by rounding effects, price stickiness, asymmetric information, etc. Kalnina and Linton (2008) introduced in their model possible non-stationarity of noise and correlation.

We also assume that the econometrician has a parametric estimator in the (non time-varying) parametric model at hand, such as the MLE. Chapter 3 aims to build general estimators $\widehat{\Theta}_n$ of (1.6), based on the parametric estimator. When parameters are time-varying, researchers frequently chop the available data into short time blocks of size $h_n$ and assume homoskedasticity within the blocks (e.g. Foster and Nelson (1996)). In Chapter 3, we propose to use the same strategy to estimate the integrated parameter (1.6): we estimate the *local parameter* on each block by using the parametric estimator on the observations of the block and take a weighted sum of the local parameter estimates, where each weight is equal to the corresponding block length. We call the obtained estimator the *local parametric estimator* (LPE). Depending on the case, the LPE can actually differ from the original parametric estimator. For this technique to work, we need that the parametric estimator is not badly-biased and that it uniformly satisfies limit theory conditions, in addition to other assumptions that can be found in Section 3.6.

Depending on the model and the estimator chosen by the econometrician, the conditions of Section 3.6 can be straightforward or not to verify. Nonetheless, our hope is that for any LPM and for a large class of estimators Chapter 3 will help the econometrician by breaking the original nonparametric problem into two easier sub-problems, a parametric problem and a control of the error between the nonparametric and parametric problem. We used exactly this strategy in Chapter 2 to find the limit distribution of the integrated covariation in the HBT model (which is contained in the LPM class), with the bias-corrected HY estimator (which can be expressed as a local parametric estimator (LPE)). In Section 3.7.1, we give another direct application of the techniques provided in Chapter 3 by verifying conditions of Section 3.2 for another problem, the estimation of the integrated volatility in a time-varying friction parameter extension of the model with uncertainty zones, using the parametric estimator of volatility provided in Robert and Rosenbaum (2012).

We assume in chapter 3 that the stochastic parameter is continuous. We prove that under the diffusion assumption we can trust the parametric model locally. The order of the block size $h_n$ to use is also given in Section 3.6. The idea that continuous local martingale parameter implies that the model is locally true builds on previous investigations by Mykland and Zhang (2009, 2011) of the maximum size of a neighborhood in which we can hold volatility of an asset constant. In the case where the observation times of the price process are endogenous, techniques were extended in Chapter 2.

Finally, from a practical point-of-view, empirical work (see Section 3.8) with Orange France Telecom stock on the CAC 40 using the time-varying friction parameter extension of the model with uncertainty zones reveal that the friction parameter is indeed time-varying and that the volatility estimates obtained with the LPE can differ very much from the ones using the original parametric estimator. In addition, they indicate that the estimates are robust to minor changes in the block size $h_n$.

To sum up, we expect the techniques of this Ph.D. thesis to provide to researchers automatic estimators of high-frequency quantities and techniques to prove consistency and the associated central limit theorem.

# CHAPTER 2
# ESTIMATING THE HIGH-FREQUENCY COVARIANCE

## 2.1 Introduction

The consistency of HY estimator was originally achieved in a non-random volatility and independence between observation times and assets' prices setting (Hayashi and Yoshida (2005)) before being extended to a general Itô-process price model, with the unique assumption that the observation times are stopping times (Hayashi and Kusuoka (2008)). The corresponding central limit theorems were investigated in Hayashi and Yoshida (2008, 2011) under strong predictability of observation times, which is a more restrictive assumption than only assuming they are stopping times but still allows some dependence between prices and observation times. Note that the second order asymptotic expansion was completed in Dalalyan and Yoshida (2011). Since then, most of the literature has been interested in suitable modifications of this estimator, or the use of a different estimator, that is robust to noisy observations (Zhang (2011), Barndorff-Nielsen et al. (2011), Aït-Sahalia et al. (2010), Christensen et al. (2010), Christensen et al. (2013) among others), with assumption of independence between the sampling times and the prices of the assets. Recently, Koike (2014, 2015) extended the pre-averaged Hayashi-Yoshida estimator first under predictability of observation times, and then under a more general endogenous setting of stopping times. Interestingly, we can read in Remark 3.5 ($i$) of Koike (2015) that under the assumptions chosen, the observation times affect the asymptotic distribution of the realized covariance estimator only through the asymptotic variance, but not through the asymptotic bias.

In a general one-dimensional endogenous model, the asymptotic behaviour of the realized volatility (1.4) has been investigated in the case of sampling times given by hitting times on a regular grid (Fukasawa (2010a)). The model with uncertainty zones, a more intricate model based on hitting times of random and time-dependent grid, was introduced and studied in Robert and Rosenbaum (2011, 2012). Also, a central limit theorem under hitting times of a non-random, non time-dependent but irregular grid was established in Fukasawa and Rosenbaum (2012). Note that due to the regularity of those three models (see the discussion in the latter paper), they don't obtain any bias in the limit distribution of the normalized error. Also, the case of strongly predictable stopping times is treated in Hayashi et al. (2011). Two general results (Fukasawa (2010b), Li and al. (2014)) showed that we can identify and estimate the asymptotic bias. In the latter work, the authors provide a new estimator of volatility that is free of asymptotic bias. Correspondingly, they give an estimator of the asymptotic variance, that variance being different from the theoretical variance when there is no endogeneity in sampling times.

The authors want to take no position on the joint distribution of the log-return and the next observation time that corresponds to an asset price change because they know that their unknown relationship is most likely contributing to the bias and the variance of the high-frequency covariance's estimate when we (wrongly) assume full independence between the price process and observation times. We want to understand how badly HY can be affected in the worst endogenous setting. For this purpose, we introduce the *hitting boundary process with time process* (HBT) model, which under weak conditions is more general than the models in the existing literature (e.g. the model with uncertainty zones or the structural autoregressive conditional duration model (Renault et al. (2014)). Even though the mixing variables (in keeping with the notations of the paper) $\tilde{\mu}_{t_i} = \mu_{t_i}(M_i)$ and $\tilde{c}_{t_i} = c_{t_i}(M_i)$ introduced in (1), (4) and (5) of the dynamic mixed hitting-time model can generate all kinds of autoregressive or log-autoregressive dynamics, it only partly accommodates for what Russell (1999) notes: "The problem is that it is difficult to model the distribution of a duration when new information can arrive within a duration." Indeed, the model doesn't capture new information arriving between two trades, because $\mu_{t_i}$ and $c_{t_i}$ must be known with information at time $t_i$. One natural way to accomodate for new information to arrive between two trades is to allow the fixed variables $\mu_{t_i}$ and $c_{t_i}$ to follow a stochastic process (within a trade), which is what we are going to do in this chapter. Note that the one dimensional HBT model is straightforwardly extended to the multidimensional case.

The HBT model is not solely a general class of models. It can provide semiparametric and nonparametric models which can be fitted to the data. The parameter estimates of those models could provide crucial information on some deep financial questions about the connection between efficient price return and observation time. As an example, we could build tests of *asymmetric information* with the model introduced in Section 2.3.2 of this work. Also, if we are able to shed light on the structure driving the endogeneity, stock market agents could trade or hedge on the basis of that information. Investigating what the endogenous structure looks like empirically is beyond the scope of Chapter 2 and will be left for further work.

As far as the authors know, no investigation of a possible bias in an endogenous model has been carried out when estimating the high-frequency covariance. This work can be considered as one of the last building blocks of the wall of the limit theory of the Hayashi-Yoshida estimator when holding the asset price continuous because we choose to work under the weakest assumptions (adapted to our proofs) regarding the observation times.

Finally, techniques developed in the proofs are innovative in the sense that they reduce the normalized error of the Hayashi-Yoshida estimator to a discrete process, which is locally a uniformly ergodic homogeneous Markov chain. Thus, the problem can be solved locally, and because we assume that the volatility of assets is continuous, the error of approximation between the local Markov structure and the real structure of the normalized error vanishes asymptotically.

Chapter 2 is organized as follows. We introduce the HBT model in Section 2.2.

Examples contained in this class are given in Section 2.3. The main theorem of this work, the limit distribution of the normalized error is given in Section 2.4. Estimators of the asymptotic bias and variance are provided in Section 2.5. We carry out numerical simulations in Section 2.6 to corroborate the theory. Proofs are developed in Appendix.

## 2.2 Definition of the HBT model

We first introduce the model in 1-dimension. We assume that $X_t$ is the efficient (log-)price of the security of interest. In practice, rather than continuously observing the asset price $X_t$ for any $t$, we can only observe it at discrete observation times $\tau_i$. In addition, we assume that the observations are noisy and that we observe $Z_{\tau_i} := X_{\tau_i} + \epsilon_{\tau_i}$ where the microstructure noise $\epsilon_{\tau_i}$ can be expressed as a known function of the observed prices $Z_0, \ldots, Z_{\tau_i}$. Thus, when observing $Z_0, \ldots, Z_{\tau_{N_n}}$, we can retrieve exactly the value of the efficient price at arrival times $X_0, \ldots, X_{\tau_{N_n}}$. As an example, Robert and Rosembaum (2012) showed in (2.3) in $p.5$ that the model with uncertainty zones can be written with that noise structure if we assume that we know the endogeneity parameter $\eta$. The model with uncertainty zones will be studied in Example 4 of Section 2.3. Note that the microstructure noise can be explained among other things by the discreteness of prices, the bid-ask mechanisms, etc.

In this work, we assume that for any positive integer $i$, $\tau_{i+1}$ is the next arrival time (after $\tau_i$) that corresponds to an actual change of price. In particular, several trades can occur at the same price $Z_{\tau_i}$ between $\tau_i$ and $\tau_{i+1}$, but no trade can occur with a price different than $Z_{\tau_i}$ before $\tau_{i+1}$. On the contrary, because $\tau_i$ corresponds to a change of price, we assume $Z_{\tau_{i+1}} \neq Z_{\tau_i}$. Additionally, if we define $\alpha > 0$ the tick size, we assume that the observed price $Z_{\tau_i}$ lays on the tick grid, i.e. there exists positive integers $m_i$ such that $Z_{\tau_i} := m_i \alpha$.

Empirically, no economical model based on rational behaviors of agents on the stock markets, that shed light on the relationship between the efficient return $\Delta X_{\tau_i}$ and time before the next price change $\Delta\tau_i = \tau_i - \tau_{i-1}$, has won unanimous support. When arrival times are independent of the asset price, it follows directly from the continuous Itô-assumption that the dependence structure is such that the return $\Delta X_{\tau_i}$ is a function of $\Delta\tau_i$. The longer we wait, the bigger the variance of the return is expected to be. In Chapter 2, we take the opposite point of view by building a model in which $\tau_i$ is defined as a function of the efficient price path. Our goal is to provide a model that allows the most general structure between $\Delta X_{\tau_i}$ and $\Delta\tau_i$. The definition of the model that follows in this section will be the starting point of further work on endogeneity because it is also possible to allow more general microstructure noise in the model. One other quick extension of the model is that it could also allow error in the arrival times. Investigating both those issues is beyond the scope of Chapter 2.

In general, the space $\mathcal{E}$ of all possible joint-distributions $\left(\Delta X_{\tau_i}, \Delta\tau_i\right)$ is too large to work with. We define the *observation time process* $X_t^{(t)}$ that will drive the observation times. We also define the *down process* $d_t(s)$ and the *up process* $u_t(s)$. Note that for any $t \geq 0$, we assume that $d_t$ and $u_t$ are functions on $\mathbb{R}^+$. We also assume that the down process takes only negative values, the up process takes only positive values and the observation time process starts from 0. A new observation time will be generated whenever one of those two processes is hit by the observation time process. Then, the observation time process will start again from 0, and the next observation time will be generated whenever it hits the up or the down process. Formally, we define $\tau_0 := 0$ and for any positive integer $i$

$$\tau_i := \inf\left\{t > \tau_{i-1} : \Delta X^{(t)}_{[\tau_{i-1},t]} \notin \left[d_t\left(t-\tau_{i-1}\right), u_t\left(t-\tau_{i-1}\right)\right]\right\}, \tag{2.1}$$

where $\Delta Y_{[a,b]} := Y_b - Y_a$. Note that if the observation time process $X_t^{(t)}$ is equal to the price process $X_t$ itself, then the price will go up (respectively go down) whenever it hits the up process (down process). Note also that if the time process, the up process and the down process are independent of the efficient price process, then the arrival times are independent of the efficient price process. We assume that the two-dimensional process $(X_t, X_t^{(t)})$ is an Itô-process.. Section 3.3.1 provides examples of the literature identifying the observation time process, the down process and the up process.

Generalizing to two dimensions is straightforward. We define $X_t^{(t,k)}$ for $k = 1, 2$ to be the observation time process associated with the $k$th price process , $u_t^{(k)}$ the up process and $d_t^{(k)}$ the down process, and the arrival times $\tau_i^{(k)}$ generated by (2.1). We also define the four dimensional process $Y_t := (X_t^{(1)}, X_t^{(2)}, X_t^{(t,1)}, X_t^{(t,2)})$, and assume $Y_t$ follows an Itô-process with volatility

$$\sigma_t := \begin{pmatrix} \sigma_t^{1,1} & \sigma_t^{1,2} & \sigma_t^{1,3} & \sigma_t^{1,4} \\ \sigma_t^{2,1} & \sigma_t^{2,2} & \sigma_t^{2,3} & \sigma_t^{2,4} \\ \sigma_t^{3,1} & \sigma_t^{3,2} & \sigma_t^{3,3} & \sigma_t^{3,4} \\ \sigma_t^{4,1} & \sigma_t^{4,2} & \sigma_t^{4,3} & \sigma_t^{4,4} \end{pmatrix}.$$

In particular, we have $dY_t = \mu_t dt + \sigma_t dW_t$, where $W_t$ is a four dimensional standard Brownian motion (for $i = 1, \ldots, 4$ and $j = 1, \ldots, 4$ such that $i \neq j$, $W_t^{(i)}$ is independent of $W_t^{(j)}$). If we set $\zeta_t = \sigma_t \sigma_t^T$, then the integrated covariance (or quadratic covariation) process is given by $\langle Y, Y \rangle_t = \int_0^t \zeta_s ds$. Let $\rho_t$ be the associated correlation process of $Y_t$, i.e. for $i = 1, \ldots, 4$ and $j = 1, \ldots, 4$ we set $\rho_t^{i,j} = \zeta_t^{i,j} (\zeta_t^{i,i})^{-1}$. Finally, it is useful sometimes to see $Y_t$ as a four dimensional vector expressed as in equations (1.1) and (1.2). For $k = 1, \ldots, 4$ we define the volatility of the $k$th process as $\sigma_t^{(k)} := (\zeta_t^{k,k})^{\frac{1}{2}}$, we can thus express $Y_t^{(k)}$ as

$$dY_t^{(k)} = \mu_t^{(k)} dt + \sigma_t^{(k)} dB_t^{(k)}$$

where $B_t^{(k)}$ is a standard Brownian motion, which typically depends on $B_t^{(l)}$ for $l = 1, \ldots, 4$.

## 2.3 Examples

We insist on the fact that estimators of covariance and associated asymptotic variance given in Chapter 2 don't require any knowledge of the structure of the observation time process, the up process and the down process. Nonetheless, for financial and economic interpretation purposes, the reader might be interested in getting an idea on how those processes behave in practice. We provide in this section several examples from the literature as well as possible extensions of the model with uncertainty zones of Robert and Rosenbaum (2011) that can be expressed as HBT models.

### *2.3.1 Endogenous models contained in the HBT class*

*Example* 1. (hitting constant boundaries) The simplest endogenous semi-parametric model we can think of is a model where the time process $X_t^{(t)}$ is equal to the price process $X_t$, and times are generated by hitting a constant barrier. Formally, it means that there exists a two-dimensional parameter $(\theta_u, \theta_d)$ such that the up process is equal to $\theta_u$ and the down process is equal to $\theta_d$. We don't assume noise in that model.

*Example* 2. (hitting constant boundaries of the tick size) One issue with Example 1 is that the efficient price $X_{\tau_i}$, which is observed because no microstructure noise is assumed in the model, is not necessarily a modulo of the tick size $\alpha$ if $\theta_u$ and $\theta_d$ are not multiples of $\alpha$. To make Example 2.3.1 feasible in practice, we assume here that the constant barriers $\theta_u$ and $\theta_d$ are respectively equal to the tick size $\alpha$ and its additive inverse $-\alpha$. We also assume that $Z_{\tau_i} := X_{\tau_i}$.

*Example* 3. (hitting constant boundaries of the jump size) The issue with Example 2 is that the absolute jump size of the observed price $Z_{\tau_i}$ is $\alpha$. On the contrary, in practice the absolute jump size can actually be bigger than the tick size $\alpha$. In the notation of Robert and Rosenbaum (2011), for any positive integer $i$, we introduce a discrete variables $L_i$ which corresponds to the observed price jump's tick number between $\tau_{i-1}$ and $\tau_i$, with $L_i \geq 1$. The arrival times are defined recursively as $\tau_0 := 0$ and for any positive integer $i$

$$\tau_i := \inf\Big\{t > \tau_{i-1} : X_t = X_{\tau_{i-1}} - L_{i-1}\alpha \text{ or } X_t = X_{\tau_{i-1}} + L_{i-1}\alpha\Big\}$$

We assume that $L_i$ are IID and independent of the other quantities. We finally assume that $Z_{\tau_i} := X_{\tau_i}$. The up and down processes are piecewise constant in $t$ and constant in $s$, defined for any $s \geq 0$ as

$$\begin{aligned} d_t(s) = & \ -L_{i-1}\alpha \quad \text{for } t \in (\tau_{i-1}, \tau_i] \\ u_t(s) = & \ \ \ L_{i-1}\alpha \quad \text{for } t \in (\tau_{i-1}, \tau_i] \end{aligned}$$

*Example* 4. (model with uncertainty zones) We go one step further Example 3 and introduce now the model with uncertainty zones of Robert and Rosenbaum (2011). In a frictionless market, we can assume that a trade with change of price $Z_{\tau_i}$ will occur whenever the efficient price process crosses one of the mid-tick values $Z_{\tau_{i-1}} + \frac{\alpha}{2}$ or $Z_{\tau_{i-1}} - \frac{\alpha}{2}$. In that case, if the efficient price process hits the former value, we would observe an increment of the observed price $Z_{\tau_i} = Z_{\tau_{i-1}} + \alpha$ and if it hits the latter value, we would observe a decrement $Z_{\tau_i} = Z_{\tau_{i-1}} - \alpha$. There are two reasons why in practice such a frictionless model is too simplistic. The first reason is that the absolute value of the increment (or the decrement) of the observed price can be bigger than the tick size $\alpha$ and was already pointed out in Example 3. We will thus keep the notation $L_i$ in this example. The second reason is that the frictions induce that the transaction will not exactly occur when the efficient process is equal to the mid-tick values. For this purpose in the notation of Robert and Rosenbaum (2012), let $0 < \eta < 1$ be a parameter that quantifies the aversion to price changes of the market participants. If we let $X_t^{(\alpha)}$ be the value of $X_t$ rounded to the nearest multiple of $\alpha$, the sampling times are defined recursively as $\tau_0 := 0$ and for any positive integer $i$

$$\tau_i := \inf\Big\{t > \tau_{i-1} : X_t = X^{(\alpha)}_{\tau_{i-1}} - \alpha\big(L_{i-1} - \frac{1}{2} + \eta\big) \text{ or } X_t = X^{(\alpha)}_{\tau_{i-1}} + \alpha\big(L_{i-1} - \frac{1}{2} + \eta\big)\Big\}$$

The observed price is equal to the rounded efficient price $Z_{\tau_i} := X_{\tau_i}^{(\alpha)}$. The time process $X_t^{(t)}$ is again equal to the price process $X_t$ itself in this model. The up and down processes are piecewise constant in $t$ and constant in $s$, defined for any $s \geq 0$ as

$$\begin{aligned} d_t(s) = & \quad -L_{i-1}\alpha\mathbf{1}_{\{X_{\tau_{i-1}} < X_{\tau_{i-2}}\}} - (2\eta + L_{i-1} - 1)\,\alpha\mathbf{1}_{\{X_{\tau_{i-1}} > X_{\tau_{i-2}}\}} \quad \text{for } t \in (\tau_{i-1}, \tau_i] \\ u_t(s) = & \quad L_{i-1}\alpha\mathbf{1}_{\{X_{\tau_{i-1}} > X_{\tau_{i-2}}\}} + (2\eta + L_{i-1} - 1)\,\alpha\mathbf{1}_{\{X_{\tau_{i-1}} < X_{\tau_{i-2}}\}} \quad \text{for } t \in (\tau_{i-1}, \tau_i] \end{aligned}$$

where $\mathbf{1}_A$ is the indicator function of A. Note that in the case where $\eta = \frac{1}{2}$, we are back to Example 3.

*Example* 5. (times generated by hitting an irregular grid model) The fourth model we are looking at is called *times generated by hitting an irregular grid model*. We follow the notation of Fukasawa and Rosenbaum (2012) and consider the irregular grid $\mathcal{G} = \{p_k\}_{k \in \mathbb{Z}}$, with $p_k < p_{k+1}$. We set $\tau_0 = 0$ and for $i \geq 1$

$$\tau_i = \inf\left\{t > \tau_{i-1} : X_t \in \mathcal{G} - \{X_{\tau_{i-1}}\}\right\},$$

where $\mathcal{G} - \{X_{\tau_{i-1}}\}$ is the set obtained by removing $\{X_{\tau_{i-1}}\}$ from $\mathcal{G}$. We can rewrite it as an element of the HBT model where the time process is equal to the price process, and for all $s \geq 0$ the up and down processes are defined as

$$\begin{aligned} d_t(s) = & \quad p_{k-1} - p_k \quad \text{for } t \in (\tau_{i-1}, \tau_i] \\ u_t(s) = & \quad p_{k+1} - p_k \quad \text{for } t \in (\tau_{i-1}, \tau_i], \end{aligned}$$

where $k$ is the (random) index such that $p_k = X_{\tau_{i-1}}$.

*Example* 6. (structural autoregressive conditional duration model) There have been several drafts for this model. We follow here a former version (Renault, van der Heijden and Werker (2009)), because we can directly express it as an element of the HBT model[1].

[1]Generating the sampling times (2.1) of the HBT model as a first hitting-time of a unique barrier instead of the first hitting time of one of two barriers as in the latter version of Renault et al. (2014)

In the structural autoregressive conditional duration model, the time $\tau_i$ when the next event occurs is given by $\tau_0 = 0$ and for $i > 0$

$$\tau_i = \inf \left\{ t > \tau_{i-1} : A_t - A_{\tau_{i-1}} = \tilde{d}_{\tau_{i-1}} \text{ or } A_t - A_{\tau_{i-1}} = \tilde{c}_{\tau_{i-1}} \right\}$$

where $A_t$ is a standard Brownian motion (not necessarily independent of $X_t$). Expressed as an element of the HBT model, we have that the time process $X_t^{(t)}$ is equal to the Brownian motion $A_t$ and for all $s \geq 0$

$$\begin{aligned} d_t(s) = &\ \tilde{d}_{\tau_{i-1}} \text{ for } \quad t \in (\tau_{i-1}, \tau_i] \\ u_t(s) = &\ \tilde{c}_{\tau_{i-1}} \text{ for } \quad t \in (\tau_{i-1}, \tau_i]. \end{aligned}$$

### *2.3.2 Possible extensions of the model with uncertainty zones*

The model with uncertainty zones of Robert and Rosenbaum (2011) introduced in Example 4, which is semi-parametric, assumes that the observed price is the efficient price rounded to the nearest tick value $Z_{\tau_i} = X_{\tau_i}^{(\alpha)}$ and thus the noise is equal to $\epsilon_i := \alpha(\frac{1}{2} - \eta)$ if the last trade increased the price and $\epsilon_i := -\alpha(\frac{1}{2} - \eta)$ if the last trade decreased the price. In particular, the noise is auto-correlated and correlated to the efficient price. Because of this specific noise distribution, it is directly possible to estimate the underlying friction parameter $\eta$ without any data pre-processing such as preaveraging (see Robert and Rosenbaum (2012)). We believe the model with uncertainty zones is a very interesting starting point, because all the endogenous and noise structure of the model is reduced to the estimation of the 1-dimensional friction parameter $\eta$. Nevertheless, as this semi-parametric model wants to be the simplest, it suffers from several issues. We will investigate two of them in the following.

First, the model doesn't allow for asymmetric information between the buyers and

---

wouldn't change much the proofs of Chapter 2, but we chose the two-boundaries setting because it seems more natural if interpretation of time processes, up processes and down processes is needed.

the sellers. Define $\eta^+$ and $\eta^-$, which are respectively the aversion to a positive price change and a negative price change. As a positive price change means that a buyer decided to put an order at the best ask price and a negative price change corresponds to a seller that puts an order at the best bid price (if we assume that cancel and repost orders are not the reason why the price changed), the difference $\eta^+ - \eta^-$ can be seen as a measure of information asymmetry. We define $\tau_0 := 0$ and recursively for $i$ any positive integer

$$\tau_i := \inf\left\{t > \tau_{i-1} : X_t = X^{(\alpha)}_{\tau_{i-1}} - \alpha\big(L_i - \frac{1}{2} + \eta^-\big) \text{ or } X_t = X^{(\alpha)}_{\tau_{i-1}} + \alpha\big(L_i - \frac{1}{2} + \eta^+\big)\right\}.$$

Note that the HBT class contains this model and that it can be directly fitted if we slightly modify $\hat{\eta}$ in Robert and Rosenbaum (2012) to estimate $\eta^+$ and $\eta^-$. One possible application would be to build a test of asymmetric information $\eta^+ := \eta^-$. This is beyond the scope of Chapter 2.

One other issue is that the authors don't do any model checking in their work. According to their empirical work (see pp.359-361 of Robert and Rosenbaum (2011)), the estimated values for $\eta$ are stable accross days for the ten French assets tested. Stability of $\eta$ favors their model but by doing so, the model doesn't allow any other structure than the full-endogeneity for the sampling times. Even if the true structure of sampling times is (mostly) independent of the asset price, we will still estimate an $\eta$ that will be stable across days. If we allow the time process to be different from the price process itself, we can estimate the correlation $\rho^{1,3}$ between them and see how endogenous the sampling times are (the bigger $\left|\rho^{1,3}\right|$ is, the more endogenous the sampling times are). We would need to add more general microstructure noise in the model, and thus this is left for further work.

## 2.4 Main result

### *2.4.1 Assumptions and Theorem*

Without loss of generality, we fix the horizon time $T := 1$, and we consider $[0, 1]$ to represent the course of an economic event, such as a trading day. We first introduce the definition of stable convergence, which is a little bit stronger than usual convergence in distribution and needed for statistical purposes of inference, such as the prediction value of the high-frequency covariance and the construction of a confidence interval at a given confidence level.

**Definition 1.** We suppose that the random processes $Y_t$, $\mu_t$ and $\sigma_t$ are adapted to a filtration $(\mathcal{F}_t)$. Let $Z_n$ be a sequence of $\mathcal{F}_1$-measurable random variables. We say that $Z_n$ converges stably in distribution to $Z$ as $n \to \infty$ if $Z$ is measurable with respect to an extension of $\mathcal{F}_1$ so that for all $A \in \mathcal{F}_1$ and for all bounded continuous[2] functions $f$, $\mathbb{E}\left[\mathbf{1}_A f\left(Z_n\right)\right] \to \mathbb{E}\left[\mathbf{1}_A f\left(Z\right)\right]$ as $n \to \infty$.

In the setting of Section 2.2, the target of inference, the integrated covariation, can be written for all $t \in [0, 1]$ as

$$\langle X^{(1)}, X^{(2)} \rangle_t := \int_0^t \sigma_s^{(1)} \sigma_s^{(2)} \rho_s^{1,2} ds.$$

We are providing now the asymptotics. We want to make the number of observations go to infinity asymptotically. The idea is to scale and thus keep the structure that drives the next return and the next observation time, while making the tick size vanish (and thus the number of observations explode on $[0, 1]$). Formally, we let the tick size

[2] Note that the continuity of $f$ refers to continuity with respect to the Skorokhod topology of $\mathbb{D}[0, 1]$. Nevertheless, we can also use continuity given by the sup-norm, because all our limits are in $\mathbb{C}[0, 1]$. One can look at Chapter $VI$ of Jacod and Shiryaev (2003) as a reference. For further definition of stable convergence, one can look at Rényi (1963), Aldous and Eagleson (1978), Chapter 3 (p. 56) of Hall and Heyde (1980), Rootzén (1980), and Section 2 (pp. 169-170) of Jacod and Protter (1998).

$\alpha > 0$ and we define the observation times $\mathrm{T}_\alpha := \{\tau_{i,\alpha}^{(k)}\}_{i\geq 0}^{k=1,2}$ such that for $k = 1, 2$ we have $\tau_{0,\alpha}^{(k)} := 0$ and for $i$ any positive integer

$$\tau_{i,\alpha}^{(k)} := \inf\Big\{t > \tau_{i-1,\alpha}^{(k)} : \Delta X_t^{(t,k)} \notin \big[\alpha d_t^{(k)}(t - \tau_{i-1,\alpha}^{(k)}), \alpha u_t^{(k)}(t - \tau_{i-1,\alpha}^{(k)})\big]\Big\}.$$

We define the HY estimator when the tick size is equal to $\alpha$

$$\langle \widehat{X^{(1)}, X^{(2)}} \rangle_{t,\alpha}^{HY} := \sum_{0<\tau_{i,\alpha}^{(1)}\ ,\ \tau_{j,\alpha}^{(2)}<t} \Delta X_{\tau_{i,\alpha}^{(1)}}^{(1)} \Delta X_{\tau_{j,\alpha}^{(2)}}^{(2)} \mathbf{1}_{\left\{[\tau_{i-1,\alpha}^{(1)},\tau_{i,\alpha}^{(1)})\cap[\tau_{j-1,\alpha}^{(2)},\tau_{j,\alpha}^{(2)})\neq\emptyset\right\}}. \tag{2.2}$$

We now give the assumptions needed to prove the central limit theorem of (2.2). We need to introduce some definitions for this purpose. In view of the different models introduced in Section 2.3, there are three different possible assumptions regarding the correlation between the time processes $X_t^{(t)}$ and the price processes $X_t$. The first possibility is that they can be equal for all $0 \leq t \leq T$. In this case we define $\lambda_t^{\min}$ the smallest eigen-value of $(\sigma_t^{(i,j)})_{i=1,2}^{j=1,2}$. The second scenario is that for one $k \in 1, 2$ we have $X_t^{(k)} := X_t^{(t,k)}$, but the other time process is different from its associated price process. In that case, we define $\lambda_t^{\min}$ the smallest eigen-value of $(\sigma_t^{(i,j)})_{i\in\{1,2,3,4\}-\{k+2\}}^{j\in\{1,2,3,4\}-\{k+2\}}$. The third possible setting is that the time process is different from its associated asset price for both assets, and we let $\lambda_t^{\min}$ the smallest eigen-value of $\sigma_t$ in that case. Assumption $(A1)$ provides conditions on the price processes $X_t^{(1)}$ and $X_t^{(2)}$, the time processes $X_t^{(t,1)}$ and $X_t^{(t,2)}$ as well as their covariance matrix $\sigma_t$. There are two types of assumptions in $(A1)$. First, we want to get rid of the drift in the proofs, and this will be done using condition $(A1)$ together with the Girsanov theorem and local arguments (see e.g. pp.158-161 in Mykland and Zhang (2012)). This is a very standard assumption in the literature of high-frequency statistics. Furthermore, we assume that the covariance matrix $\sigma_t$ is continuous.

**Assumption (A1).** The drift $\mu_t$, the volatility matrix $\sigma_t$ and the (four dimensional) Brownian motion $W_t$ are adapted to a filtration $(\mathcal{F}_t)$. Also, $\mu_t$ is integrable and locally bounded. Furthermore, $\sigma_t$ is continuous. Finally, we assume that $\inf_{t \in (0,1]} \lambda_t^{\min} > 0$ a.s.

The following condition roughly assumes that both time processes can't be equal to each other, even on a very small time interval. Specifically, we will assume that there is a constant stricly smaller than 1 such that the module of the *instantaneous high-frequency correlation* $\rho_t^{3,4}$ can't be bigger than this constant. In practice, assumption $(A2)$ is harmless.

**Assumption (A2).** For all $t \in [0, 1]$

$$\rho_t^{3,4} \in [\rho_-^{3,4}, \rho_+^{3,4}], \tag{2.3}$$

where $\max(\mid \rho_-^{3,4} \mid, \mid \rho_+^{3,4} \mid) < 1$

The next assumption deals with the down process $d_t$ and the up process $u_t$. It is clear that $d_t$ and $u_t$ have to be known with information at time $t$, which is why we assume that they are adapted to $(\mathcal{F}_t)$. The rest of assumption $(A3)$ is very technical and we only try to be as general as we can with respect to the proof techniques we will use. The reader should understand Assumption $(A3)$ as "assume the worst dependence structure possible between the return $\Delta X_{\tau_i}$ and the time increment $\Delta\tau_i$, knowing that they follow the HBT model". We insist once again on the fact that we only make the dependence structure as bad as we can in our model so that we can investigate how biased the HY estimator can be in practice, and how much the estimates of the variance assuming no endogeneity are wrong.

**Assumption (A3).** For both assets $k = 1, 2$, define the couple of the down process and the up process $g_t^{(k)} := (d_t^{(k)}, u_t^{(k)})$ and let $g_t := (g_t^{(1)}, g_t^{(2)})$. We assume that

$$\begin{array}{rccl} g^{(k)}: & \mathbb{R}^+ & \rightarrow & (\mathbb{R}^+ \rightarrow \mathbb{R}^- \times \mathbb{R}^+) \\ & t & \mapsto & g_t^{(k)} \end{array}$$

is adapted to $(\mathcal{F}_t)$. Moreover, there exists two non-random constants $0 < g^- < g^+$ such that a.s. for any $t \in [0,1]$ and for any $s \geq 0$

$$g^- \leq \min(-d_t^{(k)}(s), u_t^{(k)}(s)) \leq \max(-d_t^{(k)}(s), u_t^{(k)}(s)) \leq g^+ \tag{2.4}$$

Furthermore, there exists non-random constants $K > 0$ and $d > 1/2$ such that a.s.

$$\forall s \geq K \ , \ g_t\left(s\right) = g_t\left(K\right), \tag{2.5}$$

$$\forall t \geq 0,\ g_t \text{ is differentiable and } \forall s \geq 0,\ \max\left(|(d_t^{(k)})'(s)|, |(u_t^{(k)})'(s)|\right) \leq K, \tag{2.6}$$

$$\forall \ (u,v) \in [0,1]^2 \text{ s.t. } 0 < u < v, \quad \|g_v - g_u\|_\infty \leq K|v-u|^d, \tag{2.7}$$

where $\|(f_1, f_2)\|_\infty = \sup_{w \geq 0} \max\left(|f_1\left(w\right)|, |f_2\left(w\right)|\right)$.

*Remark* 1. Consider the space $\mathcal{C}$ of constants defined in Assumption $(A3)$

$$\mathcal{C} := \left\{(g^-, g^+, K, d) \text{ s.t. } 0 < g^- < g^+ \ , \ K > 0 \ , \ d > \frac{1}{2}\right\}.$$

For any $c \in \mathcal{C}$, we define $\mathcal{G}(c)$ to be the functional subspace of $\mathbb{R}^+ \to (\mathbb{R}^+ \to \mathbb{R}^- \times \mathbb{R}^+)^2$ such that $\forall g \in \mathcal{G}$, $g$ satisfies (2.4), (2.5), (2.6) and (2.7). When there is no room for confusion, we use $\mathcal{G}$. Assumption (A3) is equivalent to

$$\exists c \in \mathcal{C} \text{ s.t. } \forall t \in [0,1] \ , \ g_t \in \mathcal{G}(c).$$

*Remark* 2. The advised reader will have noticed that Example 3, Example 4, Example 5 and Example 6, where time processes are piecewise-constant, don't follow the assumption $(A3)$. Nonetheless, because of the IID structure (respectively Markovian structure) in the size of the jumps in Example 3 (Example 4), the proofs of this work can adapt straightforwardly, and we can still use the results of this chapter. Also, we can add Markovian conditions in Example 5 and Example 6 so that all the techniques of this chapter would apply to them too. We have made the choice not to state more general conditions to keep tractability of Assumption $(A3)$.

The last assumption is only technical, and also appears in the literature (Mykland and Zhang (2012), Li et al. (2014)).

**Assumption (A4).** The filtration $(\mathcal{F}_t)$ is generated by finitely many Brownian motions.

We can now state the main theorem.

**Theorem 1.** *Assume* $(A1)-(A4)$. *Then, there exist processes* $AB_t$ *and* $AV_t$ *adapted to* $(\mathcal{F}_t)$ *such that stably in law as the tick size* $\alpha \to 0$,

$$\alpha^{-1}\left(\widehat{\langle X^{(1)}, X^{(2)}\rangle}^{HY}_{t,\alpha} - \langle X^{(1)}, X^{(2)}\rangle_t\right) \to AB_t + \int_0^t (AV_s)^{1/2}\, dZ_s, \tag{2.8}$$

*where* $Z_t$ *is a Brownian motion independent of the underlying* $\sigma$*-field. The asymptotic bias* $AB_t$ *and the asymptotic variance* $AV_t$ *are defined in Section* 2.4.3 *and estimated in Section* 2.5.

*Remark* 3. (path-bias) Note that the asymptotic bias term $AB_t$ on the right-hand side of (2.8) doesn't mean that the Hayashi-Yoshida estimator is *biased*, but rather *path-biased*. The latter is a weaker statement which means that once we have seen a path, there is a *bias* for the HY estimator on this specific path of value $AB_t$. In practice, we only get to see one path and thus *bias* and *path-bias* can be confused easily. When doing simulations, we can observe many paths and the reader should keep in mind that the *path-bias* will be different for each path. In addition, note that if we assume that $\sigma_t$ is bounded and bounded away from 0 on $[0, T]$, there is no *bias* in Theorem 1 because $\mathbb{E}[AB_t] = 0$.

*Remark* 4. (convergence rate) At first glance, the convergence rate $\alpha^{-1}$ looks different from the optimal rate of convergence $n^{1/2}$ we obtain in the no-endogeneity case. This is merely a change of perspective because we are looking from the tick size point-of-view. Actually, if for $k = 1, 2$ we define $N^{(k)}_{t,\alpha}$ as the number of observations before $t$ of the $k$th asset and the sum of observations of both processes $N^{(S)}_{t,\alpha} := N^{(1)}_{t,\alpha} + N^{(2)}_{t,\alpha}$, we have that $N^{(S)}_{t,\alpha}$ is exactly of order $O_p(\alpha^{-2})$. Thus, if we define the expected number of observations $n := \mathbb{E}\big[N^{(S)}_{t,\alpha}\big]$, we obtain the optimal rate of convergence $n^{\frac{1}{2}}$ in (2.8).

*Remark* 5. (arbitrary number of assets) We chose for simplicity to work only with two assets, but we conjecture that this result would stay true for an arbitrary number of assets, and that our proofs would adapt to show it, at the cost of more involved notations and definitions.

### 2.4.2 Definition of the bias-corrected HY estimator

Assume that we have a consistent estimator[3] $\widehat{AB}_{t,\alpha}$ of the bias $AB_{t,\alpha} := \alpha\big(\int_0^t AB^{(1)}_s dX^{(1)}_s + \int_0^t AB^{(2)}_s dX^{(2)}_s\big)$. Such an estimator will be provided in Section 2.5. We define the new estimator $\widehat{\langle X^{(1)}, X^{(2)}\rangle}^{BC}_{t,\alpha}$ of high-frequency covariance as the estimate obtained when removing the bias estimate $\widehat{AB}_{t,\alpha}$ from the Hayashi-Yoshida estimator

$$\widehat{\langle X^{(1)}, X^{(2)}\rangle}^{BC}_{t,\alpha} := \widehat{\langle X^{(1)}, X^{(2)}\rangle}^{HY}_{t,\alpha} - \widehat{AB}_{t,\alpha}. \tag{2.9}$$

With the bias-corrected estimator $\widehat{\langle X^{(1)}, X^{(2)}\rangle}^{BC}_{t,\alpha}$, we get rid of the asymptotic bias and keep the same asymptotic variance as we can see in the following corollary.

**Corollary 2.** *Assume* $(A1) - (A4)$. *Then, stably in law as* $\alpha \to 0$,

$$\alpha^{-1}\left(\widehat{\langle X^{(1)}, X^{(2)}\rangle}^{BC}_{t,\alpha} - \langle X^{(1)}, X^{(2)}\rangle_t\right) \to \int_0^t (AV_s)^{1/2}\, dZ_s. \tag{2.10}$$

[3] $\widehat{AB}_{t,\alpha}$ is consistent means that $\alpha^{-1}\widehat{AB}_{t,\alpha} = \alpha^{-1}AB_{t,\alpha} + o_p(1)$

### *2.4.3 Computation of the theoretical asymptotic bias and asymptotic variance*

We warn the reader interested in implementing the bias-corrected estimator that this section is highly technical and we advise her to go directly to Section 2.5 and refer to this section only for the definitions. On the contrary, if the reader wants to understand the main ideas of the proofs, she should take this section as a reference. We also want to emphasize on the fact that the theoretical values of asymptotic bias and asymptotic variance found at the end of this section are rather abstract and don't shed easily light on how the change of parameters $\sigma_t$ and $g_t$ in the model would influence the asymptotic bias and asymptotic variance. The main purpose of this work is that we don't need to know the theoretical values in order to compute the estimators in Section 2.5

We need to introduce some definitions in order to compute the theoretical asymptotic bias $AB_t$ and the asymptotic variance term $AV_t$. We first need to rewrite the HY estimator (2.2) in a different way. For any positive integer $i$, consider the $i$th sampling time of the first asset $\tau^{(1)}_{i-1,\alpha}$. We define two times, $\tau^{-}_{i-1,\alpha}$ and $\tau^{+}_{i-1,\alpha}$, which are functions of $\tau^{(1)}_{i-1,\alpha}$ and all the observation times of the second asset $\{\tau^{(2)}_{j,\alpha}\}_{j\geq 0}$, and which correspond respectively to the closest sampling time of the second asset that is strictly smaller than $\tau^{(1)}_{i-1,\alpha}$[4], and the closest sampling time of the second asset that is (not necessarily strictly) bigger than $\tau^{(1)}_{i-1,\alpha}$

$$\tau^{-}_{i-1,\alpha} = \max\{\tau^{(2)}_{j,\alpha} : \tau^{(2)}_{j,\alpha} < \tau^{(1)}_{i-1,\alpha}\}, \tag{2.11}$$

$$\tau^{+}_{i-1,\alpha} = \min\{\tau^{(2)}_{j,\alpha} : \tau^{(2)}_{j,\alpha} \geq \tau^{(1)}_{i-1,\alpha}\}. \tag{2.12}$$

[4] Connoisseurs will have noticed that $\tau^{-}_{i-1,\alpha}$ is not a $\mathcal{F}_t$-stopping time, which will not be a problem in the proofs

We consider $\Delta X^{(2)}_{\tau^{-,+}_{i,\alpha}}$ the increment of the second asset between $\tau^{-}_{i-1,\alpha}$ and $\tau^{+}_{i,\alpha}$

$$\Delta X^{(2)}_{\tau^{-,+}_{i,\alpha}} := \Delta X^{(2)}_{[\tau^{-}_{i-1,\alpha},\tau^{+}_{i,\alpha}]}. \tag{2.13}$$

Rearranging the terms in (2.2) gives us (except for a few terms at the edge)

$$\langle \widehat{X^{(1)}, X^{(2)}} \rangle_{t,\alpha} = \sum_{\tau^{+}_{i,\alpha}<t} \Delta X^{(1)}_{\tau^{(1)}_{i,\alpha}} \Delta X^{(2)}_{\tau^{-,+}_{i,\alpha}}. \tag{2.14}$$

The representation in (2.14) is very useful in the sense that it gives a natural order between the terms in the sum. Nevertheless, any term of this sum is a priori correlated with the other terms. We will rearrange once again the terms in (2.14), so that each term is only correlated with the previous and the next term of the sum. In this case, we say that they are *1-correlated*. For this purpose, we need to introduce some notation. We remind the reader that $\mathrm{T}_\alpha$ is the two-dimensional vector of sampling times, where for each $k = 1, 2$ the $k$th component $\mathrm{T}^{(k)}_\alpha$ is equal to the sequence of sampling times associated with the $k$th asset. We will construct a subsequence $\mathrm{T}^{1C}_\alpha$ of $\mathrm{T}^{(1)}_\alpha$ that also depends on the observation times of the second asset $\mathrm{T}^{(2)}_\alpha$, and will be such that we can write the Hayashi-Yoshida estimator as a 1-correlated sum similar to (2.14), except the new sampling times $\tau^{1C}_{i,\alpha}$ will replace the original observation times $\tau^{(1)}_{i,\alpha}$. The new sampling times $\tau^{1C}_{i,\alpha}$ are obtained using the following algorithm. We define $\tau^{1C}_{0,\alpha} := \tau^{(1)}_{0,\alpha}$, and recursively for $i$ any nonnegative integer

$$\tau^{1C}_{i+1,\alpha} := \min\big\{\tau^{(1)}_{u,\alpha} : \text{ there exists a nonnegative integer } j \text{ such that}$$

$$\tau^{1C}_{i,\alpha} \leq \tau^{(2)}_{j,\alpha} < \tau^{(1)}_{u,\alpha}\big\}. \tag{2.15}$$

In words, if we sit at the observation time $\tau^{1C}_{i,\alpha}$ of the first asset, we wait first to hit an observation time of the second asset, and we then choose the next strictly bigger observation time of the first asset. In analogy with (2.11), (2.12) and (2.13), we define the following times

$$\tau^{1C,-}_{i-1,\alpha} := \max\{\tau^{(2)}_{j,\alpha} : \tau^{(2)}_{j,\alpha} < \tau^{1C}_{i-1,\alpha}\}, \tag{2.16}$$

$$\tau^{1C,+}_{i-1,\alpha} := \min\{\tau^{(2)}_{j,\alpha} : \tau^{(2)}_{j,\alpha} \geq \tau^{1C}_{i-1,\alpha}\}, \tag{2.17}$$

$$\Delta X^{(2)}_{\tau^{1C,-,+}_{i,\alpha}} := \Delta X^{(2)}_{[\tau^{1C,-}_{i-1,\alpha},\tau^{1C,+}_{i,\alpha}]}. \tag{2.18}$$

First, observe that, except for maybe a few terms at the edge, we can rewrite (2.14) as

$$\langle \widehat{X^{(1)}, X^{(2)}} \rangle_{t,\alpha} = \sum_{\tau^{1C,+}_{i,\alpha} < t} \Delta X^{(1)}_{\tau^{1C}_{i,\alpha}} \Delta X^{(2)}_{\tau^{1C,-,+}_{i,\alpha}}. \tag{2.19}$$

Also, we define the following compensated increments of the HY estimator

$$N_{i,\alpha} = \Delta X^{(1)}_{\tau^{1C}_{i,\alpha}} \Delta X^{(2)}_{\tau^{1C,-,+}_{i,\alpha}} - \int_{\tau^{1C}_{i-1,\alpha}}^{\tau^{1C}_{i,\alpha}} \zeta^{1,2}_s ds. \tag{2.20}$$

Note that they are compensated in the sense that they are centered (if we decompose $\Delta X^{(2)}_{\tau^{1C}_{i,\alpha}}$ into a left $(-)$, a central and a right $(+)$ part and condition the expectation, this is straightforward to show). Similarly, we can show that they are 1-correlated.

The idea of the proof is the following. If we consider the volatility matrix $\sigma_t$ and the grid function $g_t$ to be constant over time, we can express the conditional returns of the normalized error of HY as a homogeneous Markov chain (of order 1), show that the Markov chain is uniformly ergodic and thus use results in the limit theory of Markov chains (see, e.g., Meyn and Tweedie (2009)) to show that it has a stationary distribution. Then, we prove that we can approximate locally the returns of the normalized error when the volatility matrix and grid function are not constant by the returns when holding them constant on a small block. Finally, using limit theory techniques developed in Mykland and Zhang (2012) together with standard probability results of conditional distribution (see, e.g., Breiman (1992)), we can bound uniformly in time the error of the returns when holding the volatility matrix and grid function constant.

We define now all the previous quantities assuming the volatility matrix $\sigma_t$ and the grid function $g_t$ are constant. For that purpose, let $\tilde{W}_t$ be a four dimensional Wiener process, $c := (g^-, g^+, K, d)$ a four-dimensional vector such that $c \in \mathcal{C}$ and $\tilde{\sigma}$ a constant volatility matrix such that the associated $\tilde{\lambda}^{\min}$, which is the analog of $\lambda_t^{\min}$ defined in Section 2.4.1 when we replace $\sigma_t$ by $\tilde{\sigma}$, is strictly bigger than 0 and $\tilde{g} \in \mathcal{G}(c)$ a constant grid function. In analogy with the definition of the grid function $g_t$ in $(A3)$, we assume that $\tilde{g}$ can be written in terms of the down and up functions of both assets, i.e. $\tilde{g} := (\tilde{g}^{(1)}, \tilde{g}^{(2)})$ where for each $k = 1, 2$ we have $\tilde{g}^{(k)} := (\tilde{d}^{(k)}, \tilde{u}^{(k)})$. Also, we introduce $\mathcal{S}_{\tilde{g}}$ the subspace of $\mathbb{R}^2$ defined as

$$\mathcal{S}_{\tilde{g}} := \{(y, v) \in \mathbb{R} \times \mathbb{R}^+ \text{ s.t. } \tilde{d}^{(2)}(v) \leq y \leq \tilde{u}^{(2)}(v)\}.$$

If we set $\tilde{X} = \tilde{\sigma}\tilde{W}$ and the corresponding sampling times of both assets $\tilde{\mathrm{T}} := (\tilde{\mathrm{T}}^{(1)}, \tilde{\mathrm{T}}^{(2)})$, where for $k = 1, 2$ we have $\tilde{T}^{(k)} := \{\tilde{\tau}_i\}_{i \geq 0}$, we define the observation times of the first asset as $\tilde{\tau}_0^{(1)} := 0$ and recursively for $i$ any positive integer

$$\tilde{\tau}_i^{(1)} := \inf\left\{t > \tilde{\tau}_{i-1}^{(1)} : \Delta\tilde{X}_t^{(3)} \notin [\tilde{d}^{(1)}(t - \tilde{\tau}_{i-1}^{(1)}), \tilde{u}^{(1)}(t - \tilde{\tau}_{i-1}^{(1)})]\right\}.$$

These stopping times will be seen as approximations of the observation times of the first asset when we hold the volatility matrix $\sigma_t$ and the grid $g_t$ constant. We will always start our approximation at a *1-correlated* observation time $\tau_{i,n}^{1C}$, which corresponds to an observation time of the first asset. As the sampling times of the second asset are not synchronized with the ones from the first asset, we need two more quantities $(x, u) \in \mathcal{S}_{\tilde{g}}$ to approximate the observation times of the second asset. They correspond respectively to the increment of the second asset's time process $X_t^{(t,2)}$ since the last observation of the second asset occured and the time elapsed since the last observation time of the second asset. We define $\tilde{\tau}_0^{(2)} := 0$,

$$\tilde{\tau}_1^{(2)} := \inf\big\{t > 0 : x + \Delta\tilde{X}_t^{(4)} \notin [\tilde{d}_2(t+u), \tilde{u}_2(t+u)]\big\},$$

and for any integer $i \geq 2$

$$\tilde{\tau}_i^{(2)} := \inf\big\{t > \tilde{\tau}_{i-1}^{(2)} : \Delta\tilde{X}_t^{(4)} \notin [\tilde{d}_2(t-\tilde{\tau}_{i-1}^{(2)}), \tilde{u}_2(t-\tilde{\tau}_{i-1}^{(2)})]\big\}.$$

Similarly, we define the analogs of (2.11), (2.12), (2.13), (2.15), (2.16), (2.17), (2.18) and (2.20) respectively as $\tilde{\tau}_{i-1}^{-}$, $\tilde{\tau}_{i-1}^{+}$, $\Delta\tilde{X}_{\tilde{\tau}_i^{-},+}^{(2)}$, $\tilde{\tau}_{i-1}^{1C,-}$, $\tilde{\tau}_{i-1}^{1C,+}$, $\Delta\tilde{X}_{\tilde{\tau}_i^{1C,-},+}^{(2)}$ and $\tilde{N}_i$ by putting tildes on the quantities of the definitions.

We can now define the *instantaneous variance* of the normalized HY estimate's error (2.21), which depends on the volatility matrix $\tilde{\sigma}$ and the grid $\tilde{g}$. Similarly, we also define the *instantaneous covariance* between the normalized HY's error and the first asset price (2.22), and the *instantaneous covariance* between the error and the second asset price (2.23). Finally, we define the *instantaneous 1-correlated time*, which is the approximation of $\mathbb{E}_{\tau_{i,n}^{1C}}\left[\Delta\tau_{i+2}^{1C}\right]$, where if $\tau$ is a $(\mathcal{F}_t)$-stopping time, $\mathbb{E}_\tau\left[Y\right]$ is defined as the conditional distribution of $Y$ given $\mathcal{F}_\tau$.

$$\begin{aligned}
\psi^{AV}(\tilde{\sigma}, \tilde{g}, x, u) &:= \mathbb{E}\big[\tilde{N}_2^2 + 2\tilde{N}_2\tilde{N}_3\big], && (2.21)\\
\psi^{AC1}(\tilde{\sigma}, \tilde{g}, x, u) &:= \mathbb{E}\big[\tilde{N}_2\Delta\tilde{X}_{\tilde{\tau}_2^{1C}}^{(1)}\big], && (2.22)\\
\psi^{AC2}(\tilde{\sigma}, \tilde{g}, x, u) &:= \mathbb{E}\big[\tilde{N}_2\Delta\tilde{X}_{\tilde{\tau}_2^{1C,-},+}^{(2)}\big], && (2.23)\\
\psi^{\tau}(\tilde{\sigma}, \tilde{g}, x, u) &:= \mathbb{E}\big[\Delta\tilde{\tau}_2^{1C}\big]. && (2.24)
\end{aligned}$$

Set $\tilde{Z}_0 := (x, u)$ and for any positive integer $i$

$$\tilde{Z}_i := \big(\Delta\tilde{X}_{[\tilde{\tau}_i^{1C,-}, \tilde{\tau}_i^{1C}]}^{(4)}, \tilde{\tau}_i^{1C} - \tilde{\tau}_i^{1C,-}\big). \tag{2.25}$$

For any nonnegative integer $i$, we consider $\tilde{\pi}_i(\tilde{\sigma}, \tilde{g}, x, u)$ the distribution of $\tilde{Z}_i$. We also introduce the notation $\Pi(\tilde{\sigma},\tilde{g},x,u) := \{\tilde{\pi}_i(\tilde{\sigma},\tilde{g},x,u)\}_{i\geq 0}$. By the strong Markov property of Brownian motion, we can show that $\tilde{Z}_i$ is a homogeneous Markov chain (of order 1) on the state space $\mathcal{S}_{\tilde{g}}$. In the following lemma, we show that there exists a stationary distribution of $\tilde{\pi}_i(\tilde{\sigma},\tilde{g},x,u)$.

**Lemma 3.** *Let $c := (g^-, g^+, K, d)$ be a four-dimensional vector such that $c \in \mathcal{C}$ and consider $\tilde{\sigma}$ a constant volatility matrix such that $\tilde{\lambda}^{\min} > 0$ and $\tilde{g} \in \mathcal{G}(c)$ a constant grid. Then, there exists a stationary distribution $\tilde{\pi}(\tilde{\sigma},\tilde{g})$.*

The proof of Lemma 3 can be found in the Appendix (proof of Lemma 14). The next definition is the average (regarding the stationary distributions) of the instantaneous variance, covariances and 1-correlated time. For any $\theta \in \{\text{AV, AC1, AC2}, \tau\}$,

$$\phi^{\theta}\left(\tilde{\sigma},\tilde{g}\right) := \int_{\mathbb{R}^2} \psi^{\theta}\left(\tilde{\sigma},\tilde{g},y,v\right) d\tilde{\pi}\left(\tilde{\sigma},\tilde{g}\right)(y,v)\,.$$

We introduce the notation $\phi_s^{\theta} := \phi^{\theta}\left(\sigma_s, g_s\right)$ and consider the following quantities needed to compute the asymptotic bias and variance.

$$k_s^{(1)} \;:=\; \big(\sigma_s^{(1)}\big)^{-2}\phi_s^{AC1}\big(\phi_s^{\tau}\big)^{-1}, \tag{2.26}$$

$$k_s^{1,\perp} \;:=\; \big(1-(\rho_s^{1,2})^2\big)^{-1}\big((\sigma_s^{(2)})^{-2}\phi_s^{AC2} - (\sigma_s^{(1)}\sigma_s^{(2)})^{-1}\rho_s^{1,2}\phi_s^{AC1}\big)\big(\phi_s^{\tau}\big)^{-1}. \tag{2.27}$$

We express now $AV_s$ the quantity integrated to obtain the asymptotic variance.

$$AV_s \;:=\; \big(\phi_s^{AV} + 2\big(k_s^{(1)}(\sigma_s^{(1)})^{-1}\sigma_s^{(2)}\rho_s^{1,2}\phi_s^{AC1} - (k_s^{(1)}+k_s^{1,\perp})\phi_s^{AC2}\big)\big)\big(\phi_s^{\tau}\big)^{-1} \tag{2.28}$$

$$+\big(\sigma_s^{(1)}\big)^2\big(k_s^{(1)}\big)^2 + \big(\sigma_s^{(2)}\big)^2\Big(1-(\rho_s^{1,2})^2\big)\big(k_s^{1,\perp}\big)^2.$$

The asymptotic bias is defined as $AB_t := \int_0^t AB_s^{(1)} dX_s^{(1)} + \int_0^t AB_s^{(2)} dX_s^{(2)}$ where

$$AB_s^{(1)} := k_s^{(1)} - k_s^{1,\perp} \rho_s^{1,2} \sigma_s^{(2)} \big(\sigma_s^{(1)}\big)^{-1}, \tag{2.29}$$

$$AB_s^{(2)} := k_s^{1,\perp}. \tag{2.30}$$

*Remark* 6. (asymptotic bias) Looking at the expressions for $AB_s^{(1)}$ and $AB_s^{(2)}$, one can be tempted to think that because of the $\big(1-(\rho_s^{1,2})^2\big)^{-1}$ term in $k_s^{1,\perp}$, the bias will increase drastically when both assets are highly correlated. In this case, the reader should keep in mind that the second term of $AB_s^{(1)}$, when integrated with respect to $X_s^{(1)}$, and $AB_s^{(2)}$, when integrated with respect to $X_s^{(2)}$, will be roughly of the same magnitude, with opposite signs, and thus there is no explosion of asymptotic bias. We chose the above asymptotic bias' representation because it is straightforward to build estimators from it. We can also express the asymptotic bias differently. For this purpose, we can rewrite the log-price process as

$$\begin{aligned} dX_t^{(1)} &= \sigma_t^{(1)} dB_t^{(1)}, \\ dX_t^{(2)} &= \rho_t^{1,2} \sigma_t^{(2)} dB_t^{(1)} + \big(1-(\rho_t^{1,2})^2\big)^{1/2} \sigma_t^{(2)} dB_t^{1,\perp}, \end{aligned}$$

where $B_t^{(1)}$ and $B_t^{1,\perp}$ are independent Brownian motions. Let

$$dX_t^{1,\perp} = \big(1-(\rho_t^{1,2})^2\big)^{1/2} \sigma_t^{(2)} dB_t^{1,\perp} \tag{2.31}$$

be the part of $X_t^{(2)}$ that is not correlated with $X_t^{(1)}$. We can express the asymptotic bias as $AB_t = \int_0^t \tilde{AB}_s^{(1)} dX_s^{(1)} + \int_0^t \tilde{AB}_s^{(2)} dB_s^{1,\perp}$. In this case, $\tilde{AB}_s^{(1)} = k_s^{(1)}$ and

$$\tilde{AB}_s^{(2)} = \lim_{n\to\infty} \langle M^n, B^{1,\perp}\rangle_s$$

where $M^n$ is defined in the proofs. We can show that this limit exists, and does not explode when both assets are highly correlated.

## 2.5 Estimation of the bias and variance

We need to introduce some new notations. We recall that $N_{1,\alpha}^{(1)}$ is the number of observations corresponding to the first asset before 1 and we also define $N_{1,\alpha}^{1C}$ the number of 1-correlated observations before 1, i.e. $N_{1,\alpha}^{1C} := \max\{i \in \mathbb{N} \text{ s.t. } \tau_{i,\alpha}^{1C} < 1\}$. In practice, the first step is to transform the returns of the first asset

$$\big\{(\Delta X_{\tau_{i,\alpha}^{(1)}}^{(1)}, \Delta\tau_{i,\alpha}^{(1)})\big\}_{i=1}^{N_{1,\alpha}^{(1)}}$$

into 1-correlated returns

$$\big\{(\Delta X_{\tau_{i,\alpha}^{1C}}^{1C}, \Delta\tau_{i,\alpha}^{1C})\big\}_{i=1}^{N_{1,\alpha}^{1C}}$$

using algorithm (2.15). Then, for each asset, we will chop the data into $B_n$ blocks and on each block $i = 1, \dots, B_n$ we will estimate $\widehat{AV}_{i,\alpha}$, $\widehat{AB}_{i,\alpha}^{(1)}$ and $\widehat{AB}_{i,\alpha}^{(2)}$, pretending that the volatility matrix $\sigma_t$ and grid $g_t$ are block-constant.

Because there is asynchronicity in the observation times, the blocks of each asset are not exactly equal. Let $h_n$ be the block size. For the first asset, we consider *block* $1^{(1)} := [0, \tau_{h_n,\alpha}^{1C}]$, *block* $2^{(1)} := [\tau_{h_n,\alpha}^{1C}, \tau_{2h_n,\alpha}^{1C}]$, $\dots$ , *block* $B_{n-1}^{(1)} := [\tau_{(B_n-2)h_n,\alpha}^{1C}, \tau_{(B_n-1)h_n,\alpha}^{1C}]$, *block* $B_n^{(1)} := [\tau_{(B_n-1)h_n,\alpha}^{1C}, 1]$. For the second asset, we let *block* $1^{(2)} := [\tau_{0,\alpha}^{1C,+}, \tau_{h_n,\alpha}^{1C,+}]$, *block* $2^{(2)} := [\tau_{h_n,\alpha}^{1C,+}, \tau_{2h_n,\alpha}^{1C,+}]$, $\dots$ , *block* $B_{n-1}^{(2)} := [\tau_{(B_n-2)h_n,\alpha}^{1C,+}, \tau_{(B_n-1)h_n,\alpha}^{1C,+}]$, *block* $B_n^{(2)} := [\tau_{(B_n-1)h_n,n}^{1C,+}, 1]$. In the following, we will say $j \in$ block $i^{(1)}$ when $\tau_{j,\alpha}^{(1)} \in$ block $i^{(1)}$. Similarly, we say $j \in$ block $i^{(2)}$ when $\tau_{j,n}^{(2)} \in$ block $i^{(2)}$. Finally, we define $j \in$ block $i$ if $j \in \{(i-1)h_n + 1, \dots, ih_n\}$. First, we estimate the volatility of both assets using the corrected estimator in Li et al. (2014). To do this, we need to define an estimate of the spot volatility on each block for each asset $k = 1, 2$

$$\tilde{\sigma}_{i,\alpha}^{(k)} := \Big( \sum_{j \in \text{block } i^{(k)}} (\Delta X_{\tau_{j,\alpha}^{(k)}}^{(k)})^2 \Big)^{1/2}.$$

Then, we estimate the asymptotic bias of the volatility

$$\widehat{AB\sigma}_{i,\alpha}^{(k)} = \frac{2}{3(\tilde{\sigma}_{i,\alpha}^{(k)})^2} \sum_{j \in \text{block } i^{(1)}} (\Delta X_{\tau_{j,\alpha}^{(k)}}^{(k)})^3.$$

We obtain the bias-corrected estimators of volatility on each block

$$\hat{\sigma}_{i,\alpha}^{(k)} = \tilde{\sigma}_{i,\alpha}^{(k)} - \widehat{AB\sigma}_{i,\alpha}^{(k)}.$$

Then, we estimate the correlation between both assets using the *naive* HY estimator

$$\hat{\rho}_{i,\alpha}^{1,2} = \frac{1}{\hat{\sigma}_{i,\alpha}^{(1)}\hat{\sigma}_{i,\alpha}^{(2)}} \sum_{j \in \text{block } i} \Delta X_{\tau_{j,\alpha}^{1C}}^{(1)} \Delta X_{\tau_{j,\alpha}^{1C,-,+}}^{(2)}.$$

We then build an estimator of the compensated increments of the HY estimator, following the definition in (2.20),

$$\widehat{N}_{i,\alpha} = \Delta X_{\tau_{i,\alpha}^{1C}}^{(1)} \Delta X_{\tau_{i,\alpha}^{1C,-,+}}^{(2)} - \Delta\tau_{i,\alpha}^{1C} \hat{\sigma}_{i,\alpha}^{(1)} \hat{\sigma}_{i,\alpha}^{(2)} \hat{\rho}_{i,\alpha}^{1,2}.$$

The next step is to estimate the instantaneous variance (2.21), both instantaneous covariances (2.22) and (2.23) and the instantaneous 1-correlated time (2.24) on each block. This is done by taking the sample average of the corresponding estimated quantities. Note that we don't directly estimate $\psi^{AV}$, $\psi^{AC1}$, $\psi^{AC2}$ and $\psi^{\tau}$, but rather a scaling version of them, i.e. $\alpha_n^2\psi^{AV}$, $\alpha_n\psi^{AC1}$, $\alpha_n\psi^{AC2}$ and $\alpha_n\psi^{\tau}$. In practice, we can always assume $\alpha_n := 1$ by scaling $g_t$ by the tick size, and thus we match the definitions of the following estimators with (2.21)-(2.24). For the sake of simplicity, we assume that the number of 1-correlated observations of the last block $B_n$ is also $h_n$. In practice, this will be most likely different from $h_n$, and thus the denominator of (2.32)-(2.35) will have to be changed so that it is equal to the number of 1-correlated observations in this last block. The estimates are given by

$$
\begin{aligned}
\hat{\phi}^{AV}_{i,\alpha} &:= h_n^{-1} \sum_{j \in \text{block } i} \hat{N}^2_{j,\alpha} + 2\hat{N}_{j,\alpha}\hat{N}_{j+1,\alpha}, &(2.32)\\
\hat{\phi}^{AC1}_{i,\alpha} &:= h_n^{-1} \sum_{j \in \text{block } i} \hat{N}_{j,\alpha}\Delta X^{(1)}_{\tau^{1C}_{j,\alpha}}, &(2.33)\\
\hat{\phi}^{AC2}_{i,\alpha} &:= h_n^{-1} \sum_{j \in \text{block } i} \hat{N}_{j,\alpha}\Delta X^{(2)}_{\tau^{1C,-,+}_{j,\alpha}}, &(2.34)\\
\hat{\phi}^{\tau}_{i,\alpha} &:= h_n^{-1} \sum_{j \in \text{block } i} \Delta\tau^{1C}_{j,\alpha}. &(2.35)
\end{aligned}
$$

We estimate now the quantities (2.26) and (2.27) as

$$
\begin{aligned}
\hat{k}^{(1)}_{i,\alpha} &:= \big(\hat{\sigma}^{(1)}_{i,\alpha}\big)^{-2}\hat{\phi}^{AC1}_{i,\alpha}\big(\hat{\phi}^{\tau}_{i,\alpha}\big)^{-1}, &(2.36)\\
\hat{k}^{1,\perp}_{i,\alpha} &:= \big(1-(\hat{\rho}^{1,2}_{i,\alpha})^2\big)^{-1}\big((\hat{\sigma}^{(2)}_{i,\alpha})^{-2}\hat{\phi}^{AC2}_{i,\alpha} - (\hat{\sigma}^{(1)}_{i,\alpha}\hat{\sigma}^{(2)}_{i,\alpha})^{-1}\hat{\rho}^{1,2}_{i,\alpha}\hat{\phi}^{AC1}_{i,\alpha}\big)\big(\hat{\phi}^{\tau}_{i,\alpha}\big)^{-1}. &(2.37)
\end{aligned}
$$

We follow (2.29) and (2.30) to estimate the bias integrated terms $AB^{(1)}_s$ and $AB^{(2)}_s$ on each block

$$
\begin{aligned}
\widehat{AB}^{(1)}_{i,\alpha} &:= \hat{k}^{(1)}_{i,\alpha} - \hat{k}^{1,\perp}_{i,\alpha}\hat{\rho}^{1,2}_{i,\alpha}\hat{\sigma}^{(2)}_{i,\alpha}\big(\hat{\sigma}^{(1)}_{i,\alpha}\big)^{-1},\\
\widehat{AB}^{(2)}_{i,\alpha} &:= \hat{k}^{1,\perp}_{i,\alpha}.
\end{aligned}
$$

For the variance term $AV_s$, we decide not to use the direct definition in (3.55) because it can provide negative estimates. Instead, we will be using the following estimator

$$
\begin{aligned}
\widehat{AV}_{i,\alpha} := \Big(&\big(\sum_{j \in \text{block } i} \widehat{N}_{j,\alpha}\big) - \hat{k}^{(1)}_{i,\alpha}\big(X^{(1)}_{\tau^{1C}_{ih_n,\alpha}} - X^{(1)}_{\tau^{1C}_{(i-1)h_n,\alpha}}\big)\\
&-\hat{k}^{\perp}_{i,\alpha}\big((X^{(2)}_{\tau^{1C,+}_{ih_n,\alpha}} - X^{(2)}_{\tau^{1C,+}_{(i-1)h_n,\alpha}}) - \hat{\rho}^{1,2}_{i,\alpha}\hat{\sigma}^{(2)}_{i,\alpha}(\hat{\sigma}^{(1)}_{i,\alpha})^{-1}(X^{(1)}_{\tau^{1C}_{ih_n,\alpha}} - X^{(1)}_{\tau^{1C}_{(i-1)h_n,\alpha}})\big)\Big)^2.
\end{aligned}
$$

We define the final estimators of asymptotic bias and asymptotic variance as

$$\widehat{AB}_\alpha := \sum_{i=1}^{B_n} \widehat{AB}^{(1)}_{i,\alpha}\left(X^{(1)}_{\tau^{1C}_{ih_n,\alpha}} - X^{(1)}_{\tau^{1C}_{(i-1)h_n,\alpha}}\right) + \widehat{AB}^{(2)}_{i,\alpha}\left(X^{(2)}_{\tau^{1C,+}_{ih_n,\alpha}} - X^{(2)}_{\tau^{1C,+}_{(i-1)h_n,\alpha}}\right), \quad (2.38)$$

$$\widehat{AV}_\alpha := \sum_{i=1}^{B_n} \widehat{AV}_{i,\alpha}\left(\tau^{1C}_{ih_n,\alpha} - \tau^{1C}_{(i-1)h_n,\alpha}\right). \quad (2.39)$$

As a corollary of Theorem 1, we obtain the following result, which states the consistency of (2.38) and (2.39).

**Corollary 4.** *There exists a choice of the block size $h_n$*[5] *such that we have the following convergences when $\alpha \to 0$*

$$\alpha^{-1}\widehat{AB}_\alpha \xrightarrow{\mathbb{P}} AB_1 \quad (2.40)$$

$$\alpha^{-2}\widehat{AV}_\alpha \xrightarrow{\mathbb{P}} \int_0^1 AV_s ds. \quad (2.41)$$

*In particular, in view of Corollary 2, the bias-corrected estimator*

$$\widehat{\langle X^{(1)}, X^{(2)}\rangle}^{BC}_{1,\alpha} := \widehat{\langle X^{(1)}, X^{(2)}\rangle}^{HY}_{1,\alpha} - \widehat{AB}_\alpha$$

*is such that*

$$\frac{\widehat{\langle X^{(1)}, X^{(2)}\rangle}^{BC}_{1,\alpha} - \langle X^{(1)}, X^{(2)}\rangle_1}{\widehat{AV}^{1/2}_\alpha} \to \mathcal{N}(0,1). \quad (2.42)$$

*Remark* 7. (exchanging $X^{(1)}_t$ and $X^{(2)}_t$) When estimating the asymptotic bias and the asymptotic variance, we considered one specific asset to be $X^{(1)}_t$ and the other one to be $X^{(2)}_t$. We could exchange $X^{(1)}_t$ and $X^{(2)}_t$, and find new estimators $\tilde{AB}_\alpha$ and $\tilde{AV}_\alpha$ according to the previous definitions. One could then take $\frac{AB_\alpha + \tilde{AB}_{t,\alpha}}{2}$ (respectively $\frac{AV_\alpha + \tilde{AV}_{t,\alpha}}{2}$) as final estimators of asymptotic bias (asymptotic variance).

[5] the exact assumptions on $h_n$ can be found in the proofs of Theorem 1

*Remark* 8. (optimal block size) In practice, the optimal block size $h_n$ is not straightforward to choose. On the one hand, $h_n$ should be as small as possible so that the volatility matrix $\sigma_t$ and the grid $g_t$ are almost constant on each block, and thus (2.32)-(2.35) are less biased. On the other hand, we need as many observations as we can on each block, so that the variance of approximations (2.32)-(2.35) is not too big. We are facing here the usual bias-variance tradeoff.

## 2.6 Numerical simulations

We assume the same setting as the toy model described in Example 1, in two dimensions. Thus, there exists a four-dimensional parameter $\theta := (\theta_u^{(1)}, \theta_d^{(1)}, \theta_u^{(2)}, \theta_d^{(2)})$ such that for any $t \geq 0$ and any $s \geq 0$, $u_t^{(1)}(s) := \theta_u^{(1)}$, $d_t^{(1)}(s) := \theta_d^{(1)}$, $u_t^{(2)}(s) := \theta_u^{(2)}$ and $d_t^{(2)}(s) := \theta_d^{(2)}$. We assume that the two-dimensional price process $(X_t^{(1)}, X_t^{(2)})$ has a null-drift. Also, we assume that the volatility of the first process is $\sigma_t^{(1)} := 0.016$ and the volatility of the second process $\sigma_t^{(2)} := 0.02$, and that the correlation between both assets is $\rho_t^{1,2} := 0.2$. We set $\theta := \big(0.0007, 0.0001, 0.0006, 0.0001\big)$. According to this rule, a change of price occurs whenever the price of the first (respectively second) asset increases by 0.07% (0.06%) or decreases by 0.01% (0.01%). Finally, we assume that the price processes $(X_t^{(1)}, X_t^{(2)})$ and the time processes $(X_t^{(t,1)}, X_t^{(t,2)})$ are equal.

We simulate price processes and observation times for 10 years of 252 business days. We provide in Table 2.1 a summary of the comparison results between HY and the bias-corrected HY. As expected from the theory, the RMSE is improved when using the bias-corrected estimator. In addition, the sample bias is almost the same when using HY and the bias-corrected estimator which is also expected from Remark 3. Furthermore, this sample bias tends to 0, which comes from the fact that both estimators are consistent. Finally, the standardized feasible statistic (2.42) is reported in Table 2.2 and plotted in Figure 2.1.

| No. years | estim | sample bias | RMSE | % Reduced RMSE |
|---|---|---|---|---|
| 1 | HY | $5.41e-07$ | $1.36e-05$ | - |
| 1 | BCHY | $5.43e-07$ | $1.19e-05$ | 13% |
| 2 | HY | $2.73e-07$ | $1.39e-05$ | - |
| 2 | BCHY | $2.74e-07$ | $1.23e-05$ | 12% |
| 5 | HY | $1.10e-07$ | $1.42e-05$ | - |
| 5 | BCHY | $1.07e-07$ | $1.26e-05$ | 11% |
| 10 | HY | $5.54e-08$ | $1.39e-05$ | - |
| 10 | BCHY | $5.53e-08$ | $1.20e-05$ | 14% |

Table 2.1: Summary statistics based on simulated endogenous data of 1, 2, 5 and 10 years. The RMSE in the table corresponds to the square root of the squared distance between the estimated value and the true value $6.4e-05$. HY stands for the usual Hayashi-Yoshida estimator (1.5), and BCHY represents the bias-corrected estimator (2.9).

| No. years | 0.5 % | 2.5 % | 5 % | 95 % | 97.5 % | 99.5 % |
|---|---|---|---|---|---|---|
| 1 | -2.48 | -1.99 | -1.59 | 1.66 | 2.13 | 2.57 |
| 2 | - 2.57 | -1.83 | -1.60 | 1.68 | 2.13 | 2.67 |
| 5 | - 2.60 | -1.96 | -1.64 | 1.64 | 2.05 | 2.62 |
| 10 | - 2.68 | - 1.98 | -1.60 | 1.65 | 2.01 | 2.73 |

Table 2.2: In this table, we report the finite sample quartiles of the feasible standardized statistic (2.42). The benchmark quartiles are those for the limit distribution $\mathcal{N}(0,1)$.

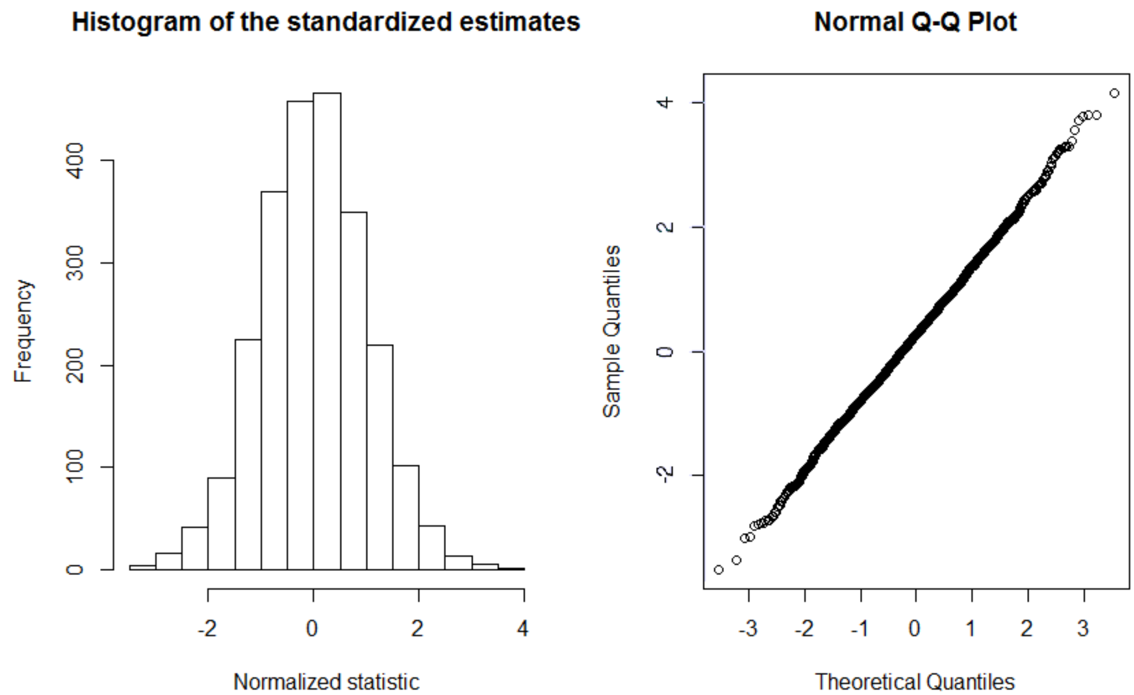


Figure 2.1: Histogram and Normal QQ-plot of the standardized estimates (2.42) on a 10-year period of observations.

## 2.7 Conclusion

We have introduced in this chapter the HBT model, and we have shown that it is more general than some of the endogenous models of the literature. This model can be extended to a model including more general noise structure in observations, and even noise in sampling times. The model with noisy observations is introduced in Chapter 3.

Under this model, we have proved the central limit theorem of the Hayashi-Yoshida estimator. Our main theorem states that there is an asymptotic bias. Accordingly, we built a bias-corrected HY estimator. We also computed the theoretical standard deviation, and we provided consistent estimates of it. Numerical simulations show that the new estimator performs better.

The techniques used for the proof of the main theorem can be applied to more general models and to other problems such as the estimation of the integrated variance of noise, integrated betas, etc. In particular, independence between the efficient price process and the noise is not needed in the model. As long as we can approximate the joint distribution of the noise and the returns by a Markov chain, ideas of our proof can be used. This is also left for Chapter 3.

## 2.8 Appendix

### *2.8.1 Preliminary lemmas*

Without loss of generality, we choose to work under the third scenario defined in Section 2.4, i.e. the asset price is different from the time process for both assets. Because we shall prove stable convergence, and because of the local boundedness of $\sigma$ (because by (A1) $\sigma$ is continuous), and that $\inf_{t\in(0,1]} \lambda_t^{\min} > 0$ we can without loss of generarality assume that for all $t \in [0,1]$ there exists some nonrandom constants $\sigma^-$ and $\sigma^+$ such that for any eigen value $\lambda_t$ of $\sigma_t$

$$0 < \sigma^- < \lambda_t < \sigma^+ \tag{2.43}$$

by using a standard localization argument such that the one used in Section 2.4.5 of Mykland and Zhang (2012). One can further supress $\mu$ as in Section 2.2 (pp. 1407-1409) of Mykland and Zhang (2009), and act as if $X$ is a martingale.

We define the subspace $\mathcal{M}$ of matrices of dimension $4 \times 4$ such that $\forall M \in \mathcal{M}$, for any eigen value $\lambda_M$ of $M$, we have

$$\sigma^- < \lambda_M < \sigma^+ \tag{2.44}$$

and $\frac{(MM^T)^{3,4}}{(MM^T)^{4,4}} \in [\rho_-^{3,4}, \rho_+^{3,4}]$. By (2.3) of $(A2)$ and (2.43), we will assume in the following that $\forall t \in [0,1]$, $\sigma_t \in \mathcal{M}$.

We define $\sigma^p$ the process (of dimension $4 \times 4$) on $\mathbb{R}^+$ such that

$$\begin{cases} \sigma_t^p &= \sigma_t \quad \forall t \in [0,1] \\ \sigma_t^p &= \sigma_1 \quad \forall t \in [1,\infty) \end{cases}$$

Define now $X^p$ the process such that for all $t \geq 0$

$$\begin{cases} dX_t^p &= \sigma_t^p dW_t \\ X_0^p &= X_0 \end{cases}$$

Because $X^p$ and $X$ have the same initial value and follow the same stochastic differential equation on $[0,1]$, they are equalt for all $t \in [0,1]$. For simplicity, we keep from now the notation $X$ for $X^p$.

In the following, $C$ will be defining a constant which does not depend on $i$ or $n$, but that can vary from a line to another. Also, we are going to use the notation $\tau_{i,n}^\theta$ as a subtitute of $\tau_{i,\alpha_n}^\theta$, where $\theta$ can take various names, such that $(1),(2)$ and so on. Let $h : \mathbb{N} \to \mathbb{N}$ a (not strictly) increasing non-random sequence such that

$$h_n \to +\infty \tag{2.45}$$

$$h_n \alpha_n \to 0 \tag{2.46}$$

To keep notation as simple as possible, we define $\tau_{i,n}^h := \tau_{ih_n,n}^{1C}$, $\tau_{i,n}^{h,-} := \tau_{ih_n,n}^{1C,-}$, $\tau_{i,n}^{h,+} := \tau_{ih_n,n}^{1C,+}$. We also let $A_n := \{i \geq 1 \text{ s.t. } \tau_{i-1,n}^h \leq t\}$, where $t \in [0,1]$. Also, we recall the notation $(X_t^{(3)}, X_t^{(4)}) := (X_t^{(t,1)}, X_t^{(t,2)})$ Finally, for $\theta \in \{(1),(2),1C,h\}$, we define $s_n^\theta = \sup_{\tau_{i,n}^\theta < T} \Delta\tau_{i,n}^\theta$. We show that these quantities tend to 0 almost surely in the following lemma.

**Lemma 5.** *We have* $s_n^\theta \overset{a.s.}{\to} 0$.

*Proof.* We can follow the proof of Lemma 4.5 in Robert and Rosenbaum (2012) to prove that for $k \in \{1,2\}$, $s_n^{(k)} \overset{a.s.}{\to} 0$. Then, we can notice that a.s. $s_n^{1C} < s_n^{(1)} + s_n^{(2)}$ to deduce that $s_n^{1C} \overset{a.s.}{\to} 0$. To show that $s_n^h \to 0$, define the process $Z$ such that $Z_0 = 0$ and $\forall i > 0$ :

$$Z_t := \begin{cases} \Delta X^{(2)}_{[\tau^{1C}_{i-1,n},t]} + Z_{\tau^{1C}_{i-1,n}} & \forall t \in [\tau^{1C}_{i-1,n}, \tau^{1C,+}_{i-1,n}] \\ \Delta X^{(1)}_{[\tau^{1C,+}_{i-1,n},t]} + Z_{\tau^{1C,+}_{i-1,n}} & \forall t \in [\tau^{1C,+}_{i-1,n}, \tau^{1C}_{i,n}] \end{cases}$$

Substituting X in Lemma 4.5 of Robert and Rosenbaum's proof by our Z, we can follow the same reasoning. The only main change will be that in their notation $M_n \leq C h_n \alpha_n$, but this tends to 0 by (2.46). □

Let $f$ be a random process, $s$ a random number, we define :

$$S(f,s) := \sup_{0 \leq u,v \leq 1, |u-v| \leq s} \left| f_u - f_v \right|$$

**Lemma 6.** *Let $f$ be a bounded random process such that for all non-random sequence $(q_n)_{n\geq 0}$, if $q_n \to 0$, then $S(f,q_n) \overset{\mathbb{P}}{\to} 0$. Let also a random sequence $(s_n)_{n\geq 0}$ such that $s_n \overset{\mathbb{P}}{\to} 0$. Then we have $\forall l \geq 1$*

$$S(f,s_n) \overset{\mathbf{L}^l}{\to} 0$$

*Proof.* As $f$ is bounded, convergence in $\mathbb{P}$ implies convergence in $\mathbf{L}^l$ for any $l \geq 1$. Hence it is sufficient to show that $S(f,s_n) \overset{\mathbb{P}}{\to} 0$. Let $\eta > 0$ and $\epsilon > 0$, we want to show that $\exists N > 0$ such that $\forall n \geq N$,

$$\mathbb{P}(S(f,s_n) > \eta) < \epsilon$$

$\exists$ non-random $\chi > 0$ such that $\mathbb{P}(S(f,\chi) > \eta) < \frac{\epsilon}{2}$. Also, $\exists N > 0$ such that $\forall n \geq N$, $\mathbb{P}(s_n \geq \chi) < \frac{\epsilon}{2}$. Thus

$$\begin{aligned} \mathbb{P}(S(f,s_n) > \eta) &= \mathbb{P}(S(f,s_n) > \eta, s_n > \chi) + \mathbb{P}(s(f,s_n) > \eta, s_n \leq \chi) \\ &\leq \mathbb{P}(s_n > \chi) + \mathbb{P}(S(f,\chi) > \eta) < \epsilon \end{aligned}$$

$\square$

We aim to define the approximations of observation times on blocks

$$\left(K_{i,n} := [\tau^h_{i,n}, \tau^h_{i+1,n}]\right)_{i\geq 0}.$$

We need some definitions first. Let $(C^{(i)}_t)_{i\geq 0}$ a sequence of independent 4-dimensional Brownian motions (i.e. for each $i$, $C^{(i)}_t$ is a 4-dimensional Brownian motion), independent of everything we have defined so far. We define $\forall i, n \geq 0$,

$$S^{i,n}_t := \begin{cases} \Delta W_{[\tau^h_{i,n}, \tau^h_{i,n}+\cdot]} & \forall t \in [0, \Delta\tau^h_{i+1,n}] \\ \Delta W_{[\tau^h_{i,n}, \tau^h_{i+1,n}]} + C^{(i)}_{t-\Delta\tau^h_{i+1,n}} & \forall t \geq \Delta\tau^h_{i+1,n} \end{cases}$$

and

$$\left(\tilde{\tau}^k_{i,j,n}\right)_{j\geq 0; k=1,2} = \tilde{\mathrm{T}}\left(S^{i,n}, \sigma_{\tau^h_{i,n}}, \alpha_n g_{\tau^h_{i,n}}, \Delta X^{(4)}_{[\tau^{h,-}_{i,n}, \tau^h_{i,n}]}, \tau^h_{i,n} - \tau^{h,-}_{i,n}\right)$$

To keep symmetry in notations, we define for all integers $i$ and $n$ positive integers, $\left(\tau^{(1)}_{i,j,n}\right)_{j\geq 0}$ consisting of the observation times of the process 1 after $\tau^h_{i,n}$, substracting the value of $\tau^h_{i,n}$, i.e. $\tau^{(1)}_{i,j,n} = \tau^{(1)}_{i^*+j,n} - \tau^{(1)}_{i^*,n}$ where $i^*$ is the (random) index on the original grid of process 1 corresponding to $\tau^h_{i,n}$ ($\tau^{(1)}_{i^*,n} = \tau^h_{i,n}$). For process 2, we define $\tau^{(2)}_{i,0,n} = 0$ and for integers $j \geq 1$, $\tau^{(2)}_{i,j,n} = \tau^{(2)}_{j^*+j-1,n} - \tau^{(1)}_{i^*,n}$, where $j^*$ is the index on the original grid of process 2 corresponding to the smallest observation time of process 2 bigger (not necessarily strictly) than $\tau^h_{i,n}$. We also define $\tau^-_{i,j,n}$, $\tau^+_{i,j,n}$, $\tau^{1C}_{i,j,n}$, $\tau^{1C,-}_{i,j,n}$, $\tau^{1C,+}_{i,j,n}$, $\tilde{\tau}^-_{i,j,n}$, $\tilde{\tau}^+_{i,j,n}$, $\tilde{\tau}^{1C}_{i,j,n}$, $\tilde{\tau}^{1C,-}_{i,j,n}$, $\tilde{\tau}^{1C,+}_{i,j,n}$ following the construction we used to define (2.11), (2.12), (2.15), (2.16) and (2.17). We also set

$$(\tilde{\pi}_{i,j,n})_{j\geq 0} = \mathit{\Pi}\left(S^{i,n}, \sigma_{\tau^h_{i,n}}, \alpha_n g_{\tau^h_{i,n}}, \Delta X^{(4)}_{[\tau^{h,-}_{i,n}, \tau^h_{i,n}]}, \tau^h_{i,n} - \tau^{h,-}_{i,n}\right)$$

**Lemma 7.** *For $\theta \in \{(1), (2), 1C\}$, any real $l > 0$, any positive integer $i$ and $n$, any* *non-negative integer $j$, we have $0 < C_l^- < C_l^+$ such that :*

$$C_l^- \alpha_n^{2l} < \mathbb{E}\left[\left(\Delta\tilde{\tau}_{i,j,n}^{\theta}\right)^l\right] \leq C_l^+ \alpha_n^{2l} \tag{2.47}$$

*where $\Delta\tilde{\tau}_{i,j,n}^{\theta} := \tilde{\tau}_{i,j,n}^{\theta} - \tilde{\tau}_{i,j-1,n}^{\theta}$ and*

$$C_l^- \alpha_n^{2l} < \mathbb{E}\left[\left(\Delta\tau_{i,n}^{(k)}\right)^l\right] \leq C_l^+ \alpha_n^{2l} \tag{2.48}$$

*Proof.* For $\theta \in \{(1),(2)\}$, because of (7) and (2.43), we can deduce (2.47) using well-known result on exit zone of a Brownian motion (see for instance Borodin and Salminen (2002)). (2.48) can be deduced using Dubins-Schwarz theorem for continuous local martingale (see, e.g. th. $V$.1.6 in Revuz and Yor (1999)). If $\theta = 1C$ writing $\Delta\tilde{\tau}_{i,j,n}^{\theta} = \left(\tilde{\tau}_{i,j-1,n}^{\theta,+} - \tilde{\tau}_{i,j-1,n}^{\theta}\right) + \left(\tilde{\tau}_{i,j,n}^{\theta,+} - \tilde{\tau}_{i,j-1,n}^{\theta,+}\right)$ and working those two terms, we can obtain (2.47) and (2.48). □

Now, we define for $\theta \in \{(1),(2),1C,h\}$ the number of observation times before $t$.

$$N_{t,n}^{\theta} = \sup\{i : \tau_{i,n}^{\theta} < t\}$$

We have the following lemma

**Lemma 8.** *For $\theta \in \{(1),(2),1C\}$, we have that the sequence $\left(\alpha_n^2 N_{t,n}^{\theta}\right)_{n\geq 1}$ is tight*

*Proof.* Here for $\theta \in \{(1),(2)\}$ we can follow the proof of Lemma 4.6 in Robert and Rosenbaum (2012) together with Lemma 5. Also, by definition we have $N_{t,n}^{1C} \leq N_{t,n}^{(1)}$ so we also deduce the tightness of $\left(\alpha_n^2 N_{t,n}^{1C}\right)_{n\geq 1}$. □

**Lemma 9.** *Let $(U_{i,n})_{i,n\geq 1}$ an array of positive random variables and $\theta \in \{(1),(2),1C\}$. If*

$$\forall u > 0, \quad \sum_{i=1}^{\llcorner u\alpha_n^{-2} \lrcorner} U_{i,n} \xrightarrow{\mathbb{P}} 0 \tag{2.49}$$

*then $\sum_{i=1}^{N_{t,n}^{\theta}} U_{i,n} \xrightarrow{\mathbb{P}} 0$. Also, if $\forall u > 0$, $\sum_{i=1}^{\llcorner u\alpha_n^{-2}h(n)^{-1} \lrcorner} U_{i,n} \xrightarrow{\mathbb{P}} 0$, then $\sum_{i=1}^{N_{t,n}^{h}} U_{i,n} \xrightarrow{\mathbb{P}} 0$*

*Proof.* Let $\epsilon > 0$ and $u > 0$.

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{N^{\theta}_{t,n}} U_{i,n} > \epsilon\right) &= \mathbb{P}\Bigg(\sum_{i=1}^{\lfloor u\alpha_n^{-2}\rfloor} U_{i,n} + \sum_{i=\lfloor u\alpha_n^{-2}\rfloor+1}^{N^{\theta}_{t,n}} U_{i,n}\mathbf{1}_{\{\lfloor u\alpha_n^{-2}\rfloor < N^{\theta}_{t,n}\}} \\
&\quad - \sum_{i=N^{\theta}_{t,n}+1}^{\lfloor u\alpha_n^{-2}\rfloor} U_{i,n}\mathbf{1}_{\{\lfloor u\alpha_n^{-2}\rfloor > N^{\theta}_{t,n}\}} > \epsilon\Bigg) \\
&\leq \mathbb{P}\left(\sum_{i=1}^{\lfloor u\alpha_n^{-2}\rfloor} U_{i,n} + \sum_{i=\lfloor u\alpha_n^{-2}\rfloor+1}^{N^{\theta}_{t,n}} U_{i,n}\mathbf{1}_{\{\lfloor u\alpha_n^{-2}\rfloor < N^{\theta}_{t,n}\}} > \epsilon\right) \\
&\leq \mathbb{P}\left(\sum_{i=1}^{\lfloor u\alpha_n^{-2}\rfloor} U_{i,n} > \frac{\epsilon}{2}\right) + \mathbb{P}\left(\sum_{i=\lfloor u\alpha_n^{-2}\rfloor+1}^{N^{\theta}_{t,n}} U_{i,n}\mathbf{1}_{\{\lfloor u\alpha_n^{-2}\rfloor < N^{\theta}_{t,n}\}} > \frac{\epsilon}{2}\right) \\
&\leq \mathbb{P}\left(\sum_{i=1}^{\lfloor u\alpha_n^{-2}\rfloor} U_{i,n} > \frac{\epsilon}{2}\right) + \mathbb{P}\left(\lfloor u\alpha_n^{-2}\rfloor < N^{\theta}_{t,n}\right)
\end{aligned}$$

We take the $\limsup_{n\to\infty}$ and uses (2.49). We obtain :

$$\limsup_{n\to\infty}\mathbb{P}\left(\sum_{i=1}^{N^{\theta}_{t,n}} U_{i,n} > \epsilon\right) \leq \limsup_{n\to\infty} \mathbb{P}\left(\lfloor u\alpha_n^{-2}\rfloor < N^{\theta}_{t,n}\right)$$

We now tend $u \to \infty$ and conclude using Lemma 8. The second statement is proved in the same way. □

**Lemma 10.** *For any $\alpha > 0$, $\sigma \in \mathcal{M}, g \in \mathcal{G}, (x,u) \in \mathcal{S}_g$, we have that*

$$\begin{aligned}
\psi^{AV}(\sigma, g, x, u) &= \alpha^{-4}\psi^{AV}\left(\sigma, \alpha g, \alpha x, \alpha^2 u\right) \\
\psi^{AC1}(\sigma, g, x, u) &= \alpha^{-3}\psi^{AC1}\left(\sigma, \alpha g, \alpha x, \alpha^2 u\right) \\
\psi^{AC2}(\sigma, g, x, u) &= \alpha^{-3}\psi^{AC2}\left(\sigma, \alpha g, \alpha x, \alpha^2 u\right) \\
\psi^{\tau}(\sigma, g, x, u) &= \alpha^{-2}\psi^{\tau}\left(\sigma, \alpha g, \alpha x, \alpha^2 u\right)
\end{aligned}$$

*Proof.* For any Brownian motion $(W_t)_{t\geq 0}$, by the scale property we have that $(W_t)_{t\geq 0} \stackrel{\mathcal{L}}{=} (\alpha^{-1}W_{\alpha^2 t})_{t\geq 0}$. Thus, if we define $\tau = \inf\{t > 0 \text{ s.t. } W_t \notin [d(t), u(t)]\}$ and $\tau_\alpha = \inf\{t > 0 \text{ s.t. } W_t \notin [\alpha d(t), \alpha u(t)]\}$, we have that

$$\tau \stackrel{\mathcal{L}}{=} \inf\{t > 0 \text{ s.t. } W_{\alpha^2 t} \notin [\alpha d(t), \alpha u(t)]\} \stackrel{\mathcal{L}}{=} \alpha^{-2}\tau_\alpha$$

We deduce that :

$$(\tau, W_\tau) \stackrel{\mathcal{L}}{=} \left(\alpha^{-2}\tau_\alpha, W_{\alpha^{-2}\tau_\alpha}\right) \stackrel{\mathcal{L}}{=} \left(\alpha^{-2}\tau_\alpha, \alpha W_{\tau_\alpha}\right) \tag{2.50}$$

We can prove the lemma based on the way we proved (2.50), at the cost of 2-dimension definitions that would be more involved and straightforward applications of Strong Markov property of Brownian motions that we won't write, so that we don't lose ourselves in the technicality of this proof. □

We introduce the number of points in the i-th block in the k-process as the following

$$N_{i,n}^{(k)} = \max\{j \geq 0 \text{ s.t. } \tau_{i,n}^h + \tau_{i,j,n}^{(k)} \leq \tau_{i+1,n}^h\}$$

We also introduce the total number of points in the i-th block $N_{i,n} = N_{i,n}^{(1)} + N_{i,n}^{(2)}$. We show now that we can control uniformly the error of the approximations of the observation times.

**Lemma 11.** *Let $l \geq 1$, we have that*

$$\sup_{i\geq 0\ ,\ 2\leq j\leq h_n} \mathbb{E}\left[\left|\Delta\tau_{i,j,n}^{1C} - \Delta\tilde{\tau}_{i,j,n}^{1C}\right|^l\right] = o_p\left(\alpha_n^{2l}\right) \tag{2.51}$$

*and*

$$\sup_{i\geq 0\ ,\ 2\leq j\leq h_n} \mathbb{E}\left[\left|\Delta\tau_{i,j,n}^{1C,-,+} - \Delta\tilde{\tau}_{i,j,n}^{1C,-,+}\right|^l\right] = o_p\left(\alpha_n^{2l}\right) \tag{2.52}$$

*Proof.* We introduce the notation $o_p^U$ where $U$ stands for "uniformly in $i \geq 0$", meaning that the sup of the rests is of the given order

First step : We define $\tilde{s}_n^h = \sup_{i \in A_n} \tilde{\tau}_{i,h_n,n}^{1C}$. We show in this step that

$$\tilde{s}_n^h \xrightarrow{\mathbb{P}} 0 \tag{2.53}$$

We define the accumulated time of approximated durations, i.e.

$$\tilde{\tau}_{i,n}^h = \sum_{l=0}^{l=i} \tilde{\tau}_{l,h_n,n}^{1C}$$

Using Lemma 7 together with Lemma 8, $\exists M > 0$ such that

$$\mathbb{P}\left(\tilde{\tau}_{N_n^h,n}^h \leq M\right) \to 1$$

We define $Z_0^n = 0$ and $\forall t \in [\tilde{\tau}_{i-1,n}^h, \tilde{\tau}_{i,n}^h]$,

$$Z_t^n = Z_{\tilde{\tau}_{i-1,n}^h}^n + S_{t-\tilde{\tau}_{i-1,n}^h}^{i-1,n}$$

A slight modification of the proof of Lemma 5 will conclude.

Second step : We show that we can do a localization in the number of observations in the i-th block, i.e. there exists a non-random $M_n$ such that

$$\mathbb{P}\left(\max\left(N_{i,n}^{(1)}, N_{i,n}^{(2)}\right) > M_n\right) \tag{2.54}$$

converges uniformly (in $i$) towards 0 and $M_n$ increasing at most linearly with $h_n$, i.e. we have $M_n \leq \beta h_n$ where $\beta > 0$.

To prove (2.54), we need some definitions. Define for $i \geq 0$ the order of observation times $O_{i,k,n}$ and the order of the approximated observation times $\tilde{O}_{i,k,n}$ in the following way. Let $\mathrm{T}_{i,n}^O := \left(\tau_{i,j,n}^O\right)_{j\geq 0}$ the sorted set of all observation times (corresponding to process 1 and 2) strictly greater than $\tau^h_{i,n}$. Then for $j \geq 1$, we will set $O_{i,j,n} = 1$ if the j-th observation time in $\mathrm{T}^O_{i,n}$ corresponds to an observation of the first process and $O_{i,j,n} = 2$ if it corresponds to an observation of the second process. Similarly, we set $\tilde{\mathrm{T}}^O_{i,n}$ the sorted set of all approximated times $\left(\tilde{\tau}^{(k)}_{i,j,n}\right)_{j\geq 0, k=1,2}$. $\tilde{O}_{i,j,n}$ are defined in the same way. There exists a $p > 0$ such that for all integers $i, j, n$ :

$$\mathbb{P}\left(O_{i,j+1,n} = 1 \Big| O_{i,j,n} = 2\right) \geq p \text{ and } \mathbb{P}\left(O_{i,j+1,n} = 2 \Big| O_{i,j,n} = 1\right) \geq p \tag{2.55}$$

Indeed, let $l$ the (random) index such that $\tau^{(1)}_{i,l,n} = \tau^O_{i,j,n}$. Conditioned on $\left\{O_{i,j,n} = 1\right\}$, we know that $O_{i,j+1,n} = 2$ if $\Delta X^{(4)}_{[\tau^h_{i,n}+\tau^l_{i,j,n},.]}$ crosses $g^+$ or $-g^+$ before $\Delta X^{(3)}_{[\tau^h_{i,n}+\tau^O_{i,j,n},.]}$ crosses $g^-$ or $-g^-$. Using (2.3) of (A2) and (2.43), we can easily bound away from 0 this probability, thus we deduce (2.55). Now, using (2.15) together with (2.55) and strong Markov property of Brownian motions, we deduce (2.54).

Third step : let $g = (d, u)$ such that $(g, g) \in \mathcal{G}$, $\sigma \in [\sigma^-, \sigma^+]$ and $\epsilon \leq \frac{g^-}{2}$. We define $\tau(g, \sigma, \epsilon) = \inf\{t > 0 : \sigma W_t = u(t) + \epsilon \text{ or } \sigma W_t = d(t) - \epsilon\}$, where $W_t$ is a standard Brownian motion. We show that

$$\mathbb{E}\left[\left|\tau(g, \sigma, \epsilon) - \tau(g, \sigma, 0)\right|^l\right] \leq \gamma^{(l)}(\epsilon) \tag{2.56}$$

where $\gamma^{(l)}(\epsilon) \overset{\epsilon \to 0}{\to} 0$.

In order to show (2.56), let

$$\tau^1(g, \sigma, \epsilon) = \inf\{t > 0 : \sigma W_{t+\tau(g,\sigma,0)} = \min\left(u(\tau(g, \sigma, 0)) + Kt + \epsilon, g^+\right)$$

$$\text{or } \sigma W_{t+\tau(g,\sigma,0)} = \max\left(d(\tau(g, \sigma, 0)) - Kt - \epsilon, g^-\right)\}.$$

By (2.4) and (2.6) of (A3), we have $\tau(g, \sigma, \epsilon) - \tau(g, \sigma, 0) \leq \tau^1(g, \sigma, \epsilon)$. Conditioned on $\left\{\tau(g, \sigma, \epsilon)\right\}$ and using strong Markov property of Brownian motions, we can show that $\mathbb{E}_{\tau(g,\sigma,\epsilon)}\left[\left|\tau^1(g, \sigma, \epsilon)\right|^l\right] \overset{\epsilon \to 0}{\to} 0$ using Theorem 2 in Potzelberger and Wang (2001) for instance.

Fourth step : let $k \in \{1,2\}$. We show here

$$\sum_{j \leq M_n} \mathbb{E}\left[\left|\tau_{i,j,n}^{(k)} - \tilde{\tau}_{i,j,n}^{(k)}\right|^l\right] = o_p^U\left(\alpha_n^{2l}\right) \tag{2.57}$$

The idea is to show that by recurrence in $j$, $\mathbb{E}\left[\left|\tau_{i,j,n}^{(k)} - \tilde{\tau}_{i,j,n}^{(k)}\right|^l\right]$ can be arbitrarily small when $n$ grows. It is then a straightforward analysis exercise to use the localization in second step and choose a different sequence $h$ if necessary, that will still be non-random increasing and following (2.45) and (2.46), so that the sum in (2.57) will be also arbitrarily small. Let's start with $j = 1$ and $k = 1$.

$$\mathbb{E}\left[\left|\tau_{i,1,n}^{(k)} - \tilde{\tau}_{i,1,n}^{(k)}\right|^l\right] = \mathbb{E}\left[\left|\tau_{i,1,n}^{(k)} - \tilde{\tau}_{i,1,n}^{(k)}\right|^l \mathbf{1}_{E_{i,n}}\right] + \mathbb{E}\left[\left|\tau_{i,1,n}^{(k)} - \tilde{\tau}_{i,1,n}^{(k)}\right|^l \mathbf{1}_{E_{i,n}^C}\right]$$

where $E_{i,n} = E_{i,n}^{(1)} \cap E_{i,n}^{(2)}$ with

$$E_{i,n}^{(1)} = \left\{\sup_{s \in [\tau_{i,n}^h, \tau_{i,n}^h + \tau_{i,1,n}^{(1)} \vee \tilde{\tau}_{i,1,n}^{(1)}]} \left|\Delta X_{[\tau_{i,n}^h, s]}^{(1)} - \Delta \tilde{X}_{[\tau_{i,n}^h, s]}^{(1)}\right| < \eta_{1,n}\right\},$$

$$E_{i,n}^{(2)} = \left\{\sup_{s \in [\tau_{i,n}^h, \tau_{i,n}^h + \tau_{i,1,n}^{(1)} \vee \tilde{\tau}_{i,1,n}^{(1)}]} \left\|g_s^{(1)} - g_{\tau_{i,n}^h}^{(1)}\right\|_\infty < \eta_{1,n}\right\},$$

$\eta_{1,n} = q_n \alpha_n$, $q_n = \max\left(\alpha_n^{d-1/2}, z_n^{1/2}\right)$ and $z_n = \sup_{1 \leq u,v \leq 4}\left(\mathbb{E}\left[\left(S\left(\sigma^{u,v}, s_n^h \vee \tilde{s}_n^h\right)\right)^2\right]\right)^{1/2}$. By (2.50) and (2.56),

$$\mathbb{E}\left[\left|\tau_{i,1,n}^{(k)} - \tilde{\tau}_{i,1,n}^{(k)}\right|^l \mathbf{1}_{E_{i,n}}\right] \leq C\alpha_n^{2l}\left(\gamma^{(l)}\left(2q_n\right) + \gamma^{(l)}\left(-2q_n\right)\right)$$

Using Cauchy-Schwarz inequality and Lemma 7,

$$\mathbb{E}\left[\left|\tau_{i,1,n}^{(k)}-\tilde{\tau}_{i,1,n}^{(k)}\right|^{l}\mathbf{1}_{E_{i,n}^{C}}\right]\leq C\alpha_{n}^{2l}\mathbb{P}\left(E_{i,n}^{C}\right)^{1/2}\leq C\alpha_{n}^{2l}\left(\mathbb{P}\left(\left(E_{i,n}^{(1)}\right)^{C}\right)+\mathbb{P}\left(\left(E_{i,n}^{(2)}\right)^{C}\right)\right)^{1/2}$$

On the one hand,

$$\begin{aligned}
\mathbb{P}\left(\left(E_{i,n}^{(1)}\right)^{C}\right) &\leq (\eta_{1,n})^{-1}\,\mathbb{E}\left[\sup_{s\in[\tau_{i,n}^{h},\tau_{i,n}^{h}+\tau_{i,1,n}^{(1)}\vee\tilde{\tau}_{i,1,n}^{(1)}]}\left|\Delta X_{[\tau_{i,n}^{h},s]}^{(1)}-\Delta\tilde{X}_{[\tau_{i,n}^{h},s]}^{(1)}\right|\right]\\
&\leq C\,(\eta_{1,n})^{-1}\max_{1\leq u,v\leq 4}\mathbb{E}\left[\left(\int_{\tau_{i,n}^{h}}^{\tau_{i,n}^{h}+\tau_{i,1,n}^{(1)}\vee\tilde{\tau}_{i,1,n}^{(1)}}\left(\sigma_{s}^{u,v}-\sigma_{\tau_{i,n}^{h}}^{u,v}\right)^{2}ds\right)^{1/2}\right]\\
&\leq C\,(\eta_{1,n})^{-1}\max_{1\leq u,v\leq 4}\mathbb{E}\left[\left(\left(\tau_{i,1,n}^{(1)}\vee\tilde{\tau}_{i,1,n}^{(1)}\right)S\left(\sigma^{u,v},s_{n}^{h}\vee\tilde{s}_{n}^{h}\right)^{2}\right)^{1/2}\right]\\
&\leq C\,(\eta_{1,n})^{-1}\left(\mathbb{E}\left[\tau_{i,1,n}^{(1)}\vee\tilde{\tau}_{i,1,n}^{(1)}\right]\right)^{1/2}z_{n}\\
&\leq Cz_{n}^{1/2}
\end{aligned}$$

where we used Markov inequality in the first inequality, conditional Burkholder-Davis-Gundy inequality in the second inequality, Cauchy-Schwarz inequality in the fourth inequality, Lemma 7 in the last inequality. On the other hand,

$$\begin{aligned}
\mathbb{P}\left(\left(E_{i,n}^{(2)}\right)^{C}\right) &\leq (\eta_{1,n})^{-1}\,\mathbb{E}\left[\sup_{s\in[\tau_{i,n}^{h},\tau_{i,n}^{h}+\tau_{i,1,n}^{(1)}\vee\tilde{\tau}_{i,1,n}^{(1)}]}\left\|g_{s}^{(1)}-g_{\tau_{i,n}^{h}}^{(1)}\right\|_{\infty}\right]\\
&\leq C\,(\eta_{1,n})^{-1}\,\mathbb{E}\left[\left(\tau_{i,1,n}^{(1)}\vee\tilde{\tau}_{i,1,n}^{(1)}\right)^{d}\right]\\
&\leq C\alpha_{n}^{d-1/2}
\end{aligned}$$

where we used Markov inequality in the first inequality, (2.7) of (A3) in the second inequality, Lemma 7 in the last inequality. In summary, we have

$$\mathbb{E}\left[\left|\tau_{i,j,n}^{(k)}-\tilde{\tau}_{i,j,n}^{(k)}\right|^{l}\right]\leq C\alpha_{n}^{2l}\left(\gamma^{(l)}\left(2q_{n}\right)+\gamma^{(l)}\left(-2q_{n}\right)+z_{n}^{1/2}+\alpha^{d-1/2}\right)$$

which we can make arbitrarily small, because $z_n \to 0$ by first step together with Lemma 6 and the continuity of $\sigma$ (A1). The case with $k = 2$ is very similar. Finally, for $j > 1$, the same kind of computation techniques, using in addition (2.6) of (A3), will work.

Fifth step : Prove that uniformly (in $i$)

$$\mathbb{P}\left(\forall j \le M_n, O_{i,j,n} = \tilde{O}_{i,j,n}\right) \to 1 \tag{2.58}$$

To show (2.58), let $j \le M_n$. We define the (random) index $v$ such that $\tau^O_{i,v,n} = \tau^{(k)}_{i,j,n}$. Modifying suitably $h$ if needed, there exists (using fourth step) a sequence $(\epsilon_n)$ such that

$$\mathbb{P}\left(\left|\tau^{(k)}_{i,j,n} - \tilde{\tau}^{(k)}_{i,j,n}\right| \le \alpha_n^2 \epsilon_n\right) \;\to\; 1 \tag{2.59}$$

$$\mathbb{P}\left(\left|\tau^{O}_{i,v+1,n} - \tau^{O}_{i,v,n}\right| \le \alpha_n^2 \epsilon_n\right) \;\to\; 0 \tag{2.60}$$

Using (2.59) and (2.60), we can verify (2.58) by recurrence.

Sixth step : We prove here (2.51) and (2.52). Using Lemma 7 and (2.58)

$$\mathbb{E}\left[\left|\Delta\tau^{1C}_{i,j,n} - \Delta\tilde{\tau}^{1C}_{i,j,n}\right|^l\right] = \mathbb{E}\left[\left|\Delta\tau^{1C}_{i,j,n} - \Delta\tilde{\tau}^{1C}_{i,j,n}\right|^l \mathbf{1}_{\{\forall j \le M_n, O_{i,j,n} = \tilde{O}_{i,j,n}\}}\right] + o_p^U\left(\alpha_n^{2l}\right)$$

The first term on the right part of the inequality can be bounded by

$$C\left(\mathbb{E}\left[\left|\tau^{1C}_{i,j,n} - \tilde{\tau}^{1C}_{i,j,n}\right|^l \mathbf{1}_{\{\forall j \le M_n, O_{i,j,n} = \tilde{O}_{i,j,n}\}}\right] + \mathbb{E}\left[\left|\tau^{1C}_{i,j-1,n} - \tilde{\tau}^{1C}_{i,j-1,n}\right|^l \mathbf{1}_{\{\forall j \le M_n, O_{i,j,n} = \tilde{O}_{i,j,n}\}}\right]\right)$$

Both terms can be treated with the same trick. Using the second step and Lemma 7, the first term is equal to

$$\sum_{v \le M_n} \mathbb{E}\left[\left|\tau^{1C}_{i,j,n} - \tilde{\tau}^{1C}_{i,j,n}\right|^l \mathbf{1}_{\{\forall j \le M_n, O_{i,j,n} = \tilde{O}_{i,j,n}\}} \mathbf{1}_{\{\tau^{1C}_{i,j,n} = \tau^{(1)}_{i,v,n}\}}\right] + o_p^U\left(\alpha_n^{2l}\right)$$

The sum is obviously bounded by

$$\sum_{v \le M_n} \mathbb{E}\left[\left|\tau^{1C}_{i,j,n} - \tilde{\tau}^{1C}_{i,j,n}\right|^l\right]$$

and using (2.57), we prove (2.51). We can deduce (2.52) with the same kind of computations. $\square$

Let $M^n$ the interpolated normalized error, i.e.

$$M^n_t = \alpha_n^{-1}\left(\sum_{i\ge 1} \Delta X^{(1)}_{[\tau^{1C}_{i-1,n}\wedge t,\tau^{1C}_{i,n}\wedge t]} \Delta X^{(2)}_{[\tau^{1C,-}_{i-1,n}\wedge t,\tau^{1C,+}_{i,n}\wedge t]} - \int_0^t \sigma^{(1)}_s \sigma^{(2)}_s \rho^{1,2}_s ds\right)$$

$M^n_t$ corresponds exactly to the normalized error of the Hayashi-Yoshida estimator if we observe the price of both assets at time $t$. We remind to the reader the definition of $N_{i,n}$ in (2.20)

$$N_{i,n} = \Delta X^{(1)}_{\tau^{1C}_{i,n}} \Delta X^{(2)}_{\tau^{1C,-,+}_{i,n}} - \int_{\tau^{1C}_{i-1,n}}^{\tau^{1C}_{i,n}} \sigma^{(1)}_s \sigma^{(2)}_s \rho^{1,2}_s ds$$

**Lemma 12.**

$$\sum_{i \in A_n} \mathbb{E}_{\tau^h_{i-1,n}}\left[\left(\Delta M^n_{\tau^h_{i,n}}\right)^2\right]$$

$$= \alpha_n^{-2} \sum_{i\in A_n} \mathbb{E}_{\tau^h_{i-1,n}}\left[\sum_{u=2}^{h_n} \left(N_{(i-1)h_n+u}\right)^2 + 2N_{(i-1)h_n+u}N_{(i-1)h_n+u+1}\right] + o_p(1)$$

*Proof.* We obtain this equality noting that $(N_{i,n})_{n\ge 0}$ are centered and 1-correlated, and that the terms left converge to 0 in probability. $\square$

We introduce the observation time at the start of a block, where "s" stands for "start"

$$\tau^s_{i,n} = \sup\{\tau^h_{j,n} \text{ s.t. } \tau^h_{j,n} < \tau^{1C}_{i,n}\}$$

**Lemma 13.**

$$\alpha_n^{-2}\sum_{i\in A_n}\mathbb{E}_{\tau^h_{i-1,n}}\left[\sum_{u=2}^{h_n}\left(N_{(i-1)h_n+u}\right)^2+2N_{(i-1)h_n+u}N_{(i-1)h_n+u+1}\right]$$

$$=\alpha_n^2\sum_{i\in A_n}\sum_{j=0}^{h_n-2}\int_{\mathbb{R}^2}\psi^{AV}\left(\sigma_{\tau^h_{i-1,n}},g_{\tau^h_{i-1,n}},\alpha_n^{-1}x,\alpha_n^{-2}v\right)d\tilde{\pi}_{i-1,j,n}\left(x,v\right)+o_p(1)$$

*Proof.* First step : approximating with holding volatility constant. Set

$$\tilde{N}_{i,n}=\left(\sigma_{\tau^s_{i-1,n}}\Delta W_{\tau^{1C}_{i,n}}\right)^{(1)}\left(\sigma_{\tau^s_{i-1,n}}\Delta W_{\tau^{1C,-,+}_{i,n}}\right)^{(2)}-\int_{\tau^{1C}_{i-1,n}}^{\tau^{1C}_{i,n}}\zeta^{1,2}_{\tau^s_{i-1,n}}ds$$

where $A^{(i)}$ is the i-th component of the vector A. we want to show that :

$$\alpha_n^{-2}\sum_{i\in A_n}\mathbb{E}_{\tau^h_{i-1,n}}\left[\sum_{u=2}^{h_n}\left(N_{(i-1)h_n+u}\right)^2+2N_{(i-1)h_n+u}N_{(i-1)h_n+u+1}\right]$$

$$=\alpha_n^{-2}\sum_{i\in A_n}\mathbb{E}_{\tau^h_{i-1,n}}\left[\sum_{u=2}^{h_n}\left(\tilde{N}_{(i-1)h_n+u}\right)^2+2\tilde{N}_{(i-1)h_n+u}\tilde{N}_{(i-1)h_n+u+1}\right]+o_p\left(1\right)$$

Noting $F_{i,n}=\left(N_{i,n}\right)^2+2N_{i,n}N_{i+1,n}$ and $\tilde{F}_{i,n}=\left(\tilde{N}_{i,n}\right)^2+2\tilde{N}_{i,n}\tilde{N}_{i+1,n}$, it is sufficient to show that

$$\alpha_n^{-2}\sum_{i\geq 1}\mathbb{E}_{\tau^s_{i-1,n}}\left[\left|F_{i,n}-\tilde{F}_{i,n}\right|\mathbf{1}_{\{\tau^s_{i-1,n}<t\}}\right]\xrightarrow{\mathbb{P}}0$$

that we can rewrite as $\alpha_n^{-2}\sum_{i\geq 1}^{N^{(1)}_{t,n}}\mathbb{E}_{\tau^s_{i-1,n}}\left[\left|F_{i,n}-\tilde{F}_{i,n}\right|\mathbf{1}_{\{\tau^s_{i-1,n}<t\}}\right]\xrightarrow{\mathbb{P}}0$. Using Lemma 9, it is sufficient to show that $\forall u>0$ :

$$\alpha_n^{-2}\sum_{i=1}^{u\alpha_n^{-2}}\mathbb{E}_{\tau^s_{i-1,n}}\left[\left|F_{i,n}-\tilde{F}_{i,n}\right|\mathbf{1}_{\{\tau^s_{i-1,n}<t\}}\right]\xrightarrow{\mathbb{P}}0$$

Thus, it is sufficient to show the convergence $\mathbf{L}^1$ of this quantity, i.e. that

$$\alpha_n^{-2}\sum_{i=1}^{u\alpha_n^{-2}}\mathbb{E}\left[\left|F_{i,n}-\tilde{F}_{i,n}\right|\mathbf{1}_{\{\tau^s_{i-1,n}<t\}}\right]\to 0$$

We have that

$$\left|F_{i,n}-\tilde{F}_{i,n}\right|\leq B^{(1)}_{i,n}+2B^{(2)}_{i,n}$$

where $B^{(1)}_{i,n}=\left|N^2_{i,n}-\tilde{N}^2_{i,n}\right|$ and $B^{(2)}_{i,n}=\left|N_{i-1,n}N_{i,n}-\tilde{N}_{i-1,n}\tilde{N}_{i,n}\right|$. We have that

$$B^{(1)}_{i,n}\leq C^{(1)}_{i,n}+C^{(2)}_{i,n}+C^{(3)}_{i,n}$$

where

$$\begin{aligned}
C^{(1)}_{i,n} &= \left|\left(\Delta X^{(1)}_{\tau^{1C}_{i,n}}\Delta X^{(2)}_{\tau^{1C,-,+}_{i,n}}\right)^2-\left(\left(\sigma_{\tau^s_{i-1,n}}\Delta W_{\tau^{1C}_{i,n}}\right)^{(1)}\left(\sigma_{\tau^s_{i-1,n}}\Delta W_{\tau^{1C,-,+}_{i,n}}\right)^{(2)}\right)^2\right| \\
C^{(2)}_{i,n} &= \left|\left(\int_{\tau^{1C}_{i-1,n}}^{\tau^{1C}_{i,n}}\zeta^{1,2}_s ds\right)^2-\left(\int_{\tau^{1C}_{i-1,n}}^{\tau^{1C}_{i,n}}\zeta^{1,2}_{\tau^s_{i-1,n}}ds\right)^2\right|
\end{aligned}$$

$$\begin{aligned}
C^{(3)}_{i,n} &= 2\left|\Delta X^{(1)}_{\tau^{1C}_{i,n}}\Delta X^{(2)}_{\tau^{1C,-,+}_{i,n}}\int_{\tau^{1C}_{i-1,n}}^{\tau^{1C}_{i,n}}\zeta^{1,2}_s ds\right. \\
&\quad\left.-\left(\sigma_{\tau^s_{i-1,n}}\Delta W_{\tau^{1C}_{i,n}}\right)^{(1)}\left(\sigma_{\tau^s_{i-1,n}}\Delta W_{\tau^{1C,-,+}_{i,n}}\right)^{(2)}\int_{\tau^{1C}_{i-1,n}}^{\tau^{1C}_{i,n}}\zeta^{1,2}_{\tau^s_{i-1,n}}ds\right|
\end{aligned}$$

Let's show that $\alpha_n^{-2}\sum_{i=1}^{u\alpha_n^{-2}}\mathbb{E}\left[C^{(1)}_{i,n}\mathbf{1}_{\{\tau^s_{i-1,n}<t\}}\right]\to 0$. We can write it as $C^{(1)}_{i,n}\leq D^{(1)}_{i,n}+D^{(2)}_{i,n}$,

where

$$\begin{aligned} D_{i,n}^{(1)} &= \left| \left( \Delta X_{\tau_{i,n}^{1C}}^{(1)} \Delta X_{\tau_{i,n}^{1C,-,+}}^{(2)} \right)^2 \left( \left( \sigma_{\tau_{i-1,n}^s} \Delta W_{\tau_{i,n}^{1C}} \right)^{(1)} \Delta X_{\tau_{i,n}^{1C,-,+}}^{(2)} \right)^2 \right| \\ D_{i,n}^{(2)} &= \left| \left( \left( \sigma_{\tau_{i-1,n}^s} \Delta W_{\tau_{i,n}^{1C}} \right)^{(1)} \Delta X_{\tau_{i,n}^{1C,-,+}}^{(2)} \right)^2 - \right. \\ & \left. - \left( \left( \sigma_{\tau_{i-1,n}^s} \Delta W_{\tau_{i,n}^{1C}} \right)^{(1)} \left( \sigma_{\tau_{i-1,n}^s} \Delta W_{\tau_{i,n}^{1C,-,+}} \right)^{(2)} \right)^2 \right| \end{aligned}$$

We want to show that $\alpha_n^{-2} \sum_{i=1}^{u\alpha_n^{-2}} \mathbb{E}\left[ D_{i,n}^{(1)} \mathbf{1}_{\{\tau_{i-1,n}^s < t\}} \right] \to 0$. We define :

$$\begin{aligned} E_{i,n}^{(1)} &= \Delta X_{\tau_{i,n}^{1C}}^{(1)} \Delta X_{\tau_{i,n}^{1C,-,+}}^{(2)} \\ E_{i,n}^{(2)} &= \left( \sigma_{\tau_{i-1,n}^s} \Delta W_{\tau_{i,n}^{1C}} \right)^{(1)} \Delta X_{\tau_{i,n}^{1C,-,+}}^{(2)} \end{aligned}$$

Using Cauchy-Schwarz inequality, we deduce :

$$\begin{aligned} \mathbb{E}\left[ D_{i,n}^{(1)} \mathbf{1}_{\{\tau_{i-1,n}^s < t\}} \right] &= \mathbb{E}\left[ \left( E_{i,n}^{(1)} + E_{i,n}^{(2)} \right) \left( E_{i,n}^{(1)} - E_{i,n}^{(2)} \right) \mathbf{1}_{\{\tau_{i-1,n}^s < t\}} \right] \\ &\leq \left( \mathbb{E}\left[ \left( E_{i,n}^{(1)} + E_{i,n}^{(2)} \right)^2 \right] \mathbb{E}\left[ \left( E_{i,n}^{(1)} - E_{i,n}^{(2)} \right)^2 \mathbf{1}_{\{\tau_{i-1,n}^s < t\}} \right] \right)^{1/2} \end{aligned}$$

Using Cauchy-Schwarz inequality together with Burkholder-Davis-Gundy inequality and Lemma 7, we obtain that :

$$\mathbb{E}\left[ \left( E_{i,n}^{(1)} + E_{i,n}^{(2)} \right)^2 \right] = O^U\left( \alpha_n^4 \right)$$

where $U$ stands for "uniformly in $1 \leq i \leq u\alpha_n^{-2}$". Another application of Cauchy-Schwarz inequality gives us

$$\mathbb{E}\left[ \left( E_{i,n}^{(1)} - E_{i,n}^{(2)} \right)^2 \mathbf{1}_{\{\tau_{i-1,n}^s < t\}} \right]$$

$$\leq \left(\mathbb{E}\left[\left(\Delta X^{(1)}_{\tau^{1C}_{i,n}} - \left(\sigma_{\tau^s_{i-1,n}}\Delta W_{\tau^{1C}_{i,n}}\right)^{(1)}\right)^4 \mathbf{1}_{\{\tau^s_{i-1,n}<t\}}\right]\mathbb{E}\left[\left(\Delta X^{(2)}_{\tau^{1C,-,+}_{i,n}}\right)^4\right]\right)^{1/2}$$

Using once again Cauchy-Schwarz inequality together with Burholder-Davis-Gundy inequality and Lemma 7, we obtain that :

$$\mathbb{E}\left[\left(\Delta X^{(2)}_{\tau^{1C,-,+}_{i,n}}\right)^4\right] = O^U\left(\alpha_n^4\right)$$

Similarly, we compute using conditional Burkholder-Davis-Gundy in first inequality, Cauchy-Schwarz in third inequality, Lemma 5, Lemma 6 and Lemma 7 together with the continuity of $\sigma$ (A1) in last equality.

$$\mathbb{E}\left[\left(\Delta X^{(1)}_{\tau^{1C}_{i,n}} - \left(\sigma_{\tau^s_{i-1,n}}\Delta W_{\tau^{1C}_{i,n}}\right)^{(1)}\right)^4 \mathbf{1}_{\{\tau^s_{i-1,n}<t\}}\right]$$

$$\begin{aligned}
&= \mathbb{E}\left[\mathbf{1}_{\{\tau^s_{i-1,n}<t\}}\mathbb{E}_{\tau^{1C}_{i-1,n}}\left[\left(\Delta X^{(1)}_{\tau^{1C}_{i,n}} - \left(\sigma_{\tau^s_{i-1,n}}\Delta W_{\tau^{1C}_{i,n}}\right)^{(1)}\right)^4\right]\right] \\
&= \mathbb{E}\left[\mathbf{1}_{\{\tau^s_{i-1,n}<t\}}\mathbb{E}_{\tau^{1C}_{i-1,n}}\left[\left(\int_{\tau^{1C}_{i-1,n}}^{\tau^{1C}_{i,n}}\left(\left(\sigma_s - \sigma_{\tau^s_{i-1,n}}\right)dW_s\right)^{(1)}\right)^4\right]\right] \\
&\leq C\sup_{1\leq j,l\leq 4}\mathbb{E}\left[\mathbf{1}_{\{\tau^s_{i-1,n}<t\}}\mathbb{E}_{\tau^{1C}_{i-1,n}}\left[\left(\int_{\tau^{1C}_{i-1,n}}^{\tau^{1C}_{i,n}}\left(\sigma^{j,l}_s - \sigma^{j,l}_{\tau^s_{i-1,n}}\right)^2 ds\right)^2\right]\right] \\
&= C\sup_{1\leq j,l\leq 4}\mathbb{E}\left[\mathbf{1}_{\{\tau^s_{i-1,n}<t\}}\left(\int_{\tau^{1C}_{i-1,n}}^{\tau^{1C}_{i,n}}\left(\sigma^{j,l}_s - \sigma^{j,l}_{\tau^s_{i-1,n}}\right)^2 ds\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&\leq C \sup_{1\leq j,l\leq 4} \mathbb{E}\left[\left(\Delta\tau_{i,n}^{1C} S\left(\sigma^{j,l}, s_n^h\right)^2\right)^2\right] + o^U\left(\alpha_n^4\right) \\
&\leq C\left(\mathbb{E}\left[\left(\Delta\tau_{i,n}^{1C}\right)^4\right]\mathbb{E}\left[\sup_{1\leq j,l\leq 4}\left(S\left(\sigma^{j,l}, s_n^h\right)\right)^8\right]\right)^{1/2} + o^U\left(\alpha_n^4\right) \\
&= O^U\left(\alpha_n^4\right)
\end{aligned}
$$

With the same kind of computations, we show that $\alpha_n^{-2}\sum_{i=1}^{u\alpha_n^{-2}} \mathbb{E}\left[D_{i,n}^{(2)}\mathbf{1}_{\{\tau_{i-1,n}^s<t\}}\right] \to 0$, and we also can show $\alpha_n^{-2}\sum_{i=1}^{u\alpha_n^{-2}} \mathbb{E}\left[C_{i,n}^{(2)}\mathbf{1}_{\{\tau_{i-1,n}^s<t\}}\right] \to 0$, $\alpha_n^{-2}\sum_{i=1}^{u\alpha_n^{-2}} \mathbb{E}\left[C_{i,n}^{(3)}\mathbf{1}_{\{\tau_{i-1,n}^s<t\}}\right] \to 0$ (thus we have also that $\alpha_n^{-2}\sum_{i=1}^{u\alpha_n^{-2}} \mathbb{E}\left[B_{i,n}^{(1)}\mathbf{1}_{\{\tau_{i-1,n}^s<t\}}\right] \to 0$) and

$$
\alpha_n^{-2}\sum_{i=1}^{u\alpha_n^{-2}} \mathbb{E}\left[B_{i,n}^{(2)}\mathbf{1}_{\{\tau_{i-1,n}^s<t\}}\right] \to 0.
$$

Second step : approximating using $(\tilde{\tau}_{i,j,n})_{i,j,n\geq 0}$ instead of $(\tau_{i,n})_{i,n\geq 0}$. We set

$$
\tilde{\tilde{N}}_{i,j,n} = \left(\sigma_{\tau_{i,n}^h}\Delta W_{\tilde{\tau}_{i,j,n}^{1C}}\right)^{(1)}\left(\sigma_{\tau_{i,n}^h}\Delta W_{\tilde{\tau}_{i,j,n}^{1C,-,+}}\right)^{(2)} - \int_{\tilde{\tau}_{i,j-1,n}^{1C}}^{\tilde{\tau}_{i,j,n}^{1C}} \zeta_{\tau_{i,n}^h}^{1,2}\, ds
$$

we want to show that :

$$
\alpha_n^{-2}\sum_{i\in A_n}\mathbb{E}_{\tau_{i-1,n}^h}\left[\sum_{u=2}^{h_n}\left(\tilde{N}_{(i-1)h_n+u}\right)^2 + 2\tilde{N}_{(i-1)h_n+u}\tilde{N}_{(i-1)h_n+u+1}\right]
$$

$$
= \alpha_n^{-2}\sum_{i\in A_n}\mathbb{E}_{\tau_{i-1,n}^h}\left[\sum_{u=2}^{h_n}\left(\tilde{\tilde{N}}_{i-1,u,n}\right)^2 + 2\tilde{\tilde{N}}_{i-1,u,n}\tilde{\tilde{N}}_{i,u+1,n}\right] + o_p(1)
$$

Using the same kind of computations as in the first step together with Lemma 11, we conclude.

Third step : express the result as a function of $\psi^{AV}$. Using Lemma 10 in last equality, we deduce for any integer u such that $2 \leq u \leq h_n$ :

$$\begin{aligned}
& \mathbb{E}_{\tau^h_{i-1,n}}\left[\left(\tilde{\tilde{N}}_{i-1,u,n}\right)^2 + 2\tilde{\tilde{N}}_{i-1,u,n}\tilde{\tilde{N}}_{i-1,u+1,n}\right] \\
= & \int_{\mathbb{R}^2} \psi^{AV}\left(\sigma_{\tau^h_{i-1,n}}, \alpha_n g_{\tau^h_{i-1,n}}, x, v\right) d\tilde{\pi}_{i,u-2,n}(x,v) \\
= & \alpha_n^4 \int_{\mathbb{R}^2} \psi^{AV}\left(\sigma_{\tau^h_{i-1,n}}, g_{\tau^h_{i-1,n}}, \alpha_n^{-1}x, \alpha_n^{-2}v\right) d\tilde{\pi}_{i,u-2,n}(x,v)
\end{aligned}$$

$\square$

**Lemma 14.** $\forall \sigma \in \mathcal{M}, g \in \mathcal{G}, \exists \pi(\sigma, g)$ *distribution such that :*

$$\alpha_n^2 \sum_{i \in A_n} \sum_{j=0}^{h_n-2} \int_{\mathbb{R}^2} \psi^{AV}\left(\sigma_{\tau^h_{i-1,n}}, g_{\tau^h_{i-1,n}}, \alpha_n^{-1}x, \alpha_n^{-2}u\right) d\tilde{\pi}_{i-1,j,n}(x,u)$$

$$= \alpha_n^2 \sum_{i \in A_n} h_n \phi^{AV}\left(\sigma_{\tau^h_{i-1,n}}, g_{\tau^h_{i-1,n}}\right) + o_p(1)$$

*Proof.* We define the transition functions of the Markov chains $\left(\tilde{Z}_i(\sigma, g)\right)_{i \geq 0}$ defined in (2.25). For $(x,u) \in \mathcal{S}_g$, $B \in \mathcal{B}(\mathcal{S}_g)$ (borelians of $\mathcal{S}_g$)

$$P(\sigma, g)((x,u), B) = \mathbb{P}\left(\tilde{Z}_1(\sigma, g) \in B \middle| \tilde{Z}_0(\sigma, g) = (x,u)\right)$$

First step : We prove that $\forall \sigma \in \mathcal{M}$, $\forall g \in \mathcal{G}$, the state space $\mathcal{S}_g$ is $\nu$-small, i.e. there exists a non-trivial measure $\nu$ on $\mathcal{B}(\mathbb{R}^2)$ such that $\forall (x,u) \in \mathcal{S}_g, \forall B \in \mathcal{B}(\mathcal{S}_g)$, $P(\sigma, g)((x,u), B) \geq \nu(B)$. Let $B = [x_a, x_b] \times [u_a, u_b]$. We are choosing $\nu$ such that $\nu = 0$ outside $[-\frac{g^-}{4}, \frac{g^-}{4}] \times [3,4]$. Thus, without loss of generality, we have that $[x_a, x_b] \times [u_a, u_b] \subset [-\frac{g^-}{4}, \frac{g^-}{4}] \times [3,4]$. We want to show that $\exists c > 0$ such that uniformly

$$P(\sigma, g)((x,u), B) \geq c(x_b - x_a)(u_b - u_a)$$

There are two useful ways to rewrite $(\tilde{X}^{(3)}, \tilde{X}^{(4)})$. The first one is :

$$\tilde{X}_t^{(3)} := \sigma^{(3)}\tilde{B}_t^{(3)} \tag{2.61}$$

$$\tilde{X}_t^{(4)} := \rho^{3,4}\sigma^{(4)}\tilde{B}_t^{(3)} + \left(1 - \left(\rho^{3,4}\right)^2\right)^{1/2}\sigma^{(4)}\tilde{B}_t^{3,\perp} \tag{2.62}$$

where $\tilde{B}^{(3)}$ and $\tilde{B}^{3,\perp}$ are independent, $\rho^{3,4} \in [\rho_-^{3,4}, \rho_+^{3,4}]$ and $\max\left(-\rho_-^{3,4}, \rho_+^{3,4}\right) < 1$ (because $\sigma \in \mathcal{M}$),

$$\delta = \left(1 - \max\left(\left(\rho_-^{3,4}\right)^2, \left(\rho_+^{3,4}\right)^2\right)\right)^{1/2} \tag{2.63}$$

The other way to rewrite it is :

$$\tilde{X}_t^{(4)} := \sigma^{(4)}\tilde{B}_t^{(4)} \tag{2.64}$$

$$\tilde{X}_t^{(3)} := \rho^{3,4}\sigma^{(3)}\tilde{B}_t^{(4)} + \left(1 - \left(\rho^{3,4}\right)^2\right)^{1/2}\sigma^{(3)}\tilde{B}_t^{4,\perp} \tag{2.65}$$

where $\tilde{B}^{(4)}$ and $\tilde{B}^{4,\perp}$ are independent. For $(B_t)_{t\geq 0}$ a standard Brownian motion, $a < x < b$, we denote the exiting-zone time of the Brownian motion

$$\tau_x^{a,b} = \inf\{t > 0 \text{ s.t. } x + B_t = a \text{ or } x + B_t = b\}$$

and $p_1(x, a, b, t)$ the density of $\tau_x^{a,b}$. We also define $p_2(x, a, b, s, y)$ the distribution of $B_s + x$ conditioned on $\{\tau_x^{a,b} \geq s\}$. Finally, let $p_3(x, a, b, t)$ the distribution of $\tau_x^{a,b}$ conditioned on $\{B_{\tau_x^{a,b}} = b\}$. All the formulas can be found in Borodin and Salminen (2002). Consider the spaces $C_1 = C_3 = \{(x, a, b, t) \in \mathbb{R}^4 \text{ s.t. } a \leq x \leq b, t > 0\}$, $C_2 = \{(x, a, b, t, y) \in \mathbb{R}^5 \text{ s.t. } a \leq x \leq b\ ,\ a < y < b\ ,\ t > 0\}$. The functions $p_i$ are continuous on $C_i$ and positive. Thus, for all compact set $K_i \subset C_i$, we have

$$\inf_{k \in K_i} p_i(k) > 0 \tag{2.66}$$

We can bound below

$$P(\sigma, g)((x,u), B) \quad \geq \quad \mathbb{P}\left(E_0 \bigcap E_1 \bigcap E_2 \bigcap E_3 \bigcap E_4 \middle| \tilde{Z}_0 = (x,u)\right)$$

where

$$\begin{aligned}
E_0 &= \left\{\sup_{0 \leq s \leq \tilde{\tau}_1^{(2)}} \left|\tilde{X}_s^{(3)}\right| < \frac{\epsilon\sigma^- \min(\sigma^-, 1)}{15\sigma^+} \ , \ \tilde{\tau}_1^{(2)} \leq K\right\} \\
E_1 &= \left\{\sup_{\tilde{\tau}_1^{(2)} \leq s \leq K+1} \left|\tilde{X}_s^{(3)}\right| < \frac{\epsilon\sigma^-}{10\sigma^+} \ , \ \sup_{\tilde{\tau}_1^{(2)} \leq s \leq K+1} \left|\Delta\tilde{B}^{3,\perp}_{[\tilde{\tau}_1^{(2)}, s]}\right| < \frac{g^-\sigma^-}{4(\sigma^+)^2}\right\} \\
E_2 &= \left\{\sup_{K+1 \leq s \leq \tilde{\tau}_2^{(2)}} \left|\tilde{X}_s^{(3)}\right| \leq \frac{\epsilon}{5} \ , \ \tilde{\tau}_2^{(2)} \in [K+2, K+3]\right\} \\
E_3 &= \left\{\forall s \in [\tilde{\tau}_2^{(2)}, K+4] \ \tilde{X}_s^{(3)} \in [d_1(K), u_1(K)] \ , \ \tilde{X}_{K+4}^{(3)} \in [u_1(K) - 2\epsilon, u_1(K) - \epsilon]\right\} \\
&\quad \bigcap \left\{\sup_{\tilde{\tau}_2^{(2)} \leq s \leq K+4} \left|\Delta\tilde{X}^{(4)}_{[\tilde{\tau}_2^{(2)}, s]}\right| < \frac{g^-}{12}\right\} \\
E_4 &= \left\{\tilde{\tau}_1^{(1)} \in [u_a + \tilde{\tau}_2^{(2)}, u_b + \tilde{\tau}_2^{(2)}] \ , \ \inf_{K+4 \leq s \leq \tilde{\tau}_1^{(1)}} \Delta\tilde{X}^{(3)}_{[K+4, s]} > -2\epsilon\right\} \\
&\quad \bigcap \left\{\sup_{K+4 \leq s \leq \tilde{\tau}_1^{(1)}} \left|\Delta\tilde{X}^{(4)}_{[\tilde{\tau}_2^{(2)}, s]}\right| < g^- \ , \ \Delta\tilde{X}^{(4)}_{[\tilde{\tau}_2^{(2)}, \tilde{\tau}_1^{(1)}]} \in [x_a, x_b]\right\}
\end{aligned}$$

where $\epsilon = \frac{g^-\sigma^-}{24\sigma^+}$. Using extensively Bayes formula, we can rewrite

$$\mathbb{P}\left(E_0 \bigcap E_1 \bigcap E_2 \bigcap E_3 \bigcap E_4 \bigcap \{\tilde{Z}_1 \in B\} \middle| \tilde{Z}_0 = (x,u)\right) = I \times II \times III \times IV \times V$$

where $I = \mathbb{P}\left(E_0 \middle| \{\tilde{Z}_0 = (x,u)\}\right)$, $II = \mathbb{P}\left(E_1 \middle| E_0 \bigcap \{\tilde{Z}_0 = (x,u)\}\right)$, and also $III = \mathbb{P}\left(E_2 \middle| E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\}\right)$, $IV = \mathbb{P}\left(E_3 \middle| E_2 \bigcap E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\}\right)$ and $V = \mathbb{P}\left(E_4 \middle| E_3 \bigcap E_2 \bigcap E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\}\right)$.

We prove that $I$ is uniformly bounded away from 0. Using (2.44), (2.61), (2.62) and

(2.63), we deduce that $E_0^{(1)} \bigcap E_0^{(2)} \subset E_0$ where

$$\begin{aligned} E_0^{(1)} &= \left\{ \sup_{0 \le s \le K} \left| \tilde{B}_s^{(3)} \right| < \frac{\epsilon \sigma^- \min(\sigma^-, 1)}{15 \left(\sigma^+\right)^2} \right\} \\ E_0^{(2)} &= \left\{ \sup_{0 \le s \le K} \left| \frac{x}{\sigma^{(4)} \left(1 - (\rho^{3,4})^2\right)^{1/2}} + \tilde{B}_s^{3,\perp} \right| \ge \frac{g^+}{\delta \sigma^-} + \frac{\epsilon \sigma^- \min(\sigma^-, 1)}{15 \left(\sigma^+\right)^2} \right\} \end{aligned}$$

Conditionally on $\{\tilde{Z}_0 = (x,u)\}$, $E_0^{(1)}$ and $E_0^{(2)}$ are independent. Thus, we deduce

$$I \ge \mathbb{P}\left( E_0^{(1)} \middle| \{\tilde{Z}_0 = (x,u)\} \right) \mathbb{P}\left( E_0^{(2)} \middle| \{\tilde{Z}_0 = (x,u)\} \right)$$

Using Markov property of Brownian motions, we obtain that the right part of the inequality is equal to

$$\left( 1 - \int_0^K p_1 \left( 0, -\frac{\epsilon \sigma^- \min(\sigma^-, 1)}{15 \left(\sigma^+\right)^2}, \frac{\epsilon \sigma^- \min(\sigma^-, 1)}{15 \left(\sigma^+\right)^2}, t \right) dt \right) \int_0^K p_1 \left( y_0^{(1)}, -y_0^{(2)}, y_0^{(2)}, t \right) dt$$

where $y_0^{(1)} = \frac{x}{\sigma^{(4)}\left(1-(\rho^{3,4})^2\right)^{1/2}}$, $y_0^{(2)} = \frac{g^+}{\delta\sigma^-} + \frac{\epsilon\sigma^- \min(\sigma^-,1)}{15(\sigma^+)^2}$, which is uniformly (in $x$, $\sigma$ and $g$) bounded away from 0 using (2.44) and (2.66).

We prove that $II$ is uniformly bounded away from 0. Conditionally on $E_0 \bigcap \{\tilde{Z}_0 = (x,u)\}$, the two quantities of $E_1$ are independent. Thus, we bound below $II$ (the same way we did for $I$) by :

$$\left( 1 - \int_{\tilde{\tau}_1^{(2)}}^{K+1} p_1 \left( \tilde{B}^{(3)}_{\tilde{\tau}_1^{(2)}}, -\frac{\epsilon \sigma^-}{10 \sigma^+ \sigma^{(3)}}, \frac{\epsilon \sigma^-}{10 \sigma^+ \sigma^{(3)}}, t \right) dt \right)$$

$$\left( 1 - \int_{\tilde{\tau}_1^{(2)}}^{K+1} p_1 \left( 0, -\frac{g^- \sigma^-}{4 \sigma^+ \sigma^{(4)}}, \frac{g^- \sigma^-}{4 \sigma^+ \sigma^{(4)}}, t \right) dt \right)$$

which is uniformly bounded away from 0 using (2.44) together with (2.66).

We prove that $III$ is uniformly bounded away from 0. Using (2.44), (2.61), (2.62) and (2.63), we deduce that $E_2^{(1)} \bigcap E_2^{(2)} \subset E_2$ where

$$
\begin{aligned}
E_2^{(1)} &= \left\{ \sup_{K+1 \le s \le K+3} \left| \tilde{B}_s^{(3)} \right| \le \frac{\epsilon}{5\sigma^+} \right\} \\
E_2^{(2)} &= \left\{ \sup_{K+1 \le s \le K+2} \left| \Delta \tilde{B}^{3,\perp}_{[\tilde{\tau}_1^{(2)}, s]} \right| < \frac{g^-}{2\sigma^+} \,, \sup_{K+2 \le s \le K+3} \left| \Delta \tilde{B}^{3,\perp}_{[\tilde{\tau}_1^{(2)}, s]} \right| \ge \frac{g^+}{\delta\sigma^-} + \frac{\epsilon}{5\sigma^+\delta} \right\}
\end{aligned}
$$

Conditionally on $E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\}$, $E_2^{(1)}$ and $E_2^{(2)}$ are independent. Thus, we deduce

$$
III \ge \mathbb{P}\left( E_2^{(1)} \middle| E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\} \right) \mathbb{P}\left( E_2^{(2)} \middle| E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\} \right)
$$

Using Markov property of Brownian motions, we obtain that the right part of the inequality conditioned on $\{\tilde{B}^{(3)}_{K+1} \,, \Delta \tilde{B}^{3,\perp}_{[\tilde{\tau}_1^{(2)}, K+1]} \big| E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\}\}$ is equal to

$$
\left(1 - \int_0^2 p_1 \left( \tilde{B}^{(3)}_{K+1}, -\frac{\epsilon}{5\sigma^+}, \frac{\epsilon}{5\sigma^+}, t \right) dt \right) \left(1 - \int_0^1 p_1 \left( \Delta \tilde{B}^{3,\perp}_{[\tilde{\tau}_1^{(2)}, K+1]}, -\frac{g^+}{2\sigma^+}, \frac{g^+}{2\sigma^+}, t \right) dt \right)
$$

$$
\times \int_{-\frac{g^-}{2\sigma^+}}^{\frac{g^-}{2\sigma^+}} \int_1^2 p_1 \left( y, -\left( \frac{g^+}{\delta\sigma^-} + \frac{\epsilon}{5\sigma^+\delta} \right), \frac{g^+}{\delta\sigma^-} + \frac{\epsilon}{5\sigma^+\delta}, t \right) dt dq(y)
$$

where $q$ is the (conditional) distribution of $\Delta \tilde{B}^{3,\perp}_{[\tilde{\tau}_1^{(2)}, K+1]} + B_1$ conditioned on

$$
\left\{ \tau^{-\frac{g^-}{2\sigma^+}, \frac{g^-}{2\sigma^+}}_{\Delta \tilde{B}^{3,\perp}_{[\tilde{\tau}_1^{(2)}, K+1]}} \ge 1 \right\}.
$$

Using the definition of $E_1$ together with (2.44) and (2.66), we have $III$ which is uniformly bounded away from 0.

We prove that $IV$ is uniformly bounded away from 0. Using (2.64) and (2.65), we deduce that $E_3^{(1)} \bigcap E_3^{(2)} \subset E_3$ where

$$\begin{aligned} E_3^{(1)} &= \left\{ \sup_{\tilde{\tau}_2^{(2)} \le s \le K+4} \left| \Delta \tilde{B}^{(4)}_{[\tilde{\tau}_2^{(2)}, s]} \right| < \frac{\epsilon \sigma^-}{5\sigma^+ \sigma^{(4)}} \right\} \\ E_3^{(2)} &= \left\{ \forall s \in [\tilde{\tau}_2^{(2)}, K+4]\ \Delta \tilde{B}^{4,\perp}_{[\tilde{\tau}_2^{(2)}, s]} \in [y_3^{(1)}, y_3^{(2)}]\ ,\ \Delta \tilde{B}^{4,\perp}_{[\tilde{\tau}_2^{(2)}, K+4]} \in [y_3^{(3)}, y_3^{(4)}] \right\} \end{aligned}$$

with $y_3^{(1)} = \frac{d_1(K)+2\epsilon/5}{\sigma^{(4)}\left(1-(\rho^{3,4})^2\right)^{1/2}}$, $y_3^{(2)} = \frac{u_1(K)-2\epsilon/5}{\sigma^{(4)}\left(1-(\rho^{3,4})^2\right)^{1/2}}$, $y_3^{(3)} = \frac{u_1(K)-8\epsilon/5}{\sigma^{(4)}\left(1-(\rho^{3,4})^2\right)^{1/2}}$, as well as $y_3^{(4)} = \frac{u_1(K)-7\epsilon/5}{\sigma^{(4)}\left(1-(\rho^{3,4})^2\right)^{1/2}}$. Conditionally on $E_2 \bigcap E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\}$, $E_3^{(1)}$ and $E_3^{(2)}$ are independent. Thus, we deduce

$$IV \ge \mathbb{P}\left( E_3^{(1)} \middle| E_2 \bigcap E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\} \right)$$

$$\mathbb{P}\left( E_3^{(2)} \middle| E_2 \bigcap E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\} \right)$$

Using Markov property of Brownian motions, we obtain that the right part of the inequality conditioned on $\{\tilde{\tau}_2^{(2)} \Big| E_2 \bigcap E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\}\}$ is equal to

$$\left(1 - \int_0^{K+4-\tilde{\tau}_2^{(2)}} p_1\left(0, -\frac{\epsilon\sigma^-}{5\sigma^+\sigma^{(4)}}, \frac{\epsilon\sigma^-}{5\sigma^+\sigma^{(4)}}, t\right) dt\right) \left(1 - \int_0^{K+4-\tilde{\tau}_2^{(2)}} p_1\left(0, y_3^{(1)}, y_3^{(2)}, t\right) dt\right)$$

$$\times \int_{y_3^{(3)}}^{y_3^{(4)}} p_2\left(0, y_3^{(1)}, y_3^{(2)}, K+4-\tilde{\tau}_2^{(2)}, y\right) dy$$

which is uniformly bounded away from 0 using (2.44), (2.63) and (2.66).

We prove that $V > c(x_b - x_a)(u_b - u_a)$. Using (2.61) and (2.62), we deduce that $E_4^{(1)} \bigcap E_4^{(2)} \subset E_4$ where

$$\begin{aligned} E_4^{(1)} &= \left\{ \tilde{\tau} \in [u_a + \tilde{\tau}_2^{(2)}, u_b + \tilde{\tau}_2^{(2)}]\ ,\ \tilde{X}^{(3)}_{\tilde{\tau}} = u_1(K) \right\} \\ E_4^{(2)} &= \left\{ \sup_{K+4 \le s \le \tilde{\tau}} \left| \Delta \tilde{B}^{3,\perp}_{[K+4,s]} \right| < \frac{5g^-}{6\sigma^{(4)}\left(1-(\rho^{3,4})^2\right)^{1/2}}\ ,\ \Delta \tilde{B}^{3,\perp}_{[L+4,\tilde{\tau}]} \in [y_4^{(1)}, y_4^{(2)}] \right\} \end{aligned}$$

$\tilde{\tau} = \inf\{t > K+4 : \tilde{X}^{(3)}_t = u_1(K) \text{ or } \Delta\tilde{X}^{(3)}_{[K+4,t]} = -2\epsilon\}$,

$$y_4^{(1)} = \frac{x_a - \Delta\tilde{X}^{(4)}_{[\tilde{\tau}_2^{(2)},K+4]} - \rho^{3,4}\sigma^{(4)}\left(\sigma^{(3)}\right)^{-1}\left(u_1(K) - \tilde{X}^{(3)}_{K+4}\right)}{\sigma^{(4)}\left(1-(\rho^{3,4})^2\right)^{1/2}},$$

and $y_4^{(2)} = \frac{x_b - \Delta\tilde{X}^{(4)}_{[\tilde{\tau}_2^{(2)},K+4]} - \rho^{3,4}\sigma^{(4)}\left(\sigma^{(3)}\right)^{-1}\left(u_1(K) - \tilde{X}^{(3)}_{K+4}\right)}{\sigma^{(4)}\left(1-(\rho^{3,4})^2\right)^{1/2}}$. We have

$$V = \mathbb{P}\left(\tilde{X}^{(3)}_{\tilde{\tau}} = u_1(K)\right)$$

$$\mathbb{P}\left(E_4^{(1)} \bigcap E_4^{(2)} \middle| E_3 \bigcap E_2 \bigcap E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\} \bigcap \{\tilde{X}^{(3)}_{\tilde{\tau}} = u_1(K)\}\right)$$

The first term on the right part of the equation is uniformly bounded away from 0 (Borodin and Salminen (2002)). Because $\tilde{\tau}$ is a function of $\tilde{X}^{(3)}$ and $\tilde{B}^{3,\perp}$ is independent with $\tilde{X}^{(3)}$, $\tilde{\tau}$ and $\tilde{B}^{3,\perp}$ are independent. Thus the second term on the right conditioned on

$$\{y_4^{(1)}, y_4^{(2)}, X^{(3)}_{K+4}, \tilde{\tau}_2^{(2)} \big| E_3 \bigcap E_2 \bigcap E_1 \bigcap E_0 \bigcap \{\tilde{Z}_0 = (x,u)\}\}$$

can be expressed as :

$$\int_{u_a+\tilde{\tau}_2^{(2)}-(K+4)}^{u_b+\tilde{\tau}_2^{(2)}-(K+4)} \int_{y_4^{(1)}}^{y_4^{(2)}} p_3\left(\frac{X^{(3)}_{K+4}}{\sigma^{(3)}}, \frac{X^{(3)}_{K+4} - 2\epsilon}{\sigma^{(3)}}, \frac{u_1(K)}{\sigma^{(3)}}, t\right) p_2\left(0, -\frac{5g^-}{y_4^{(3)}}, \frac{5g^-}{y_4^{(3)}}, t, y\right) dtdy$$

where $y_4^{(3)} = 6\sigma^{(4)}\left(1-(\rho^{3,4})^2\right)^{1/2}$. We have that $y_4^{(1)}$ and $y_4^{(2)}$ are dominated by $\frac{3g^-}{4\sigma^{(4)}\left(1-(\rho^{3,4})^2\right)^{1/2}}$. Using this together with (2.44), (2.63) and (2.66), we deduce that $V \geq c(x_b - x_a)(u_b - u_a)$.

Second step : We prove that $\left\|\psi^{AV}\right\|_\infty := \sup_{\sigma\in\mathcal{M}, g\in\mathcal{G}, (x,u)\in\mathcal{S}_g} \left|\psi^{AV}(\sigma, g, x, u)\right| < \infty$. To show this :

$$\mathbb{E}\left[\left(\Delta\tilde{X}^{(1)}_{\tilde{\tau}_2^{1C}}\Delta\tilde{X}^{(2)}_{\tilde{\tau}_2^{1C,-,+}}-\tilde{\zeta}^{1,2}\Delta\tilde{\tau}_2^{1C}\right)^2\right]\leq 2\mathbb{E}\left[\left(\Delta\tilde{X}^{(1)}_{\tilde{\tau}_2^{1C}}\Delta\tilde{X}^{(2)}_{\tilde{\tau}_2^{1C,-,+}}\right)^2+\left(\tilde{\zeta}^{1,2}\Delta\tilde{\tau}_2^{1C}\right)^2\right]$$

The second term in the right part of the inequality is uniformly bounded using (2.44) and Lemma 7. Using successively Cauchy-Schwarz and Burholder-Davis-Gundy inequality, (2.44) and Lemma 7, we can also bound uniformly the first term. The other term of (2.21) can be bounded in the same way.

Third step : Define $q=(\sigma,g,x,u)$ and

$$\mathcal{Q}=\{(\sigma,g,x,u)\text{ s.t. }\sigma\in\mathcal{M},g\in\mathcal{G},(x,u)\in\mathcal{S}_g\}.$$

Prove that $\forall q\in\mathcal{Q}$, there exists a measure $\tilde{\pi}(\sigma,g)$ such that

$$\sup_{q\in\mathcal{Q}}\left|\sum_{l=0}^{n-1}\int_{\mathbb{R}^2}\psi^{AV}(\sigma,g,y,v)\,d\tilde{\pi}_l(\sigma,g,x,u)(y,v)-n\int_{\mathbb{R}^2}\psi^{AV}(\sigma,g,y,v)\,d\tilde{\pi}(\sigma,g)(y,v)\right|$$

$$=no_p(1)$$

To show this, we use first step together with $Th.16.0.2$ $(v)$ (Meyn and Tweedie (2009)). We obtain that there exists $\tilde{\pi}(\sigma,g)$ where

$$\left\|P^n(\sigma,g)((x,u),.)-\tilde{\pi}(\sigma,g)\right\|_{TV}\leq 2r^n$$

where $r=1-\nu(\mathbb{R}^2)$. Thus, we deduce :

$$\left|\int_{\mathbb{R}^2}\psi^{AV}(\sigma,g,y,v)\,d\tilde{\pi}_l(\sigma,g,x,u)(y,v)-\int_{\mathbb{R}^2}\psi^{AV}(\sigma,g,y,v)\,d\tilde{\pi}(\sigma,g)(y,v)\right|$$

$$\leq\left\|\psi^{AV}\right\|_\infty\left\|\tilde{\pi}_l(\sigma,g,x,u)-\tilde{\pi}(\sigma,g)\right\|_{TV}\leq 2\left\|\psi^{AV}\right\|_\infty r^l \qquad (2.67)$$

We want to show that $\forall\epsilon > 0$, $\exists N > 0$ such that $\forall n \geq N$ :

$$\left| \sum_{l=0}^{n-1} \int_{\mathbb{R}^2} \psi^{AV}(\sigma, g, y, v)\, d\tilde{\pi}_l(\sigma, g, x, u)(y, v) - n \int_{\mathbb{R}^2} \psi^{AV}(\sigma, g, y, v)\, d\tilde{\pi}(\sigma, g)(y, v) \right|$$

$$< \epsilon n \tag{2.68}$$

The rest is a straightforward analysis exercise. Let $\epsilon > 0$. $\exists N_1 > 0$ such that $r^{N_1} < \frac{\epsilon}{2}$. Choosing $N > 8N_1\epsilon^{-1}\|\psi^{AV}\|_\infty^{-1}$, we first use the triangular inequality, and then split the sum of the left part of (2.68) in two parts, one up to $N_1$ and the other one up to N. We use (2.67) in the second part to obtain (2.68).

Fourth step : Proving the Lemma. Let $w > 0$. From Lemma 9, we just have to show that

$$\alpha_n^2 \sum_{i=1}^{\llcorner w\alpha_n^{-2}h(n)^{-1}\lrcorner} \left| \sum_{j=0}^{h_n-2} \int_{\mathbb{R}^2} \psi^{AV}\left(\sigma_{\tau_{i-1,n}^h}, g_{\tau_{i-1,n}^h}, \alpha_n^{-1}y, \alpha_n^{-2}v\right) d\tilde{\pi}_{i-1,j,n}(y, v) \right.$$

$$\left. -h_n\phi^{AV}\left(\sigma_{\tau_{i-1,n}^h}, g_{\tau_{i-1,n}^h}\right) \right|$$

tends to 0 in probability. Using third step together with standard results on regular conditional distributions (see for instance Breiman (1992)), we prove the lemma.

□

**Lemma 15.**

$$\alpha_n^2 \sum_{i\in A_n} \mathbb{E}_{\tau_{i-1,n}^h}\left[\left(\sigma^{(1)}_{\tau_{i-1,n}^h}\right)^2 \left(\sigma^{(2)}_{\tau_{i-1,n}^h}\right)^2 h_n\phi^{AV}\left(\sigma_{\tau_{i-1,n}^h}, g_{\tau_{i-1,n}^h}\right)\Delta\tau_{i,n}^h\left(\mathbb{E}_{\tau_{i-1}^h}\left[\Delta\tau_{i,n}^h\right]\right)^{-1}\right]$$

$$= \sum_{i\in A_n} \mathbb{E}_{\tau_{i-1,n}^h}\left[\phi^{AV}\left(\sigma_{\tau_{i-1,n}^h}, g_{\tau_{i-1,n}^h}\right)\Delta\tau_{i,n}^h\left(\phi^{\tau}_{\tau_{i-1,n}^h}\right)^{-1}\right] + o_p(1)$$

*Proof.* First step : Defining

$$\begin{cases} u_{i,n} & := & \sum_{j=0}^{h_n-2} \int_X \psi^\tau \left(\sigma_{\tau^h_{i-1,n}}, g_{\tau^h_{i-1,n}}, x, u\right) d\tilde{\pi}_{i-1,j,n}(x,u) \\ A_0 & := & \alpha_n^2 \sum_{i \in A_n} \mathbb{E}_{\tau^h_{i-1,n}} \left[ h_n \phi^{AV} \left(\sigma_{\tau^h_{i-1,n}}, g_{\tau^h_{i-1,n}}\right) \Delta\tau^h_{i,n} \left(\mathbb{E}_{\tau^h_{i-1}} \left[\Delta\tau^h_{i,n}\right]\right)^{-1} \right] \\ A_1 & := & \alpha_n^2 \sum_{i \in A_n} \mathbb{E}_{\tau^h_{i-1,n}} \left[ h_n \phi^{AV} \left(\sigma_{\tau^h_{i-1,n}}, g_{\tau^h_{i-1,n}}\right) \Delta\tau^h_{i,n} \left(u_{i,n}\right)^{-1} \right] \end{cases}$$

we have that $A_0 = A_1 + o_p(1)$. To show this, in light of Lemma 11, we have that

$$\left| \mathbb{E}_{\tau^h_{i-1,n}} \left[\Delta\tau^h_{i,n}\right] - u_{i,n} \right| \leq h(n)\, C_n$$

where $C_n$ tends to 0 in probability. From this, we can easily show that $A_0 = A_1 + o_p(1)$.

Second step : We have that

$$A_1 = \sum_{i \in A_n} \mathbb{E}_{\tau^h_{i-1,n}} \left[ \phi^{AV} \left(\sigma_{\tau^h_{i-1,n}}, g_{\tau^h_{i-1,n}}\right) \Delta\tau^h_{i,n} \left(\phi^\tau_{\tau^h_{i,n}}\right)^{-1} \right] + o_p(1)$$

To prove it, we can mimic the proof of Lemma 14, together with Lemma 11. □

### *2.8.2 Computation of the limits of $\langle M^n \rangle_t$, $\langle M^n, X^{(1)} \rangle_t$ and $\langle M^n, X^{(2)} \rangle_t$*

$$\begin{aligned} \langle M^n \rangle_t &= \sum_{i \in A_n} \mathbb{E}_{\tau^h_{i-1,n}} \left[ \left(\Delta M^n_{\tau^h_{i,n}}\right)^2 \right] + o_p(1) \\ &= \alpha_n^{-2} \sum_{i \in A_n} \mathbb{E}_{\tau^h_{i-1,n}} \left[ \sum_{u=2}^{h_n} \left(N_{(i-1)h_n+u}\right)^2 + 2N_{(i-1)h_n+u} N_{(i-1)h_n+u+1} \right] + o_p(1) \\ &= \alpha_n^2 \sum_{i \in A_n} \sum_{j=0}^{h_n-2} \int_{\mathbb{R}^2} \psi^{AV} \left(\sigma_{\tau^h_{i-1,n}}, g_{\tau^h_{i-1,n}}, \alpha_n^{-1} x, \alpha_n^{-2} u\right) d\tilde{\pi}_{i-1,j,n}(x,u) + o_p(1) \end{aligned}$$

where we used Lemma 2.2.11 of Jacod and Protter (2012) in first equality, Lemma 12 in second equality, Lemma 13 in third equality.

We deduce (using Lemma 14 in first equality and Lemma 15 in third equality)

$$\begin{aligned}
\langle M^n \rangle_t &= \alpha_n^2 \sum_{i \in A_n} h_n \phi^{AV}_{\tau^h_{i-1,n}} + o_p(1) \\
&= \alpha_n^2 \sum_{i \in A_n} \mathbb{E}_{\tau^h_{i-1,n}} \left[ h_n \phi^{AV}_{\tau^h_{i-1,n}} \Delta\tau^h_{i,n} \left( \mathbb{E}_{\tau^h_{i-1}} \left[ \Delta\tau^h_{i,n} \right] \right)^{-1} \right] + o_p(1) \\
&= \sum_{i \in A_n} \mathbb{E}_{\tau^h_{i-1,n}} \left[ \phi^{AV}_{\tau^h_{i-1,n}} \Delta\tau^h_{i,n} \left( \phi^{\tau}_{\tau^h_{i,n}} \right)^{-1} \right] + o_p(1)
\end{aligned}$$

Using Lemma 2.2.11 of Jacod and Protter (2012) again, we deduce :

$$\langle M^n \rangle_t = \sum_{i \in A_n} \phi^{AV}_{\tau^h_{i-1,n}} \Delta\tau^h_{i,n} \left( \phi^{\tau}_{\tau^h_{i,n}} \right)^{-1} + o_p(1)$$

Using Lemma 5 together with Prop. $I.4.44$ (page 51) in Jacod and Shiryaev (2003), we obtain

$$\langle M^n \rangle_t \to \int_0^t \phi^{AV}_s \left( \phi^{\tau}_s \right)^{-1} ds \tag{2.69}$$

Using the same approximations and computations, we also compute :

$$\langle M^n, X^{(1)} \rangle_t \to \int_0^t \phi^{AC1}_s \left( \phi^{\tau}_s \right)^{-1} ds \tag{2.70}$$

$$\langle M^n, X^{(2)} \rangle_t \to \int_0^t \phi^{AC2}_s \left( \phi^{\tau}_s \right)^{-1} ds \tag{2.71}$$

### *2.8.3 Computation of the asymptotic bias and variance*

We follow the idea in 1-dimension in pp. 155-156 of Mykland and Zhang (2012), and define an auxiliary martingale

$$\tilde{M}^n_t = M^n_t - \int_0^t k^{(1)}_s dX^{(1)}_s - \int_0^t k^{1,\perp}_s dX^{1,\perp}_s$$

where $X_t^{1,\perp}$ is defined in (2.31). Using (2.70), we deduce :

$$\begin{aligned}\langle \tilde{M}^n, X^{(1)}\rangle_t &= \langle M^n, X^{(1)}\rangle_t - \int_0^t k_s^{(1)} d\langle X^{(1)}\rangle_s \\ &\xrightarrow{\mathbb{P}} \int_0^t \phi_s^{AC1} \left(\phi_s^{\tau}\right)^{-1} ds - \int_0^t k_s^{(1)} \left(\sigma_s^{(1)}\right)^2 ds\end{aligned}$$

Hence, we choose

$$k_s^{(1)} = \left(\sigma_s^{(1)}\right)^{-2} \phi_s^{AC1} \left(\phi_s^{\tau}\right)^{-1}$$

By the same techniques that we used to compute (2.70), we have that :

$$\langle M^n, \int_0^{\cdot} \rho_s^{1,2} \sigma_s^{(2)} dB_s^{(1)}\rangle_t \to \int_0^t \left(\sigma_s^{(1)}\right)^{-1} \sigma_s^{(2)} \rho_s^{1,2} \phi_s^{AC1} \left(\phi_s^{\tau}\right)^{-1} ds \tag{2.72}$$

Using (2.71) and (2.72) we compute :

$$\begin{aligned}\langle \tilde{M}^n, X^{1,\perp}\rangle_t &= \langle M^n, X^{1,\perp}\rangle_t - \int_0^t k_s^{1,\perp} d\langle X^{1,\perp}\rangle_s \\ &= \langle M^n, X^{(2)} - \int_0^{\cdot} \rho_s \sigma_s^{(2)} dB_s^{(1)}\rangle_t - \int_0^t k_s^{1,\perp} d\langle X^{1,\perp}\rangle_s \\ &= \langle M^n, X^{(2)}\rangle - \langle M^n, \int_0^{\cdot} \rho_s \sigma_s^{(2)} dB_s^{(1)}\rangle_t - \int_0^t k_s^{1,\perp} d\langle X^{1,\perp}\rangle_s \\ &\xrightarrow{\mathbb{P}} \int_0^t \left(\phi_s^{AC2} - \left(\sigma_s^{(1)}\right)^{-1} \sigma_s^{(2)} \rho_s^{1,2} \phi_s^{AC1}\right) \left(\phi_s^{\tau}\right)^{-1} ds \\ &\quad - \int_0^t k_s^{1,\perp} \left(1 - \left(\rho_s^{1,2}\right)^2\right) \left(\sigma_s^{(2)}\right)^2 ds\end{aligned}$$

Hence, we choose

$$k_s^{1,\perp} = \left(1 - \left(\rho_s^{1,2}\right)^2\right)^{-1} \left(\left(\sigma_s^{(2)}\right)^{-2} \phi_s^{AC2} - \left(\sigma_s^{(1)} \sigma_s^{(2)}\right)^{-1} \rho_s^{1,2} \phi_s^{AC1}\right) \left(\phi_s^{\tau}\right)^{-1}$$

By $(A4)$, there exists $S > 0$ such that the $S$ Brownian motions $\{D^{(1)}, ..., D^{(S)}\}$ generate the filtration $(\mathcal{F}_t)_{t \geq 0}$. To show that $\langle \tilde{M}^n, D^{(s)}\rangle_t$ tends to 0 in probability, we decompose $D^{(s)} = D^{s,1} + D^{s,2}$ where $D^{s,1}$ belongs to the space spanned by $\{X^{(1)}, X^{(2)}\}$, $D^{s,2}$ is orthogonal to this space. By what precedes, we have clearly $\langle \tilde{M}^n, D^{s,1}\rangle_t$ tends to 0 in probability. Also, $D^{s,2}$ is a martingale that is, conditionally on the observations times of both processes, independent of $\tilde{M}^n$. Thus we also deduce that $\langle \tilde{M}^n, D^{s,2}\rangle_t$ converges to 0 in probability.

We can now compute :

$$\begin{aligned}
\langle \tilde{M}^n \rangle_t &= \langle M^n - \int_0^{\cdot} k_s^{(1)} dX_s^{(1)} - \int_0^{\cdot} k_s^{1,\perp} dX_s^{1,\perp} \rangle_t \\
&= \langle M^n \rangle_t + \int_0^t \left(\sigma_s^{(1)}\right)^2 \left(k_s^{(1)}\right)^2 ds + \int_0^t \left(\sigma_s^{(2)}\right)^2 \left(1 - \left(\rho_s^{1,2}\right)^2\right) \left(k_s^{1,\perp}\right)^2 ds \\
&- 2\int_0^t k_s^{(1)} d\langle X^{(1)}, M^n\rangle_s - 2\int_0^t k_s^{1,\perp} d\langle X^{1,\perp}, M^n\rangle_s \\
&\xrightarrow{\mathbb{P}} \int_0^t \left(\phi_s^{AV} + 2\left(k_s^{(1)}\left(\sigma_s^{(1)}\right)^{-1} \sigma_s^{(2)} \rho_s^{1,2} \phi_s^{AC1} - \left(k_s^1 + k_s^{1,\perp}\right)\phi_s^{AC2}\right)\right)\left(\phi_s^{\tau}\right)^{-1} \\
&+ \left(\sigma_s^{(1)}\right)^2 \left(k_s^{(1)}\right)^2 + \left(\sigma_s^{(2)}\right)^2 \left(1 - \left(\rho_s^{1,2}\right)^2\right)\left(k_s^{1,\perp}\right)^2 ds
\end{aligned}$$

By letting

$$\begin{aligned}
AV_s &= \left(\phi_s^{AV} + 2\left(k_s^{(1)}\left(\sigma_s^{(1)}\right)^{-1} \sigma_s^{(2)} \rho_s^{1,2} \phi_s^{AC1} - \left(k_s^1 + k_s^{1,\perp}\right)\phi_s^{AC2}\right)\right)\left(\phi_s^{\tau}\right)^{-1} \\
&\quad + \left(\sigma_s^{(1)}\right)^2 \left(k_s^{(1)}\right)^2 + \left(\sigma_s^{(2)}\right)^2 \left(1 - \left(\rho_s^{1,2}\right)^2\right)\left(k_s^{1,\perp}\right)^2
\end{aligned}$$

we deduce using Theorem 2.28 in Mykland and Zhang (2012) that stably in law as $\alpha_n \to 0$, :

$$\alpha_n^{-1}\left(\widehat{RCV}_{t,n} - RCV_t\right) \to \int_0^t k_s^{(1)} dX_s^{(1)} + \int_0^t k_s^{1,\perp} dX_s^{1,\perp} + \int_0^t \left(AV_s\right)^{1/2} d\tilde{W}_s$$

Now, we can express the asymptotic bias $AB_t = \int_0^t k_s^{(1)} dX_s^{(1)} + \int_0^t k_s^{1,\perp} dX_s^{1,\perp}$ differently as

$$\begin{aligned}
AB_t &= \int_0^t k_s^{(1)} dX_s^{(1)} + \int_0^t k_s^{1,\perp} (1 - \left(\rho_s^{1,2}\right)^2)^{1/2} \sigma_s^{(2)} dB_s^{1,\perp} \\
&= \int_0^t k_s^{(1)} dX_s^{(1)} - \int_0^t k_s^{1,\perp} \rho_s^{1,2} \sigma_s^{(2)} dB_s^{(1)} + \int_0^t k_s^{1,\perp} \rho_s^{1,2} \sigma_s^{(2)} dB_s^{(1)} \\
&\quad + \int_0^t k_s^{1,\perp} \left(1 - \left(\rho_s^{1,2}\right)^2\right)^{1/2} \sigma_s^{(2)} dW_s^{1,\perp} \\
&= \int_0^t \left(k_s^{(1)} - k_s^{1,\perp} \rho_s^{1,2} \sigma_s^{(2)} \left(\sigma_s^{(1)}\right)^{-1}\right) dX_s^{(1)} + \int_0^t k_s^{1,\perp} dX_s^{(2)}
\end{aligned}$$

We thus deduce the expression of $AB_s^{(1)}$ and $AB_s^{(2)}$.

# CHAPTER 3
# ESTIMATING THE INTEGRATED PARAMETER IN THE LOCALLY PARAMETRIC MODEL

## 3.1 Introduction

### *3.1.1 Time-varying parameter models*

Modeling dynamics is very important in various fields, such as finance, economics, physics, environmental engineering, geology or sociology. (Semi)parametric time dependent models can deal with one type of dynamics, the temporal evolution of systems. This was extensively studied in Baillie and Bollerslev (1992), Robins and Tsiatis (1992), Fan et al. (2003) and Chen and Fan (2006) among others. By definition, semiparametric approaches come with the strong assumption that an underlying parameter is non time-varying. Au contraire, as time goes by, the structure driving the observations is most likely evolving. Thus, questions about the constancy of the parameter, which would stay the same through thick and thin, are raised. To corroborate this natural skepticism, it can even be the case that empirical work strongly suggests that the non time-varying assumption is too restrictive. To acknowledge the issue, one can for instance build a varying-coefficient model as in Fan and Gijbels (1996), Hastie and Tibshirani (1993) or Fan and Zhang (1999) when regression and generalized regression models are involved, locally stationary processes following the work of Dahlhaus (1997, 2000), Dahlhaus and Rao (2006), or any other time-varying parameter model, e.g. Stock and Watson (1998) and Kim and Nelson (2006).

Models of the variance of the return terms followed exactly this path. Originally, it was assumed that the returns $R_i := Y_i - Y_{i-1}$ of a time-series $\{Y_1, Y_2, \dots, Y_n\}$ observed at regular times $\{\tau_1, \tau_2, \ldots, \tau_n\}$ were conditionally homoskedastic with variance parameter $\sigma$, and they were split as $R_i = \sigma z_i$ where $z_i$ was the stochastic piece, typically a Gaussian with a time-increment variance $\Delta\tau_i := \tau_i - \tau_{i-1}$. However, on the parametric side, Engle (1982) allowed conditional heteroskedasticity, the variance component following a moving-average (MA) model. He was soon followed by Bollerslev (1986) and many other authors like Nelson (1991) and Engle and Ng (1993) who allowed respectively an auto-regressive-moving-average (ARMA) and more general models for the evolution in time of the variance parameter $\sigma$. Taking a nonparametric approach, the analysis of high-frequency financial data gave rise to the Stochastic Volatility model where log prices follow a local martingale and returns are of the form

$$R_i := \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t. \tag{3.1}$$

The object of interest, which used to be the fixed volatility parameter, became the *integrated volatility* (IV), defined as

$$IV_t = \int_0^t \sigma_s^2 ds. \tag{3.2}$$

Time-varying volatility is of substantial importance in modeling for options pricing (see Hull and White (1987), Stein and Stein (1991), Heston (1993) and Ball and Roma (1994)). In order to price using the formula (3.2), one can first estimate it as in Andersen and Bollerslev (1998a,b), Andersen, Bollerslev, Diebold and Labys (2001,2003), Barndorff-Nielsen and Shephard (2001,2002), Barndorff-Nielsen (2004), Jacod and Protter (1998), Zhang (2001) and Mykland and Zhang (2006).

### *3.1.2 The serious statistical implications of assuming that a parameter is non time-varying when it is*

We remind the reader that assuming that $\theta^*$ is constant when it's not can raise serious estimation issues. It means that we are using the estimator with the wrong model. In likelihood theory, ever since Fisher (1922, 1925) introduced the method of maximum likelihood, a significant body of the literature has taken an interest in the asymptotic behavior of the MLE when the model is misspecified. This was pioneered by Berk (1966, 1970) for the Bayesian approach and Huber (1967), who took the classical perspective.

More recently, White (1982), among numerous other authors, also investigated the issue and showed that the *Quasi Maximum Likelihood Estimator* (QMLE) is no longer necessarily consistent with the object of interest, in our case the integrated parameter (1.6). The author showed that the QMLE can converge to a value, but this is not necessarily the one the econometrician has in mind. Also, the estimated standard deviation can be wrong.

Consequently, this has lead to wrong inferences, building confidence intervals of the wrong size, rejecting or accepting hypothese with a probability different from the acceptance rate, setting wrong forecast intervals and so on. This is much of the same problem as the one we face when we are fitting a *general linear model* (GLM) and we encounter over-dispersion (see p.124 of McCullagh and Nelder (1989)). This issue is easily and very often overlooked, even if it seems to be the norm in practice.

## 3.2 The setup

### *3.2.1 Data-Generating Mechanism*

We assume that we infer from the $d_r$-dimensional vectors $\{R_{1,n}, \ldots, R_{N_n,n}\}$, which are functions of the observations where $N_n$ can be random, the observation times are such that $\tau_{0,n} := 0 < \tau_{1,n} < \ldots < \tau_{N_n,n} \leq T$ and $\left[\tau_{i-1,n}, \tau_{i,n}\right]$ is the time block corresponding to $R_{i,n}$. As an example, $R_{i,n}$ can be defined as the (possibly log) *returns* of the original observations. As such, we will refer abusively to $R_{i,n}$ as *returns* in the rest of the chapter. The reader should keep in mind that it is not always the case that $\{R_{1,n}, \ldots, R_{N_n,n}\}$ can be expressed as returns from the original observations. We assume that the returns depend on the underlying parameter process $\theta_t^*$.

### *3.2.2 The parameter*

We will assume that the $p$-dimensional parameter process $\theta_t^* := \left((\theta_t^*)^{(1)}, \ldots, (\theta_t^*)^{(p)}\right)$, which is restricted to lie in $K$, a (not necessarily compact) subset of $\mathbb{R}^p$, is a continuous local martingale of the form

$$\theta_t^* := \int_0^t \sigma_s^\theta dW_s^\theta \tag{3.3}$$

where $\sigma_t^\theta$ is a random nonnegative process (of dimension $p \times p$), and $W_t^\theta$ a standard $p$-dimensional Brownian motion. The parameter $\theta_t^*$ can be for example equal to the volatility, the high-frequency covariance, the betas, etc.

We don't assume any independence between $\theta_t^*$ and the other quantities driving the observations, such as the Brownian motion $W_t$ of the efficient price process. In particular, there can be leverage effect (see e.g. Wang and Mykland (2014), Aït-Sahalia et al. (2014)). Also, the arrival times $\tau_{i,n}$ and the parameter $\theta_t^*$ can be correlated, i.e. there is endogeneity in sampling times.

In the rest of this chapter, for any $u$-dimensional vector $\nu := (\nu^{(1)}, \ldots, \nu^{(u)})$, we will use the notation $\mid \nu \mid := \sum_{k=1}^u |\nu^{(k)}|$ when referring to the Manhattan norm. Also, for any $\theta \in K$, we define the subvector $\theta^+$ of 1-dimensional parameters restricted to be positive. We assume that $\mid \theta_t^* \mid$ is locally bounded and $(\theta_t^*)^+$ is locally bounded away from 0. Furthermore, we assume that the volatility of the parameter, $\sigma_t^\theta$, is locally bounded.

### *3.2.3 Asymptotics*

There are commonly two choices of asymptotics in the literature: the *high-frequency* asymptotics, which makes the number of observations explode on $[0, T]$, and the *low-frequency* asymptotics, which takes $T$ to infinity. We chose the former one. Investigating the low-frequency implementation case is beyond the scope of this chapter[1].

Furthermore, when assuming high-frequency asymptotics, there are several ways to make the number of observations explode. Zhang et al. (2005) considered the fixed noise case. Consequently, the rate of convergence to estimate volatility is no longer $n^{\frac{1}{2}}$, but $n^{\frac{1}{4}}$. In the model with uncertainty zones, Robert and Rosenbaum (2011, 2012) made the *tick size* vanish. The volatility's convergence rate is preserved in this case. We follow the latter idea by scaling (and thus keeping) the structure which drives the observations asymptotically. In particular, the variance of the efficient returns and the variance of the microstructure noise goes to 0 at the same speed $n$.

## 3.3 Estimation

We need first to fix the vocabulary for the rest of this chapter. *Parametric model* will refer to the (non time-varying) parametric model of the econometrician, and not to the time-varying parameter model. Correspondingly, *parametric estimator* stands for the parametric estimator (in the parametric model).

Since the stochastic parameter $\theta_t^*$ is continuous, the parametric model is not too far from the time-varying parameter model locally. We recall that $h_n$ corresponds to the block size. If we let $\tilde{\Theta}_{1,n} := \theta_0^*$ be the initial parameter value and we define on the first block for $j = 1, \cdots, h_n$ the returns approximations $\tilde{R}_{j,n}$ which follows the parametric model with mixture of parameters $\tilde{\Theta}_{1,n}$, then $\tilde{R}_{1,n}^j$ and $R_{j,n}$ are very close to each other

[1]If we set down the asymptotic theory in the same way as in p.3 in Dalhaus (1997), we conjecture that the results of this chapter would stay true.

since the observation times $\tau_{1,n}, \cdots, \tau_{h_n,n}$ are in a small neighborhood of 0. Thus, because true and approximated returns are approximately the same, one can apply the parametric estimator to the observed returns $R_{j,n}$, even though they are not following the parametric model. We thus obtain an estimate $\widehat{\Theta}_{1,n}$ of the initial parameter value $\tilde{\Theta}_{1,n}$. We also define the spot parameter's average on the first block as

$$\Theta_{1,n} := \frac{\int_{\mathrm{T}_{0,n}}^{\mathrm{T}_{1,n}} \theta_s^* ds}{\Delta \mathrm{T}_{1,n}}, \tag{3.4}$$

where the first block inital time $\mathrm{T}_{0,n} := 0$ and final time $\mathrm{T}_{1,n} := \tau_{h_n,n}$. Since the first block length $\Delta \mathrm{T}_{1,n}$ is very small, the value of $\tilde{\Theta}_{1,n}$ is approximately equal to the average of the spot parameter on the first block $\Theta_{1,n}$. Thus, $\widehat{\Theta}_{1,n}$ can be used to estimate $\Theta_{1,n}$.

More generally, for any $i = 2, \cdots, B_n$ we define the spot parameter's average on the $ith$ block as

$$\Theta_{i,n} := \frac{\int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}} \theta_s^* ds}{\Delta \mathrm{T}_{i,n}}, \tag{3.5}$$

where $\mathrm{T}_{i,n} := \min(\tau_{ih_n}, T)$. Let $B_n := \ulcorner N_n h_n^{-1} \urcorner$ be the number of blocks. For any $i = 2, \cdots, B_n$ we estimate in the same way $\Theta_{i,n}$ with initial value $\tilde{\Theta}_{i,n} := \theta^*_{\mathrm{T}_{i-1,n}}$ using the parametric estimator on the $i$th block $\widehat{\Theta}_{i,n}$. Then, we take the weighted sum of $\widehat{\Theta}_{i,n}$ and obtain an estimator of the integrated spot process

$$\widehat{\Theta}_n := \frac{1}{T} \sum_{i=1}^{B_n} \widehat{\Theta}_{i,n} \Delta \mathrm{T}_{i,n}. \tag{3.6}$$

Note that each block includes exactly $h_n$ observations, except for the last one, which might include fewer observations. We call (3.6) the *local parametric estimator* (LPE), since we are are estimating with the parametric estimator on each block.

## 3.4 Outline of the problem

### *3.4.1 A simple model*

We focus on a simple setting in this section. First, we work with $d_r := 2$. Also, we assume that the observations occur at equidistant time intervals $\Delta\tau_n := \frac{T}{n}$, so that $\tau_{i,n} = \frac{i}{n}T$ and thus $R_{i,n}^{(2)} := \Delta\tau_n$. For the rest of this section, we will forget about the second component of the returns $R_{i,n}$, which doesn't provide us any further information, and pretend that $R_{i,n}$ is real-valued. Furthermore, the parametric model is assumed to be very simple. It assumes that there exists a parameter $\theta^* \in K$ such that $R_{i,n}$ are independent and identically distributed (IID) random functions of $\theta^*$. If we introduce $U_{i,n}$ an adequate IID sequence of random variables, we can express the returns as

$$R_{i,n} \quad := \quad F_n\big(U_{i,n}, \theta^*\big), \tag{3.7}$$

where $F_n(x, y)$ is a non-random function. In Equation (3.7), $U_{i,n}$ can be seen as the *random innovation*. Since $\theta_t^*$ can in fact be time-varying, $R_{i,n}$ don't necessarily follow (3.7). A formal time-varying generalization of (3.7) will be given in (3.10). 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 $\theta_t^*$, as we can see in the following two toy examples.

*Example* 7. (estimating volatility) Consider when $\theta_t^* := \sigma_t^2$ (the volatility is thus assumed to follow (3.3)), 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 $i$th observation time and $\sigma$ is the fixed volatility. Under that assumption, the returns are IID. Under the time-varying parameter model, $R_{i,n}$ are clearly not IID, and they are also not conditionally independent given the whole volatility process $\sigma_t^*$ if there is a leverage effect.

*Example* 8. (estimating the rate of a Poisson process) Suppose an econometrician 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 $\lambda$. Because she doesn't 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 $Y_{i,n} = N^-_{\tau_{i,n}} - N_{\tau_{i-1,n}}$. This is an example where the series is already stationary under the parametric model assumption, and thus $Y_{i,n} = R_{i,n}$. If the econometrician's assumption of homogeneity is true, the returns are IID. In case of heterogeneity, the parameter rate $\lambda_t$ will be assumed to follow (3.3), $N_t$ will be a nonhomogenous Poisson process, and the returns $R_{i,n}$ will be neither identically distributed nor independent.

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

We are now providing a generalization of the parametric model (3.7) as well as an expression of the approximated returns. To deal with the former, we assume that in general

$$R_{i,n} \quad := \quad F_n\big(U_{i,n}, \{\theta^*_s\}_{\tau_{i-1,n} \leq s \leq \tau_{i,n}}\big). \tag{3.8}$$

The time-varying parameter model in (3.8) is a natural extension of the parametric model (3.7) because the return $R_{i,n}$ can depend on the parameter process path from the previous sampling time $\tau_{i-1,n}$ and up to the current sampling time $\tau_{i,n}$. As $R_{i,n}$

depends on the whole parameter path, it seems natural to allow $U_{i,n}$ to be itself a process path. For example, when assuming that the parameter is the volatility process, we will assume that $U_{i,n}$ is equal to the underlying Brownian motion $W_t$ path (see Example 9 for more details). Also, as $U_{i,n}$ is the random innovation, it should be independent of the parameter process path, 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 (3.8) is

$$R_{i,n} \quad := \quad F_n\big(U_{i,n}, \theta^*_{\tau_{i-1,n}}\big), \tag{3.9}$$

i.e. we fix the path parameter equal to its initial value. The time-varying friction parameter of the model with uncertainty zones is defined as a mix of (3.8) and (3.9) of the model with uncertainty zones (see Section 3.7.1 for details). Finally, the approximated returns $\tilde{R}_{i,n}$ follow a mixture of the parametric model (3.7) with initial block parameter value. We are now providing a formal definition of our intuition. We assume that

$$R^j_{i,n} \quad := \quad F_n\big(U^j_{i,n}, \{\theta^*_s\}_{\tau^{j-1}_{i,n} \leq s \leq \tau^j_{i,n}}\big) \tag{3.10}$$

$$\tilde{R}^j_{i,n} \quad := \quad F_n\big(U^j_{i,n}, \tilde{\Theta}_{i,n}\big) \tag{3.11}$$

where the random quantity $U^j_{i,n}$ takes values on a space $\mathcal{U}_n$ that can be functional[2] and that can depend on $n$, $U^j_{i,n}$ are IID for a fixed $n$ but the distribution can depend on $n$, and $F_n(x,y)$ is a non-random function[3]. Note that (3.10) is a mere expression of

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

[3] Let $\mathcal{C}(\mathbb{R}^+)$ be the space of continuous paths parametrized by time $t \in \mathbb{R}^+$, which is a Borel space. Consequently, $\mathcal{U}_n \times \mathcal{C}(\mathbb{R}^+)$ is also a Borel space. We assume that $F_n(x,y)$ is jointly measurable real-valued function on $\mathcal{U}_n \times \mathcal{C}(\mathbb{R}^+)$. Note that the advised reader will have seen that a priori $\{\theta^*_s\}_{\tau^{j-1}_{i,n} \leq s \leq \tau^j_{i,n}}$ is defined on $\mathcal{C}[0, \tau_n]$ (after according translation) in (3.10) and $\tilde{\Theta}_{i,n}$ is a vector in (3.11), whereas they should be defined on the space $\mathcal{C}(\mathbb{R}^+)$ according to the definition. We match the definitions by extending them as continuous paths on $\mathbb{R}^+$. Formally, if $\theta_t \in \mathcal{C}[0, \tau_n]$, 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 $t \geq 0$.

(3.8) using a different notation. For any block $i = 1, \ldots, B_n$ and for any observation time $j = 0, \ldots, h_n$ of the $ith$ block, we define $\mathcal{I}^j_{i,n}$[4] the information up to time $\tau^j_{i,n}$. The crucial assumption is that $U^j_{i,n}$ has to be independent of the past information[5] (and in particular of $\tilde{\Theta}_{i,n}$). $U^j_{i,n}$ can be seen as the "random innovation" between $\tau^{j-1}_{i,n}$ and $\tau^j_{i,n}$. Note that we don't assume any independence between the "random innovation" $U^j_{i,n}$ and the parameter process $\{\theta^*_s\}_{\tau^{j-1}_{i,n} \leq s \leq \tau^j_{i,n}}$. We provide directly the definitions of $F_n$ and $U^j_{i,n}$ in the two toy examples.

*Example* 9. (estimating volatility) In this case, $\mathcal{U}_n$ is defined as the space $\mathcal{C}[0, \Delta\tau_n]$ of continuous paths parametrized by time $t \in [0, \tau_n]$, $U^j_{i,n} := \{\Delta W_{[\tau^{j-1}_{i,n}, s]}\}_{\tau^{j-1}_{i,n} \leq s \leq \tau^j_{i,n}}$ are the Brownian motion increment path processes between two consecutive observation times. If we assume that $(W^\theta_t, W_t)$ is jointly a standard 2-dimensional Brownian motion, then the random innovation $U^j_{i,n}$ is indeed independent of the past in view of the Markov property of Brownian motions. We also define $F_n(u_t, \theta_t) := \int_0^{\tau_n} \theta_s du_s$. We thus obtain that the returns are defined as $R^j_{i,n} := \int_{\tau^{j-1}_{i,n}}^{\tau^j_{i,n}} \sigma^*_s dW_s$ and the approximated returns $\tilde{R}^j_{i,n} := \sigma^*_{\tau^0_{i,n}} \Delta W_{[\tau^{j-1}_{i,n}, \tau^j_{i,n}]}$ are the same quantity when holding the volatility constant on the block.

*Example* 10. (estimating the rate of a Poisson process) We assume that the rate of the inhomogeneous Poisson process is $\beta_n \lambda_t$, where $\beta_n$ is a non time-varying and non-random quantity such that $\beta_n \Delta\tau_n := 1$. 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 $\mathbb{R}^+$. $U^j_{i,n}$ can be defined as standard Poisson processes $\{N^{i,j,n}_t\}_{t \geq 0}$, independent of each other. We also have $F_n(u_t, \theta_t) := u_{\int_0^{\tau_n} \beta_n \theta_s du_s}$. Thus, if we let

---

[4]In this chapter, we will be using the term *information* to refer to the mathematical object of *filtration*. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Define the *sorted information* $\{\mathcal{I}_{k,n}\}_{k \geq 0}$ such that for any non-negative integer $k$ that we can decompose as $k = (i-1)h_n + j$ where $i \in \{1, \ldots, B_n\}$ and $j \in \{0, \ldots, h_n\}$, $\mathcal{I}_{k,n} := \mathcal{I}^j_{i,n}$. We assume that $\mathcal{I}_{k,n}$ is a (discrete-time) filtration on $(\Omega, \mathcal{F}, P)$. In addition, we assume that $\{\theta^*_s\}_{0 \leq s \leq \tau^j_{i,n}}$ and $U^j_{i,n}$ are $\mathcal{I}^j_{i,n}$-measurable.

[5]past information means up to time $\tau^{j-1}_{i,n}$

$t^j_{i,n} := \int_{\tau^{j-1}_{i,n}}^{\tau^j_{i,n}} \beta_n \lambda^*_s ds$, the returns are the time-changed Poisson processes

$$R^j_{i,n} = N^{i,j,n}_{t^j_{i,n}}, \tag{3.12}$$
$$\tilde{R}^j_{i,n} = N^{i,j,n}_{\beta_n \Delta\tau^j_{i,n} \lambda^*_{\tau^0_{i,n}}}. \tag{3.13}$$

### 3.4.2 Consistency

In the following of this section and Section 3.6, we will make the block size $h_n$ go to infinity

$$h_n \to \infty. \tag{3.14}$$

Furthermore, we will make the block length $\Delta \mathrm{T}_{i,n}$ vanish asymptotically. Because we assumed observations occur at equidistant time, this can be expressed as

$$h_n n^{-1} \to 0. \tag{3.15}$$

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

$$\sum_{i=1}^{B_n} \big(\widehat{\Theta}_{i,n} - \Theta_{i,n}\big)\Delta \mathrm{T}_{i,n} \xrightarrow{\mathbb{P}} 0. \tag{3.16}$$

where the formal definition of $\widehat{\Theta}_{i,n}$ can be found in (3.22). In order to show (3.16), 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

$$\widehat{\Theta}_{i,n} - \Theta_{i,n} = \big(\widehat{\Theta}_{i,n} - \widehat{\tilde{\Theta}}_{i,n}\big) + \big(\widehat{\tilde{\Theta}}_{i,n} - \tilde{\Theta}_{i,n}\big) + \big(\tilde{\Theta}_{i,n} - \Theta_{i,n}\big), \tag{3.17}$$

where $\widehat{\tilde{\Theta}}_{i,n}$, which is defined formally in (3.23), is the parametric estimator used on the underlying non-observed approximated returns. It is not a feasible estimator and appears in (3.17) only to shed light on the way we can obtain the consistency of the estimator. We first deal with the last error term of (3.17), which is due to the non-constancy of the spot parameter $\theta_t^*$. Note that

$$\sum_{i=1}^{B_n} \big(\tilde{\Theta}_{i,n} - \Theta_{i,n}\big)\Delta \mathrm{T}_{i,n} = \sum_{i=1}^{B_n} \Big(\theta^*_{\mathrm{T}_{i-1,n}} \Delta \mathrm{T}_{i,n} - \int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}} \theta_s^* ds\Big) \tag{3.18}$$

and thus we deduce from Riemann-approximation[6] that

$$\sum_{i=1}^{B_n} \big(\tilde{\Theta}_{i,n} - \Theta_{i,n}\big)\Delta \mathrm{T}_{i,n} \xrightarrow{\mathbb{P}} 0 \tag{3.19}$$

To deal with other terms of (3.17), we make another key asymptotics assumption for this section. We further assume that there exists a sequence $\alpha_n$ with for all $\theta \in K$

$$\alpha_n F_1(U_{i,1}^j, \theta) \overset{\mathcal{D}}{=} F_n(U_{i,n}^j, \theta) \tag{3.20}$$

The equality (3.20) means that we scale and keep the structure that drives the next return asymptotically. This is a relatively strong assumption, and in particular it prevents us from allowing different returns distribution shape when $n$ is varying, except for the scaling constant. We provide examples of $\alpha_n$.

*Example* 11. (estimating volatility) Because we assume that $F_n(U_{i,n}^j, \theta)$ are normally distributed with null-mean and variance $\theta^2 \Delta\tau_{i,n}$ and since the normal distribution is scale invariant, we have (3.20) with $\alpha_n = n^{-\frac{1}{2}}$.

*Example* 12. (estimating the rate of a Poisson process) (3.20) is satisfied with $\alpha_n = 1$.

Furthermore, we assume that for any positive integer $k$, the practitioner has at

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

hand an estimator $\hat{\theta}_{k,n} := \hat{\theta}_{k,n}(r_{1,n};\ \ldots;\ r_{k,n})$, which depends on the input of returns $\{r_{1,n}; \ldots; r_{k,n}\}$, and also asymptotically of $n$ in the following sense

$$\hat{\theta}_{k,n}(r_{1,n}; \ldots; r_{k,n}) = \hat{\theta}_{k,1}(\alpha_n^{-1} r_{1,n}; \ldots; \alpha_n^{-1} r_{k,n}). \tag{3.21}$$

This makes sense to use the same estimator on a scaled version (which in practice depends of the sampling frequency) of the returns in light of (3.20). In other words, the estimator $\hat{\theta}_{k,n}$ is scale invariant. We provide the estimators in the couple of toy examples.

*Example* 13. (estimating volatility) The estimator is the scaled usual RV, i.e. $\hat{\theta}_{k,n}(r_{1,n}; \ldots; r_{k,n}) := T^{-1}k^{-1}n\sum_{j=1}^{k} r_{j,n}^2$. Note that $\hat{\theta}_{k,n}$ can also be asymptotically seen as the MLE (see the discussion pp. 112-115 of Mykland and Zhang (2012)). We can verify (3.21) easily.

*Example* 14. (estimating the rate of a Poisson process) The estimator to be used is the return mean $\hat{\theta}_{k,n}(r_{1,n}; \ldots; r_{k,n}) := k^{-1}\sum_{j=1}^{k} r_{j,n}$. It is straightforward to see (3.21).

On each block $i = 1, \ldots, B_n$ we estimate the local parameter as

$$\widehat{\Theta}_{i,n} := \hat{\theta}_{h_n,n}(R_{i,n}^1; \ldots; R_{i,n}^{h_n}). \tag{3.22}$$

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

$$\widehat{\tilde{\Theta}}_{i,n} := \hat{\theta}_{h_n,n}(\tilde{R}_{i,n}^1; \ldots; \tilde{R}_{i,n}^{h_n}). \tag{3.23}$$

Note that (3.23) is infeasible because the approximated returns $\tilde{R}_{i,n}^j$ are non-observable quantities. For any $M > 0$, we define $K_M := \{\theta \in K :\ \mid \theta \mid \leq M \text{ and } \theta^+ \geq M^{-1}\}$ a subset of $K$. In order to obtain the consistency of (3.6), we make the assumption that the parametric estimator is $\mathbf{L}^1$-convergent, locally uniformly in the model parameter $\theta$ if we actually observes returns coming from the parametric model. This can be expressed in the following condition.

**Condition (C1).** For any $M > 0$,

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

*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.

Under Condition (C1), standard results on regular conditional distributions[7] give us that the error made on the estimation of the underlying non-observed returns tends to 0, i.e.

$$\sum_{i=1}^{B_n} \big(\widehat{\tilde{\Theta}}_{i,n} - \tilde{\Theta}_{i,n}\big)\Delta \mathrm{T}_{i,n} \xrightarrow{\mathbb{P}} 0. \tag{3.24}$$

To deal with the first term of (3.17), we need to make sure that we can control the discrepancy between the estimate made on the observed returns and the estimate made on the underlying approximated returns, uniformly in the initial parameter and in the future path of the parameter process. This can be expressed as the new following assumption. For any $M > 0$, we define $\mathcal{E}_M$ the space of all null-drift continuous $p$-dimensional Itô-process $\theta_t$ such that the initial value $\theta_0$ is non-random, $\theta_t \in K_M$ for all $0 \leq t \leq T$, the volatility of $\theta_t$ is bounded by $M$ for all $0 \leq t \leq T$, and for any $j = 1, \cdots, h_n$ we have $U_{1,n}^j$ independent of the past path $\{\theta_s\}_{0 \leq s \leq \tau_{1,n}^{j-1}}$.

**Condition (C2).** We have

$$\sup_{\theta_t \in \mathcal{E}_M} \mathbb{E}\Bigg[\Bigg|\hat{\theta}_{h_n,n}\big(F_n(U_{1,n}^1,\theta_0),\ldots,F_n(U_{1,n}^{h_n},\theta_0)\big)$$

---

[7]see for instance Leo Breiman (1992), see Appendix for more details.

$$-\hat{\theta}_{h_n,n}\Big(F_n(U^1_{1,n},\{\theta_s\}_{0\leq s\leq \tau^1_{1,n}}),\dots,F_n(U^{h_n}_{1,n},\{\theta_s\}_{\tau^{h_n-1}_{1,n}\leq s\leq \tau^{h_n}_{1,n}})\Big)\Big|\Big] \to 0.$$

Condition (C2) implies[8] that the error due to the local model approximation vanishes in the limit, i.e.

$$\sum_{i=1}^{B_n}(\widehat{\Theta}_{i,n}-\widehat{\tilde{\Theta}}_{i,n})\Delta \mathrm{T}_{i,n} \xrightarrow{\mathbb{P}} 0. \tag{3.25}$$

We can now summarize the theorem on consistency in this very simple case where observations occur at equidistant time intervals and returns are IID under the parametric model.

**Theorem (Consistency).** *Under Condition (C1) and Condition (C2), we have the consistency of (3.6), i.e.*

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

We obtain the consistency in the couple of toy examples[9].

*Remark* 10. (adding a drift) In Example 7, by Girsanov's theorem, together with local arguments (see, e.g., pp.158-161 in Mykland and Zhang (2012)), we can weaken the price and volatility local-martingale assumption by allowing them to follow an Itô-process (of dimension 2), with volatility matrix locally bounded and locally bounded away from 0, and drift locally bounded.

*Remark* 11. (LPE equal to the parametric estimator) The advised reader will have noticed that in the couple of examples, the parametric estimator is equal to the LPE. This is because in those very basic examples, the parametric estimator is linear, i.e. for any positive integer $k$ and $l = 1, \dots, k-1$

$$\hat{\theta}_{k,n}(r_{1,n};\dots;r_{k,n}) = \frac{l}{k}\hat{\theta}_{l,n}(r_{1,n};\dots;r_{l,n}) + \frac{k-l}{k}\hat{\theta}_{k-l,n}(r_{l+1,n};\dots;r_{k,n}) \tag{3.26}$$

[8] see Appendix for more details

[9] see Appendix for proofs

In more general examples of Section 3.7, (3.26) will break, and we will obtain two distinct estimators.

### *3.4.3 Challenges*

There are three empirical reasons why the presentation above is too simple. In practice, the observed returns can be *autocorrelated*, *noisy* and there can be *endogeneity* in sampling times. Accordingly, we will build the general LPM in Section 3.5. Also, we will investigate the limit distribution in Section 3.6.

## 3.5 The LPM

Let $m$ be a nonnegative integer (which can be infinite) which stands for the order of memory in the model. We assume that under the parametric model, $R_{i,n}$ is an homogeneous partially observed ($m$-dependent) Markov chain of order m, i.e. there exists a $d_q$-dimensional vector $Q_{i,n}$ such that $(Q_{i,n}, R_{i,n})$ is an homogeneous Markov chain of order $m$. Under the time-varying parameter model, it is not necessarily even a nonhomogeneous Markov chain. Nonetheless, we will see in (3.28) that it is almost a nonhomogeneous Markov chain (of order $m$), except for the evolution of the parameter $\theta_t^*$ part which is not necessarily Markovian. In particular, if we assume that the volatility $\sigma_t^\theta$ of the parameter $\theta_t^*$ is not time-varying, $(Q_{i,n}, R_{i,n})$ is a nonhomogeneous Markov chain (of order $m$). In the following, we will use the expression "Markov chain" but the reader should understand "Markov chain under the parametric model" or even "locally Markov chain". We also drop the part "of order $m$".

The interpretation of $Q_{i,n}$ is straightforward: it can be seen as the quantities we don't observe. In particular, it can include the efficient price return, whereas $R_{i,n}$ would include the noisy version of the efficient return. The reader should look at Section 3.7 to see how we identify $Q_{i,n}$ and $R_{i,n}$ to the quantities of different models.

Note that the simple model introduced in Section 3.4.1 is a particular case with $m = 0$. If we assume that we have regular observation times and that the microstructure noise is uncorrelated with the efficient price and IID normally distributed, the parametric model of the observed returns is an MA(1) (see Aït-Sahalia et al. (2005) and Xiu (2010)). Thus, we have $m = 1$ in that example. The details are to be found in Section 3.7.3.

We introduce the notations $d := d_q + d_r$ as well as $M_{i,n} := (Q_{i,n}, R_{i,n})$ for the Markov chain quantities, and assume that $M_{i,n}$ takes values on the space $\mathcal{M}_n$, which is a subset of $\mathbb{R}^d$. Also, we define the $m$ initial values $M_{-(m-1),n}, \dots, M_{0,n}$ of the Markov chain. Finally, we introduce the $m$-dimensional "memory" vector of Markov chain quantities $\mathbf{M}_{i,n} := (M_{i,n}, \dots, M_{i-(m-1),n})$, which takes values on a space $\mathcal{M}_{m,n}$ (subset of $\mathcal{M}_n^m$). The parametric model can be expressed as

$$M_{i,n} := F_n(\mathbf{M}_{i-1,n}, U_{i,n}, \theta^*), \tag{3.27}$$

where $F_n(x, y, z)$ is a $\mathbb{R}^d$-valued non-random function[10], $U_{i,n}$ are IID for a fixed $n$ but the distribution can depend on $n$. The parametric model (3.27) can be compared to the simple parametric model (3.7): the only change is that we allow past-correlation in the model. In analogy with (3.8), we assume that the time-varying parameter model can be expressed as

$$M_{i,n} = F_n(\mathbf{M}_{i-1,n}, U_{i,n}, \{\theta_s^*\}_{\tau_{i-1,n} \leq s \leq \tau_{i,n}}). \tag{3.28}$$

[10] We assume that for $i$ any positive integer, $U_{i,n} \in \mathcal{U}_n$ where $\mathcal{U}_n$ is a Borel space and that $F_n(x, y, z)$ is defined on $\mathcal{M}_{m,n} \times \mathcal{U}_n \times \mathcal{X}_n$. Additionally, we assume that $F_n(x, y, z)$ is a jointly measurable $\mathbb{R}^d$-valued function such that for any $(\mathbf{M}_n, U_n, \chi_n) \in \mathcal{M}_{m,n} \times \mathcal{U}_n \times \mathcal{X}_n$, we have $\mathbb{E} \mid F_n(\mathbf{M}_n, U_n, \chi_n) \mid < \infty$

## 3.6 Asymptotic properties

In analogy with (3.20) of Section 3.4.2, we keep asymptotically the structure of the returns by scaling the distribution of $M_{i,n}$. Formally, we need to introduce some definitions. For any $\mathbf{M} \in \mathcal{M}_{m,1}$, we can write

$$\mathbf{M} := \begin{pmatrix} M^{1,1} & \dots & M^{1,m} \\ \vdots & \ddots & \vdots \\ M^{d,1} & \dots & M^{d,m} \end{pmatrix}.$$

For any d-dimensional vector $\alpha := (\alpha^{(1)}, \dots, \alpha^{(d)})$, we let

$$\alpha * \mathbf{M} := \begin{pmatrix} \alpha^{(1)} M^{1,1} & \dots & \alpha^{(1)} M^{1,m} \\ \vdots & \ddots & \vdots \\ \alpha^{(d)} M^{d,1} & \dots & \alpha^{(d)} M^{d,m} \end{pmatrix}.$$

Also, if $\alpha$ and $\beta$ are $u$-dimensional vectors, we define $\alpha * \beta := (\alpha^{(1)}\beta^{(1)}, \dots, \alpha^{(u)}\beta^{(u)})$. Finally, we consider $\alpha_n * \mathcal{M}_{m,1} := \{\alpha_n * \mathbf{M} \text{ s.t. } \mathbf{M} \in \mathcal{M}_{m,1}\}$. We assume that there exists a $d$-dimensional $\alpha_n$ such that $\mathcal{M}_{m,n} := \alpha_n * \mathcal{M}_{m,1}$ and we have

$$\alpha_n * F_1(\mathbf{M}, U_{1,1}, \theta) \stackrel{\mathcal{D}}{=} F_n(\alpha_n * \mathbf{M}, U_{1,n}, \theta). \tag{3.29}$$

We insist on the fact that $\alpha_n$ has the same signification as in (3.20) (the returns are scale-invariant) and that the notations are more involved only because the model is past-correlated. The reader should go to Section 3.7 to see how $\alpha_n$ is defined on different models. Equation (3.29) is a key assumption for the proofs.

We investigate in the following the limit distribution. Formally, for a $l > 0$ (with corresponding rate of convergence $n^{\frac{1}{l}}$), we aim to find the limit ditribution of

$$n^{\frac{1}{l}} \sum_{i=1}^{B_n} \big(\widehat{\Theta}_{i,n} - \Theta_{i,n}\big)\Delta \mathrm{T}_{i,n}. \tag{3.30}$$

Specifically, we want to show that (3.30) converges *stably* to a limit distribution. We remind the reader that we defined stable convergence in Chapter 2.4.1. Since the stable convergence needs a corresponding information $\mathcal{J}$ to be defined with, we need to be more specific about how to obtain $\mathcal{J}$. We will be needing the following technical assumption, which turns out to be easily verified on examples of Section 3.7. The idea goes back to Heath (1977). We define $\mathcal{I}_{i,n}$[11] the information up to time $\tau_{i,n}$.

**Condition (E0).** $\mathcal{I}_{i,n}$ can be extended into $\mathcal{J}_{i,n}$[12], where $\mathcal{J}_{i,n}$ is the interpolated information of a continuous information $\mathcal{J}_t^{(c)}$, i.e. $\mathcal{J}_{i,n} = \mathcal{J}_{\tau_{i,n}}^{(c)}$

In the following of Chapter 3, when using the conditional expectation $\mathbb{E}_\tau\big[Z\big]$[13], we will refer to the conditional expecation of $Z$ knowing $\mathcal{J}_\tau^{(c)}$. Finally, we consider $\mathcal{J} := \mathcal{J}_T^{(c)}$ the information to go with stable convergence.

We are now more specific about the form of the parametric estimator. As in Section 3.4.2, the parametric estimator will include the returns $\{r_{1,n}, \dots, r_{k,n}\}$ as inputs. Moreover, because the Markov chain is past-correlated, we allow the parametric estimator to possibly depend on the $m$-dimensional vector of initial returns $\mathbf{r}_{0,n}$. To sum up, the parametric estimator takes the following form

$$\hat{\theta}_{k,n} := \hat{\theta}_{k,n}(r_{1,n}; \dots; r_{k,n}; \mathbf{r}_{0,n}). \tag{3.31}$$

[11] We assume that $\mathcal{I}_{i,n}$ is a (discrete-time) filtration of $(\Omega, \mathcal{F}, P)$ such that $\{\theta_s^*\}_{0 \leq s \leq \tau_{i,n}}$ and $U_{i,n}$ are adapted to $\mathcal{I}_{i,n}$. Also, we assume that the initial $m$-dimensional vector $\mathbf{M}_{0,n}$ is $\mathcal{I}_{0,n}$-measurable.

[12] It means that $J_{i,n}$ is a discrete filtration and for any $i$ nonnegative integer $\mathcal{I}_{i,n} \subset \mathcal{J}_{i,n}$

[13] $\tau$ has to be a $\mathcal{J}_t^{(c)}$-stopping time

We keep the same asymptotic property as in (3.21) and we adapt it to the multi-dimensional case. We need some notations for this purpose. Define the part $\alpha_{R,n} := (\alpha_n^{(d_q+1)}, \dots, \alpha_n^{(d)})$ related to the observed returns $R_{i,n}$ of $\alpha_n$. Let $\boldsymbol{\alpha}_{R,n} := (\alpha_{R,n}, \dots, \alpha_{R,n})$ consisting of m $\alpha_{S,n}$ appened together. Also, for any matrix $\boldsymbol{\alpha}$ with non-zero entries, we define $\boldsymbol{\alpha}^{-1}$ the matrix of same dimension with each entry equal to the inverse of the same entry of $\boldsymbol{\alpha}$. In analogy with (3.21) in Section 3.4.2, we assume that

$$\hat{\theta}_{k,n}(r_{1,n}; \dots; r_{k,n}; \mathbf{r}_{0,n}) = \hat{\theta}_{k,1}(\alpha_{R,n}^{-1} r_{1,n}; \dots; \alpha_{R,n}^{-1} r_{k,n}; \boldsymbol{\alpha}_{R,n}^{-1} \mathbf{r}_{0,n}) \tag{3.32}$$

Let $i$ be any positive integer. In analogy with the "block notations" of Section 3.4.1, we define the Markov chain elements on the $i$th block $M_{i,n}^j := M_{(i-1)h_n+j,n}$ for $j = 1, \dots, h_n$. We also define the initial vector of the $i$th block as $(M_{i,n}^0, \dots, M_{i,n}^{-(m-1)}) := \mathbf{M}_{(i-1)h_n+j,n}$. For $\mathbf{M} \in \mathcal{M}_{m,n}$ and $j = -(m-1), \dots, h_n$ we let $\tilde{M}_{i,n}^{j,\mathbf{M}}$ be the approximations of the Markov chain on the $i$th block with starting vector $\mathbf{M}$. As such, we first define the initial vector of approximation as the vector $\mathbf{M}$ itself, i.e. $(\tilde{M}_{i,n}^{0,\mathbf{M}}, \dots, \tilde{M}_{i,n}^{-(m-1),\mathbf{M}}) := \mathbf{M}$. Second, we define the $m$-dimensional "memory" vector of approximation as $\tilde{\mathbf{M}}_{i,n}^{j,\mathbf{M}} := (\tilde{M}_{i,n}^{j,\mathbf{M}}, \dots, \tilde{M}_{i,n}^{j-(m-1),\mathbf{M}})$. Thirdly, for any positive integer $j$, we define the approximated returns by the recurrence relation similar to (3.28)

$$\tilde{M}_{i,n}^{j,\mathbf{M}} := F_n(\tilde{\mathbf{M}}_{i,n}^{j-1,\mathbf{M}}, U_{i,n}^j, \tilde{\Theta}_{i,n}). \tag{3.33}$$

The approximated returns $\tilde{M}_{i,n}^{j,\mathbf{M}}$ follow a mixture of the parametric model with parameter $\tilde{\Theta}_{i,n}$. Finally, the infeasible estimator on the $i$th block $\hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}}$ with initial vector $\mathbf{M} \in \mathcal{M}_{m,n}$ is defined as

$$\hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}} := \hat{\theta}_{h_n,n}(\tilde{R}_{i,n}^{1,\mathbf{M}}; \dots; \tilde{R}_{i,n}^{h_n,\mathbf{M}}; \tilde{\mathbf{R}}_{i,n}^{0,\mathbf{M}}). \tag{3.34}$$

The spot parameter estimator on the $i$th block $\widehat{\Theta}_{i,n}$ is defined with the same parametric estimator, but with observed returns as input

$$\widehat{\Theta}_{i,n} := \hat{\theta}_{h_n,n}(R^1_{i,n}; \dots ; R^{h_n}_{i,n}; \mathbf{R}^0_{i,n}). \tag{3.35}$$

Let $\mathbf{M}^*_1 \in \mathcal{M}_{m,1}$ and $\mathbf{M}^*_n := \alpha_n * \mathbf{M}^*_1$. We can decompose $\big(\widehat{\Theta}_{i,n} - \Theta_{i,n}\big)$ as

$$\big(\hat{\Theta}_{i,n} - \hat{\tilde{\Theta}}^{\mathbf{M}^0_{i,n}}_{i,n}\big) + \big(\hat{\tilde{\Theta}}^{\mathbf{M}^0_{i,n}}_{i,n} - \hat{\tilde{\Theta}}^{\mathbf{M}^*_n}_{i,n}\big) + \big(\hat{\tilde{\Theta}}^{\mathbf{M}^*_n}_{i,n} - \tilde{\Theta}_{i,n}\big) + \big(\tilde{\Theta}_{i,n} - \Theta_{i,n}\big) \tag{3.36}$$

where the first term is the error in estimation due to the use of the approximated model (3.33) instead of the time-varying parameter model (3.28), the second term is the error made when taking $\mathbf{M}^*_n$ instead of $\mathbf{M}^0_{i,n}$ as initial value of the block in the mixture of the parametric model (3.33), the third term corresponds to the error of the estimation of the constant parameter by the underlying approximations starting with a fixed initial value $\mathbf{M}^*_n$ and the last term is the error made by holding the process parameter constant on each block. Note that $\hat{\tilde{\Theta}}^{\mathbf{M}^0_{i,n}}_{i,n}$ is a mixture of the parametric model with parameter $\tilde{\Theta}_{i,n}$ and a mixture of starting value $\mathbf{M}^0_{i,n}$.

It is instructive to consider (3.36) when we assume that the time-varying parameter model is equal to the parametric model. In that simple case, the first term and the fourth term are equal to 0. Additionally, we can hope that under right conditions

$$n^{\frac{1}{l}} \sum_{i=1}^{B_n} \big(\hat{\tilde{\Theta}}^{\mathbf{M}^0_{i,n}}_{i,n} - \hat{\tilde{\Theta}}^{\mathbf{M}^*_n}_{i,n}\big)\Delta \mathrm{T}_{i,n} \approx 0.$$

Finally, if we assume that we know the convergence rate $n^{\frac{1}{2}}$ and the limit distribution $\mathcal{N}(0, V_{\theta^*})$ of the parametric estimator, for any $i = 1, \dots, B_n$ we have that $h_n^{\frac{1}{2}}\big(\hat{\Theta}_{i,n} - \theta^*\big)\Delta \mathrm{T}_{i,n} \approx \mathcal{N}(0, V_{\theta^*})\Delta \mathrm{T}_{i,n}$ and thus we can hope that

$$n^{\frac{1}{2}} \sum_{i=1}^{B_n} \big(\hat{\Theta}_{i,n} - \theta^*\big)\Delta \mathrm{T}_{i,n} \approx T^{-1}\mathcal{N}(0, V_{\theta^*})$$

under right assumptions (in particular on the block size $h_n$).

The time-varying parameter model case will be very similar to the parametric model case. Formally, we will be providing in the following conditions such that

$$n^{\frac{1}{l}} \sum_{i=1}^{B_n} \left(\tilde{\Theta}_{i,n} - \Theta_{i,n}\right) \quad \xrightarrow{\mathbb{P}} \quad 0, \tag{3.37}$$

$$n^{\frac{1}{l}} \sum_{i=1}^{B_n} \left(\hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}_n^*} - \tilde{\Theta}_{i,n}\right) \quad \xrightarrow{\mathcal{D}} \quad T^{-1}\Big(\int_0^T V_{\theta_s^*} ds\Big)^{\frac{1}{2}} \mathcal{N}(0,1), \tag{3.38}$$

$$n^{\frac{1}{l}} \sum_{i=1}^{B_n} \left(\hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}_{i,n}^0} - \hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}_n^*}\right) \quad \xrightarrow{\mathbb{P}} \quad 0, \tag{3.39}$$

$$n^{\frac{1}{l}} \sum_{i=1}^{B_n} \left(\hat{\Theta}_{i,n} - \hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}_{i,n}^0}\right) \quad \xrightarrow{\mathbb{P}} \quad 0. \tag{3.40}$$

We make the first assumption, which is on observation times.

**Condition (E1).** The observation times are such that for $k = 1, 2, 4, 8$

$$\inf_{1 \leq i \leq N_n} \mathbb{E}_{\tau_{i-1,n}}\big[(\Delta\tau_{i,n})^k\big], \sup_{1 \leq i \leq N_n} \mathbb{E}_{\tau_{i-1,n}}\big[(\Delta\tau_{i,n})^k\big] \text{ are exactly of order } O_p(n^{-k}). \tag{3.41}$$

*Remark* 12. Note that condition $(E1)$ is satisfied by the HBT model introduced in Chapter 2.

We make a second assumption on the observation times, which is due to endogeneity. Indeed, when approximating the returns on a block holding the parameter $\theta_t^*$ constant, we also induce a change in the observation times. In the following assumption, we make sure that we can bound the difference in length between the approximated block and the true block, uniformly in the initial value parameter $\theta_0$, the path of parameter process $\theta_t$ and the initial Markov-chain $\mathbf{M}$. For that reason, we introduce the following notations. Let $M > 0$, $\mathbf{M} \in \mathcal{M}_{m,n}$ and $\theta_t \in \mathcal{E}_M$. For any $i = -(m-1), \cdots, h_n$ we define $M_{i,n}^{\mathbf{M},\theta}$ and $M_{i,n}^{\mathbf{M},\theta_t}$ the Markov chain with initial vector $\mathbf{M}$ and fixed parameter

$\theta \in K_M$ (respectively with time-varying parameter process $\theta_t$). The initial vectors are defined as $(M_{-(m-1),n}^{\mathbf{M},\theta}, \ldots, M_{0,n}^{\mathbf{M},\theta}) := \mathbf{M}$ and $(M_{-(m-1),n}^{\mathbf{M},\theta_t}, \ldots, M_{0,n}^{\mathbf{M},\theta_t}) := \mathbf{M}$. Also, we define the $m$-dimensional "memory" vectors as $\mathbf{M}_{i,n}^{\mathbf{M},\theta} := (M_{i,n}^{\mathbf{M},\theta}, \ldots, M_{i-(m-1),n}^{\mathbf{M},\theta})$ and $\mathbf{M}_{i,n}^{\mathbf{M},\theta_t} := (M_{i,n}^{\mathbf{M},\theta_t}, \ldots, M_{i-(m-1),n}^{\mathbf{M},\theta_t})$. Finally, for any positive integer $i$, the $i$th element of the Markov chains are obtained by the same recurrence relations as (3.27) and (3.28).

$$M_{i,n}^{\mathbf{M},\theta} := F_n\big(\mathbf{M}_{i-1,n}^{\mathbf{M},\theta}, U_{i,n}, \theta\big), \tag{3.42}$$

$$M_{i,n}^{\mathbf{M},\theta_t} := F_n\big(\mathbf{M}_{i-1,n}^{\mathbf{M},\theta_t}, U_{i,n}, \{\theta_s\}_{\tau_{i-1,n} \leq s \leq \tau_{i,n}}\big). \tag{3.43}$$

We define now the lengths of the first block $\mathrm{T}_{1,n}^{\mathbf{M},\theta} := \sum_{i=1}^{h_n} (M_{i,n}^{\mathbf{M},\theta})^{(d)}$ and $\mathrm{T}_{1,n}^{\mathbf{M},\theta_t} := \sum_{i=1}^{h_n} (M_{i,n}^{\mathbf{M},\theta_t})^{(d)}$.

**Condition (E2).** For any $M > 0$, we define $\mathcal{D}_M := \mathcal{E}_M \times \mathcal{M}_{m,1}$ and we have

$$\sup_{(\theta_t, \mathbf{M}_1) \in \mathcal{D}_M} \mathbb{E}\left[ (\mathrm{T}_{1,n}^{\alpha_n * \mathbf{M}_1, \theta_0} - \mathrm{T}_{1,n}^{\alpha_n * \mathbf{M}_1, \theta_t})^4 \right] = o(h_n^4 n^{-4}). \tag{3.44}$$

*Remark* 13. Condition $(E2)$ is also satisfied by the HBT model.

The following assumption provides the existence of $l' > 0$ such that the convergence rate of the parametric estimator is $n^{\frac{1}{l'}}$. In most examples, such as the MLE under regular conditions, we have $l' = 2$. Condition $(E3)$ also assumes that the parametric estimator is not too biased. For that purpose, we introduce the definition of the bias on the first block $B_{1,n}^{\mathbf{M},\theta} := \mathbb{E}\big[\big(\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M},\theta}; \ldots; R_{h_n,n}^{\mathbf{M},\theta}; \mathbf{M}) - \theta\big)\mathrm{T}_{1,n}^{\mathbf{M},\theta}\big]$ for any $\mathbf{M} \in \mathcal{M}_{m,n}$ and any $\theta \in K$.

**Condition (E3).** For any parameter $\theta \in K$, we assume that there exists a covariance matrix $V_\theta > 0$ such that for any $M > 0$, we have uniformly in $\theta \in K_M$

$$\begin{aligned}
\mathrm{Var}\left[h_n^{\frac{1}{l'}}\big(\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M}_n^*,\theta};\ldots;R_{h_n,n}^{\mathbf{M}_n^*,\theta};\mathbf{M}_n^*)-\theta\big)\mathrm{T}_{1,n}^{\theta}\right] &= V_\theta \mathrm{T}_{1,n}^{\theta} h_n n^{-1} \qquad (3.45)\\
&\quad + o_p(h_n^2 n^{-2})\\
\mathbb{E}\left[\left(\big(\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M}_n^*,\theta};\ldots;R_{h_n,n}^{\mathbf{M}_n^*,\theta};\mathbf{M}_n^*)-\theta\big)\mathrm{T}_{1,n}^{\theta}\right)^4\right] &= O(h_n^{4-\frac{4}{l'}} n^{-4}) \qquad (3.46)\\
B_{1,n}^{\mathbf{M}_n^*,\theta} &= O(h_n n^{-1-\frac{1}{l}}) \qquad (3.47)
\end{aligned}$$

*Remark* 14. (regular observations case) The reader might get confused at first reading with the block length term $\mathrm{T}_{1,n}^{\theta}$ showing up in (3.45), (3.46) and (3.47). One should keep in mind that in the simple case where observations are regular, we have $\mathrm{T}_{1,n}^{\theta} := T h_n n^{-1}$. In that case, (3.45) only assumes that the variance of the normalized error of the parametric estimator converges uniformly, (3.46) assumes that the fourth moment of the normalized error has bounded expectation and (3.47) that the parametric estimator used on a block of $h_n$ observations has a bias of order at most $O(n^{-\frac{1}{l}})$. If we assume that $l = 2$, as we know that $h_n$ can be of order up to $n^{\frac{1}{2}}$ in view of (3.48), we obtain that the bias must be of order $o(h_n)$. In the case where observations are not regular, the presence of the random block length term $\mathrm{T}_{1,n}^{\theta}$ doesn't seem to make Condition $(E3)$ harder to verify than in the regular case, at least in the model with uncertainty zones (See Appendix).

*Remark* 15. Condition $(E3)$ is used to show (3.38).

We make a fourth assumption, which is on the block size $h_n$. In practice, Condition $(E4)$ provides us with the maximum block size $h_n$ to use for constant approximation of parameter. Note that in the most common case when $l = 2$ and $l' = 2$, (3.49) is automatically verified. Also in that case, (3.48) can be written as $h_n = o(n^{\frac{1}{2}})$, which is the same block size's order found in Mykland and Zhang (2011), who investigated how constant we can hold volatility in a small neighborhood in the case of regular observations of the price following a continuous Itô-process.

**Condition (E4).** The block size $h_n$ is such that

$$n^{\frac{1}{l}-1}h_n = o(1) \tag{3.48}$$
$$n^{\frac{2}{l}-1}h_n^{1-\frac{2}{l'}} \to 1 \tag{3.49}$$

*Remark* 16. (3.48) is in particular used to show (3.37).

The next condition assumes that uniformly in the parameter value and future parameter path, we can bound the discrepancy on the time-varying parameter model starting with two different initial vectors. The reason why Condition ($E5$) doesn't assume also uniformity in the starting vector is that in the MA(1) example in Section 3.7.3 we don't have such uniformity. Consequently, we need a weaker assumption. For any $M > 0$, any $\theta_t \in \mathcal{E}_M$ and any $\mathbf{M} \in \mathcal{M}_{m,n}$ we introduce the bias with parameter $\theta_t$ as $B_{1,n}^{\mathbf{M},\theta_t} := \mathbb{E}\big[\big(\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M},\theta_t};\dots;R_{h_n,n}^{\mathbf{M},\theta_t};\mathbf{M}) - \theta\big)\mathrm{T}_{1,n}^{\mathbf{M},\theta_t}\big]$. For any $\mathbf{N} \in \mathcal{M}_{m,n}$, we also define the difference in bias as

$$B_{1,n}^{\mathbf{M},\mathbf{N},\theta_t} := \mathbb{E}\big[\big(\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M},\theta_t},\dots,R_{k,n}^{\mathbf{M},\theta_t};\mathbf{M}) - \hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{N},\theta_t},\dots,R_{k,n}^{\mathbf{N},\theta_t};\mathbf{N})\big)\mathrm{T}_{1,n}^{\mathbf{N},\theta_t}\big].$$

We finally define the variance of the estimated difference

$$E_{1,n}^{\mathbf{M},\mathbf{N},\theta_t} := \mathrm{Var}\,\big[\big(\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M},\theta_t},\dots,R_{k,n}^{\mathbf{M},\theta_t};\mathbf{M}) - \hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{N},\theta_t},\dots,R_{k,n}^{\mathbf{N},\theta_t};\mathbf{N})\big)\mathrm{T}_{1,n}^{\mathbf{N},\theta_t}\big].$$

**Condition (E5).** For any $M > 0$ we have

$$\sum_{i=1}^{B_n} \sup_{\theta_t \in \mathcal{E}_M} \big|B_{1,n}^{\mathbf{M}_{i,n}^0,\mathbf{M}_n^*,\theta_t}\big| = o_p(n^{-\frac{1}{l}}), \tag{3.50}$$
$$\sum_{i=1}^{B_n} \sup_{\theta_t \in \mathcal{E}_M} E_{1,n}^{\mathbf{M}_{i,n}^0,\mathbf{M}_n^*,\theta_t} = o_p(n^{-\frac{2}{l}}). \tag{3.51}$$

*Remark* 17. Condition ($E5$) is used to show (3.39).

The last condition is very similar to Condition ($E5$), except that we are not looking at the discrepancy induced by different initial vectors but rather the difference between the time-varying parameter model and the parametric model when initial vectors are equal. For that reason, we introduce the difference in bias

$$B_{1,n}^{\mathbf{M},\theta,\theta_t} := \mathbb{E}\big[\big(\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M},\theta_t},\dots,R_{k,n}^{\mathbf{M},\theta_t};\mathbf{M}) - \hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M},\theta},\dots,R_{k,n}^{\mathbf{M},\theta};\mathbf{M})\big)\mathrm{T}_{1,n}^{\mathbf{M},\theta_t}\big],$$

as well as the variance of the difference

$$E_{1,n}^{\mathbf{M},\theta,\theta_t} := \mathrm{Var}\left[\big(\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M},\theta_t},\dots,R_{k,n}^{\mathbf{M},\theta_t};\mathbf{M}) - \hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M},\theta},\dots,R_{k,n}^{\mathbf{M},\theta};\mathbf{M})\big)\mathrm{T}_{1,n}^{\mathbf{M},\theta_t}\right].$$

**Condition (E6).** For any $M > 0$, we have

$$\sum_{i=1}^{B_n} \sup_{\theta_t \in \mathcal{E}_M} \left|B_{1,n}^{\mathbf{M}_{i,n}^0,\theta,\theta_t}\right| = o_p(n^{-\frac{1}{l}}) \tag{3.52}$$

$$\sum_{i=1}^{B_n} \sup_{\theta_t \in \mathcal{E}_M} E_{k,n}^{\mathbf{M}_{i,n}^0,\theta,\theta_t} = o_p(n^{-\frac{2}{l}}) \tag{3.53}$$

*Remark* 18. Condition (E6) is used to show (3.39).

We now state the main theorem of Chapter 3, which investigates the limit distribution of (3.36).

**Theorem (Central Limit Theorem).** *Assume conditions* $(E0)-(E6)$*. Then, stably in law as* $n \to \infty$*,*

$$n^{\frac{1}{l}}\big(\widehat{\Theta}_n - \Theta\big) \to T^{-1}\big(\int_0^T V_{\theta_s^*} ds\big)^{\frac{1}{2}}\mathcal{N}(0,1). \tag{3.54}$$

*Remark* 19. (convergence rate and asymptotic variance) In most examples, we have that $l = 2$, which is the best convergence rate in the parametric model. In view of (3.49) in Condition $(E4)$, in that case we also have $l' = 2$ and thus the convergence rate in (3.54) is the best attainable (Gloter and Jacod (2001)). If we also assume that we have a parametric estimator which achieves the Cramér-Rao bound of efficiency, we conjecture that the variance in the right term of (3.59) is the nonparametric efficient bound, at least in the case where there is no endogeneity.

*Remark* 20. (MLE) The MLE is a natural and popular estimator. In order for the practitioner to be able to apply the local MLE and show the associated central limit theorem, one has to keep in mind that in finite sample, the MLE is biased, with a bias magnitude of the order of the number of observations inverse. To verify bias conditions (3.47), (3.50) and (3.52), one has to apply a bias correction to the MLE discussed for instance in Firth (1993).

*Remark* 21. (block size) Condition ($E4$) provides the asymptotic order to use for the block size $h_n$. Thus, it gives a rule basis to use on finite sample, but it is left to the practitioner to ultimitalely choose $h_n$. If the parametric estimator is badly biased, the practitioner should increase the value of $h_n$. Also, if the parameter process $\theta_t^*$ seems not to be moving a lot, $h_n$ can be chosen to be bigger. In Section 3.8, we show on the model with uncertainty zones that the estimated volatility is not $h_n$-dependent if we choose $h_n \approx N_n^{\frac{1}{2}}$.

*Remark* 22. (subset parameter estimation) In practice, one can be interested in estimating only a $p'$-dimensional integrated parameter $\Xi$, where $p' \in \{1, \dots, p\}$ and $\Xi$ is a subset of $\Theta$. If we define in analogy with the quantities depending on $\theta_t^*$ the $p'$-dimensional subquantities $\xi_t^*$, $\hat{\xi}_{k,n}$, $\Xi_{i,n}$, $\widehat{\Xi}_{i,n}$ $\widehat{\Xi}_n$ as well as the $p' \times p'$-dimensional matrix $V_\xi$, the Central Limit Theorem still holds in that case under the same assumptions

$$n^{\frac{1}{l}}\big(\widehat{\Xi}_n - \Xi\big) \to T^{-1}\Big(\int_0^T V_{\theta_s^*} ds\Big)^{\frac{1}{2}}.$$

*Remark* 23. (Estimating the Asymptotic Variance) If the practitioner doesn't have a (parametric) variance estimator at hand and that her parametric estimator can be written as in Mykland and Zhang (2014), one can use the techniques of the Chapter 3 to obtain a variance estimate. Investigating if such techniques would work in our setting is beyond the scope of this Chapter 3. If she has an estimator, then for any $i = 1, \dots, B_n$

she can estimate the $i$th block covariance $\widehat{V}_{i,n}$ as $\widehat{V}_{i,n} := \hat{v}_{h_n,n}(R^1_{i,n}; \ldots; R^{h_n}_{i,n}; \mathbf{R}^0_{i,n})$, and she can estimate the asymptotic variance as the weighted sum

$$\widehat{V}_n = \frac{1}{T^2} \sum_{i=1}^{B_n} \widehat{V}_{i,n} \Delta \mathrm{T}_{i,n}. \tag{3.55}$$

Under conditions similar to the ones of Section 3.4.2 and Section 3.6, we can obtain the consistency of (3.55).

*Remark* 24. (Estimator with unobservable returns as input) As a function of several non-observable quantities can be observed, we can define the parametric estimator with a slight difference $\hat{\theta}_{k,n} := \hat{\theta}_{k,n}(M_{1,n}; \ldots; M_{k,n}; \mathbf{M}_{0,n})$. The central Limit Theorem still holds in that more general case. An example of such $\hat{\theta}_{k,n}$ will be given with the HY in Section 3.7.2.

*Remark* 25. (asymptotic assumption on returns) As discussed in Section 3.4.2, the assumption (3.29) is relatively strong. We conjecture that under a more general assumption, such as the uniform convergence in distribution of $F_n(\alpha_n * \mathbf{M}, U_{1,n}, \theta)$ to a non-degenerate limit, the Central Limit Theorem would still hold. In particular, that would permit the noise to have a different convergence rate and the observation times to follow more general point processes.

## 3.7 Examples

We have two purposes in this section. First, we want to document that assumptions of the LPM are widely satisfied by models of the literature used to estimate high-frequency quantities. In addition, in Section 3.7.4, we introduce a new model where efficient price follows a continuous local martingale and the correlation structure between efficient price, noise and arrival times is very general and then show that this model is contained in the LPM class.

Moreover, we aim to show that theoretical conditions provided in Section 3.6 can be satisfied for specific problems. As a first example, to estimate high-frequency covariance in the HBT model, Chapter 2 provided a bias-corrected HY estimator and used techniques similar to the ones given in Chapter 3 to find the limit distribution. We provide insight into their work in Section 3.7.2. As a second example, we introduce a time-varying friction parameter extension of the model with uncertainty zones introduced in Robert and Rosenbaum (2011) to estimate volatility and show that conditions of Section 3.6 are easily satisfied in that case. This can be found in Section 3.7.1.

### 3.7.1 *Volatility in the HBT model*

The HBT model was introduced in Chapter 2 as a general multidimensional endogenous model which can possibly include microstructure noise of a specific form. In this section, we consider the one-dimensional case and notation of Chapter 2 are "in force". We recall the definition of the two-dimensional process $Y_t := (X_t, X_t^{(t)})$, where $X_t^{(t)}$ is the time process and assume that $Y_t$ follows a null-drift Itô-process with (matrix) volatility $\sigma_t^Y$. In order to embed the HBT model into the LPM, we need to assume that $d_t$ and $u_t$ depend on a multidimensional parameter $\mu_t$ such that there exists a partially observed Markov chain following (3.28) with $\theta_t^* := (\sigma_t^Y (\sigma_t^Y)^T, \mu_t)$. The parameter thus includes information on the volatility and also on observation times (which in particular depend heavily on $\mu_t$). We propose in the following to have a look at $\mu_t$ and other LPM quantities on a couple of HBT examples, Example 1 in Chapter 2 and the model with uncertainty zones. Furthermore, we investigate an associated LPE and the Central Limit Theorem in the latter model.

#### Hitting Constant Boundaries

In Example 1 of Chapter 2, we assume no noise in observations and we have $\mu_t := (\theta_u, \theta_d)$, and because $X_t = X_t^{(t)}$, the information contained in $\sigma_t^Y (\sigma_t^Y)^T$ can be expressed as the one-dimensional volatility $\sigma_t^2$ of $X_t$. Thus, we obtain that $\theta_t^* := (\sigma_t^2, \theta_u, \theta_d)$. With the technology of Chapter 3, we can assume that $\theta_u$ and $\theta_d$ are continuous local martingale parameters in the model where arrival times are defined recursively as $\tau_{0,n} := 0$ and for any positive integer $i$

$$\tau_{i,n} := \inf\left\{t > \tau_{i-1,n} : \Delta X_{[\tau_{i-1,n},s]} = \theta_{u,\tau_{i-1,n}} \text{ or } \Delta X_{[\tau_{i-1,n},s]} = \theta_{d,\tau_{i-1,n}}\right\}. \tag{3.56}$$

In (3.56), the boundaries are piece-wise constant equal to the initial-parameter value. We can write (3.56) as a LPM with $M_{i,n} := R_{i,n} := (\Delta X_{\tau_{i,n}}, \Delta\tau_{i,n})$, in particular no unobserved quantity $Q_{i,n}$ is needed. Also, we have that the order of memory is $m := 0$. The "random innovation" is defined as the future path of the Brownian motion $U_{i,n} := \left(\Delta X_{[\tau_{i-1,n},t]}\right)_{t\geq\tau_{i-1,n}}$ and since $m = 0$, the returns can be written as $R_{i,n} := F_n(U_{i,n}, \{\theta_s^*\}_{\tau_{i-1,n}\leq s\leq\tau_{i,n}})$. If we define

$$v := \inf\left\{t > 0 : \int_0^t \sigma_s du_s = \theta_{u,0} \text{ or } \int_0^t \sigma_s du_s = \theta_{d,0}\right\}$$

we can express $F_n(u_t, (\sigma_t, \theta_{u,t}, \theta_{d,t})) := (\int_0^v \sigma_s du_s, v)$. Thus, Example 1 can be written as a LPM.

## Model with uncertainty zones

The model with uncertainty zones assumes that $\mu_t := (\eta, \chi_t)$ where $\chi_t$ is the $M$-dimensional time-varying parameter driving the conditional distribution of the jump sizes in ticks, see p.5 in Robert and Rosenbaum (2012). Thus, we have that $\theta_t^* := (\sigma_t, \eta, \chi_t)$. We are not interested in estimating $\chi_t$ and thus we follow Remark 22 and consider the subparameter $\xi_t^* = (\sigma_t, \eta)$ to be estimated. Following the setting of this Chapter 3, we will assume that $\eta$ is a time-varying parameter $\eta_t$ and we will extend the model with uncertainty zones in (3.57). We call this extension the time-varying model with uncertainty zones. We define $\alpha_n^{(tick)}$ (in lieu of $\alpha_n$ in the cited paper) the tick size (which vanishes asymptotically). The sampling times are defined recursively for any positive integer $i$

$$\tau_{i,n} := \inf\Big\{t > \tau_{i-1,n} : X_t = X^{(\alpha_n^{(tick)})}_{\tau_{i-1,n}} - \alpha_n^{(tick)}\big(L_{i,n} - \frac{1}{2} + \eta_{\tau_{i-1,n}}\big)$$

$$\text{or } X_t = X^{(\alpha_n^{(tick)})}_{\tau_{i-1,n}} + \alpha_n^{(tick)}\big(L_{i,n} - \frac{1}{2} + \eta_{\tau_{i-1,n}}\big)\Big\}. \tag{3.57}$$

We want to estimate the integrated parameter $\Xi := (\int_0^T \sigma_t^2 dt, \int_0^T \eta_t dt)$. We define the unobserved return as $Q_{i,n} := \Delta X_{\tau_{i,n}}$, the observed returns as $R_{i,n} := (\Delta Z_{\tau_{i,n}}, \Delta\tau_{i,n})$, where the observed price is defined in Example 4 in Chapter 2. The order of memory is $m := 1$. The "random innovation" is defined as the two-dimensional path $U_{i,n} := ((\Delta X_{[\tau_{i-1,n},t]})_{t\geq\tau_{i-1,n}}, (\Delta W'_{[\tau_{i-1,n},t]})_{t\geq\tau_{i-1,n}})$ where $W'$ is a Brownian motion independent of the other quantities which deals with the jump size and which is defined in p.11 of Robert and Rosenbaum (2012). We also define

$$w^+ := \inf\Big\{t > 0 : \int_0^t \sigma_s du_s^{(1)} = -\alpha_n^{(tick)}\big(\phi(\chi_0, u_t^{(2)}) - 1 + 2\eta_0\big)$$

$$\text{or } \int_0^t \sigma_s du_s^{(1)} = \alpha_n^{(tick)}\phi(\chi_0, u_t^{(2)})\Big\}$$

and

$$w^- := \inf\Big\{t > 0 : \int_0^t \sigma_s du_s^{(1)} = -\alpha_n^{(tick)}\phi(\chi_0, u_t^{(2)})$$

$$\text{or } \int_0^t \sigma_s du_s^{(1)} = \alpha_n^{(tick)}\big(\phi(\chi_0, u_t^{(2)}) - 1 + 2\eta_0\big)\Big\},$$

where $\phi(\chi_0, u_t^{(2)})$ corresponds to the size of the next jump (in absolute number of ticks) and can be found in p.11 of Robert and Rosenbaum (2012). Furthermore, we define

$$w := w^+\mathbf{1}_{\{r^{(1)}>0\}} + w^-\mathbf{1}_{\{r^{(1)}<0\}}.$$

We can express

$$F_n\big((q, r^{(1)}, r^{(2)}), (u_t^{(1)}, u_t^{(2)}), (\sigma_t^2, \eta_t, \chi_t)\big)$$
$$:= \Big(\int_0^w \sigma_s du_s, \alpha_n^{(tick)}\phi(\chi_0, u_t^{(2)}) \text{ sign } \big(\int_0^w \sigma_s du_s\big), w\Big).$$

We estimate on each block the volatility and the friction parameter using respectively the estimator $\widehat{RV}_{\alpha_n^{(tick)},t}$ and a slight modification of $\hat{\eta}_{\alpha_n^{(tick)},t}$ (see p.8 in Robert and Rosenbaum (2012), we use $w$ in place of $m$) defined as

$$\hat{\eta}_{\alpha,t}^{(m)} := \sum_{k=1}^{w} \lambda_{\alpha,t,k}^{(m)} u_{\alpha,t,k}^{(m)} \tag{3.58}$$

with

$$\lambda_{\alpha,t,k}^{(m)} := \frac{N_{\alpha,t,k}^{(a)} + N_{\alpha,t,k}^{(c)}}{\sum_{j=1}^{w}\big(N_{\alpha,t,j}^{(a)} + N_{\alpha,t,j}^{(c)}\big)} \text{ and } u_{\alpha,t,k}^{(m)} := \max\Big(0, \min\Big(1, \frac{1}{2}\big(k\big(\frac{N_{\alpha,t,k}^{(c)}}{N_{\alpha,t,k}^{(a)}} - 1\big) + 1\big)\Big)\Big),$$

where we assume that $\frac{C}{0} := \infty$, and in particular $u_{\alpha,t,k} = 1$ when $N_{\alpha,t,k}^{(a)} = 0$. We choose to work with this modified estimator for both reasons because the definition of $\hat{\eta}_{\alpha,t}$ was slightly unclear when $N_{\alpha,t,k}^{(a)} = 0$ in p.8 of Robert and Rosenbaum (2012) and because the finite sample bias of $\hat{\eta}_{\alpha,t}^{(m)}$ is slightly smaller with the modified estimator. We have $\boldsymbol{\alpha}_n := (\alpha_n^{(tick)}, \alpha_n^{(tick)}, (\alpha_n^{(tick)})^2)$. We obtain the following theorem as an application of the Central Limit Theorem.

**Theorem (Time-varying friction parameter model with uncertainty zones).** *Stably in law as $n \to \infty$,*

$$(\alpha_n^{(tick)})^{-1}\big(\widehat{\Xi}_n - \Xi\big) \to T^{-1}\Big(\int_0^T V_{\theta_s^*} ds\Big)\mathcal{N}(0, 1), \tag{3.59}$$

*where $V_\theta$ can be straightforwardly inferred from the definition of Lemma 4.19 in p.26 of Robert and Rosenbaum (2012).*

The proof, checking that conditions (E0)-(E6) are satisfied, can be found in Appendix.

*Remark* 26. We can add a drift to the volatility process following the techniques in Remark 10.

*Remark* 27. Note that equivalently $\boldsymbol{\alpha}_n := (n^{-\frac{1}{2}}, n^{-\frac{1}{2}}, n^{-1})$ and the convergence rate is $n^{\frac{1}{2}}$, see Remark 4 for more details.

### 3.7.2 High-frequency covariance in the HBT model

We recall the definition of the four-dimensional process $Y_t := (X_t^{(1)}, X_t^{(2)}, X_t^{(t,1)}, X_t^{(t,2)})$ in Chapter 2 and assume that $Y_t$ follows a non-drift Itô-process with volatility $\sigma_t^Y$. We also assume that $d_t^{(k)}$ and $u_t^{(k)}$ can be embedded in a multidimensional parameter $\mu_t$ such that there exists a partially observed Markov chain with two-dimensional returns following (3.28) with $\theta_t^* := (\sigma_t^Y(\sigma_t^Y)^T, \mu_t)$. We are concerned with the estimation of the high-frequency covariance and thus $\xi_t^* := (\sigma_t^Y(\sigma_t^Y)^T)^{1,2}$. We are exactly in the setting of Remark 22 with $p' = 1$. We choose to work with the bias-corrected HY introduced in Section 2.4.2.

We first show that the 2-dimensional HBT model is contained in the LPM class. For that purpose, we identify the local Markov chain quantities used in the proof of Chapter 2. Following the notation of Chapter 2, we define the non-observed part as equal to

$$Q_{i,n} := \left(\Delta X^{(t,2)}_{[\tau^{1C}_{i-1,n}, \tau^{1C,+}_{i-1,n}]}, \Delta X^{(t,2)}_{[\tau^{1C,-}_{i,n}, \tau^{1C}_{i,n}]}\right)$$

and the observed-part as equal to

$$R_{i,n} := \left(\tau^{1C,+}_{i-1,n} - \tau^{1C}_{i-1,n}, \tau^{1C}_{i,n} - \tau^{1C,-}_{i,n}, \Delta X^{(1)}_{\tau^{1C}_{i,n}}, \Delta \tau^{1C}_{i,n}\right).$$

In view of Lemma 10 in Appendix of Chapter 2 and Remark 4, we have that $\boldsymbol{\alpha}_n := (n^{-\frac{1}{2}}, n^{-\frac{1}{2}}, n^{-1}, n^{-1}, n^{-\frac{1}{2}}, n^{-1})$.

We now point to the parts of the proof in Chapter 2 that can be used to verify the conditions of Section 3.6. Condition (C0) is straightforwardly satisfied by the assumptions of Chapter 2. Condition (C1) is satisfied by Lemma 7 and Condition (C2) can be proven using the proof of Lemma 11. In Condition (C3), Equation (3.45) can be satisfied using the same techniques than in the first three steps of the proof in Lemma 14. Equation (3.46) is straightforward to show. Finally, as the bias-corrected HY estimator is unbiased for a fixed $\theta \in K$, we show directly (3.47). We obtain Condition (E4) with $l = \frac{1}{2}$, $l' = \frac{1}{2}$ and $h_n = o(n^{\frac{1}{2}})$. Equation (3.50) in Condition (C5) comes from the unbiasedness of the estimator. Proof techniques of the first three steps of the proof in Lemma 14 can be used to prove (3.51) in Condition (C5). In Condition (C6), (3.52) also comes from the unbiasedness of the estimator and (3.53) can be satisfied using a similar proof to Lemma 13.

*Remark* 28. Because the sum of two unobserved variables from consecutive Markov chain element $\Delta X^{(t,2)}_{[\tau^{1C,-}_{i-1,n},\tau^{1C}_{i-1,n}]}$ and $\Delta X^{(t,2)}_{[\tau^{1C}_{i-1,n},\tau^{1C,+}_{i-1,n}]}$ is actually the increment of the second asset (and thus observed), we are actually in the case of Remark 24.

We could add noise in the model extending the one-dimensional model introduced in Section 3.7.4. We conjecture that a local parametric (bias-corrected) MLE would satisfy the conditions of Section 3.6, and it would then be interesting to compare it to the pre-averaged Hayashi-Yoshida estimator of Christensen et al. (2010, 2013) and Koike (2014), two scales covariance estimator in Zhang (2011), the multivariate realised kernel in Barndorff-Nielsen et al. (2011) and the high-frequency covariance estimator of Aït-Sahalia et al. (2010).

### *3.7.3 Volatility in an asset price with non auto-correlated noise not correlated with the efficient price*

We assume that the observations occur at regular times $\tau_{i,n} = n^{-1}T$. We consider the noised-return model

$$R_{i,n} := \Delta X_{\tau_{i,n}} + \epsilon_{\tau_{i,n}} - \epsilon_{\tau_{i-1,n}}, \tag{3.60}$$

where the noise $\epsilon_t := n^{-\frac{1}{2}}(v_t^*)^{\frac{1}{2}}\chi_t$ is time-varying, $\chi_t$ are standard independent normally identically distributed and independent of the (null-drift Itô-process with volatility $\sigma_t^*$) efficient process $X_t := \int_0^T \sigma_s^* dW_s$. The parameter process is defined as the two-dimensional volatility and noise process $\theta_t^* := ((\sigma_t^*)^2, v_t^*)$. We are interested in the estimation of $\Theta := \big(\int_0^T (\sigma_t^*)^2 dt, \int_0^T v_t^* dt\big)$. First, note that the model (3.60) can be easily written as an LPM of order $m = 1$ with

$$\begin{aligned}
Q_{i,n} &:= (\Delta X_{\tau_{i,n}}, \epsilon_{\tau_{i,n}}),\\
R_{i,n} &:= \Delta X_{\tau_{i,n}} + \epsilon_{\tau_{i,n}} - \epsilon_{\tau_{i-1,n}},\\
U_{i,n} &:= (\{\Delta W_{[\tau_{i-1,n},s]}\}_{\tau_{i-1,n}\leq s\leq\tau_{i,n}}, \chi_{\tau_{i,n}}).
\end{aligned}$$

Define $G : \mathbb{R}_*^- \to \mathbb{R}$ such that $G(x) = -\frac{1}{x} - x^2$. If we assume that the parameters are constant equal to $((\sigma^*)^2, v^*)$, we can rewrite (3.60) as

$$R_{i,n} := \eta_{i,n} + \beta\eta_{i-1,n},$$

where $\beta := G^{-1}\big(\frac{(\sigma^*)^2 T}{v^*} + 2\big)$ and $\eta_{i,n}$ are IID normal with variance $u_n := -\frac{v^*}{n\beta}$ and thus the returns follow locally an MA(1) process, as Aït-Sahalia et al. (2005) and Xiu (2010) pointed out when the volatility and noise were non time varying parameters. In the latter paper, the author showed that we can still use the MLE of MA(1) when volatility is a time-varying quantity pretending that it is constant on the whole dataset. He showed the consistency and the associated limit distribution of the QMLE in a slightly different asymptotic setting than in Chapter 3 (with the noise component not shrinking to 0 asymptotically). Nonetheless, the author assumed non time-varying noise. The techniques of Chapter 3 allows us to go one step further by allowing time-varying noise. We define the parametric estimator $\hat{\theta}_{k,n}$ as the bias-corrected MLE of the MA(1) process. We obtain $\boldsymbol{\alpha}_n := (n^{-\frac{1}{2}}, n^{-\frac{1}{2}}, n^{-\frac{1}{2}})$. We conjecture that we can verify the assumptions of Section 3.6 in this model with the local MLE.

### *3.7.4 Volatility in an extended noisy HBT model where noise can be auto-correlated and correlated with the efficient price*

We go one step further the HBT model by allowing noise in the model as well. We keep the structure generating the sampling times (2.1), but we observe now a noisy return

$$R_{i,n} := (\Delta X_{\tau_{i,n}} + \epsilon_{\tau_{i,n}} - \epsilon_{\tau_{i-1,n}}, \Delta\tau_{i,n}). \tag{3.61}$$

We also define the unobserved return $Q_{i,n} := (\Delta X_{\tau_{i,n}}, \epsilon_{\tau_{i,n}})$ and we assume that there exists $\theta_t^*$ and $U_{i,n}$ such that they satisfy (3.28) for a positive integer order of memory $m$. Note that the assumption (3.28) together with the assumption (3.61) allow for a very general model. In particular, the noise and the efficient price can be correlated with each other and the noise auto-correlated. In the simplest (non realistic) case, we can imagine that the noise follows the same assumption as in Section 3.7.3. Because observations occur on the tick grid, a realistic model assumes that the observed price $Z_{\tau_{i,n}} := X_{\tau_{i,n}} + \epsilon_{\tau_{i,n}}$ takes only modulo of the tick size values. One possible extension of the model introduced in Section 3.7.3 assumes that the price is rounded, i.e. $Z_{\tau_{i,n}} := (X_{\tau_{i,n}} + \epsilon_{\tau_{i,n}})^{(\alpha_n^{(tick)})}$. It can be written as an LPM of order $m = 1$ with

$$\begin{aligned}
Q_{i,n} &:= (\Delta X_{\tau_{i,n}}, \epsilon_{\tau_{i,n}}, X_{\tau_{i,n}} \bmod \alpha_n^{(tick)}),\\
R_{i,n} &:= \Delta X_{\tau_{i,n}} + \epsilon_{\tau_{i,n}} - \epsilon_{\tau_{i-1,n}},\\
U_{i,n} &:= (\{\Delta W_{[\tau_{i-1,n},s]}\}_{\tau_{i-1,n} \leq s \leq \tau_{i,n}}, \chi_{\tau_{i,n}}).
\end{aligned}$$

Furthermore, we can show that one other simple model, the floor rounding with probability $\frac{1}{2}$ and ceiling rounding with probability $\frac{1}{2}$ of the efficient price model, which is decribed in p.7 of Dahlaus and Neddermeyer (2014), can be expressed as a LPM. We insist on the fact that those are basic examples and that the LPM class is much broader.

We conjecture here again that a local MLE would satisfy the conditions of Section 3.6, and that we could then infer in particular about the integrated volatalitity, observation times parameter and integrated microstructure noise.

### 3.7.5 High Frequency Regression and ANOVA

We are interested in systems of the form $dV_t = \beta_t dX_t + dZ_t$, where the high-frequency correlation between $X_t$ and $Z_t$ is null, i.e. $\langle X, Z\rangle_t = 0$, we can observe the two processes $V_t$ and $X_t$ and $X_t$ can be multidimensional. We can see $\beta_t$ as the beta from portfolio optimization and $Z_t$ the idiosyncratic noise, or $\beta_t$ can be the hedging delta of an option, with $Z_t$ the error. There are two different objects of interest. First, the regression problem seeks to infer about the integrated beta (Mykland and Zhang (2009, Section 4.2, pp. 1424-1426), Kalnina (2012), Zhang (2012, Section 4, pp. 268-273), Reiss, Todorov, and Tauchen (2014)). The ANOVA problem seeks to estimate the $\langle Z, Z\rangle_T := \int_0^T (\sigma_s^Z)^2 ds$ (Zhang (2001) and Mykland and Zhang (2006)). We define $Y_t := (V_t, X_t)$ and assume that $Y_t$ is a null-drift Itô-process with volatility $\sigma_t^Y$. We can define the multidimensional parameter $\theta_t^* := (\beta_t, \sigma_t^Y, (\sigma_t^Z)^2)$. We are interested in the estimation of the integrated sub-parameter $\xi_t^* := (\beta_t, (\sigma_t^Z)^2)$.

In the case where there is no microstructure noise in observations and the observations occur at regular times, it is easy to show that the LPM class contains the model and that the estimator to be used locally is the usual least squares estimator, together with the residual variance estimator. Furthermore, the assumptions of Section 3.6 are easily satisfied. In the more general case where there can be microstructure noise, we can use a model similar to Section 3.7.4. We conjecture in this case again that a local MLE (or possibly another estimator) would satisfy the conditions of Section 3.6.

### 3.7.6 Limit Order Book (LOB)

A LOB is a multidimensional queuing system, that gathers for any $t \geq 0$ the total volume of non-executed orders for every price level. Each 1-dimensional Order Book can be seen as a stochastic process of its inter-arrival times between two consecutive events. Consequently, we can describe a LOB as a high-dimensional point process, where each component counts the number of orders of a given type and a given price level. Very often it is assumed in the literature that the LOB process is a function of a non time-varying parameter $\theta^*$ as in Ogihara and Yoshida (2011). This constancy assumption is seldom checked properly. Using the techniques of Chapter 3, The LOB user could allow for a time-varying parameter. First, she would need to investigate if the LPM class contains her parametric model, build a time-varying parametric model and then check the conditions of Section 3.6 to use a LPE. We saw in Example 8 that a 1-dimensional Poisson-process is a LPM. In particular, we conjecture that we can build a time-varying parameter extension of the Hawkes process introduced in Hawkes (1971) and used in Bacry et al. (2013) and Aït-Sahalia et al. (2015) which is contained into the LPM class and that the LPE of the QMLE in Clinet and Yoshida (2015) would satisfy the conditions of Section 3.6.

### *3.7.7 Leverage effect*

The leverage effect describes the (usually) negative relation between stock returns and their volatility (see e.g. Wang and Mykland (2014), Aït-Sahalia et al. (2014)). In that case, the parameter of interest can be defined as $\xi_t^* := d\langle X, F(\sigma^*)\rangle_t/dt$ where conditions on the nonrandom function $F$ can be found in p.199 of Wang and Mykland (2014). Note that the convergence rate is $l = \frac{1}{4}$ if there is no microstructure noise. In light of Condition (E4), this example would require probably extrawork, and provide a new estimator with a convergence rate not as good as the parametric estimator. If we assume that the model is the same as in Section 3.7.4, a LPE could work but a parametric estimator would first need to be given.

### 3.7.8 *Volatility of volatility*

This is similar to the leverage effect, one would have to investigate first a parametric estimator. In that case, the parameter of interest is defined as $\xi_t^* := d\langle(\sigma^*)^2, (\sigma^*)^2\rangle_t/dt$. The convergence rate is also $l = \frac{1}{4}$. We can find results on this inference problem in Vetter (2011) and Mykland, Shephard and Sheppard (2012).

## 3.8 Empirical work

In this section, we are interested in estimating the integrated parameter in the time-varying friction parameter model with uncertainty zones introduced in Section 3.7.1. We remind the reader that the parameter of interest is defined as $\xi_t^* = (\sigma_t^2, \eta_t)$. We are looking at Orange France Telecom stock price on the CAC 40 on Monday March 4th, 2013. The number of returns between 9am and 4pm corresponding to a "change of price" is equal to $N_n = 3306$, and the tick size $\alpha_n^{(tick)} = 0.001$ euro. We assume that $T := 1$, consider that $t = 0$ corresponds to 9am and that $t = T$ corresponds to 4pm.

When assuming the non time-varying model with friction parameter equal to $\eta$, we can estimate the non time-varying endogeneity parameter $\hat{\eta}_{\alpha,T}^{(m)}$. Note that in finite sample, $\hat{\eta}_{\alpha,T}^{(m)}$ is biased and that this bias can be estimated following the estimate $\hat{B}$ in Appendix 3.10.7. Also, we define the standard deviation $s_n(\eta)$ and a standard deviation estimate $\hat{s}_n(\eta)$ where the expression is also provided in Appendix 3.10.7. We will estimate the standard deviation as $\hat{s}_n := \hat{s}_n(\eta)$. We find empirically $\hat{\eta}_{\alpha,T}^{(m)} := 0.155$ and $\hat{s}_n := 0.008$.

We define $\hat{\eta}_{i,n}$ the estimate of $\eta$ on the $i$th block following (3.35). We also define the standard deviation estimate of $\hat{\eta}_{i,n}$ as $\hat{s}_{i,h_n} := s_{i,h_n}(\hat{\eta}_n)$. Note that $\hat{s}_{i,h_n}$ is not block-dependent except for the last block which is thus removed from the analysis in the following. Consequently we define $\hat{s}_{h_n} := \hat{s}_{1,h_n}$. Figure 3.1 shows the evolution of $\hat{\eta}_{i,n}$ for different values of $h_n$. Based on those estimates and the standard deviation estimate

$s_n$, we compute the associated chi-square statistic

$$\chi_n^2 := \sum_{i=1}^{B_n-1} \left(\frac{\hat{\eta}_{i,n} - \hat{\eta}}{s_n}\right)^2.$$

Under the null hypothesis which states that $\eta_t$ is non time-varying, $\chi_n^2$ follows approximately a chi-square distribution with $B_n - 1$ degrees of freedom. We report $\chi_n^2$ for different values of $h_n$ in Table 3.1. The obtained values indicate that we have strong evidence against the null hypothesis, thus it provides us very good reasons to use the techniques of Chapter 3.

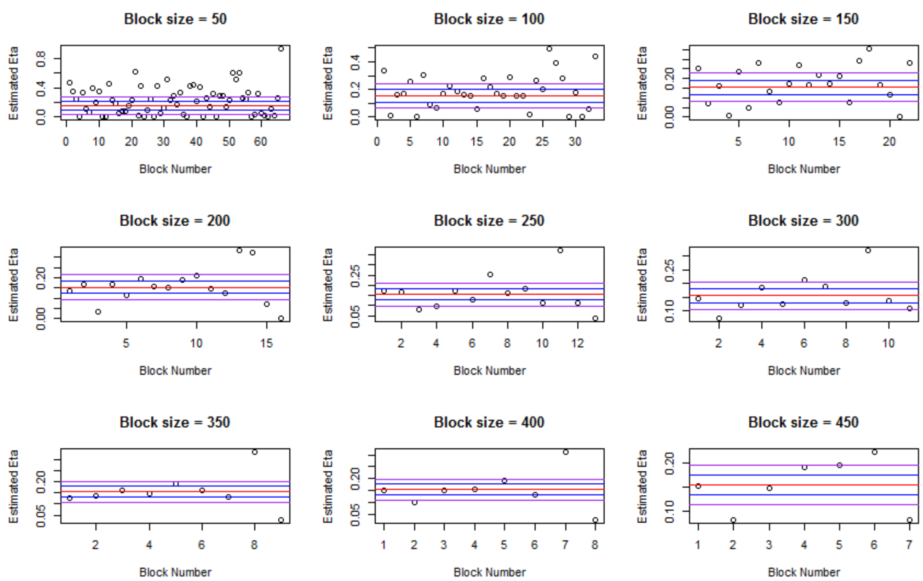


Figure 3.1: Evolution of $\hat{\eta}_{i,n}$ for different values of $h_n$. The red line corresponds to $\hat{\eta}$. The blue lines are one standard deviation $\hat{s}_{h_n}$ away from $\hat{\eta}$. The purple lines are two standard deviations away from $\hat{\eta}$.

| $h_n$ | $B_n$ | 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 | $6e^{-7}$ |

Table 3.1: Summary chi-square statistics based on the block size $h_n$. Note that since the number of observations of the last block is arbitrary, the last block estimate $\hat{\eta}_{B_n,n}$ is not used to compute the chi-square statistic.

We now compute $\widehat{\Theta}_n$ for different values of $h_n$ following the time-varying friction parameter model with uncertainty zones. In view of (3.49), because $N_n^{\frac{1}{2}} \approx 57.5$ we will choose to work with $h_n = 43, \ldots, 63$. We can see in Figure 3.2 that we obtain different estimates of volatility using the techniques of this work compared to the estimates of the model with uncertainty zones, which is one reason why it is crucial to use a proper time-varying model for $\eta_t$. The estimates of the model with uncertainty zones seems to underestimate the integrated volatility. In addition, the RV estimator, which doesn't take account of the microstructure noise, seems to be overestimating the integrated volatility. This is what to be expected and thus indicates that the correction made to the estimated volatility by $\widehat{\Theta}_n$ is reasonable. Finally, the estimates are very similar for different values of $h_n$, which seems to indicate that the method is robust to small variation of the block size.

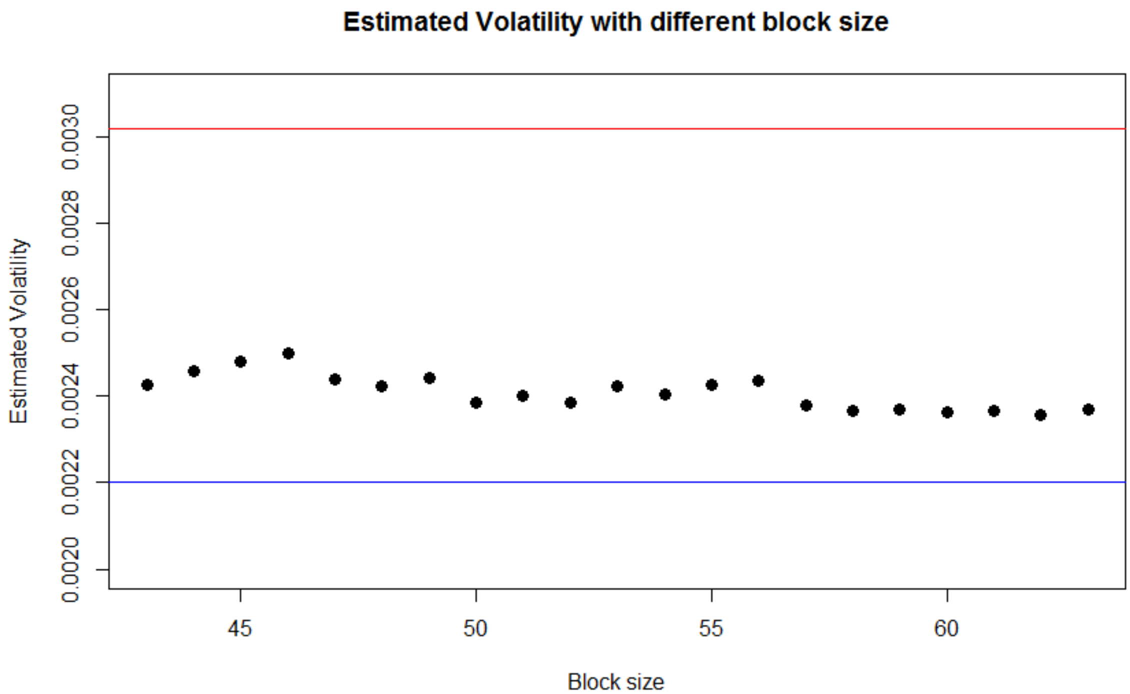


Figure 3.2: Evolution of the estimated volatility $\hat{\sigma}^2_{i,n}$ for different values of $h_n$. The red line corresponds to the RV estimator. The blue line stands for the estimated volatility if we use the volatility estimator $\widehat{RV}_{\alpha_n^{(tick)},T}$ of the model with uncertainty zones.

## 3.9 Conclusion

We have introduced in this chapter a general class of time-varying parameter models called LPM. In particular, if the asset price is assumed to follow a stochastic process, the LPM allows for auto-correlated time-varying noise and correlation between the efficient process, the microstructure noise and the sampling times. If the econometrician has a specific LPM and a parametric estimator at hand, we provided an estimator of the integrated parameter. We also gave conditions under which the econometrician can show the limit distribution.

Depending on the problem, verifying those conditions is not necessarily straightforward. Nonetheless, Chapter 3 simplifies consequently the work of the econometrician because she can solve a nonparametric problem using only a parametric estimator. As a matter of example, we showed that we were able to estimate the integrated volatility when assuming a time-varying model with uncertainty zones.

In the future, we would like to verify the conditions of Section 3.6 with the model introduced in Section 3.7.3 and Section 3.7.4. Also, most examples of this chapter are coming from the efficient price-process point-of-view, but the LPM class is also conjectured to contain numerous point process models used to model the LOB, such as self-exciting point processes. For that purpose, as discussed in Remark 25, we would need to weaken the asymptotic returns assumption.

Chapter 3 was focused on the estimation of the integrated parameter of the LPM. Condition (E6) states that the normalized discrepancy between the estimator on the observed returns and the approximated returns with the same starting value vanishes asymptotically. Roughly speaking, it means that we can see the LPM as a block-constant parameter model. Thus, we conjecture that parametric tests can also be used. As an example, the log-ratio statistic could test for nested models and provide us evidence on the structure driving the returns. This would most likely enable us to investigate the question of presence of noise in, say, 5-minute returns, the question of correlation between efficient price and noise, the asymmetric information problem using the extension of the model of uncertainty zones in Section 2.3.2, the presence of endogeneity etc.

One other possible application of the fact that the LPM can be seen as a block-constant parameter model is in *model selection*. Given data and a set of candidate LPM, on the one hand we could sum up their block-local maximum likelihood functions. Because of the Markov property of the LPM, we would obtain an estimate of their maximum likelihood function on the whole interval $[0, T]$. On the other hand, we could build a measure of the integrated volatility of the parameter $\langle\theta^*, \theta^*\rangle_t$, based on techniques used in Mykland and Zhang (2014). Then, we could generalize the Akaike Information Criterion (AIC) and LPM could be compared on the basis of their maximum likelihood function and a penalization which would include the number of parameters and the volatility of the parameter.

Finally, we also believe that the ideas of Chapter 3 can be useful to *forecasting* at time $T$. The question of estimating the parameter of a time series or more generally parametric model which admit a LPM time-varying extension in order to "plug-in" the value into the forecast model could be modified to the problem of estimating the spot parameter $\theta_T^*$ and using this estimate instead. One idea could be to use a weighted sum of $\hat{\Theta}_n$ with increasing weights so that the information contained in the most recent observations will be used more that the information contained in older observations.

## 3.10 Appendix

### *3.10.1 Preliminaries*

Since $|\theta_t^*|$ is locally bounded and $(\theta_t^*)^+$ is locally bounded away from 0, we can follow standard localisation arguments (see, e.g., pp. $160-161$ of Mykland and Zhang (2012)) and assume without loss of generality that there exists $M > 0$ such that $\theta_t^* \in K_M$ for all $0 \leq t \leq T$. Furthermore, because we assume that the volatility of the parameter $\sigma_t^\theta$ is locally bounded, we can use the same techniques and assume without loss of generality that there exists $\sigma^+ > 0$ such that $\sigma_t^\theta \leq \sigma^+$ for all $0 \leq t \leq T$.

Finally, we fix some notation. In the following of Chapter 3, we will be using $C$ for any constant $C > 0$, where the value can change from one line to the next.

### *3.10.2 Proof of Theorem (consistency)*

Proof $(C1) \Rightarrow$(3.24)

It suffices to show that $(C1)$ implies that

$$\sup_{i\geq 0} \mathbb{E}\left[\left|\widehat{\tilde{\Theta}}_{i,n} - \tilde{\Theta}_{i,n}\right|\right] = o_p(1) \tag{3.62}$$

By (3.11) and (3.23), we can write

$$\left|\widehat{\tilde{\Theta}}_{i,n} - \tilde{\Theta}_{i,n}\right| = g_n(U^1_{i,n}, \ldots, U^{h_n}_{i,n}, \tilde{\Theta}_{i,n})$$

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

$$\mathbb{E}\left|g_n(U^1_{i,n}, \ldots, U^{h_n}_{i,n}, \tilde{\Theta}_{i,n})\right| < \infty.$$

$$\begin{aligned}
\mathbb{E}\left[g_n(U^1_{i,n}, \ldots, U^{h_n}_{i,n}, \tilde{\Theta}_{i,n})\right] &= \mathbb{E}\left[\mathbb{E}\left[g_n(U^1_{i,n}, \ldots, U^{h_n}_{i,n}, \tilde{\Theta}_{i,n})\middle|\tilde{\Theta}_{i,n}\right]\right] \\
&= \mathbb{E}\left[\int g_n(u, \tilde{\Theta}_{i,n})\mu_\omega(du)\right]
\end{aligned}$$

where $\mu_\omega(du)$ is a regular conditional distribution for $(U^1_{i,n}, \ldots, U^{h_n}_{i,n})$ given $\tilde{\Theta}_{i,n}$ (see, e.g., Breiman (1992)). From Condition $(C1)$, we obtain (3.62).

Proof $(C2) \Rightarrow$(3.25)

It is sufficient to show that $(C2)$ implies that

$$\sup_{i\geq 0} \mathbb{E}\left[\left|\hat{\Theta}_{i,n} - \hat{\tilde{\Theta}}_{i,n}\right|\right] = o_p(1) \tag{3.63}$$

By (3.10), (3.11), (3.22) and (3.23), we can write

$$\left|\hat{\Theta}_{i,n} - \hat{\hat{\Theta}}_{i,n}\right| = g_n^{(2)}(U_{i,n}^1, \ldots, U_{i,n}^{h_n}, \{\theta_s^*\}_{\tau_{i-1,n}^0 \leq s \leq \tau_{i,n}^0}, \tilde{\Theta}_{i,n})$$

$$\begin{aligned}
\mathbb{E}\left[|\hat{\Theta}_{i,n} - \hat{\hat{\Theta}}_{i,n}|\right] &= \mathbb{E}\left[\mathbb{E}\left[g_n^{(2)}(U_{i,n}^1, \ldots, U_{i,n}^{h_n}, \{\theta_s^*\}_{\tau_{i-1,n}^0 \leq s \leq \tau_{i,n}^0}, \tilde{\Theta}_{i,n})|\tilde{\Theta}_{i,n}\right]\right] \\
&= \mathbb{E}\left[\int g_n^{(2)}(v, \tilde{\Theta}_{i,n})\mu_\omega(dv)\right] \\
&= o_p(1)
\end{aligned}$$

where $\mu_\omega(dv)$ is a regular conditional distribution for $(U_{i,n}^1, \ldots, U_{i,n}^{h_n}, \{\theta_s^*\}_{\tau_{i-1,n}^0 \leq s \leq \tau_{i,n}^0})$ given $\tilde{\Theta}_{i,n}$ and where we used Condition $(C2)$ in the last equality.

### *3.10.3 Proof of Consistency in Example 7*

Let's show Condition $(C1)$ first. For any $M > 0$, the quantity

$$\left|\hat{\sigma}_{h_n,n}^2\big(F_n(U_{1,n}^1, \sigma^2); \ldots; F_n(U_{1,n}^{h_n}, \sigma^2)\big) - \sigma^2\right|$$

can be uniformly in $\{\sigma^2 \in K_M\}$ bounded by

$$C\Big|\sum_{j=1}^{h_n}(\Delta W_{[\tau_{1,n}^{j-1}; \tau_{1,n}^j]})^2 T^{-1} - 1\Big| \tag{3.64}$$

We can prove that (3.64) tends to 0 in probability as a straightforward consequence of Theorem $I.4.47$ of p.52 in Jacod and Shiryaev (2003).

To show Condition $(C2)$, let $M > 0$ and $\theta_t \in \mathcal{E}_M$. It is sufficient to show that the following quantity goes to 0.

$$nh_n^{-1}\sum_{j=1}^{h_n}\mathbb{E}\left[\left|\big(\theta_0\Delta W_{[\tau_{1,n}^{j-1}; \tau_{1,n}^j]}\big)^2 - \big(\int_{\tau_{1,n}^{j-1}}^{\tau_{1,n}^j}\theta_s dW_s\big)^2\right|\right] \tag{3.65}$$

Using Conditional Burkholder-Davis-Gundy inequality (BDG, see inequality (2.1.32) of p. 39 in Jacod and Protter (2012)), (3.65) can be bounded by

$$Ch_n^{-1}\sum_{j=1}^{h_n}\mathbb{E}\left[\left|\theta_{\tau_{i,n}^0}\Delta W_{[\tau_{i,n}^{j-1};\tau_{i,n}^j]}-\int_{\tau_{i,n}^{j-1}}^{\tau_{i,n}^j}\theta_s dW_s\right|\right] \tag{3.66}$$

We can also bound (3.66) by

$$Ch_n^{-1}\sum_{j=1}^{h_n}\underbrace{(\Delta\tau_{1,n}^j)^{1/2}}_{O(n^{-1/2})}\underbrace{\mathbb{E}\left[\left|\sup_{s\in[\tau_{1,n}^{j-1},\tau_{1,n}^j]}|\theta_0-\theta_s|\right|\right]}_{o_p(n^{-1/2})}$$

where we used BDG another time to obtain $o_p(n^{-1/2})$.

### *3.10.4 Proof of Consistency in Example 8*

$(C1)$ can be shown easily. Similarly $(C2)$ is a direct consequence of the definition in (3.12), (3.13) together with (3.15).

### *3.10.5 Proof of Theorem (central limit theorem)*

## We show (3.37)

We aim at showing that

$$E_{i,n} := \sum_{i=1}^{B_n}\underbrace{n^{\frac{1}{l}}\big(\tilde{\Theta}_{i,n}-\Theta_{i,n}\big)\Delta\mathrm{T}_{i,n}}_{e_{i,n}} \xrightarrow{\mathbb{P}} 0 \tag{3.67}$$

Note that $\mathbb{E}_{\mathrm{T}_{i-1,n}}\big[e_{i,n}^2\big] = 0$ and thus that $E_{i,n}$ is a discrete martingale. We compute the limit of $n^{\frac{2}{l}} \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[e_{i,n}^2\big]$

$$\begin{aligned}
n^{\frac{2}{l}} \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[e_{i,n}^2\big] &= n^{\frac{2}{l}} \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\Big[\Big(\int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}} (\theta_u^* - \theta_{\mathrm{T}_{i-1,n}}^*) du\Big)^2\Big] \\
&\leq C n^{\frac{2}{l}} \sum_{i=1}^{B_n} \underbrace{\Big(\mathbb{E}_{\mathrm{T}_{i-1,n}}\big[(\Delta \mathrm{T}_{i,n})^4\big]\Big)^{\frac{1}{2}}}_{O_p(h_n^2 n^{-2})} \\
&\quad \underbrace{\Big(\mathbb{E}_{\mathrm{T}_{i-1,n}}\Big[\sup_{\mathrm{T}_{i-1,n} \leq s \leq \mathrm{T}_{i,n}} (\theta_s^* - \theta_{\mathrm{T}_{i-1,n}}^*)^4\Big]\Big)^{\frac{1}{2}}}_{O_p(h_n n^{-1})} \\
&= o_p(1)
\end{aligned}$$

where we used Conditional Cauchy-Schwarz in the inequality, (3.41) of Condition $(E1)$ together with BDG inequality to obtain the big taus, and the last equality by (3.48) of Condition $(E4)$. Because we showed that $n^{\frac{2}{l}} \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[e_{i,n}\big]$ tends to 0 in probability, we obtain (3.67).

## We show (3.38)

Without loss of generality, we can assume that the prameter $\theta_t^*$ is a 1-dimensional process. Because the parametric estimator can be biased, in all generality, $A_{i,n} := n^{\frac{1}{l}}\big(\hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}_n^*} - \tilde{\Theta}_{i,n}\big)\Delta \mathrm{T}_{i,n}$ is not the increment term of a discrete martingale. Thus, we need first to compensate it in order to apply usual discrete martingale limit theorems. Let $B_{i,n} = A_{i,n} - \mathbb{E}_{\mathrm{T}_{i-1,n}}[A_{i,n}]$. We want to use Corollary 3.1 of pp. $58-59$ in Hall and Heyde (1980). First, note that by Condition $(E0)$ condition (3.21) in p.58 of Hall and Heyde (1980) is satistfied. We turn now to the two other conditions of the corollary in the two following steps.

*First condition :* We will show in this step that for all $\epsilon > 0$,

$$\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[ B_{i,n}^2 \mathbf{1}_{\{B_{i,n} > \epsilon\}} \right] \xrightarrow{\mathbb{P}} 0 \tag{3.68}$$

The conditional Cauchy-Schwarz inequality gives us that each term of the sum in (3.68) can be bounded by

$$\left( \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[ B_{i,n}^4 \right] \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[ \mathbf{1}_{\{B_{i,n} > \epsilon\}} \right] \right)^{\frac{1}{2}} \tag{3.69}$$

Consider the approximated time $\Delta \tilde{\mathrm{T}}_{i,n} := \sum_{j=1}^{h_n} (\tilde{R}_{i,n}^j)^{(d)}$. Define $\tilde{A}_{i,n} := n^{\frac{1}{l}} \big( \hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}_n^*} - \tilde{\Theta}_{i,n} \big) \Delta \tilde{\mathrm{T}}_{i,n}$ and the compensated quantity $\tilde{B}_{i,n} = \tilde{A}_{i,n} - \mathbb{E}_{\tau_{i-1,n}} \big[ \tilde{A}_{i,n} \big]$. By (3.44), we have

$$\sum_{i=1}^{B_n} \left( \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[ B_{i,n}^4 \right] \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[ \mathbf{1}_{\{B_{i,n} > \epsilon\}} \right] \right)^{\frac{1}{2}}$$

$$= \sum_{i=1}^{B_n} \left( \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[ \tilde{B}_{i,n}^4 \right] \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[ \mathbf{1}_{\{B_{i,n} > \epsilon\}} \right] \right)^{\frac{1}{2}} + o_p(1) \tag{3.70}$$

To show that the right term of the sum in the right-hand side of (3.70) vanishes uniformly, we use regular condition together with (3.41), (3.45) and (3.48). We apply regular conditional distribution together with (3.46) on the left term of the sum in the right-hand side of (3.70). Then, we use the block assumption (3.49) when taking the sum of all the terms in the right-hand side of (3.70) and we can prove (3.68).

*Second condition :* We will prove that

$$\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[ B_{i,n}^2 \right] \xrightarrow{\mathbb{P}} \int_0^T V_{\theta_s^*} ds \tag{3.71}$$

By regular conditional ditstribution and (3.44), we have

$$\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[B_{i,n}^2\right] = \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[\tilde{B}_{i,n}^2\right] + o_p(1)$$

By regular conditional distribution and (3.45), we have that

$$\sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}} [\tilde{B}_{i,n}^2] \quad = \quad h_n^{1-\frac{2}{l'}} n^{\frac{2}{l}-1} \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}} [V_{\theta^*_{\mathrm{T}_{i-1,n}}} \Delta\tilde{\mathrm{T}}_{i,n}] + o_p(1)$$

Using Lemma 2.2.11 of Jacod and Protter (2012) with (3.44), we obtain

$$h_n^{1-\frac{2}{l'}} n^{\frac{2}{l}-1} \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[V_{\theta^*_{\mathrm{T}_{i-1,n}}} \Delta\tilde{\mathrm{T}}_{i,n}\right] = h_n^{1-\frac{2}{l'}} n^{\frac{2}{l}-1} \sum_{i=1}^{B_n} V_{\theta^*_{\mathrm{T}_{i-1,n}}} \Delta\mathrm{T}_{i,n} + o_p(1)$$

We can apply now Proposition I.4.44 (p. 51) in Jacod and Shiryaev (2003) and (3.49) and we get

$$h_n^{1-\frac{2}{l'}} n^{\frac{2}{l}-1} \sum_{i=1}^{B_n} V_{\theta^*_{\mathrm{T}_{i-1,n}}} \Delta\mathrm{T}_{i,n} \quad \to \quad \int_0^T V_{\theta^*_s} ds$$

We are interested in the stable convergence of the sum of $A_{i,n}$ terms, but by using Corollary 3.1 of pp. $58-59$ in Hall and Heyde (1980), we only obtain the stable convergence of the increment martingale terms $B_{i,n}$. We will show now that the sum of the conditional means $S_n := \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[A_{i,n}\right]$ tends to 0 in probability. First, by (3.44), we have $S_n = \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}} \left[\tilde{A}_{i,n}\right] + o_p(1)$. Then, an application of (3.47) together with regular conditional distribution will give us the convergence to 0 of $S_n$.

## We show (3.39)

We want to prove

$$n^{\frac{1}{l}} \sum_{i=1}^{B_n} \big(\hat{\hat{\Theta}}_{i,n}^{\mathbf{M}^0_{i,n}} - \hat{\hat{\Theta}}_{i,n}^{\mathbf{M}^*_n}\big) \xrightarrow{\mathbb{P}} 0 \tag{3.72}$$

We define the conditional expectation of the terms in the sum of (3.72)

$$C_{i,n} := \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[\hat{\hat{\Theta}}_{i,n}^{\mathbf{M}^0_{i,n}} - \hat{\hat{\Theta}}_{i,n}^{\mathbf{M}^*_n}\big].$$

In analogy with the previous part, we can rewrite the left term of (3.72) as

$$n^{\frac{1}{l}}\big(\sum_{i=1}^{B_n} \underbrace{\big(\hat{\hat{\Theta}}_{i,n}^{\mathbf{M}^0_{i,n}} - \hat{\hat{\Theta}}_{i,n}^{\mathbf{M}^*_n}\big) - C_{i,n}}_{D_{i,n}} + \sum_{i=1}^{B_n} C_{i,n}\big)$$

Note that $\sum_{i=1}^{k} D_{i,n}$ is a discrete martingale, and thus to show that it vanishes asymptotically, it is sufficient to show

$$n^{\frac{2}{l}} \sum_{i=1}^{B_n} \mathbb{E}_{\mathrm{T}_{i-1,n}}\big[D_{i,n}^2\big] \xrightarrow{\mathbb{P}} 0 \tag{3.73}$$

Regular conditional distribution, (3.44) and (3.51) implies (3.73). Similarly, regular conditional distribution together with (3.44) and (3.50) enables us to deduce $\sum_{i=1}^{B_n} C_{i,n} \xrightarrow{\mathbb{P}} 0$. Thus, we proved (3.72).

## We show (3.40)

The proof is very similar to the one of Second term, using the Condition ($E6$) instead of Condition ($E5$).

### *3.10.6 Proof of Theorem (Time-varying friction parameter model with uncertainty zones)*

We verify Conditions $(E0)-(E6)$ of Section 3.2.

*Condition* $(E0)$ *:* The continuous information can be defined in this problem such that $X_t, \sigma_t^2, \eta_t, \chi_t$, $W_t'$ are adapted to $\mathcal{J}_t^{(c)}$, where $W_u'$ is an Brownnian motion independent of the other quantities, which was defined in p.11 of Robert and Rosenbaum (2012) and which is used to define the absolute size of the next jump price $L_{i,n}$.

*Condition* $(E1)$ *:* This follows exactly Corollary 4.4 in Robert and Rosenbaum (2012).

*Condition* $(E2)$ *:* This can be shown exactly the same way as Lemma 11.

*Condition* $(E3)$ *:* In this case, we have $l'=2$. Let $M>0$. Because in the model with uncertainty zones $\mathrm{T}_{1,n}^{\theta}$ can be written as a sum of IID variables with a finite fourth moment, we have uniformly in $\theta \in K_M$

$$\mathrm{Var}\left[h_n^{\frac{1}{l'}}\left(\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M}_n^*,\theta};\ldots;R_{h_n,n}^{\mathbf{M}_n^*,\theta};\mathbf{M}_n^*)-\theta\right)\mathrm{T}_{1,n}^{\theta}\right]$$

$$\begin{aligned}
&= \mathrm{Var}\left[h_n^{\frac{1}{l'}}\left(\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M}_n^*,\theta};\ldots;R_{h_n,n}^{\mathbf{M}_n^*,\theta};\mathbf{M}_n^*)-\theta\right)\right]\left(\mathbb{E}\left[\mathrm{T}_{1,n}^{\theta}\right]\right)^2+o_p(h_n^2n^{-2})\\
&= \mathrm{Var}\left[h_n^{\frac{1}{l'}}\left(\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M}_n^*,\theta};\ldots;R_{h_n,n}^{\mathbf{M}_n^*,\theta};\mathbf{M}_n^*)-\theta\right)\right]\mathrm{T}_{1,n}^{\theta}h_nn^{-1}+o_p(h_n^2n^{-2})
\end{aligned}$$

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 $V_\theta$ such that (3.45) is satisfied. Also, (3.46) and (3.47) are straightforward to verify.

*Condition* $(E4)$ *:* We choose $l=2$ and $l'=2$, and $h_n$ which satisfies (3.48).

*Condition* $(E5)$ *:* Because of the symmetry in the model with uncertainty zones, note that for any $\mathbf{M},\mathbf{N}\in\mathcal{M}_{m,n}$, we have

$$\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{M},\theta};\ldots;R_{h_n,n}^{\mathbf{M},\theta};\mathbf{M})=\hat{\theta}_{h_n,n}(R_{1,n}^{\mathbf{N},\theta};\ldots;R_{h_n,n}^{\mathbf{N},\theta};\mathbf{N}).$$

Thus, Condition $(E5)$ is satisfied.

*Condition* $(E6)$ *:* This proof is similar to the proof of (3.67).

### *3.10.7 Estimation of the friction parameter bias and standard deviation in the model with uncertainty zones*

The notation of Section 3.7.1 are *in force*. We define $\hat{s}_n(\eta) := \hat{V}$, where an expression of $\hat{V}$ will be provided at the end of this section. First, we assume that the absolute jump size is constant equal to the tick size, i.e. $L_{i,n} := 1$. In view of (3.58), we have

$$\hat{\eta}_{\alpha,t}^{(m)} := \min\Big(1, \frac{N_{\alpha,t,k}^{(c)}}{2N_{\alpha,t,k}^{(a)}}\Big).$$

We also have by definition that the number of alternations is $N_{\alpha,t,k}^{(a)} = N_n - N_{\alpha,t,k}^{(c)}$. If we assume that $N_n$ is non-random, then

$$N_{\alpha,t,1}^{(c)} \sim Bin(N_n, \frac{2\eta}{2\eta+1}), \tag{3.74}$$

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

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

and the variance as

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

$B$ and $V$ can be computed easily numerically.

If $N_n$ is random, we can work conditionally on $N_n$. Nonetheless, as the sampling times are endogenous, (3.74) is not true in that case. We can still approximate $N_{\alpha,t,1}^{(c)}$ by $Bin(N_n, \frac{2\eta}{2\eta+1})$ if the number of observations is large enough.

# BIBLIOGRAPHY

[1] Aït-Sahalia, Y., J. Cacho-Diaz and R.J.A. Laeven (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics* 117(3), 585-606.

[2] Aït-Sahalia, Y., J. Fan and D. Xiu (2010). High-frequency covariance estimates with noisy and asynchronous financial data, *Journal of the American Statistical Association* 105, 1504-1517.

[3] Aït-Sahalia, Y., J. Fan, C.D. Wang and X. Yang (2014). *The Estimation of continuous and discontinuous leverage effect*, Available at SSRN.

[4] Aït-Sahalia, Y., P.A. Mykland and L. Zhang (2005). How often to sample a continuous-time process in the presence of market microstructure noise, *Review of Financial Studies* 18, 351-416.

[5] Aldous, D.J. and G.K. Eagleson (1978). On mixing and stability of limit theorems, *Annals of Probability* 6, 325-331.

[6] Andersen, T.G. and T. Bollerslev (1998a). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts, *International Economic Review*, 39, 885-905.

[7] Andersen, T.G. and T. Bollerslev (1998b). Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies, *Journal of Finance*, 53, 219-265.

[8] Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001). The Distribution of Realized Exchange Rate Volatility *Journal of the American Statistical Association*, 96, 42-55.

[9] Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003). Modeling and Forecasting Realized Volatility *Econometrica*, 71, 579-625.

[10] Bacry, E., S. Delattre, M. Hoffmann, and J.-F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance* 13(1), 65-77.

[11] Baillie, R. T. and T. Bollerslev (1992). Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics* 52(1), 91-113.

[12] Ball, C. A. and A. Roma (1994). Stochastic Volatility Option Pricing, *The Journal of Financial and Quantitative Analysis*, 29, 589-607.

[13] Barndorff-Nielsen, O.E. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics* 2, 1-37.

[14] Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the co-variation of equity prices with noise and non-synchronous trading, *Journal of Econometrics* 162, 149-169.

[15] Barndorff-Nielsen, O.E. and N. Shephard (2001). Non-gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society* B 63, 167-241.

[16] Barndorff-Nielsen, O.E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society* B 64, 253-280.

[17] Berk, R. H. (1966). Limiting Behavior of Posterior Distributions When the Model Is Incorrect, *Annals of Mathematical Statistics*, 37, 51-58.

[18] Berk, R. H. (1970). Consistency a Posteriori, *Annals of Mathematical Statistics*, 41, 894-906.

[19] Black, Fischer, and Myron Scholes (1973). The pricing of options and corporate liabilities. *The journal of political economy*. 637-654.

[20] Bollerslev, T. (1986). Generalised autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327.

[21] Borodin, A.N. and P. Salminen (2002). *Handbook of Brownian Motion - Facts and Formulae*. Probability and Its applications, Basel : Birkhauser.

[22] Brandt, M.W. and F.X. Diebold (2003). A no-arbitrage approach to range-based estimation of covariances and correlations. Penn Institute for Economic Research Working Paper 9664.

[23] Breiman, L. (1992). *Probability*. Classics in Applied Mathematics 7, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[24] Chen, R. and P.A. Mykland (2015). *Discerning Non-Stationary Market Microstructure Noise and Time-Varying Liquidity in High Frequency Data*. Available at SSRN: http://ssrn.com/abstract=2699927 or http://dx.doi.org/10.2139/ssrn.2699927

[25] Chen, X. and Y. Fan (2006). Estimation of copula-based semiparametric time series models, *Journal of Econometrics*, 130(2), 307-335.

[26] Christensen, K., S. Kinnebrock and M. Podolskij (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data, *Journal of Econometrics* 159, 116-133.

[27] Christensen, K., M. Podolskij and M. Vetter (2013). On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes, *Journal of Multivariate Analysis*, 120, 59-84.

[28] Clinet, S. and N. Yoshida (2015). *Statistical Inference for Ergodic Point Processes and Limit Order Book*. arXiv preprint arXiv:1512.01899.

[29] Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models, *Mathematical Finance* 8, 291-323.

[30] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes, *The Annals of Statistics* 25(1), 1-37.

[31] Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes, *The Annals of Statistics*, 1762-1794.

[32] Dahlhaus, R. and S.S. Rao (2006). Statistical inference for time-varying ARCH processes, *The Annals of Statistics* 34(3), 1075-1114.

[33] Dahlhaus, R. and J.C. Neddermeyer (2014). Online Spot Volatility-Estimation and Decomposition with Nonlinear Market Microstructure Noise Models *Journal of Financial Econometrics* 12(1), 174-212.

[34] Dalalyan, A. and N. Yoshida (2011). Second-order asymptotic expansion for a non-synchronous covariation estimator. *Ann. Inst. Henri Poincaré Probab. Stat.* 47, 748-789.

[35] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1007.

[36] Engle, R.F. and V.K. Ng (1993). Measuring and Testing the Impact of News on Volatility, *Journal of Finance*, 48, 1749-1778.

[37] Epps, T.W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association* 74, 291-298.

[38] Fan, J. and I. Gijbels (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*. Vol. 66. CRC Press.

[39] Fan, J., J. Jiang, C. Zhang and Z. Zhou (2003). Time-dependent diffusion models for term structure dynamics. *Statistica Sinica*, 13(4), 965-992.

[40] Fan, J. and W. Zhang (1999). Statistical estimation in varying coefficient models. *Annals of Statistics*, 1491-1518.

[41] Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* 80 (1), 27-38.

[42] Fisher, R.A. (1922). On the Mathematical Foundations of Theoretical Statistics. Philosophical Transactions of the Royal Society of London, Series A, 222, 309-368.

[43] Fisher, R.A. (1925). Theory of Statistical Estimation. Proceedings of the Cambridge Philosophical Society, 22, 700-725.

[44] Foster, D. and D. Nelson (1996). Continuous record asymptotics for rolling sample variance estimators, *Econometrica* 64, 139-174.

[45] Fukasawa, M. (2010a). Central limit theorem for the realized volatility based on tick time sampling. *Finance and Stochastics* 14, 209-233.

[46] Fukasawa, M. (2010b). Realized volatility with stochastic sampling. *Stochastic Processes and Their Applications* 120, 829-552.

[47] Fukasawa, M. and M. Rosenbaum (2012). Central limit theorems for realized volatility under hitting times of an irregular grid, *Stochastic Processes and Applications* 122 (12), 3901-3920.

[48] Genon-Catalot, V. and J. Jacod (1993). On the estimation of the diffusion coefficient for multidimensional diffusion. *Annales Institut Henri Poincaré Probab. Stat.* 29, 119-151

[49] Gloter, A. and J. Jacod (2001). Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM: Probability and Statistics* 5, 225-242.

[50] Guillaume, D.M., M.M. Dacorogna , R. Dave, U.A. Müller, R. B. Olsen, and O. V. Pictet (1997). From the Bird's Eye to the Microscope : A Survey of New Stylized Facts of the Intradaily Foreign Exchange Markets. *Finance and Stochastics* 1, 95-129.

[51] Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and Its Application.* Academic Press.

[52] Hansen, P.R. and A. Lunde (2006). Realized Variance and Market Microstructure Noise, *Journal of Business and Economic Statistics*

[53] Hastie T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society.* Series B (Methodological), 757-796.

[54] Hayashi, T., J. Jacod and N. Yoshida (2011). Irregular sampling and central limit theorems for power variations: the continuous case. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques* 47, 1197-1218.

[55] Hayashi, T. and S. Kusuoka (2008). Consistent estimation of covariation under nonsynchronicity. *Stat. Inference Stoch. Process.* 11, 93-106.

[56] Hayashi, T. and N. Yoshida (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11, 359-379.

[57] Hayashi, T. and N. Yoshida (2008). Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *Ann. Inst. Statist. Math.* 60, 357-396.

[58] Hayashi, T. and N. Yoshida (2011). Nonsynchronous covariation process and limit theorems. *Stochastic Processes and Applications* 121, 2416-2454.

[59] Heath, D. (1977). Interpolation of martingales. *The Annals of Probability*, 804-806.

[60] Heston, S. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options, *Review of Financial Studies*, 6, 327-343.

[61] Huber,P. J. (1967). The Behavior of Maximum Likelihood Estimates Under Non-standard Conditions. Proceedings of the Fifth Berkeley Symposiumin Mathematical Statistics and Probability. Berkeley: University of California Press.

[62] Hull, J. and A. White (1987). The Pricing of Options on Assets with Stochastic Volatilities, *Journal of Finance*, 42, 281-300.

[63] Jacod, J. (1994). Limit of Random Measures Associated with the Increments of a Brownian Semi-martingale. Technical report, Université de Paris VI

[64] Jacod, J. and P. Protter (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability* 26, 267-307.

[65] Jacod, J. and P. Protter (2012). *Discretization of Processes*. Springer.

[66] Jacod, J. and A. Shiryaev (2003). *Limit Theorems For Stochastic Processes* (2nd ed.). Berlin: Springer-Verlag.

[67] Kalnina, I. and O. Linton (2008). Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *Journal of Econometrics* 147(1), 47-59.

[68] Kalnina, I. (2012). *Nonparametric tests of time variation in betas*. Technical report, University of Montreal, 2012.

[69] Kim, C. J. and Nelson, C. R. (2006). Estimation of a forward-looking monetary policy rule: A time-varying parameter model using ex post data. *Journal of Monetary Economics*, 53(8), 1949-1966.

[70] Koike, Y. (2014). An estimator for the cumulative covolatility of asynchronously observed semimartingales with jumps. *Scandinavian Journal of Statistics* 41, 460-481.

[71] Koike, Y. (2014). Limit theorems for the pre-averaged Hayashi-Yoshida estimator with random sampling. *Stochastic Processes and Applications* 124 (8), 2699-2753.

[72] Koike, Y. (2015). Time endogeneity and an optimal weight function in pre-averaging covariance estimation. Preprint, Available at arXiv: http://arxiv.org/abs/1403.7889v2.

[73] Li, Y., P.A. Mykland, E. Renault, L. Zhang and X. Zheng (2014). Realized volatility when sampling times are possibly endogenous. *Econometric Theory* 30, 580-605.

[74] Lundin, M.C., M.M. Dacorogna and U.A. Müller (1999). Correlation of high frequency financial time series. In P. Lequex (eds), *The Financial Markets Tick by Tick*, pp. 91-126. New York: Wiley.

[75] McCullagh, P. and J.A. Nelder (1989). *Generalized linear models.* London: Chapman & Hall.

[76] Meyn, S.P. and R.L. Tweedie (2009). *Markov Chains and Stochastic Stability.* Cambridge University Press.

[77] Muthuswamy, J., S. Sarkar, A. Low and E. Terry (2001). Time Variation in the Correlation Structure of Exchange Rates: High Frequency Analysis *Journal of Futures Markets* 21, 127-144.

[78] Mykland, P.A., N. Shephard and K. Sheppard (2012). *Efficient and feasible inference for the components of financial variation using blocked multipower variation.* Technical Report, University of Oxford.

[79] Mykland, P.A., and L. Zhang (2006). ANOVA for Diffusions and Itô Processes. *Annals of Statistics* 34, 1931-1963.

[80] Mykland, P.A. and L. Zhang (2008). Inference for Volatility Type Objects and Implications for Hedging, *Statistics and its Interface* 1, 255-278.

[81] Mykland, P.A. and L. Zhang (2009). Inference for Continuous Semimartingales Observed at High Frequency. *Econometrica* 77, 1403-1445.

[82] Mykland, P.A. and L. Zhang (2011). The Double Gaussian Approximation for High Frequency Data. *Scandinavian Journal of Statistics* 38, 215-236.

[83] Mykland, P.A. and L. Zhang (2012). The econometrics of High Frequency Data. In M. Kessler, A. Lindner and M. Sørensen (eds.), *Statistical Methods for Stochastic Differential Equations*, pp. 109-190. Chapman nad Hall/CRC Press.

[84] Mykland, P.A. and L. Zhang (2014). Assessment of Uncertainty in High Frequency Data: The Observed Asymptotic Variance, Working paper.

[85] Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, 59, 347-370.

[86] Ogihara, T. and N. Yoshida. Quasi-likelihood analysis for the stochastic differential equation with jumps. *Statistical Inference for Stochastic Processes* 14 (3), 189-229.

[87] Pötzelberger, K. and L. Wang (2001). Boundary crossing probability for Brownian motion. *Journal of Applied Probability* 38, 152-164.

[88] Renault, E., T. van der Heijden and B.J. Werker (2009). A structural autoregressive conditional duration model. Presented at the 2010 Winter Meetings of the Econometric Society in Atlanta.

[89] Renault, E., T. van der Heijden and B.J. Werker (2014). The dynamic mixed hitting-time model for multiple transaction prices and times. *Journal of Econometrics* 180, 233-250.

[90] Reiss, M., V. Todorov and G. Tauchen (2014). Nonparametric Test for a Constant Beta over a Fixed Time Interval. Discussion paper.

[91] Rényi, A. (1963). On stable sequences of events. *Sanky Series* A 25, 293-302.

[92] Revuz, D. and M. Yor (1999). *Continuous Martingales and Brownian motion*. 3rd ed., GermanyÂ : Springer.

[93] Robert, C.Y. and M. Rosenbaum (2011). A new approach for the dynamics of ultra-high-frequency data: the model with uncertainty zones. *Journal of Financial Econometrics* 9, 344-366.

[94] Robert, C.Y. and M. Rosenbaum (2012). Volatility and covariation estimation when microstructure noise and trading times are endogenous. *Mathematical Finance* 22 (1), 133-164.

[95] Robins, J. and A. A. Tsiatis (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika*, 79(2), 311-319.

[96] Rootzén, H. (1980). Limit distributions for the error in approximations of stochastic integrals. *Annals of Probability* 8, 241-251.

[97] Russell, J.R. (1999). Econometric modeling of multivariate irregularly-spaced high-frequency data. University of Chicago Graduate School of Business Working Paper.

[98] Scholes, M. and J. Williams (1977). Estimating betas from nonsynchronous data. *Journal of Financial Economics* 5, 309-327.

[99] Stein, E. M. and J. C. Stein (1991). Stock Price Distributions with Stochastic Volatility: An Analytic Approach. *Review of Financial Studies*, 4, 727-752.

[100] Stock, J. H. and M. W. Watson (1998). Median unbiased estimation of coefficient variance in a time-varying parameter model. *Journal of the American Statistical Association*, 93(441), 349-358.

[101] Vetter, M. (2011). Estimation of Integrated Volatility of Volatility with Applications to Goodness-of-fit Testing, Discussion paper.

[102] Wang, C.D. and P.A. Mykland (2014). The Estimation of Leverage Effect with High Frequency Data. *Journal of the American Statistical Association*, 109, 197-215.

[103] White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica* 50, 1-25.

[104] Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. Almquist and Wicksell.

[105] Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics* 159, 235-250

[106] Zhang, L. (2001). From Martingales to ANOVA : Implied and Realized Volatility. Ph.D. Thesis, The University of Chicago, Department of Statistics.

[107] Zhang, L. (2011). Estimation covariation: Epps effect, microstructure noise. *Journal of Econometrics* 160, 33-47.

[108] Zhang L. (2012). Implied and realized volatility: Empirical model selection. *Annals of Finance*, 8, 259-275.

[109] Zhang, L., P.A. Mykland and Y. Aït-Sahalia (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* 100, 1394-1411.