The locally parametric model: a new class of models in high frequency data

Yoann Potiron

November 2, 2015

Abstract

This paper proposes mixed parametric and nonparametric statistical techniques for the analysis of high frequency data. It gives a general model, which can be discrete or continuous in time depending on the point-of-view. This model can be seen as a parametric model which allows its multidimensional parameter to follow a local martingale. As such, we call it the *locally parametric model* (LPM). The quantity of interest is defined as the *uniformly weighted value over time* (UWV) of the (discrete or continuous) parameter process. We provide estimators of UWV and conditions under which we can show the consistency and the corresponding central limit theorem. Those estimators are based on estimators of the parametric model when parameters are fixed. Since the estimator is obtained by chopping the data into small blocks, estimating the parameter on each block pretending it is constant locally and taking a weighted mean of the estimates on each block, we call it the *locally parametric quasi-estimator* (LPQE). We show that under conditions, some discrete standard time series models of the literature (for instance ARMA or GARCH models with MLE estimator) as well as continuous semiparametric models (for example a semimartingale asset price model with IID noise component in the observations) of the high-frequency financial econometrics literature belongs to the LPM class of models. This paper thus builds a bridge between various perspectives, parametric, semiparametric and nonparametric as well as discrete and continuous in time models. In addition, statistics to test whether the parameter's constancy hypothesis is true are provided. We also discuss model selection and provide statistics to test for nested models: as an example, this allows us to test if there is noise in observations. Based respectively on the estimate of UWV, we give a new input to use in the prediction model. Finally, an empirical study on S & P 500 daily returns, using ARMA and models is carried out. It shows that the parameters are not constant over time for both models and that we obtain better statistical inference using the new prediction's input of the model.

1 Introduction

1.1 The dilemma of assuming parameter's constancy over time in a parametric model

Modeling dynamics is very important in various fields, such as finance, economics, physics, environmental engineering, geology or even sociology. Parametric time dependent models are tools meant to deal with one type of dynamics, the temporal evolution of systems. There has been an explosion of research in the area in the last decades. We can identify two main reasons why parametric models are very attractive and popular, both for researchers and practitioners. First, by estimating an underlying (possibly multidimensional) parameter, they provide crucial information on the mechanisms of the system of interest. As an example, the fitted parameter of *autoregressive moving average* (ARMA) models (Whittle (1951)) give us insight on the correlation structure of the observations. Also, parametric models usually allow for inference such as prediction of future observations together with confidence intervals, as a function of the data. In particular, if we choose an adequate model, we can predict tomorrow's temperature.

By definition, parametric approaches come with the strong assumption that there exists an underlying parameter, who drives the structure of the observations, and which is fixed over time. In practice, the parametric model user usually tries different types of models, or has a specific class of models in mind, and she fits the models to the data. It means that she estimates the parameter of the model with the observations. Nonetheless, as time goes by, the structure driving the observations is most likely evolving as well. Thus, questions about the constancy of the parameter, that would stay the same through thick and thin, are to be raised. To corroborate this natural skepticism, it can even be the case that empirical work strongly suggests that the assumption of constancy is too restrictive. To acknowledge the issue, one has to build an extended model, that can be either parametric but typically with some more parameters, semiparametric or nonparametric .

Models of the variance of the return terms (also called error terms) followed exactly this path. Originally, we assumed that the returns $R_i := Y_i - Y_{i-1}$ of a time-series $\{Y_1, Y_2, \ldots, Y_n\}$ observed at times $\{\tau_1, \tau_2, \ldots, \tau_n\}$ where $\tau_i = i\tau$ were conditionally homoskedastic with variance parameter σ, and we split them as $R_i = \sigma z_i$ where z_i was a stochastic piece (typically a gaussian with a time-increment variance $\Delta\tau_i := \tau_i - \tau_{i-1}$). On the parametric side, Engle (1982) allowed conditional heteroskedasticity, the variance component following itself a moving-average (MA) model. He was soon imitated by Bollerslev (1986) and many other authors (see, e.g., Nelson (1991), Engle and Ng (1993)), who allowed respectively an ARMA and more general models for the evolution in time of the variance parameter. Taking a nonparametric approach, the analysis of high-frequency financial data gave rise to the model where log prices follow a semimartingale and returns are of the form

$$R_i = \int_{\tau_{i-1}}^{\tau_i} \mu_t dt + \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t \tag{1}$$

where the drift μ_t and the volatility σ_t are random processes. The object of interest, which used to be the fixed volatility parameter, became the *integrated volatility* (IV), defined as

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

This was studied extensively 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). *Generalized autoregressive conditional heteroskedasticity* (GARCH) type models and diffusion models have a significant difference in their point-of-view: the former is discrete whereas the former is a continuous-time model.

The nonparametric approach allows for a very general structure driving the returns. In the past couple of decades, the assumptions on the drift and volatility processes of (1) has been weakened as much as possible. This provides us very robust nonparametric estimators of the RV, but as far as the forecast of the future returns and confidence intervals are concerned, a model not as general as (1) on the future path of drift and volatility is further needed. This is one reason why parametric models are very convenient, they provide prediction values and intervals straightforwardly. As such, the GARCH model is very appealing and behaves very well in practice. Nonetheless, it actually involves a second level of nesting. It still assumes the existence of a fixed parameter over time, which is driving the ARMA structure of the variance term. Can we believe in such a model, which has a non time dependent input, when all the quantities of the “real World” are changing over time ? (see, i.e., Foster and Nelson (1996)) If not, what is an alternative more general model ?

We believe that when facing the problem of having a suspect parametric model, building a new parametric model will only push the issue to the next order. Eventually, doubts about the constancy of those new parameters will rise, and thus the problem is only solved partially.

In this paper, we propose to shed light on this thought by building statistical tests and giving the opportunity to test whether the constancy of the underlying parameter's assumption holds on data for a specific model.

Also, we take a stand in this paper by assuming the parameter to be locally constant. We prove that under this assumption we can trust the parametric model locally. Thus, inference can still be performed (at least locally) using the parametric model, if we can estimate the spot parameter. As a consequence, methods of parameter's estimation has to be updated to take proper account of the nonconstancy over time of the parameter. The idea that locally constant parameter implies the model to be 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 endogeneous, techniques were extended in Potiron and Mykland (2015).

For the sake of simplicity, assume a null-drift in (1). The nonparametric Itô-process model will be shown to be an element of LPM if we consider that volatility is constant locally. Similarly, a zero-mean GARCH model where we allowed the parameters to follow a local-martingale also belongs to the class of LPM. This paper builds a bridge between some time series models, parametric, semiparametric and nonparametric models, with discrete or continuous time setting, by seeing them all as LPM.

When one looks at a model, there is commonly two ways of assessing the quality of it. One can take the engineering approach and believe that the data follows the model, in the sense that it is very robust to various tests to reject it. An alternative approach is, as Faraway (2005) points out for regression problems[1], to recognize that the parameter and its estimate is a fictional quantity in most situations. The "true" value may never be known (if it even exist in the first place). Instead concentrate on predicting future values, these may actually be observed and success can then be measured in terms of how good the predictions were. That you take either point-of-view, this paper will show that the new approach performs better. On top of describing in a more complete way the mechanisms of the system of interest, it will provide directly better forecast values and confidence intervals in the close future. We emphasize that this work is not solely focused on the very specific problem of estimation of volatility or even more general examples available in the financial econometrics literature. It applies to any (possibly semi or non) parametric model, as long as an estimator of the

[1] see the discussion on interpretation of parameter estimates of Chapter 3.10

underlying parameter is known and that they verify the conditions provided in Section 3 of this paper.

1.2 The serious statistical implications of assuming that parameter is constant over time when it's not

We remind to the lector that assuming that θ^* is constant when it's not can rise serious estimation issues. It means that we are using the estimator with the wrong model. In the likelihood theory, ever since Fisher (1922, 1925) has introduced the method of maximum likelihood, a significant body of the literature has taken an interest in the asymptotic behavior of the *maximum likelihood estimator* (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. A MLE computed on data not following the "true model" is called *Quasi Maximum Likelihood Estimator* (QMLE).

Using the analogy, we will refer to *Quasi Estimator* (QE) when an estimator is performed on a misspecified model. In particular, the QMLE is a type of QE. As the previous cited authors showed in their work on QMLE, a QE is no longer necessarily consistent to the new object of interest (see (2) or (5) for examples of such objects of interest). It might converge to a value, but this is not necessarily the one the parametric model user has in mind. Even if the procedure is consistent in the value of interest, the estimated standard deviation will most likely be wrong. Consequently, this leads to wrong inferences, building interval confidences of the wrong size, rejecting or accepting hypothesis with a probability different from the acceptance rate, setting wrong forecast intervals and so on. This is much of the same problem that 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.

Needless to add that QMLE is still a very popular estimator nowadays (see Xiu (2010) among many other papers). Usually, the approach taken is to identify and estimate the eventual bias, obtain a consistent estimator by removing the bias to the QMLE, and then investigate the error's magnitude. This seems like a lot of work, and very often there is still the assumption of underlying constant parameters in the procedure, and their non time dependency is not checked properly. This paper will provide a basis of work for anyone willing to use MLE on wrong models which are following the assumptions of our work.

We insist on the fact that this paper is not solely focused on QMLE. The philosophy is to advocate anyone who has at hands a parametric model and estimator to chop the data into several blocks and to use its estimator locally on each block, where the constancy of the parameters is much less questionable because of our locally constant parameter's assumption.

1.3 The solution: allowing the parameter to follow a local martingale

One goal of this paper is to take a proper account of the possible nonconstancy of parameters in a parametric model. We will allow the parameter to follow a local martingale. We don't include a drift in the structure because if this is the case, it should properly be included in the original parametric model.

We first take the discrete time series perspective. We observe $\{Y_{0,n}, Y_{1,n}, \dots, Y_{n_Y,n}\}$. We assume that $\{R_{1,n}, R_{2,n}, \dots R_{n,n}\}$, which is a function of the observations, of size n, follows the assumptions of the LPM defined in Section 3. In particular, $R_{i,n}$ doesn't need to be stationary or ergodic. The corresponding observation times are $\{\tau_{1,n}, \tau_{2,n}, \dots, \tau_{n,n}\}$, where $\tau_{i,n} := iT/n$ with $T > 0$ the horizon time. In a significant number of applications, $R_{i,n}$ will be defined as the (possibly log) *returns* of the original time-series. As such, we will refer to them as returns in the rest of the paper. Nonetheless, bear in mind that the order of differencing needed to stationarize the time series can be well different from 1, and $R_{i,n}$ can thus be equal to the original time series or an approximate of the differential of order 2 for instance. We assume that the moving parameter $\{\theta^*_{1,n}, \dots, \theta^*_{n,n}\}$ is a local martingale. Instead of the constant parameter θ^*, the object of interest is

$$\Theta_n^{(dis)} := \frac{1}{n}\sum_{i=1}^{n} \theta_i^* \tag{3}$$

We keep the discrete time model point-of-view, but assume now that observation times can be random, and dependent of the other quantities, we call it *endogeneity.* We observe $\{(R_{1,n}, \Delta\tau_{1,n}), \dots, (R_{1,n}, \Delta\tau_{N_n,n})\}$ coming from a parametric model, where N_n is the (random) number of observations. We will seek to infer

$$\Theta_n^{(dis)} := \frac{1}{T}\sum_{i=1}^{n} \theta_i^* \Delta\tau_{i,n} \tag{4}$$

If the parametric model user believes in a continuous time setting, then the parameter process θ_t^* can be defined for all time $0 \leq t \leq T$. Our object of interest is the estimation of the integral of *parameter* spot process

$$\Theta := \frac{1}{T} \int_0^T \theta_s^* ds \tag{5}$$

If we assume that in the discrete time model θ_i^* is the interpolation of a continuous parameter process at its time of observation $\theta_{\tau_i}^*$, we have

$$\Theta_n^{(dis)} = \Theta + o_p(1) \tag{6}$$

and thus regardless of the point-of-view taken, our object of interest will be Θ. The *uniformly weighted value over time* (UWV) or integrated parameter (5) is to be compared to the IV defined in (2). If the underlying parameter has a physical interpretation such the volatility has, it is natural to have an interest in (5). On the other hand, even if the parameter doesn't have a direct physical interpretation, it still makes sense to infer about the quantity (5): one reason is that a slight modification of its estimate $\widehat{\Theta}_n$ will provide better inference about the spot parameter θ_T^* and thus allow us to make better predictions.

By choosing a continuous local martingale parameter process in our model, we will provide a parameter's estimate $\widehat{\Theta}_n$ which is more robust to observed variation over time in data. The model could be more general by including jumps, but this is already a substantial step in the parametric model literature, where researchers have been assuming that the parameter is constant over time.

1.4 Applications

1.4.1 Theoretical Applications

The consistency of (5) under very simple assumptions is done in Section 2. The main result of this paper, a central limit theorem of (5) under the assumption that the parameter process is a null-drift continuous Itô-process can be found in Section 3. A (non-exhaustive) list of examples is provided in Section 7, where we give an extended model allowing nonconstancy of parameter as well as an estimator of the integrated parameter, the asymptotic variance and its corresponding rate of convergence.

1.4.2 Practical Applications

We give here the practical route to follow in the rest of the paper. First of all, the LPM user should try to implement the tests of Section 2 on various models she has in mind. If she gets suspicious at the constancy of parameter's assumption based on the tests, she can use the techniques of this paper. She will have the possibility to compare models in Section 6. In Section 7, a prediction model is given. It is exactly the same model as under the constancy of parameter's assumption, except that we use estimates of the mean (5) as input instead of estimates of the constant parameter. Note that even if the paremeter is constant, the criteria of comparison and the forecasts of this paper should be roughly as good as the ones used on the original models. In addition, it is very straightforward to implement. We provide constancy of a subset of the p-dimensional parameter's test in Section 8. Numerical simulations on forecasting are carried out in Section 10. Section 11 gives empirical results for S & P 500 daily returns, fitting ARMA and GARCH models.

2 Outline of the problem

2.1 Asymptotics : High-frequency and Low-frequency and bridge between two point-of-views

In our asymptotics, we are assuming that $\Delta\tau_n := \sup_{0<\tau_{i,n}\leq T} \Delta\tau_{i,n}$ tends to 0. This looks like a setting of high frequency inference. Nevertheless, this paper embraces all kind of time-series, such as the month-to-month GDP also (or something else), as long as we have enough points (to be more specific about this). Because we have only observed the data from 0 to T (0 can be year 1950 and T year 2015), we are only interested in the estimation of the integrated θ_t^* from 0 to T with T beeing fixed. Using the techniques of this paper, if we find that θ_t^* seems constant over time, then it makes sense to use a low-frequency central limit theorem by sending T to ∞. But if θ_t^* seems to have a non-zero volatility, then there is no real reason to send T to ∞ and to try to estimate something like the normalized integrated value because we have no idea what the future of the volatility of σ_t^* will look like.

For instance, if we are forecasting, we should only take account of the estimation of θ at time t, and its volatility to get the next confidence interval.

Think about p. 25 of Hamilton's book, why this would be a bit different in this case, why we want the integrated of the spot process.

Also think about : if θ^* is fixed, is there equivalence between our asymptotic and the usual asymptotic ?

2.2 Set-up

To fix ideas, for Section 3 and 4, we will assume that the parameter process θ_t^* is a local martingale of the form

$$d\theta_t^* = \sigma_t^\theta dW_t^\theta \tag{7}$$

where σ_t^θ is a random process (of dimension $p \times p$), and W_t^θ a standard p-dimensional Brownian motion. Furthermore, θ_t^* is locally bounded and restricted to lie in K. The lector which is interested in knowing the general continuous semimartingale theory can look at Section 5.

We focus on simple setting in this section. First, we work with real-valued returns $R_{i,n}$. Also, we assume that the observations occur at equidistant time interval $\Delta\tau_n := \frac{T}{n}$, so that $\tau_{i,n} = \frac{i}{n}T$. Furthermore, the parametric model's user mistakenly thinks that given the true parameter θ^*, $R_{i,n}$ are independent and identically distributed (IID). She is wrong because there is no true fixed parameter θ^*, only a true moving parameter θ_t^* that drives the returns. Since θ_t^* is in fact unfixed, $R_{i,n}$ are neither identically distributed nor independent. There are not even necessarily conditionnally independent given the true parameter process θ_t^*, as we can see on the following three toy examples.

Example 1. *(estimating volatility with time series model) Consider that $\theta_{i,n}^* = (\sigma_{i,n}^2)^*$ where $\sigma_{i,n}$ is the interpolation of a null-drift Itô-process as in ??. The real model (which allows the parameter to move over time) assumes that $R_{i,n} = \sigma_{i,n} N_{i,n}$, where $N_{i,n}$ is an IID sequence of normal distribution with mean zero and variance $\Delta\tau_{i,n}$. The time series user mistakenly believes that the distribution of returns is $R_{i,n} = \sigma^* N_{i,n}$, with σ^* the true volatility. This is a toy example because the time series user's model doesn't assume any movement in the volatility parameter. $R_{i,n}$ are IID if we trust the time series user's model. Under the more general model, they are neither independent nor identically distributed.*

Example 2. *(estimating volatility with continuous time model) Consider that $\theta_t^* = (\sigma_t^2)^*$ (the volatility is thus assumed to follow (7)), $R_{i,n} = \int_{\tau_{i-1,n}}^{\tau_{i,n}} \sigma_s^* dW_s$, where W_t is a standard Brownian motion. Also, the parameter is restricted to taking positive values, i.e. the parameter space $K = \mathbb{R}_*^+$. This is the so-called estimation of volatility in the no-noise case. The econometrician mistakenly thinks that the distribution of the returns is $R_{i,n} = \sigma^* \Delta W_{\tau_{i,n}}$, where σ^* is the fixed volatility. Under her assumption, the returns are IID. Under the real model, $R_{i,n}$ are clearly not IID, and they are also not conditionally independent given the whole volatility process if there is leverage-effect (see i.e. Wang and Mykland (2014), Aït-Sahalia et al. (2013))*

Example 3. *(estimating the rate of a Poisson process) Suppose a sociologist observes data on crimes commited in a given city, and thinks that in first approximation, N_t follows a homogeneous Poisson process with rate λ^*, where N_t is the number of crimes commited between* 0 *and t. Because she doesn't have access to the raw data, she can't observe directly the exact time of each crime. Instead, she only observes the number of crimes commited on a period (for instance a day) $[\tau_{i-1,n}, \tau_{i,n}]$, that is $Y_{i,n} = N_{\tau_{i,n}} - N_{\tau_{i-1,n}}$. This is an example where the series is already stationary, and thus $Y_{i,n} = R_{i,n}$. If the assumption on homogeneousity of the sociologist is true, the returns are IID. In case of heterogeneity, the parameter rate λ_t^* will be assumed to follow (7), N_t will be a nonhomogenous Poisson process, and the returns $R_{i,n}$ will be neither identically distributed nor independent.*

We believe that the parametric model's user had the good intuition in the sense that locally, her model of the returns is not too far from the real model. Formally, it means that if we know the parameter's initial value θ_0^*, then there exists an approximation for $i = 1, \cdots, h_n$ of the returns $R_{i,n}$, denoted $\tilde{R}_{1,i,n}$, that is conditionally IID given θ_0^*, and very close to $R_{i,n}$ 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, the parametric model's user can apply her estimator to the observed returns $R_{i,n}$. We thus obtain an estimate $\widehat{\Theta}_{1,n}$ of the value of the parameter at time 0 (that we call $\tilde{\Theta}_{1,n} := \theta_0^*$). We define the spot parameter's average on the $i-$th block as

$$\Theta_{i,n} := \frac{\int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}} \theta_s^* ds}{\Delta \mathrm{T}_{i,n}} \tag{8}$$

where $\mathrm{T}_{i,n} := \min\{\tau_{ih_n}, T\} = \min\{\frac{ih_n}{n}T, T\}$. Since the block size is very small, the value of $\tilde{\Theta}_{1,n}$ is approximately equal to the average of the spot parameter on the first block $\Theta_{1,n}$. Let $B_n := \ulcorner n h_n^{-1} \urcorner$ be the number of blocks. For $i = 2, \cdots, B_n$, we estimate in the same way the $i - th$ block's initial value of the parameter $\tilde{\Theta}_{i,n} := \theta^*_{\mathrm{T}_{i-1,n}}$ and we call this estimator $\widehat{\Theta}_{i,n}$. Then, we take the weighted sum of $\widehat{\Theta}_{i,n}$ and obtain an estimator of the integrated spot process

$$\widehat{\Theta}_n = \frac{1}{T} \sum_{i=1}^{B_n} \widehat{\Theta}_{i,n} \Delta \mathrm{T}_{i,n} \tag{9}$$

Note that each block includes exactly h_n observations, except for the last one who might include less of them. We call our new estimator (9) the *locally parametric quasi-estimator* (LPQE) of the estimator of the parametric model's user, since we are mistakenly specifying the distribution of the returns on small blocks and we are estimating with her parametric estimator on each of them.

2.3 Consistency of the estimator

In the following of this paper, we will make the number of observations inside a block go to infinity

$$h_n \to \infty \tag{10}$$

Furthermore, we will make the size of each block vanish asymptotically. Because we assumed observations occur at equidistant time, this can be expressed as (this is valid only for this section)

$$h_n n^{-1} \to 0 \tag{11}$$

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

$$\sum_{i=1}^{B_n} \big(\widehat{\Theta}_{i,n} - \Theta_{i,n}\big)\Delta \mathrm{T}_{i,n} \overset{\mathbb{P}}{\to} 0 \tag{12}$$

The basic insight to show (12) is that 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

$$\widehat{\Theta}_{i,n} - \Theta_{i,n} = \big(\widehat{\Theta}_{i,n} - \widehat{\tilde{\Theta}}_{i,n}\big) + \big(\widehat{\tilde{\Theta}}_{i,n} - \tilde{\Theta}_{i,n}\big) + \big(\tilde{\Theta}_{i,n} - \Theta_{i,n}\big) \tag{13}$$

where $\widehat{\tilde{\Theta}}_{i,n}$ is the estimator of the parametric model's user used on the underlying non-observed approximated time-series. It is not a feasible estimator and appears in (13) only to help compute the consistency of the estimator. We first cope with the last error term of (13), which is due to the non-constancy of the spot parameter θ_t^*. Note that

$$\sum_{i=1}^{B_n} \big(\tilde{\Theta}_{i,n} - \Theta_{i,n}\big)\Delta \mathrm{T}_{i,n} = \sum_{i=1}^{B_n} \Big(\theta^*_{\mathrm{T}_{i-1,n}} \Delta \mathrm{T}_{i,n} - \int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}} \theta_s^* ds\Big) \tag{14}$$

and thus we deduce from Riemann-approximation[2] that

$$\sum_{i=1}^{B_n} \big(\tilde{\Theta}_{i,n} - \Theta_{i,n}\big)\Delta \mathrm{T}_{i,n} \overset{\mathbb{P}}{\to} 0 \tag{15}$$

To deal with other terms of (13), we need to introduce some definitions. On a given block $i = 1, \cdots, B_n$ the observed returns will be called $R_{i,n}^1, \cdots, R_{i,n}^{h_n}$. Formally, it means that

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

$R^j_{i,n} = R_{(i-1)h_n+j,n}$ 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 ith 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 increment of time between the $(j-1)$th return and the jth return of the ith block as $\Delta\tau^j_{i,n} = \tau^j_{i,n} - \tau^{j-1}_{i,n}$. We assume that

$$R^j_{i,n} = F_n\big(U^j_{i,n}, \{\theta^*_s\}_{\tau^{j-1}_{i,n} \leq s \leq \tau^j_{i,n}}\big) \tag{16}$$

$$\tilde{R}^j_{i,n} = F_n\big(U^j_{i,n}, \tilde{\Theta}_{i,n}\big) \tag{17}$$

where $U^j_{i,n}$ take values on a space that can be functional[3] and that can depend on n, $U^j_{i,n}$ are IID for a fixed n but the distribution can depend on n, $F_n(x,y)$ is a non-random function[4]. For a block $i = 1, \ldots, B_n$ and for the observation time $j = 0, \ldots, h_n$ of the $i-th$ block, we define $\mathcal{I}_{i,j,n}$[5] the information up to time $\tau^j_{i,n}$. Also, we assume that $U^j_{i,n}$ is independent of the past information[6] (and in particular of $\tilde{\Theta}_{i,n}$). Furthermore, the asymptotics (explain that we take a stand by scaling the structure of the returns) is such that there exists a sequence α_n with for all $\theta \in K$

$$\alpha_n F_1(U_{i,j,1}, \theta) \overset{\mathcal{D}}{=} F_n(U_{i,j,n}, \theta) \tag{18}$$

For any positive integer k, the parametric model's user has at her hands 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 of n in the following sense

$$\hat{\theta}_{k,n}(r_{1,n}; \ldots; r_{k,n}) \overset{\mathcal{D}}{=} \hat{\theta}_{k,1}(\alpha_n^{-1} r_{1,n}; \ldots; \alpha_n^{-1} r_{k,n}) \tag{19}$$

This makes perfect sense to use the same estimator on a scaled version (which depends of the sampling frequency) of the returns in light of (18). As we can only look at the h_n observed returns on each block $i = 1, \ldots, B_n$, we estimate the local parameter as

$$\widehat{\Theta}_{i,n} := \hat{\theta}_{h_n,n}\big(R^1_{i,n}; \ldots; R^{h_n}_{i,n}\big) \tag{20}$$

[3] $U^j_{i,n}$ take values on a Borel space

[4] $F_n(x,y)$ is jointly measurable real-valued function such that $\mathbb{E} \mid R^j_{i,n} \mid < \infty$ and $\mathbb{E} \mid \tilde{R}^j_{i,n} \mid < \infty$. Moreover, the advised lector will have seen that à priori $\{\theta^*_s\}_{\tau^{j-1}_{i,n} \leq s \leq \tau^j_{i,n}}$ is a process in (16) whereas $\tilde{\Theta}_{i,n}$ is only a vector in (17). We match the definitions by seeing $\tilde{\Theta}_{i,n}$ as a process with constant values.

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

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

The non-feasible estimator $\hat{\tilde{\Theta}}_{i,n}$ is defined as the parametric model's user's estimator, with approximated returns as input instead of observed returns.

$$\hat{\tilde{\Theta}}_{i,n} := \hat{\theta}_{h_n,n}\big(\tilde{R}_{i,1,n}; \dots; \tilde{R}_{i,h_n,n}\big) \tag{21}$$

(21) is unfeasible because the approximated returns are non-observable quantities. Let $M > 0$. We define $K_M = \{x \in K \ , \ \|x\|_1 \leq M\}$ the subset of K dominated by M. In order to obtain the consistency of (9), we make the assumption that the parametric model's user's estimator is $\mathbf{L}^1$-convergent, locally uniformly in the parameter of the model θ if she actually observes returns coming from the parametric model, that we can express as

Condition (C1). *For any $\epsilon > 0$ and any $M > 0$*

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

Remark 1. Note that $\mathbf{L}^1$-convergence is slightly stronger than the consistency. Nonetheless, in most applications, we will observe both.

Under $(C1)$, 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.

$$\sum_{i=1}^{B_n} \big(\hat{\tilde{\Theta}}_{i,n} - \tilde{\Theta}_{i,n}\big)\Delta \mathrm{T}_{i,n} \xrightarrow{\mathbb{P}} 0 \tag{22}$$

To deal with the first term of (13), we need to make sure that we can control the discrepancy between the estimate made on the returns of the observed time series and the estimate made on the underlying approximation, uniformly in the parameter θ and in the information that we have. Indeed, the returns we observe depend not only on the parameter $\tilde{\Theta}_{i,n}$, but also on the past information. For instance, it depends on the volatility of the spot parameter at the initial point of the $i-th$ block $\theta^*_{\tau^0_{i,n}}$. We make the new following assumption.

Condition (C2). *For any $M > 0$, define $\mathcal{E}_M$ the product space of information, initial parameter value, and parameter process value $(\mathcal{J}_{i,j,n}, \theta, \chi_t)$[8], where $\theta \in K_M$, χ_t is a null-drift continuous Itô-process with $\chi_t \in K_M$ for all $0 \leq t \leq T$, and its initial value equal to θ ($\chi_0 = \theta$). We have*

$$\sup_{(\mathcal{J}^j_{i,n}, \theta, \chi_t) \in \mathcal{E}_M} \mathbb{E}_{\mathcal{J}^0_{1,n}}\Big[\big|\hat{\theta}_{h_n,n}(F_n(U^1_{1,n}, \theta), \dots, F_n(U^{h_n}_{1,n}, \theta))$$

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

[8] $\mathcal{J}^j_{i,n}$ can be any filtration of $(\Omega, \mathcal{F}, P)$, sorted as $\mathcal{J}_{1,0,n}, \dots, \mathcal{J}_{1,h_n,n}, \mathcal{J}_{2,1,n}, \dots, \mathcal{J}_{2,h_n,n}, \dots$, such that $U^j_{i,n}$ and $\{\chi_s\}_{0 \leq s \leq \tau^j_{i,n}}$ are adapted to $\mathcal{J}^j_{i,n}$, and $U^j_{i,n}$ is independent of the past.

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

$(C2)$ implies[9] that the error due to the local model's approximation of the returns vanishes in the limit, i.e.

$$\sum_{i=1}^{B_n}(\widehat{\Theta}_{i,n}-\widehat{\widetilde{\Theta}}_{i,n})\Delta \mathrm{T}_{i,n} \xrightarrow{\mathbb{P}} 0 \tag{23}$$

We can now give the theorem on consistency in this very simple case where observations occur at equidistant time intervals and returns of the approximated time-series are IID on each block.

Theorem (consistency). *Under $(C1)$, $(C2)$ and all the definitions of this section, we have the consistency of (9), i.e.*

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

We obtain the consistency in the two toy examples[10]

Example 4. *We continue Example 1. We identify the quantities $U^j_{i,n} = N_{ih_n+j,n}$ as the value of the normal variables. The returns are defined as $R^j_{i,n} = \sigma^*_{ih_n+j,n}N_{ih_n+j,n}$ and the approximated returns as $\tilde{R}^j_{i,n} = \sigma^*_{ih_n,n}N_{ih_n+j,n}$. The approximated returns use the volatility at the starting time of the ith block, whereas the returns use the current volatility. We can construct easily F_n easily as in (16) and (17). Also, because we assume that $U^j_{i,n}$ are normally distributed with null-mean and variance $\Delta\tau_{i,n}$, we have (18) with $\alpha_n = n^{-\frac{1}{2}}$. The estimator is the scaled usual RV, i.e. $\hat{\theta}_{k,n}(r_{1,n};\dots;r_{k,n}) = T^{-1}k^{-1}n\sum_{j=1}^k r^2_{j,n}$. Note that it can be seen also as the MLE (see the discussion pp. 112-115 of Mykland and Zhang (2012))*

Example 5. *We are back to Example 1. In this case, $U^j_{i,n} = \{\Delta W_{[\tau^{j-1}_{i,n},s]}\}_{\tau^{j-1}_{i,n}\le s\le \tau^j_{i,n}}$ are the Brownian motion increment process between two consecutive observation times. The returns are naturally 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, holding the variance constant on each block. It is clear that we can choose a unique F_n that generates the returns and the approximated returns from the parameter and $U^j_{i,n}$ as in (16) and (17). Because of the scaling property of Brownian motions, we obtain (18) with $\alpha_n = n^{-\frac{1}{2}}$. The estimator is the scaled usual RV, i.e. $\hat{\theta}_{k,n}(r_{1,n};\dots;r_{k,n}) = T^{-1}k^{-1}n\sum_{j=1}^k r^2_{j,n}$*

[9] see Appendix for more details
[10] see Appendix for proofs

Example 6. *This is a continuation of Example 2. To be able to estimate the integrated value of parameter (5), we assume that in the asymptotic the rate of the inhomogeneous Poisson process is $\beta_n \lambda_t$, where β_n is a non-time dependent and non-random quantity, that tends to infinity. $U^j_{i,n}$ can be defined as standard Poisson processes $\{N^{i,j,n}_t\}_{t \geq 0}$, independent of each other. The returns are the time-changed Poisson processes*

$$R^j_{i,n} = N^{i,j,n}_{\int_{\tau^{j-1}_{i,n}}^{\tau^j_{i,n}} \beta_n \lambda^*_s ds} \quad (24)$$

$$\tilde{R}^j_{i,n} = N^{i,j,n}_{\beta_n \Delta \tau^j_{i,n} \lambda^*_{\tau^0_{i,n}}} \quad (25)$$

where $\beta_n \Delta \tau^j_{i,n} = 1$, so that (18) is satisfied with $\alpha_n = 1$. The estimator to be used is the mean of returns $\hat{\theta}_{k,n}(r_{1,n}; \dots; r_{k,n}) = k^{-1} \sum_{j=1}^k r_{j,n}$

Remark 2. By Girsanov's theorem, together with local arguments (see, i.e., 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 3. The advised lector will have noticed that in both examples, if the parametric model's user trusts constant-volatility model (resp. homogeneous Poisson process model), she will end up with the same estimator as (9). This is because in those very basic examples, the estimator is linear, i.e. for any positive integer k and $l = 1, \dots, k-1$

$$\hat{\theta}_{k,n}(r_{1,n}; \dots; r_{k,n}) = \hat{\theta}_{l,n}(r_{1,n}; \dots; r_{l,n}) + \hat{\theta}_{k-l,n}(r_{l+1,n}; \dots; r_{k,n}) \quad (26)$$

In more general examples of Section ??, (26) will break, and we will obtain two distinct estimators.

2.4 How to build stochastic-parameter models from any parametric model ?

We've seen in the volatility example that there is a natural way of extending the model bla bla. How do we do that in the case of a general model. One easy way is to use rewrite the equation with F_n in the following way bla bla...

2.5 Challenges

There are several provisos to the presentation above. The first is that the structure of the approximate returns need not be IID. If the approximate serie follows an ARMA process or any other general kth-step Markov chain, we will still obtain consistency of (9), under conditions that are stated in Section 3. Also, we will deal with noisy observations. It means that we will observe

$$Y_{i,n} = X_{i,n} + \epsilon_{i,n}$$

(bla bla bla to be more specific about it). Moreover, we will allow for endogeneity in observation times.

Consistency is a very good limit property but as we want to know the magnitude of the error, we will also investigate the associated central limit theorem. Conditions similar to $(C1)$ and $(C2)$ will be stated, so that the parametric model's user can verify by himself if he can use the techniques of this paper on his favorite estimator. bla bla...

3 Estimation of the integral of parameter process in general case

block assumption (to add somewhere in the text)

$$n^{\frac{2}{l}-1} h_n^{1-\frac{2}{l'}} = O(1) \tag{27}$$

assumption on observation times (to add somewhere in the text: actually, make its own assumption with any type of volatility on parameters)

$$\sup_{\theta \in K} | \operatorname{Var} \left(h_n^{1/l'} (\hat{\theta}_{h_n,n}(\tilde{R}_{1,n}^{\mathbf{q}^*,\mathbf{r}^*,\theta}; \dots ; \tilde{R}_{h_n,n}^{\mathbf{q}^*,\mathbf{r}^*,\theta}; \mathbf{q}^*; \mathbf{r}^*) - \theta) n^{-1} h_n^{-1} \sum_{i=1}^{h_n} R_{i,n}^{(d_r)} \right) \tag{28}$$

$$\tag{29}$$

$$- \operatorname{Var} \left(h_n^{1/l'} (\hat{\theta}_{h_n,n}(\tilde{R}_{1,n}^{\mathbf{q}^*,\mathbf{r}^*,\theta}; \dots ; \tilde{R}_{h_n,n}^{\mathbf{q}^*,\mathbf{r}^*,\theta}; \mathbf{q}^*; \mathbf{r}^*) - \theta) n^{-1} h_n^{-1} \sum_{i=1}^{h_n} \tilde{R}_{i,n}^{(d_r)} \right) | \quad \to \quad 0$$

Technical assumption to add at the end (Assumption 5 ??) : see P216.

In this section, the returns are d_r-dimensional vectors $R_{i,n} := (R_{i,n}^{(1)}, \dots, R_{i,n}^{(d_r)})$ where d_r can be any positive integer. We assume that the last component of any return is the difference between the last two sampling times, i.e. $R_{i,n}^{(d_r)} = \Delta\tau_{i,n}$. Let m be a nonnegative integer, the general assumption of this section is that $R_{i,n}$ is an homogeneous partially observed 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. Specifically, the parametric model user supposedly thinks that $(Q_{i,n}, R_{i,n})$ is an homogeneous Markov chain, whereas under the true model it is not even a nonhomogeneous Markov chain. Nonetheless, we will see in (31) that it is almost a nonhomogeneous Markov chain, except for the evolution of the parameter θ_t^* part which is not necessarily Markovian. As an example, if we fix the volatility σ_t^θ of θ_t^* to be a constant, $(Q_{i,n}, R_{i,n})$ is an nonhomogeneous Markov chain. In the following, we will use the expression "Markov chain" but the lector should understand "Markov chain under the (wrong) parametric model" or even "locally Markov chain".

$Q_{i,n}$ can be seen as the efficient return we can't see and $R_{i,n}$ is the noisy corresponding version we get to observe. It can well be the case that the integrated parameter Θ we want to infer about is only related to the evolution of $Q_{i,n}$ and that the noise $\epsilon_{i,n} = R_{i,n} - Q_{i,n}$ is independent of the efficient return $Q_{i,n}$. In this case, it would be preferable to observe directly $Q_{i,n}$, but we can only use the noised returns $R_{i,n}$ to compute $\widehat{\Theta}$. We can also imagine a situation where the parameter θ_t^* is only dependent of the noise, dependent of both the efficient return $Q_{i,n}$ and the noise, or even where $Q_{i,n}$ is not the efficient return but something else.

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 elements, and assume that $M_{i,n}$ takes values on the space $\mathcal{M}$, which is a subset of $\mathbb{R}^d$. Also, we define the m-dimensional vector of Markov chain elements $\mathbf{M}_{i,n} := (M_{i,n}, \ldots, M_{i-(m-1),n})$ and the m initial values $M_{(m-1),n}, \ldots, M_{0,n}$ of the Markov chain. We consider for any i nonnegative integer $\mathbf{N}_{i,n} := \mathbf{M}_{im,n}$ the Markov chain of dimension $d * m$, which consists of blocks of m elements of the original Markov chain $\mathbf{M}_{i,n}$. By construction, $\mathbf{N}_{i,n}$ is a Markov chain of order 1, which takes values on a space $\mathcal{M}_m$ (subset of $\mathcal{M}^m$). We assume that the parametric model user thinks that given the true parameter value θ^*

$$M_{i,n} = F_n(\mathbf{M}_{i-1,n}, U_{i,n}, \theta^*) \tag{30}$$

where $F_n(x, y, z)$ is a $\mathbb{R}^d$-valued non-random function[11], $U_{i,n}$ are IID for a fixed n but the distribution can depend on n. The truth is that

$$M_{i,n} = F_n(\mathbf{M}_{i-1,n}, U_{i,n}, \{\theta_s^*\}_{\tau_{i-1,n} \leq s \leq \tau_{i,n}}) \tag{31}$$

In analogy with (18) of Section 2, we keep asymptotically the structure of the returns by scaling the distribution of $M_{i,n}$. Formally, it means that there exists a d-dimensional α_n such

[11]Consider $\mathcal{X}_n$ the space on which the third component of $F_n(x, y, z)$ is defined. We assume that $F_n(x, y, z)$ is a jointly measurable $\mathbb{R}^d$-valued function such that for any $\mathbf{M} \in \mathcal{M}_m$ and any $\chi_n \in \mathcal{X}_n$, we have $\mathbb{E} \mid F_n(M, U_{i,n}, \chi_n) \mid < \infty$

that for all $\mathbf{M} \in \mathcal{M}_m$ and $\theta \in K$, we have

$$\alpha_n * F_1(\mathbf{M}, U_{1,1}, \theta) \stackrel{\mathcal{D}}{=} F_n(\mathbf{M}, U_{1,n}, \theta) \tag{32}$$

where the operation $*$ defines the component-multiplication, i.e. $\alpha * \beta := (\alpha^{(1)}\beta^{(1)}, \ldots, \alpha^{(m)}\beta^{(m)})$. We are investigating the normalized error's behavior. For a $l > 0$ (with corresponding rate of convergence $n^{\frac{1}{l}}$), we want to find the limit ditribution of

$$n^{\frac{1}{l}} \sum_{i=1}^{B_n} \big(\widehat{\Theta}_{i,n} - \Theta_{i,n}\big) \Delta \mathrm{T}_{i,n} \tag{33}$$

Specifically, we want to show that (33) converges *stably*[12] to a limit distibution. This mode of convergence, which is a bit stronger than the regular convergence in distribution, is due for statistical purposes. Because we will obtain in the variance limit of (33) random quantities, we need the stable convergence to infer the same way we would do it if the variance limit was nonrandom. 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 all the examples of this paper. We define $\mathcal{I}_{i,n}$[13] the information up to time $\tau_{i,n}$.

Condition (E0). *$\mathcal{I}_{i,n}$ can be extended into $\mathcal{J}_{i,n}$[14], 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 all the following of this paper, when using the conditional expectation $\mathbb{E}_\tau\big[Z\big]$[15], 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.

It is clear that the parametric model user's estimator will include the returns $\{r_{1,n}, \ldots, r_{k,n}\}$ as inputs. Also, because the Markov chai is of order m, she needs a m-dimensional vector of initial returns $\mathbf{r}_{0,n}$. Finally, since the Markov chain is partially observed, she can also use an estimate of the unobserved part of the Markov chain $\widehat{\mathbf{Q}}_{0,n}$.

$$\hat{\theta}_{k,n} := \hat{\theta}_{k,n}(r_{1,n}; \ldots; r_{k,n}; \widehat{\mathbf{Q}}_{0,n}; \mathbf{r}_{0,n}) \tag{34}$$

[12]One can look at definitions of stable convergence in 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).

[13]We assume that $\mathcal{I}_{i,n}$ is a (discrete) filtration of $(\Omega, \mathcal{F}, P)$ such that $\{\theta_s^*\}_{0 \leq s \leq \tau_{i,n}}$ and $U_{i,n}$ are adapted to $\mathcal{I}_{i,n}$.

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

[15]τ has to be a $\mathcal{J}_t$-stopping time

We keep the same asymptotic property as in (19) of Section 2 adapted to the multidimension case. We need some notations for this purpose. Define the part $\alpha_{Q,n} := (\alpha_n^{(1)}, \dots, \alpha_n^{d_q})$ related to the unobserved returns $Q_{i,n}$ and the part $\alpha_{R,n} := (\alpha_n^{(d_q+1)}, \dots, \alpha_n^{(d)})$ related to the observed returns $R_{i,n}$ of α_n. For $S \in \{Q, R\}$, let $\boldsymbol{\alpha}_{S,n} := (\alpha_{S,n}, \dots, \alpha_{S,n})$ consisting of m $\alpha_{S,n}$ appened together. Also, for any vector β of dimension d_β, $\beta^{-1} := \big((\beta^{(1)})^{-1}, \dots, (\beta^{(d_\beta)})^{-1}\big)$. Finally, $\boldsymbol{\alpha}_{Q,n}^{-1}\widehat{\mathbf{Q}}_{0,n}$ is the scaled distribution (if $\widehat{\mathbf{Q}}_{0,n}$ is the distribution of a random vector A, $\boldsymbol{\alpha}_{Q,n}^{-1}\widehat{\mathbf{Q}}_{0,n}$ will be realized by $\alpha_{Q,n}^{-1}A$).

$$\hat{\theta}_{k,n}(r_{1,n}; \dots; r_{k,n}; \widehat{\mathbf{Q}}_{0,n}; \mathbf{r}_{0,n}) \overset{\mathcal{D}}{=} \hat{\theta}_{k,1}(\alpha_{R,n}^{-1} r_{1,n}; \dots; \alpha_{R,n}^{-1} r_{k,n}; \boldsymbol{\alpha}_{Q,n}^{-1}\widehat{\mathbf{Q}}_{0,n}; \boldsymbol{\alpha}_{R,n}^{-1}\mathbf{r}_{0,n}) \tag{35}$$

Let i be a positive integer. In analogy with Section 2, we define the Markov chain elements on the ith 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 ith block as $(M_{i,n}^0, \dots, M_{i,n}^{-(d-1)}) := \mathbf{M}_{(i-1)h_n+j,n}$. For $\mathbf{M} \in \mathcal{M}_m$, we let $\tilde{M}_{i,n}^{j,\mathbf{M}}$ be the approximations of the Markov chain on the ith block with starting vector $\mathbf{M}$ for $j = -(d-1), \dots, h_n$. The initial vector is defined as $(\tilde{M}_{i,n}^{-(m-1),\mathbf{M}}, \dots, \tilde{M}_{i,n}^{-(m-1),\mathbf{M}}) := \mathbf{M}$. The m-dimensional vector of approximation is defined as $\tilde{\mathbf{M}}_{i,n}^{j,\mathbf{M}} := (\tilde{M}_{i,n}^{j,\mathbf{M}}, \dots, \tilde{M}_{i,n}^{j-(d-1),\mathbf{M}})$. We obtain the approximated returns by the recurrence relation similar to (31)

$$\tilde{M}_{i,n}^{j,\mathbf{l}} = F_n(\tilde{\mathbf{M}}_{i,n}^{j-1,\mathbf{l}}, U_{i,n}^j, \tilde{\Theta}_{i,n}) \tag{36}$$

The unfeasible estimator $\hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}}$ with initial vector $\mathbf{l}$ is defined as

$$\hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}} := \hat{\theta}_{h_n,n}(\tilde{R}_{i,n}^{1,\mathbf{M}}; \dots; \tilde{R}_{i,n}^{h_n,\mathbf{M}}; \tilde{\mathbf{M}}_{i,n}^{0,\mathbf{M}}) \tag{37}$$

The spot parameter's estimator on the ith block $\widehat{\Theta}_{i,n}$ is defined with observed returns and $\widehat{\mathbf{Q}}_{i,n}^0$ an estimate of the initial distribution of the unobserved part of the Markov chain as input. In particular, it doesn't rely directly on unobservable quantities.

$$\widehat{\Theta}_{i,n} := \hat{\theta}_{h_n,n}(R_{i,n}^1; \dots; R_{i,n}^{h_n}; \widehat{\mathbf{Q}}_{i,n}^0; \mathbf{R}_{i,n}^0) \tag{38}$$

Let $\mathbf{M}^* \in \mathcal{M}_m$. We can decompose $\big(\widehat{\Theta}_{i,n} - \Theta_{i,n}\big)$ as

$$\big(\hat{\Theta}_{i,n} - \hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}_{i,n}^0}\big) + \big(\hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}_{i,n}^0} - \hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}^*}\big) + \big(\hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}^*} - \tilde{\Theta}_{i,n}\big) + \big(\tilde{\Theta}_{i,n} - \Theta_{i,n}\big) \tag{39}$$

where the first term is the error in estimation due to the use of the approximated model (36) instead of the true model (31), the second term is the error made when taking $\mathbf{M}^*$ instead of $M_{i,n}^0$ as initial value of the block, the third term corresponds to the error of the estimation of the constant parameter by the underlying approximations starting with a fixed initial value $\mathbf{M}^*$ and the last term is the error made by holding the process parameter constant on each block. We make the first assumption, which is on observation times.

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

$$\inf_{1 \leq i \leq N_n} \mathbb{E}_{\tau_{i-1,n}} \left[(\Delta \tau_{i,n})^k \right] \text{ and } \sup_{1 \leq i \leq N_n} \mathbb{E}_{\tau_{i-1,n}} \left[(\Delta \tau_{i,n})^k \right] \text{ are exactly of order } O_p(n^{-k}) \quad (40)$$

Note that the observation times generated by the *hitting boundary process with time process model* of Potiron and Mykland (2015), which includes very general endogeneous settings, verifies condition $(E1)$. Thus, it seems that $(E1)$ is harmless. We make a second assumption, this time on the size of a block h_n. We define $l' > 0$ such that under the assumption that the parametric model is true, the rate of convergence of the parametric model user's estimator is $n^{\frac{1}{l'}}$. To fix ideas, under regular models, $l' = 2$ for the MLE (see, e.g., Firth (1993) bias reduction of max likelihood TO ADD!).

Condition (E2). *The block size* h_n *is such that*

$$n^{\frac{2}{l}-2} h_n^2 = o(1) \quad (41)$$

$$n^{\frac{2}{l}-1} h_n^{1-\frac{2}{l'}} \to 1 \quad (42)$$

(add that (42) is used in proof of third term) In practice, this assumption provides us the maximum block size h_n to use for constant approximation of parameter. (41) will be useful to prove that bla bla and the last term of (52) tends to 0, and (42) will be used to prove bla bla. The next assumption is on the parametric model user's estimator. (to change the text from here !!) Roughly, it assumes that under the assumption that the parametric model is true, we have a central limit theorem (at the speed $n^{\frac{1}{l'}}$), and we can bound uniformly in the parameter the $\mathbb{L}^1$-convergence of the scaled error as well as the difference between the variance of the scaled error and the variance obtained in the central limit theorem.

Condition (E3). *For any parameter* $\theta \in K$*, define the expectation of the normalized error when holding the model constant* $E^{\theta}_{h_n,n} := \mathbb{E}\big[h_n^{\frac{1}{l'}}\big(\hat{\theta}_{h_n,n}(R^{\mathbf{M}^*,\theta}_{1,n}; \ldots ; R^{\mathbf{M}^*,\theta}_{h_n,n}; \mathbf{M}^*) - \theta\big)\big]$*. For any* $M > 0$*, we need the following uniform convergence*

$$\sup_{\theta \in K_M} \left| \mathbb{E} \left[h_n^{\frac{1}{l'}} \big(\hat{\theta}_{h_n,n}(R^{\mathbf{M}^*,\theta}_{1,n}; \ldots ; R^{\mathbf{M}^*,\theta}_{h_n,n}; \mathbf{M}^*) - \theta \big) - E_\theta \right] \right| = o(h_n^{1-\frac{1}{l'}} n^{\frac{1}{l}-1}) \quad (43)$$

We also consider $C^{\theta}_{h_n,n} := \mathrm{Var}\left[h_n^{\frac{1}{l'}} \big(\hat{\theta}_{h_n,n}(R^{\mathbf{M}^*,\theta}_{1,n}; \ldots ; R^{\mathbf{M}^*,\theta}_{h_n,n}; \mathbf{M}^*) - \theta \big) \sum_{j=1}^{h_n} (R^{\mathbf{M}^*,\theta}_{j,n})^{(d_r)} \right]$*. We assume that there exists* $V_\theta > 0$ *such that for any* $M > 0$*, we have uniformly in* $\theta \in K_M$

$$C^{\theta}_{h_n,n} = V_\theta \sum_{j=1}^{h_n} (R^{\mathbf{M}^*,\theta}_{j,n})^{(d_r)} h_n n^{-1} + o_p(h_n n^{-1}) \quad (44)$$

Define $D^{\theta}_{h_n,n} := \mathbb{E}\left[h_n^{\frac{1}{l'}}\left(\hat{\theta}_{h_n,n}(R^{\mathbf{M}^*,\theta}_{1,n};\dots;R^{\mathbf{M}^*,\theta}_{h_n,n};\mathbf{M}^*)-\theta\right)\sum_{j=1}^{h_n}(R^{\mathbf{M}^*,\theta}_{j,n})^{(d_r)}\right]$. *For any* $M>0$, *we have uniformly in* $\theta \in K_M$

$$D^{\theta}_{h_n,n} = o_p(\sum_{j=1}^{h_n}(R^{\mathbf{M}^*,\theta}_{j,n})^{(d_r)}) \tag{45}$$

Finally, for any $M>0$, *we need that*

$$\sup_{\theta\in K_M}\mathbb{E}\left[(h_n^{\frac{1}{l'}}\left(\hat{\theta}_{h_n,n}(R^{\mathbf{M}^*,\theta}_{1,n};\dots;R^{\mathbf{M}^*,\theta}_{h_n,n};\mathbf{M}^*)-\theta\right)-E^{\theta}_{h_n,n})^8\right] = 0(1) \tag{46}$$

$\mathrm{T}^{\mathbf{M}^*,\theta}_{h_n,n}$ and $\mathrm{T}^{\mathbf{M}^*,\chi}_{h_n,n}$ to be defined

Condition (E4). *For any* $M>0$, *define* $\mathcal{E}_M$ *the product space of information, initial parameter value, and parameter process value* $(\mathcal{J}_{i,j,n},\theta,\chi_t)$[16], *where* $\theta\in K_M$, χ_t *is a null-drift continuous Itô-process with* $\chi_t\in K_M$ *for all* $0\le t\le T$, *and its initial value equal to* θ ($\chi_0=\theta$). *We have*

$$\sup_{(\mathcal{J}^j_{i,n},\theta,\chi_t)\in\mathcal{E}_M}\mathbb{E}_{\mathcal{J}^0_{1,n}}\left[(\mathrm{T}^{\mathbf{M}^*,\theta}_{h_n,n}-\mathrm{T}^{\mathbf{M}^*,\chi}_{h_n,n})^2\right]\to 0 \tag{47}$$

(to change the text !) The next assumption is purely a model assumption of Markov chain's ergodicity. We need to introduce some definitions first. For $\mathbf{M}\in\mathcal{M}$, $\theta\in K$ and i any integer greater than $-(d-1)$, we define $\tilde{M}^{\mathbf{M},\theta}_i$ the Markov chain with initial vector $\mathbf{M}$ and parameter θ. The initial vector is such that $(M^{\mathbf{M},\theta}_{-(m-1)},\dots,\tilde{M}^{\mathbf{M},\theta}_0):=\mathbf{M}$. The m-vector is defined as $\mathbf{M}^{\mathbf{M},\theta}_i := (M^{\mathbf{M},\theta}_i,\dots,M^{\mathbf{M},\theta}_{i-(d-1)})$. The rest of the Markov chain is obtained by the same recurrence relation as (30)

$$M^{\mathbf{M},\theta}_i = F_1(\mathbf{M}^{\mathbf{M},\theta}_{i-1},U_{i,1},\theta) \tag{48}$$

We consider $\mathbf{N}^{\mathbf{M},\theta}_i := \mathbf{M}^{\mathbf{M},\theta}_{id}$ the Markov chain of dimension $d*m$, which consists of block of d-elements of the original Markov chain $\mathbf{M}^{\mathbf{M},\theta}_i$. By construction, $\mathbf{N}^{\mathbf{M},\theta}_i$ is a Markov chain of order 1. We note P^k_θ its transition probability, i.e. for any $B\in B(\mathcal{F}^m)$ (set of borelians of $\mathcal{M}^m$)

$$P^k_\theta(\mathbf{M},B) = \mathbb{P}(\mathbf{N}^{\mathbf{M},\theta}_k\in B) \tag{49}$$

[16] $\mathcal{J}^j_{i,n}$ can be any filtration of $(\Omega,\mathcal{F},P)$, sorted as $\mathcal{J}_{1,0,n},\dots,\mathcal{J}_{1,h_n,n},\mathcal{J}_{2,1,n},\dots,\mathcal{J}_{2,h_n,n},\dots$, such that $U^j_{i,n}$ and $\{\chi_s\}_{0\le s\le\tau^j_{i,n}}$ are adapted to $\mathcal{J}^j_{i,n}$, and $U^j_{i,n}$ is independent of the past.

Condition (E5). *We have*

$$\sup_{\theta\in K_M\ ,\ M_1\in\mathcal{M}^n\ ,\ M_2\in\mathcal{M}^n} |\hat{\theta}(\tilde{R}_i^{\mathbf{M}_1,\theta},\ldots,\tilde{R}_{h_n}^{\mathbf{M}_1,\theta};\mathbf{M}_0^{\mathbf{M}_1,\theta}) - \hat{\theta}(\tilde{R}_1^{\mathbf{M}_2,\theta},\ldots,R_{h_n}^{\mathbf{M}_2,\theta};\mathbf{M}_0^{\mathbf{M}_2,\theta})| = o_p(h_n^{-1/l'}) \quad (50)$$

(First two parts of this assumption should go as a remark because we don't need them, we only need the third part) For all $\theta \in K$, $\mathbf{N}_i^{\mathbf{M},\theta}$ is uniformly ergodic, i.e. there exists π_θ such that

$$\sup_{\mathbf{M}\in\mathcal{M}^m} ||P_\theta^k(\mathbf{M},.) - \pi_\theta|| \to 0$$

Furthermore, the convergence is locally uniform, i.e. for all $M > 0$

$$\sup_{\theta\in K_M} \sup_{\mathbf{M}\in\mathcal{M}^m} ||P_\theta^k(\mathbf{M},.) - \pi_\theta|| \to 0$$

There exists a d-dimensional sequence $\alpha_n := (\alpha_n^{(1)},\ldots,\alpha_n^{(d)})$

$$F_1(\{(q_k,r_k)\}_{j-m\leq k\leq j-1}, U_{i,j,n}, \tilde{\Theta}_{i,n}) = \alpha_n * F_n(\{(q_k,r_k)\}_{j-m\leq k\leq j-1}, U_{i,j,n}, \tilde{\Theta}_{i,n})$$

The estimator of the time series user includes k returns, the previous m returns and an estimate of the previous m vectors of Q $\mathbf{q}$ (which can be a distribution)

$$\hat{\theta}_{k,n}^{\mathbf{q}} := \hat{\theta}_{k,n}(r_{1,n};\ldots;r_{k,n};\mathbf{q};r_{0,n};\ldots;r_{-(m-1),n})$$

$$:= \hat{\theta}_k(\alpha_n^R * r_{1,n};\ldots;\alpha_n^R * r_{k,n};\alpha_n^Q\mathbf{q};\alpha_n^R * r_{0,n};\ldots;\alpha_n^R * r_{-(m-1),n})$$

The unfeasible estimators are defined as

$$\hat{\tilde{\Theta}}_{i,n}^{\mathbf{q},\mathbf{r}} := \hat{\theta}_{h_n,n}(\tilde{R}_{i,1,n}^{\mathbf{q},\mathbf{r}};\ldots;\tilde{R}_{i,h_n,n}^{\mathbf{q},\mathbf{r}};\mathbf{q};\mathbf{r})$$

such that on each block, the approximations $(\tilde{Q}_{i,j,n}^{\mathbf{q},\mathbf{r}}, \tilde{R}_{i,j,n}^{\mathbf{q},\mathbf{r}}) \in \mathcal{Q}\times\mathcal{R}$ starts at $(\mathbf{q},\mathbf{r})$, i.e.

$$\{(\tilde{Q}_{i,j,n}^{\mathbf{q},\mathbf{r}}, \tilde{R}_{i,j,n}^{\mathbf{q},\mathbf{r}})\}_{j=-m+1}^{j=0} = (\mathbf{q},\mathbf{r})$$

and $(\tilde{Q}_{i,j,n}^{\mathbf{q},\mathbf{r}}, \tilde{R}_{i,j,n}^{\mathbf{q},\mathbf{r}})$ is an homogeneous Markov chain of order m with

$$(\tilde{Q}_{i,j,n}^{\mathbf{q},\mathbf{r}}, \tilde{R}_{i,j,n}^{\mathbf{q},\mathbf{r}}) = F_n(\{(\tilde{Q}_{i,k,n}^{\mathbf{q},\mathbf{r}}, \tilde{R}_{i,k,n}^{\mathbf{q},\mathbf{r}})\}_{j-m\leq k\leq j-1}, U_{i,j,n}, \tilde{\Theta}_{i,n}) \quad (51)$$

For a $l > 0$, we are interested in the asymptotic behavior of

$$n^{\frac{1}{l}}\sum_{i=1}^{B_n}(\hat{\Theta}_{i,n} - \Theta_{i,n})\Delta\mathrm{T}_{i,n}$$

We can decompose $(\hat{\Theta}_{i,n} - \Theta_{i,n})$ as

$$(\widehat{\Theta}_{i,n} - \widehat{\tilde{\Theta}}_{i,n}^{\mathbf{q}_{i,n},\mathbf{r}_{i,n}}) + (\widehat{\tilde{\Theta}}_{i,n}^{\mathbf{q}_{i,n},\mathbf{r}_{i,n}} - \widehat{\tilde{\Theta}}_{i,n}^{\mathbf{q}^*,\mathbf{r}^*}) + (\widehat{\tilde{\Theta}}_{i,n}^{\mathbf{q}^*,\mathbf{r}^*} - \tilde{\Theta}_{i,n}) + (\tilde{\Theta}_{i,n} - \Theta_{i,n}) \tag{52}$$

where

$$\begin{aligned}
\mathbf{q}_{i,n} &:= (Q_{ih_n-1,n}, \dots, Q_{ih_n-m,n}) \\
\mathbf{r}_{i,n} &:= (R_{ih_n-1,n}, \dots, R_{ih_n-m,n})
\end{aligned}$$

and $(\mathbf{q}^*, \mathbf{r}^*)$ are vectors that can be taken for any θ (i.e. its density is strictly positive for any θ).

We also define $(\tilde{Q}_{i,n}^{\mathbf{q},\mathbf{r},\theta}, \tilde{R}_{i,n}^{\mathbf{q},\mathbf{r},\theta})$ as

$$(\tilde{Q}_{i,n}^{\mathbf{q},\mathbf{r},\theta}, \tilde{R}_{i,n}^{\mathbf{q},\mathbf{r},\theta}) \quad = \quad F_n(\{(\tilde{Q}_{k,n}^{\mathbf{q},\mathbf{r},\theta}, \tilde{R}_{k,n}^{\mathbf{q},\mathbf{r},\theta})\}_{i-m \leq k \leq i-1}, U_{i,j,n}, \theta)$$

and for χ a process (be more specific about $\chi_{i,n}$)

$$(\tilde{Q}_{i,n}^{\mathbf{q},\mathbf{r},\chi}, \tilde{R}_{i,n}^{\mathbf{q},\mathbf{r},\chi}) \quad = \quad F_n(\{(\tilde{Q}_{k,n}^{\mathbf{q},\mathbf{r},\chi}, \tilde{R}_{k,n}^{\mathbf{q},\mathbf{r},\chi})\}_{i-m \leq k \leq i-1}, U_{i,j,n}, \chi_{i,n})$$

We define the Markov chain (of order 1)

$$(\tilde{\mathbf{Q}}_{i,n}^{\mathbf{q},\mathbf{r},\theta}, \tilde{\mathbf{R}}_{i,n}^{\mathbf{q},\mathbf{r},\theta}) := \{(\tilde{Q}_{im+k,n}^{\mathbf{q},\mathbf{r},\theta}, \tilde{R}_{im+k,n}^{\mathbf{q},\mathbf{r},\theta})\}_{k=1}^{m}$$

and $\forall (\mathbf{q}, \mathbf{r}) \in (\mathcal{Q} \times \mathcal{R})^m$ $P_\theta^k((\mathbf{q}, \mathbf{r}), .)$ its transition probability, i.e. for any $B \in B((\mathcal{Q} \times \mathcal{R})^m)$ (set of borelians of $(\mathcal{Q} \times \mathcal{R})^m$)

$$P_\theta^k((\mathbf{q}, \mathbf{r}), B) = P((\tilde{\mathbf{Q}}_{k,n}^{\mathbf{q},\mathbf{r},\theta}, \tilde{\mathbf{R}}_{k,n}^{\mathbf{q},\mathbf{r},\theta}) \in B)$$

We define the following assumptions needed for the central limit theorem

Assumptions.

- *(E1) $h_n \to +\infty$ and $h_n n^{\frac{1}{l}-1} \to 0$. Furthermore, the observation times are such that*

$$\begin{aligned}
\sup_{i \geq 0} \mathbb{E}_{\tau_{i-1,n}} \left[(\Delta \tau_{i,n})\right] &= O_p(n^{-1}) \\
\sup_{i \geq 0} \mathbb{E}_{\tau_{i-1,n}} \left[(\Delta \tau_{i,n})^2\right] &= O_p(n^{-2})
\end{aligned}$$

- *(E2) for all parameter $\theta \in K$, we assume that*

$$h_n^{\frac{1}{l}}\left(\hat{\theta}_{h_n,n}(\tilde{R}_{1,n}^{\mathbf{q}^*,\mathbf{r}^*,\theta};\ldots;\tilde{R}_{h_n,n}^{\mathbf{q}^*,\mathbf{r}^*,\theta};\mathbf{q}^*;\mathbf{r}^*)-\theta\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,V_\theta)$$

Furthermore, we need the following uniform convergences

$$\begin{aligned}\sup_{\theta\in K}\mid\mathbb{E}\left[h_n^{1/l}\left(\hat{\theta}_{h_n,n}(\tilde{R}_{1,n}^{\mathbf{q}^*,\mathbf{r}^*,\theta};\ldots;\tilde{R}_{h_n,n}^{\mathbf{q}^*,\mathbf{r}^*,\theta};\mathbf{q}^*;\mathbf{r}^*)-\theta\right)\right]\mid &\to 0\\ \sup_{\theta\in K}\mid\mathrm{Var}\left(h_n^{1/l}(\hat{\theta}_{h_n,n}(\tilde{R}_{1,n}^{\mathbf{q}^*,\mathbf{r}^*,\theta};\ldots;\tilde{R}_{h_n,n}^{\mathbf{q}^*,\mathbf{r}^*,\theta};\mathbf{q}^*;\mathbf{r}^*)-\theta)\right)-V_\theta\mid &\to 0\end{aligned}$$

- *(E3) For all $\theta\in K$, $(\tilde{\mathbf{Q}}_{i,n}^{\mathbf{q},\mathbf{r},\theta},\tilde{\mathbf{R}}_{i,n}^{\mathbf{q},\mathbf{r},\theta})$ is uniformly ergodic, i.e. there exists π_θ such that*

$$\sup_{(\mathbf{q},\mathbf{r})\in(\mathcal{Q}\times\mathcal{R})^m}||P_\theta^k((\mathbf{q},\mathbf{r}),.)-\pi_\theta||\to 0$$

Furthermore, for all $M>0$, the convergence is uniform in $\{|\theta|\le M\}$, i.e.

$$\sup_{\theta\in K,|\theta|\le M}\sup_{(\mathbf{q},\mathbf{r})\in(\mathcal{Q}\times\mathcal{R})^m}||P_\theta^k((\mathbf{q},\mathbf{r}),.)-\pi_\theta||\to 0$$

- *(E4) For any $M>0$, define $\mathcal{E}_M$ the product space of information, initial parameter value, and parameter process value $(\mathcal{J}_{i,j,n},\theta,\chi_t)$[17], where $\chi_t\in K$ is a null-drift continuous Ito-process with its volatility bounded by M, and its initial value equal to θ ($\chi_0=\theta$). We have*

$$\begin{aligned}\sup_{(\mathcal{J}_{i,j,n},\theta,\chi_t,\mathbf{M},q_d)\in\mathcal{E}_M}\mathbb{E}_{\mathcal{J}_{1,0,n}}\Big[&\Big|\hat{\theta}_{h_n,n}(\tilde{R}_{1,n}^{\mathbf{M},\theta};\ldots;\tilde{R}_{h_n,n}^{\mathbf{M},\theta};q_d;r)\\ &-\hat{\theta}_{h_n,n}(\tilde{R}_{1,n}^{\mathbf{M},r,\chi};\ldots;\tilde{R}_{h_n,n}^{\mathbf{M},r,\chi};q_d;r)\Big|\Big]=o_p(1)\end{aligned}$$

Rq : There are several possibilities to prove (ACLT3), this is equivalent to uniform ergodicity and continuity of something (see Meyn and Tweedie!). Also it can be : For all $\theta\in K$ and $x\in\mathcal{Q}\times\mathcal{R}$,

$$P_\theta^m(x,.)\ge\nu_\theta^{k_\theta}(B)$$

where k_θ is a bounded integer and $\nu_\theta^{k_\theta}(\mathcal{Q}\times\mathcal{R})$ is uniformly bounded away from 0.

It can be the case (because of h_n or p) that (ACLT4) is not verified, i.e. $n^{\frac{1}{l}-\frac{p}{2}}h_n^{\frac{p}{2}}$ does not go to 0. In this case, we will not obtain a CLT, but only the order of convergence, which is $n^{\frac{1}{l'}}$, where

$$l'=\inf\{l'':n^{\frac{1}{l}-\frac{p}{2}}h_n^{\frac{p}{2}}\}$$

[17] $\mathcal{J}_{i,j,n}$ can be any filtration of $(\Omega,\mathcal{F},P)$ such that $U_{i,j,n}$ and $\{\chi_s\}_{0\le s\le\tau_{i,j,n}}$ are adapted to $\mathcal{J}_{i,j,n}$, and $U_{i,j,n}$ is independent of the past ($\mathcal{J}_{i,j-1,n}$ if $j\neq 0$ or $\mathcal{J}_{i-1,h_n-1,n}$ if $j=0$).

To obtain this order of convergence, needless to verify (ACLT2), we only need to have

$$h_n^{\frac{1}{\nu'}}(\hat{\theta}_{h_n,n}(\theta,\nu)-\theta)\xrightarrow{\mathbb{P}}0$$

(to look at it more though)

4 Can we trust the parameter's constancy in a model ?

4.1 The parametric case: testing if $\sigma_t^\theta = 0$

This section is written for the parametric model user who uses (and trusts) her model. The semiparametric model user, who already believes in the nonconstancy of some parameters over time, can go directly to the next section. The point of this section is to provide to the parametric model user a (non-exhaustive) list of statistical tests, which can be implemented easily, that will shed light on the parameter's behavior over time. Specifically, the null-hypothesis throughout the rest of this section will be

(H_0) The parameter of the model is constant over time equal to θ^*

In the following, we will be describing four types of tests, that will provide statistics to the lector, together with their distribution under (H_0). Accordingly, associated p-values can be computed. The way we advise the parametric model user to use results of this section is to look at the first type of tests, because it is very straightforward to implement. If (some) p-values are very small, we have enough information to reject (H_0), and there is no need to look at the other type of tests. If not, then she should use the second type of tests, and so on. The fourth type of tests will be more involved to implement, so it is better to avoid it if we can. Note that if she obtains not-significant statistic's value for the four types of tests, then it doesn't mean that (H_0) is true. A closer look at the data is further needed in this case. We insist on the fact that she shall never trust (H_0) without proper model checkings, and that using the technology of this paper will not harm her even if (H_0) is indeed true (see the simulations in Section 7).

The idea behind the statistical tests of constancy of the underlying parameter is the following. Assume that the parametric model user has a specific model in mind (such as ARMA(p,q)) and that she wants to estimate the underlying parameter of her model with an estimator (typically MLE or least squares). For this section, we assume that she is first choosing p and q by an ad-hoc criteria, and then she sticks to this choice. Thus, the number of one-dimensional underlying parameters to be estimated is equal to $p+q+2$ (or $p+q+1$ if we restrict the model to have a zero mean). When she fits the model to the whole dataset, she obtains estimated values of the parameters, together with standard errors' estimates. Under (H_0), if she fits the model to a subblock of the data, she will obtain estimates of the same parameter. Note that the standard errors will be larger than the ones we obtained by fitting the whole dataset, because the procededure is only using part of the data. We repeat the block-fitting for a moving block (with a constant size), and look at the pattern of the fitted parameter in time.

We keep the same notations as in previous sections, and also assume that the parametric user has at hands an estimator $s_{k,n}$ of the standard deviation, which depends on the same quantities as the parameter's estimator. The standard deviation's estimates are defined as $\widehat{S}_{1,n}, \ldots, \widehat{S}_{B_n,n}$. Define

$$\widehat{\Theta}_n^A = \hat{\theta}_{N_n,n}\big(R_{1,n}; \ldots; R_{N_n,n}; \widehat{\mathbf{Q}}_{0,n}; \mathbf{R}_{0,n}\big) \tag{53}$$

the parameter's estimate using the whole dataset, and $\hat{S}_n^{(1)}$ a standard error's estimate of $\widehat{\Theta}_n - \widehat{\Theta}_n^A$. The test statistic is defined as

$$T^{(1)} := \frac{\widehat{\Theta}_n - \widehat{\Theta}_n^A}{\widehat{S}_n^{(1)}} \tag{54}$$

Under (H_0), $T^{(1)}$ is approximately a standard normal distribution. We are now giving one way to compute $\widehat{S}_n^{(1)}$ in practice. Let S_n (respectively S_n^A) the theoretical standard deviation of $\widehat{\Theta}_n$ $(\widehat{\Theta}_n^A)$. Because $\hat{\Theta}_{i,n}$ are asymptotically uncorrelated under the assumptions of Section 3, S_n can be estimated by

$$\widehat{S}_n = T^{-1}\Big(\sum_{i=1}^{B_n} (\Delta \mathrm{T}_{i,n})^2 \widehat{S}_{i,n}^2\Big)^{\frac{1}{2}} \tag{55}$$

An estimate of the estimator using the whole dataset is directly available to the parametric model user by

$$\widehat{S}_n^A = \hat{s}_{N_n,n}\big(R_{1,n}; \ldots; R_{N_n,n}; \widehat{\mathbf{Q}}_{0,n}; \mathbf{R}_{0,n}\big) \tag{56}$$

The variance of the numerator of the test statistic $\big(\widehat{\Theta}_n - \widehat{\Theta}_n^A\big)$ is equal to

$$S_n^2 + (S_n^A)^2 - 2\,\mathrm{cov}\big(\widehat{\Theta}_n, \widehat{\Theta}_n^A\big)$$

We expect $\widehat{\Theta}_n$ and $\widehat{\Theta}_n^A$ to be positively correlated, and thus we can be conservative by setting

$$\widehat{S}_n^{(1)} = \big(\widehat{S}_n^2 + (\widehat{S}_n^A)^2\big)^{\frac{1}{2}} \tag{57}$$

The second test looks at the variation of $\widehat{\Theta}_{i,n}$ the parameter's estimate over time, compared to its estimated mean $\widehat{\Theta}_n$. It looks at over-dispersion or under-dispersion. We define the test statistic

$$T^{(2)} := B_n^{-1} \sum_{i=1}^{B_n} \frac{(\widehat{\Theta}_{i,n} - \widehat{\Theta}_n)^2}{\widehat{S}_{i,n}} \tag{58}$$

Under (H_0), $T^{(2)}$ is approximately a standard normal. We warn the lector that this test should be used only if $\widehat{S}_n$ is small compared to the standard errors on each block $\widehat{S}_{i,n}$. Alternative estimators of the standard deviation can be used in the denominator of (58), but as this is not the main purpose of our work, we will leave it to the lector.

The third type of tests will try to investigate if block-constant models behave significantly better than the null model. Specifically, we assume the alternative hypothesis

(H_1) The parameter of the model is constant on each block of size h_n equal to $(\theta^*_{1,n}, \ldots, \theta^*_{B_n,n})$

We define $\mathcal{L}_n$ the log-likelihood of the model based on the whole dataset. Let $\mathcal{L}_{i,n}$ the log-likelihood. Because of the Markov nature of our model, it is straightforward to verify that $\mathcal{L}_n^{(1)}$, the log likelihood of model $(H1)$, is equal to the sum of block log-likelihoods, i.e.

$$\mathcal{L}_n^{(1)} = \sum_{i=1}^{B_n} \mathcal{L}_{i,n} \tag{59}$$

The likelihood-ratio test statistic is defined as

$$T^{(3)} := 2(\mathcal{L}_n^{(1)} - \mathcal{L}_n) \tag{60}$$

Under the null hypothesis, $T^{(3)}$ follows a chi-quared distribution with $\nu := p * B_n$ degrees of freedom. The issue with model following $(H1)$ is that it allows parameter to evolve over time, but the associated number of degrees of freedom ν can be very large. Thus, in a lot of cases, $T^{(3)}$ will not be significant because of ν being too large. Our final type of tests will be similar to the third one, but we don't want the difference in log-likelihood of models to have a small nmber of degrees freedom.

Test is a test of ratio of log likelihoods ! We can do it with different size of blocks, so we will end up with a lot of correlated tests! Extract the right information from them! The first test would be to suppose that constant on blocks. The second one would be the one where we use an extra parameter for the variance of the parameter.

4.2 The semiparametric case: testing if ??

for some k $\sigma_t^{\theta,(k)} = 0$ (be more specific about it) same, but this time we only test some parameters, knowing that the others are moving over time...

5 Forecasting

- Suppose that we are at time $\tau_{i,n}$, and that we want to predict the observed returns in the next periods in time (not too far into the future). For simplicity in notations, we assume that $\tau_{i,n}$ is the starting point of a block.
- We will use the parametric model as a prediction model, but with different input $\hat{\theta}^*_{\tau_{i,n}}$
- The forecasting window is defined as $W := [\tau_{j,n}, \tau_{i,n}]$ for a $j < i$.
- We estimate the spot parameter with

$$\hat{\theta}^*_{\tau_{i,n}} := \sum_{\tau^0_{i,n} \in W} \alpha_{i,n} \widehat{\Theta}_{i,n}$$

  where $\sum_{\tau^0_{i,n} \in W} \alpha_{i,n} = 1$.
- This is just one idea, a general estimator will be investigated in further work

Look at data if we obtain intervals closer to 95%.

For econometricians looking at forecasting a few periods ahead, this paper gives a prediction model, which is not more parametrized that the one they have, because it has the same degree of freedom, only a different "spot parameter", that in theory is more accurate than the one the econometrician would have found using his estimator. This is much of the same than the estimation of the speed of a car, where we need a small window of time to estimate the spot (instantaneous) speed, or the estimation of the spot volatility. [18]

If they forecast at more periods ahead, they have the choice to still keep their model or a model including the volatility of the "spot parameter", that we consistently estimate in Section 4, that will remove p degrees of freedom. Of course, by seeing that their model is time-dependent, they also have the choice to change it, and choose a new model with less degrees of freedoms instead. Then, they could check on their new model if the new parameters are still moving a lot over time using the techniques of this paper.

Nothing too much about it either!

[18] The way we estimate the spot volatility is bla bla (to be more accurate about this, we might need to change the formula). See Foster and Nelson (1996), Comte and Renault (1998) and Mykland and Zhang (2008).

6 Model selection

6.1 Nested models

Assume that the practitioner has a *locally parametric model* in mind. He wants to test if some of the parameters are null bla bla (to continue). If he finds something significant here, it means that he should use the more general semiparametric model. But then he should go back to Section semiparametric constancy's trust, to see if he can trust the constancy of his new parameter

The way the practitioner should use this Section is the following.

For instance, testing if there is noise in a model

add the example of noise in returns !!!

6.2 A general criteria

we can do a sum of log-likelihood on each block, and we could imagine that we penalize if the parameter is moving a lot (think more about how to penalize it though). For now, we penalize it because we are not interested in estimating the integrated covariation of the parameters. But in next paper... In practise, do that with different h_n to see if we obtain the same kind of results. Nothing too much about it (only one page to write!)

(the section and the following ones would be both for discrete time and continuous time model)

7 Examples on statistical models

7.1 ARMA model

7.2 GARCH model

7.3 IID noise independent of the return process

7.4 Hitting boundary process with time process model

7.5 Autoregressive Conditional Duration model

7.6 Hawkes process

7.7 Linear Model

When looking at a dataset, the usual simplest model that comes into the picture is the linear model. Suppose we observe $\{y_i, x_{i1}, \ldots, x_{ip}\}_{i=1}^n$ of n statistical units, a linear regression model assumes that the relationship between the dependent variable y_i and the $p-$vector of regressors x_i is linear, i.e. there exists real parameters $\beta_1, \ldots, \beta_p$ s.t.

$$y_i = \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \epsilon_i \tag{61}$$

where ϵ_i is the noise of the $i-th$ observation. It is often the case that observations are time-dependent, and thus the statistician can add new predictors in (61) by fitting a polynomial of degree d to take a proper account of the time component: in practise she can often end up with a high-degree d. As far as predictions of future values are concerned, this is probably a bad idea because in most cases, the future pattern of the time effect is uncertain.

If we are assuming that parameters are moving with time, an alternative model to (61) would be

$$y_i = \beta_{1,t_i} x_{i1} + \ldots + \beta_{p,t_i} x_{ip} + \epsilon'_i \tag{62}$$

where t_i corresponds to the time of observation of the $i-$th unit. The statistician could use the technology of this paper to chop the observations into short time-blocks, estimate the value of parameters on each block (where the time-effect would be roughly the same for each observation). Formally, it corresponds to fitting a model with a block effect for each time-block, with interactions. The results of Section 4 gives us insight on how to choose the size of each block h_n. By taking the weighted sum of estimated parameters (9), we provide to the statistician a straightforward prediction's model (for future observations in time), which is effectively a linear model, but with different input parameters than (61). Furthermore, by estimating the time-effect with this nonparametric technique, the statistician has gained d degrees of freedom compared to the prediction's model she would have used instead. We insist on the fact that if there is an obvious linear trend of the time, that will most likely continue, the statistician should first build an extended model of (61) that takes a proper account of the linear trend, and then use the techniques of this paper to estimate more accurately the inputs of her model.

Finally, it would be beyond the scope of this paper to define precisely the underlying assumptions needed to obtain the Central Limit Theorem of Theorem 2, as well as all the statistical implications for tests, but we would like to get the attention of the lector on the fact that $(E4)$ holds in this example. Similarly, we could apply the same machinerie to generalized linear models (see, i.e., McCullagh and Nelder (1989)). We think that in cases where overdispersion is very high, using (62) might help the statistician understand where it comes from.

8 Application: fitting ARMA and GARCH models on the S & P 500

To illustrate the application of our results, we take the S&P5000 stock index daily close returns data from the beginning of January 1990 to the end of June 1997, which makes 1895 observation points. Since we are looking at index of returns of stock market prices, it is often the case that the best model is a unit root or random walk (see, i.e., Bachelier (1900) or Fama (1965)), but a quick check of the auto-correlation function estimation of the returns (see Figure 1), which shows significant lag-1 correlation at the 5 % confidence level, suggests us to use a different model. The Ljung-Box statistic is examined to formally test the null hypothesis that the first autocorrelation is 0. Under the null hypothesis, the test statistic is distributed as a χ_1^2. We obtain a chi-squared statistic of 8.26 and its associated p-value of 0.004 This provides us enough reasons to fit an ARMA model (with null-mean) to the returns, instead of a unit root. We follow closely Box and Jekins (1976) for the model selection (of the class $ARMA(p, q)$), parameter estimation and model checking. We find that $AR(1)$ is the best candidate. In Figure 2, we can see a plot of the residuals over time. The variance seems to be roughly constant in time, except for observations from 1996, where it increased a bit. It is due to the fact that the volatility of the returns itself increased during that period. The auto-

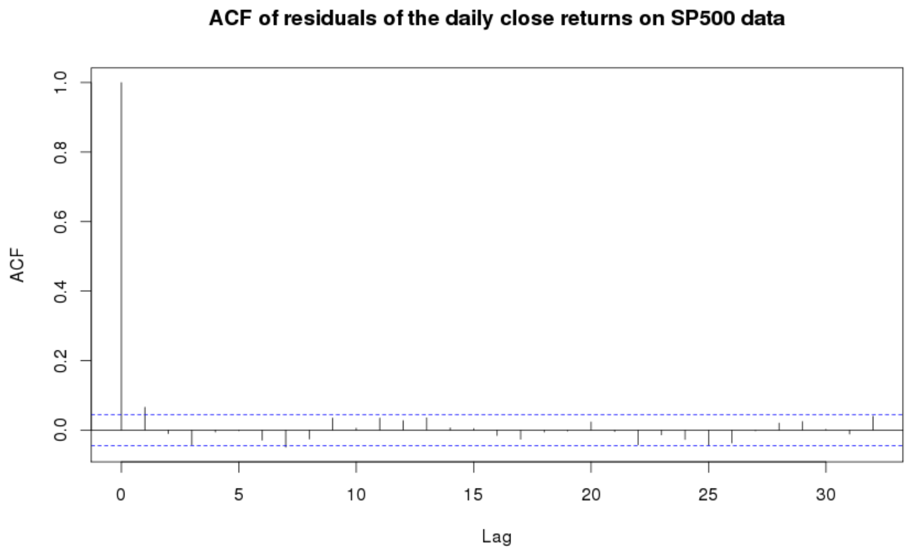

Figure 1: bla bla bla...

correlation function estimation of the residuals in figure 3 shows no significant autocorrelation effect for the first 32 lags, once again at the 5% confidence interval level. We obtain bla bla Ljung-Box statistic... bla bla... Based on those two plots, we can say that the model fits the data very well.

When fitting the 1985 data points to the model, we obtain the following estimates We make ten blocks of 190 observations (except for the last one, which is only 185 observation. On each block, we fit an $AR(1)$ model. (to continue !)

9 Conclusion

9.1 Allowing for a general structure

i.e. semimartingale with jumps... probably next paper !

9.2 Estimation of the integrated covariation

Will be the next paper, so much to say from this !

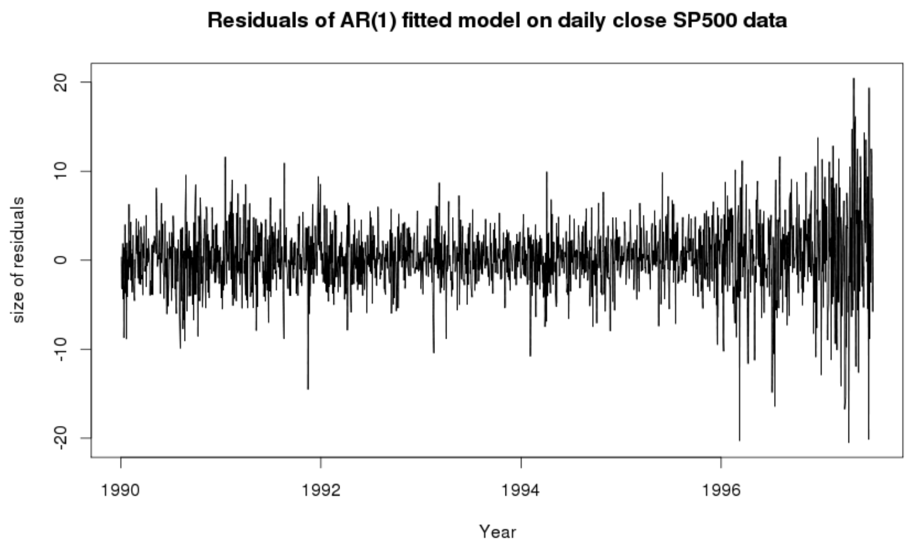

Figure 2: In this figure, bla bla bla...

9.3 Multidimension

Look at Ogihara and Yoshida paper (2014) for instance... probably next paper

9.4 Low-frequency setting

with Simon, probably next paper.

9.5 In generality how do you choose h_n ?

Look at Kalnina and Linton (2007) for ideas... In finite sample data, we should always choose n as big as possible because we will be closer to the approximation, but we have to be careful that θ is not moving too much... Think about it more... Also think about it more in the forecasting

9.6 How do you choose the window of forecasting ?

i.e. how do you estimate the spot parameter

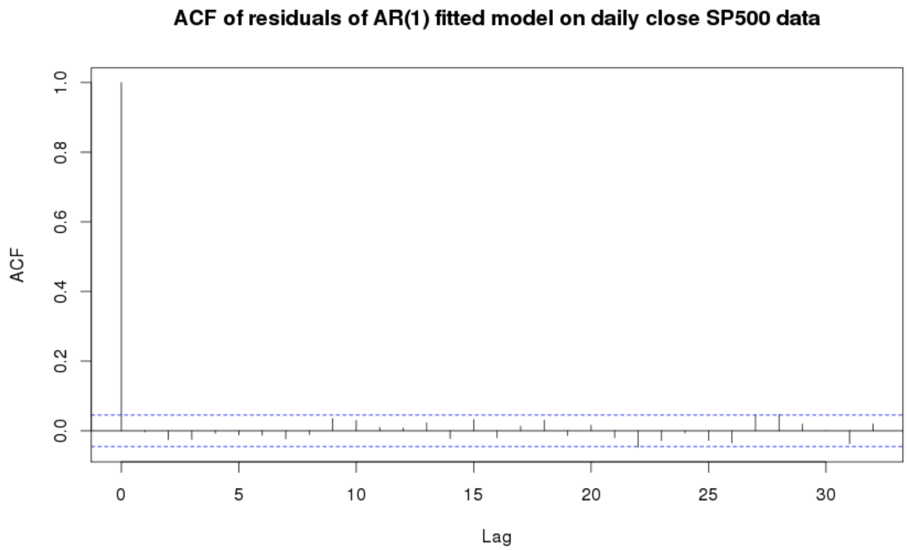

Figure 3: In this figure, bla bla bla...

9.7 Spatial time series ?

9.8 Forecasting

For the parameters to use into the forecasting, instead of the mean, use something else (a function of the coefficients ?). Which size of window to use ? Why is it better ? Also, build a new model of forecasting from the integrated covariation

9.9 To read

Look at the paper Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models

9.10 Noise, endogeneity and returns of the efficient price

9.11 Simon's model

10 Appendix

For proofs of Section 3 and 4, since the parameter process θ_t^* is locally bounded, we can follow standard localisation arguments (see, i.e., 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 $0 \leq t \leq T$.

10.1 Proof of Theorem (consistency)

Proof $(C1) \Rightarrow$**(22)**

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

$$\sup_{i\geq 0} \mathbb{E}\left[\left|\hat{\tilde{\Theta}}_{i,n} - \tilde{\Theta}_{i,n}\right|\right] = o_p(1) \tag{63}$$

By (17), we can write

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

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 condition for $(U^1_{i,n}, \ldots, U^{h_n}_{i,n})$ given $\tilde{\Theta}_{i,n}$ (see, i.e., Breiman (1992)). From $(C1)$ together with assumptions of Section 3, we obtain (63).

Proof $(C2) \Rightarrow$**(23)**

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

$$\sup_{i\geq 0} \mathbb{E}\left[\left|\hat{\Theta}_{i,n} - \hat{\tilde{\Theta}}_{i,n}\right|\right] = o_p(1) \tag{64}$$

Similarly to the previous proof, we can write

$$\left|\hat{\Theta}_{i,n} - \hat{\tilde{\Theta}}_{i,n}\right| = g^{(2)}_n(U^1_{i,n}, \ldots, U^{h_n}_{i,n}, \{\theta^*_s\}_{\tau^0_{i-1,n} \leq s \leq \tau^0_{i,n}}, \tilde{\Theta}_{i,n})$$

$$\begin{aligned} \left|\mathbb{E}\left[\hat{\Theta}_{i,n} - \hat{\tilde{\Theta}}_{i,n}\right]\right| &= \left|\mathbb{E}\left[\mathbb{E}\left[g^{(2)}_n(U^1_{i,n}, \ldots, U^{h_n}_{i,n}, \{\theta^*_s\}_{\tau^0_{i-1,n} \leq s \leq \tau^0_{i,n}}, \tilde{\Theta}_{i,n})|\tilde{\Theta}_{i,n}\right]\right]\right| \\ &= \left|\mathbb{E}\left[\int g^{(2)}_n(v, \tilde{\Theta}_{i,n})\mu_\omega(dv)\right]\right| \\ &= o_p(1) \end{aligned}$$

where $\mu_\omega(dv)$ is a regular condition for $(U^1_{i,n}, \ldots, U^{h_n}_{i,n}, \{\theta^*_s\}_{\tau^0_{i-1,n} \leq s \leq \tau^0_{i,n}})$ given $\tilde{\Theta}_{i,n}$ and where we used $(C2)$ in the last equality.

10.2 Proof of Consistency in Example 1/3

Let's show $(C1)$ first. For any $M > 0$, the quantity

$$\left|\hat{\sigma}^2_{h_n,n}\big(F_n(U^1_{1,n},\sigma^2);\ldots;F_n(U^{h_n}_{1,n},\sigma^2)\big)-\sigma^2\right|$$

can be uniformly in $\{\sigma^2 \in K_M\}$ bounded by

$$C\Big|\sum_{j=1}^{h_n}(\Delta W_{i,j,n})^2 nT^{-1}h_n^{-1}-1\Big| \tag{65}$$

We can prove that (65) tends to 0 in probability using Theorem $I.4.47$ of p.52 in Jacod and Shiryaev (2003) together with strong Markov property of Brownian motions.

To show $(C2)$, let $M > 0$, as well as $(\mathcal{J}^j_{i,n},\theta,\chi_t)$ defined as in $(C2)$. By Lemma 2.2.11 of Jacod and Protter (2012), it is sufficient to show that the following quantity goes to 0.

$$nh_n^{-1}\sum_{j=1}^{h_n}\mathbb{E}_{\mathcal{J}^{j-1}_{i,n}}\Big|\big(\chi_{\tau^0_{i,n}}\Delta W_{[\tau^{j-1}_{i,n};\tau^j_{i,n}]}\big)^2-\Big(\int_{\tau^{j-1}_{i,n}}^{\tau^j_{i,n}}\chi_s dW_s\Big)^2\Big| \tag{66}$$

(66) can be bounded by

$$Ch_n^{-1}\sum_{j=1}^{h_n}\mathbb{E}_{\mathcal{J}^{j-1}_{i,n}}\Big|\chi_{\tau^0_{i,n}}\Delta W_{[\tau^{j-1}_{i,n};\tau^j_{i,n}]}-\int_{\tau^{j-1}_{i,n}}^{\tau^j_{i,n}}\chi_s dW_s\Big| \tag{67}$$

Using Conditional Burkholder-Davis-Gundy inequality (BDG, see, i.e. inequality (2.1.32) of p. 39 in Jacod and Protter (2012)), we can bound (67) by

$$Ch_n^{-1}\sum_{j=1}^{h_n}\underbrace{(\Delta_{\tau_{i,j,n}})^{1/2}}_{O(n^{-1/2})}\underbrace{\mathbb{E}_{\mathcal{J}_{i,j-1,n}}\Big|\sup_{s\in[\tau_{i,0,n},\tau_{i+1,0,n}]}|\chi_{\tau_{i,0,n}}-\chi_s|\Big|}_{o_p(n^{-1/2})}$$

where we used BDG to obtain $o_p(n^{-1/2})$.

10.3 Proof of Consistency in Example 2/4

$(C1)$ can be shown easily. Similarly $(C2)$ is a direct consequence of the definition in (24), (25) together with (11).

10.4 Proof of Theorem (central limit theorem)

Last term of (52) tends to 0 in probability

We aim at showing that

$$\sum_{i=1}^{B_n} \underbrace{n^{\frac{1}{l}} \big(\tilde{\Theta}_{i,n} - \Theta_{i,n}\big) \Delta \mathrm{T}_{i,n}}_{e_{i,n}} \xrightarrow{\mathbb{P}} 0 \tag{68}$$

We first need to show the following lemma.

Lemma 1. *Let $\mathcal{F}_t$ be a filtration, χ_t a null-drift bounded continuous Itô-process adapted to $\mathcal{F}_t$, and τ_1, τ_2, τ_3 bounded $\mathcal{F}_t$-stopping times such that $0 \leq \tau_1 \leq \tau_2 \leq \tau_3$. We have*

$$\mathbb{E}_{\tau_1}\left[\int_{\tau_2}^{\tau_3} \chi_u du\right] = 0$$

Proof. Define $A_t = \int_0^t \chi_u du$. Since χ_t is bounded, A_t is a martingale.

$$\begin{aligned} \mathbb{E}_{\tau_1}\left[\int_{\tau_2}^{\tau_3} \chi_u du\right] &= \mathbb{E}_{\tau_1}\left[A_{\tau_3} - A_{\tau_2}\right] \\ &= \mathbb{E}_{\tau_1}\left[\mathbb{E}_{\tau_2}\left[A_{\tau_3} - A_{\tau_2}\right]\right] \\ &= 0 \end{aligned}$$

□

We define the following scaled continuous interpolation of the left term in (68)

$$E_t^n = \int_0^t \big(\theta_u^* - \theta_{\mathrm{T}_{u,n}^-}^*\big) du \tag{69}$$

where $\mathrm{T}_{u,n}^-$ is the starting point of the current block at time u., i.e. $\mathrm{T}_{u,n}^- = \sup\{\mathrm{T}_{i,n} : \mathrm{T}_{i,n} \leq s\}$.

First step of the proof of (68) In this step, we will show that E_t^n is a continuous martingale. For τ any stopping time, any $s > 0$ and $t > 0$ such that $s < t$, we define $P(\tau, s, t)$ the projection of τ onto the segment $[s, t]$, i.e.

$$P(\tau, s, t) = \begin{cases} s & \text{if } \tau \leq s \\ \tau & \text{if } s \leq \tau \leq t \\ t & \text{if } \tau \geq t \end{cases}$$

For any $s>0$ and $t>0$ such that $s<t$, we have that

$$\begin{aligned}\mathbb{E}_s\left[E_t^n-E_s^n\right] &= \mathbb{E}_s\left[\sum_{i\geq 1}\int_{P(\mathrm{T}_{i-1,n},s,t)}^{P(\mathrm{T}_{i,n},s,t)}\left(\theta_u^*-\theta_{\mathrm{T}_{i-1,n}}^*\right)du\right]\\ &= \sum_{i\geq 1}\mathbb{E}_s\left[\int_{P(\mathrm{T}_{i-1,n},s,t)}^{P(\mathrm{T}_{i,n},s,t)}\left(\theta_u^*-\theta_{\mathrm{T}_{i-1,n}}^*\right)du\right]\end{aligned}$$

By lemma 1, each term of the sum is equal to 0.

Second step of the proof of (68) We compute the limit of $n^{\frac{2}{l}}\mathbb{E}\left[(E_t^n)^2\right]$

$$\begin{aligned}n^{\frac{2}{l}}\mathbb{E}\left[(E_t^n)^2\right] &= \sum_{i=1}^{B_n}\mathbb{E}\left[\mathbb{E}_{\mathrm{T}_{i-1,n}}\left[e_{i,n}^2\right]\right]\\ &= n^{\frac{2}{l}}\sum_{i=1}^{B_n}\mathbb{E}\left[\mathbb{E}_{\mathrm{T}_{i-1,n}}\left[\left(\int_{\mathrm{T}_{i-1,n}}^{\mathrm{T}_{i,n}}(\theta_u^*-\theta_{\mathrm{T}_{i-1,n}}^*)du\right)^2\right]\right]\\ &\leq Cn^{\frac{2}{l}}\sum_{i=1}^{B_n}\mathbb{E}\left[\underbrace{\mathbb{E}_{\mathrm{T}_{i-1,n}}\left[(\Delta\mathrm{T}_{i,n})^2\right]}_{O_p(h_n^2n^{-2})}\underbrace{\mathbb{E}_{\mathrm{T}_{i-1,n}}\left[\sup_{\mathrm{T}_{i-1,n}\leq s\leq\mathrm{T}_{i,n}}(\theta_s^*-\theta_{\mathrm{T}_{i-1,n}}^*)^2\right]}_{O_p(h_nn^{-1})}\right]\\ &\to 0\end{aligned}$$

where we used that E_t^n is a martingale in the first equality, Conditional Cauchy-Schwarz in the inequality, $(E1)$ assumption together with BDG inequality to obtain the big taus. We can concludze that it tends to 0 by (41) of condition $(E2)$. Thus, because $\mathbb{L}^2$-convergence implies convergence in probability, we show (68).

Third term of (52) tends stably in distribution to the asymptotic variance $\int_0^t V_{\theta_s^*}$

In all generality, $A_{i,n}:=n^{\frac{1}{l}}\big(\hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}^*}-\tilde{\Theta}_{i,n}\big)\Delta\mathrm{T}_{i,n}$ 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 $B_{i,n}=A_{i,n}-\mathbb{E}_{\tau_{i-1,n}}\left[A_{i,n}\right]$. We want to use Corollary 3.1 of pp. $58-59$ in Hall and Heyde (1980). We will thus show that the two conditions of the corollary are verified in the two following steps.

First step : We will show in this step that for all $\epsilon>0$,

$$\sum_{i=1}^{B_n}\mathbb{E}_{\tau_{i-1,n}}\left[B_{i,n}^2\mathbf{1}_{\{B_{i,n}>\epsilon\}}\right]\xrightarrow{\mathbb{P}}0 \qquad (70)$$

The conditional Cauchy-Schwarz inequality gives us that each term of the sum in (70) can be bounded by

$$\left(\mathbb{E}_{\tau_{i-1,n}}\left[B_{i,n}^4\right]\mathbb{E}_{\tau_{i-1,n}}\left[\mathbf{1}_{\{B_{i,n}>\epsilon\}}\right]\right)^{\frac{1}{2}} \tag{71}$$

We apply another conditional Cauchy-Schwarz inequality on the left term of (71), then (40) and (46). To deal with the right term of (71), we use (43). Then, we use the block assumption (42) when taking the sum of all the terms in (71) and we can prove (70).

Second step : We will prove that

$$\sum_{i=1}^{B_n}\mathbb{E}_{\tau_{i-1,n}}\left[B_{i,n}^2\right]\xrightarrow{\mathbb{P}}\eta \tag{72}$$

Consider the approximated time $\Delta\tilde{\mathrm{T}}_{i,n} := \sum_{j=1}^{h_n}(\tilde{R}_{i,n}^j)^{(d)}$. Define $\tilde{A}_{i,n} := n^{\frac{1}{l}}\big(\hat{\tilde{\Theta}}_{i,n}^{\mathbf{M}^*} - \tilde{\Theta}_{i,n}\big)\Delta\tilde{\mathrm{T}}_{i,n}$ and the compensated quantity $\tilde{B}_{i,n} = \tilde{A}_{i,n} - \mathbb{E}_{\tau_{i-1,n}}\left[\tilde{A}_{i,n}\right]$. By (47), we have

$$\sum_{i=1}^{B_n}\mathbb{E}_{\tau_{i-1,n}}\left[B_{i,n}^2\right] = \sum_{i=1}^{B_n}\mathbb{E}_{\tau_{i-1,n}}\left[\tilde{B}_{i,n}^2\right] + o_p(1)$$

By assumption (44), we have that

$$\sum_{i=1}^{B_n}\mathbb{E}_{\tau_{i-1,n}}\left[\tilde{B}_{i,n}^2\right] \quad = \quad h_n^{1-\frac{2}{l'}}n^{\frac{2}{l}-1}\sum_{i=1}^{B_n}\mathbb{E}_{\tau_{i-1,n}}\left[V_{\theta^*_{\tau_{i-1,n}}}\Delta\tilde{\mathrm{T}}_{i,n}\right] + o_p(1)$$

Using Lemma 2.2.11 of Jacod and Protter (2012) with (47), we obtain

$$h_n^{1-\frac{2}{l'}}n^{\frac{2}{l}-1}\sum_{i=1}^{B_n}\mathbb{E}_{\tau_{i-1,n}}\left[V_{\theta^*_{\tau_{i-1,n}}}\Delta\tilde{\mathrm{T}}_{i,n}\right] = h_n^{1-\frac{2}{l'}}n^{\frac{2}{l}-1}\sum_{i=1}^{B_n}V_{\theta^*_{\tau_{i-1,n}}}\Delta\mathrm{T}_{i,n} + o_p(1)$$

We can apply now Proosition I.4.44 (p. 51) in Jacod and Shiryaev (2003) and (42) and we get

$$h_n^{1-\frac{2}{l'}}n^{\frac{2}{l}-1}\sum_{i=1}^{B_n}V_{\theta^*_{\tau_{i-1,n}}}\Delta\mathrm{T}_{i,n} \quad\to\quad \int_0^T V_{\theta^*_s}ds$$

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 $S_n := \sum_{i=1}^{B_n}\mathbb{E}_{\tau_{i-1,n}}\left[A_{i,n}\right]$ tends to 0 in probability. An application of (42), (45) together with regular conditional distribution will give us the convergence to 0 of S_n by (**??**).

Second term of (52) tends to 0

This is a straightforward consequence of (50).

first term of (52) tends to 0

We prove it the same way using bla bla of Assumption ($E4$).

References

[1] Aït-Sahalia, Y., J. Fan, C.D. Wang and X. Yang (2013) The Estimation of continuous and discontinuous leverage effect, Discussion paper.

[2] 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.

[3] Andersen, T.G. and T. Bollerslev (1998a) Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts, *International Economic Review*, 39, 885-905.

[4] 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.

[5] 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.

[6] Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001) Modeling and Forecasting Realized Volatility *Econometrica*, 71, 579-625.

[7] Bachelier, L. (1900) *Théorie de la spéculation.* Paris: Gauthier-Villars.

[8] 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.

[9] 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.

[10] Barndorff-Nielsen, O.E. (2004) Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics* 2, 1-37.

[11] Berk, R. H. (1966) Limiting Behavior of Posterior Distributions When the Model Is Incorrect, *Annals of Mathematical Statistics*, 37, 51-58.

[12] Berk, R. H. (1970) Consistency a Posteriori, *Annals of Mathematical Statistics*, 41, 894-906.

[13] Bollerslev, T. (1986) Generalised autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327.

[14] Box, G. and G. Jenkins (1976) *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

[15] Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models, *Mathematical Finance* 8, 291-323.

[16] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1007.

[17] Engle, R.F. and V.K. Ng (1993). Measuring and Testing the Impact of News on Volatility, *Journal of Finance*, 48, 1749-1778.

[18] Fama, E.F. (1965). Random Walks In Stock Market Prices. *Financial Analysts Journal* 21 (5), 55-59.

[19] Faraway, J.J. (2005) *Linear Models with R.* Chapman & Hall/CRC, Boca Raton, FL.

[20] Fisher, R.A. (1922) On the Mathematical Foundations of Theoretical Statistics. Philosophical Transactions of the Royal Society of London, Series A, 222, 309-368.

[21] Fisher, R.A. (1925) Theory of Statistical Estimation. Proceedings of the Cambridge Philosophical Society, 22, 700-725.

[22] Foster, D. and D. Nelson (1996). Continuous record asymptotics for rolling sample variance estimators, *Econometrica* 64, 139-174.

[23] M. Fukasawa 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.

[24] Hall, P. and C.C. Heyde, 1980. *Martingale Limit Theory and its Application.* Academic Press, Boston.

[25] Huber,P. J. (1967) The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions. Proceedings of the Fifth Berkeley Symposiumin Mathematical Statistics and Probability. Berkeley: University of California Press.

[26] Jacod, J. and P. Protter (1998) Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability* 26, 267-307.

[27] Jacod, J. and P. Protter (2012) *Discretization of Processes.* Springer.

[28] Jacod, J. and A. Shiryaev (2003) *Limit Theorems For Stochastic Processes* (2nd ed.). Berlin: Springer-Verlag.

[29] 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.

[30] McCullagh, P. and J.A. Nelder (1989) *Generalized linear models.* London: Chapman & Hall.

[31] Meyn, S.P. and R.L. Tweedie (2009) *Markov Chains and Stochastic Stability.* Cambridge University Press.

[32] Mykland, P.A., and L. Zhang (2006) ANOVA for Diffusions and Itô Processes. *Annals of Statistics* 34, 1931-1963.

[33] Mykland, P.A. and L. Zhang (2008) Inference for Volatility Type Objects and Implications for Hedging, *Statistics and its Interface* 1, 255-278.

[34] Mykland, P.A. and L. Zhang (2009) Inference for Continuous Semimartingales Observed at High Frequency. *Econometrica* 77, 1403-1445.

[35] Mykland, P.A. and L. Zhang (2011) The Double Gaussian Approximation for High Frequency Data. *Scandinavian Journal of Statistics* 38, 215-236.

[36] 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.

[37] Mykland, P.A. and L. Zhang (2014) Assessment of Uncertainty in High Frequency Data: The Observed Asymptotic Variance, Working paper.

[38] Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, 59, 347-370.

[39] Ogihara, T. and N. Yoshida (2014). Quasi-likelihood analysis for nonsynchronously observed diffusion processes. *Stochastic Processes and Applications* 124, 2954-3008.

[40] Potiron, Y. and P.A. Mykland (2015) Estimation of integrated quadratic covariation between two assets with endogenous sampling times. Working paper.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] White, H. (1982) Maximum Likelihood Estimation of Misspecified Models. *Econometrica* 50, 1-25.

[47] Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis.* Almquist and Wicksell.

[48] Xiu, D. (2010) Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics* 159, 235-250

[49] Zhang, L. (2001) From Martingales to ANOVA : Implied and Realized Volatility. Ph.D. Thesis, The University of Chicago, Department of Statistics.