

STATISTICAL PROPERTIES OF MICROSTRUCTURE NOISE

JEAN JACOD

Institut de Mathématiques de Jussieu, CNRS-UMR7586 and Université Pierre et Marie Curie–P6

YINGYING LI

Department of Information Systems, Business Statistics and Operations Management and
Department of Finance, Hong Kong University of Science and Technology

XINGHUA ZHENG

Department of Information Systems, Business Statistics and Operations Management,
Hong Kong University of Science and Technology

We study the estimation of (joint) moments of microstructure noise based on high frequency data. The estimation is conducted under a nonparametric setting, which allows the underlying price process to have jumps, the observation times to be irregularly spaced, and the noise to be dependent on the price process and to have diurnal features. Estimators of arbitrary orders of (joint) moments are provided, for which we establish consistency as well as central limit theorems. In particular, we provide estimators of autocovariances and autocorrelations of the noise. Simulation studies demonstrate excellent performance of our estimators in the presence of jumps, irregular observation times, and even rounding. Empirical studies reveal (moderate) positive autocorrelations of microstructure noise for the stocks tested.

KEYWORDS: Market microstructure noise, high frequency data, joint moments, autocovariance, autocorrelation.

1. INTRODUCTION

IT HAS LONG BEEN RECOGNIZED THAT market microstructure noise plays a profound role in financial markets; see the seminal paper of Black (1986) and comprehensive reviews of Madhavan (2000), O'Hara (1995), Stoll (2003), and Hasbrouck (2007), among others. The market microstructure noise is induced by various frictions in the trading process. Examples of such frictions include bid-ask spread, the discreteness of price, etc. Based on these considerations, various models have been proposed for market microstructure noise; see, for example, Roll (1984) and Glosten and Harris (1988) for modeling the bid-ask spread, and Harris (1990a and 1990b) for the discreteness of price.

With the increasing availability of high frequency data, market microstructure noise has received growing attention. While typically the market microstructure noise is small, it accumulates at high frequency and affects badly the inferences about the efficient price processes such as the estimation of volatilities. One prominent example is the volatility signature plots of Andersen, Bollerslev, Diebold, and Labys (2000a) which show clear noise accumulation effect at high frequency.

Jean Jacod: jean.jacod@gmail.com

Yingying Li: yyli@ust.hk

Xinghua Zheng: xhzheng@ust.hk

We are very grateful to the co-editor and several anonymous referees for their very valuable comments and useful suggestions which led to great improvements of this paper. Financial support from the Research Grants Council (RGC) of Hong Kong under Grants GRF 602710, GRF 606811, GRF 16305315, GRF 16502014, and GRF 16518716 is also gratefully acknowledged.

In the context of volatility estimation, several methods to de-noise the data have been proposed. Widely used methods include the two-scale method (Zhang, Mykland, and Aït-Sahalia (2005)), the multi-scale method (Zhang (2006)), the realized kernel method (Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008)), the pre-averaging method (Jacod, Li, Mykland, Podolskij, and Vetter (2009) and Podolskij and Vetter (2009)), and the quasi-maximum likelihood method (Xiu (2010)). These methods are shown to be effective when the noise is an additive white noise, or admits some kind of independence between successive observations.

Empirically, however, several articles including Hansen and Lunde (2006), Ubukata and Oya (2009), and Aït-Sahalia, Mykland, and Zhang (2011) have found evidence of dependence of microstructure noise in financial markets. Their findings, as well as the research on market microstructure, motivated us to ask the following questions:

Can we understand better the statistical properties of the noise? Specifically, for a particular security price process, how is the microstructure noise distributed and what is the dependence structure?

In this article, we answer these questions by studying how to estimate the moments and joint moments of the noise, based on high frequency data. More specifically, under a general nonparametric setting where the underlying price process can have jumps, the observation times can be irregularly spaced, and the noise can be dependent on the price process and have diurnal features, we propose estimators for (joint) moments of arbitrary orders of the noise. We establish consistency as well as central limit theorems (CLTs) for our estimators under mild mixing conditions on the noise (see Assumption (N) below for precise statements).

Our approach is nonparametric and does not depend on particular modeling of the noise. While a specific model of noise could provide insight into the source of the noise and greatly facilitate the estimation, there is always the risk of model misspecification. In contrast, our method works in relatively general situations, can be directly applied to tick-by-tick data, and enables one to understand the marginal distribution, in particular, the size of noise, as well as the dependence structure. This information is valuable in various studies such as volatility estimation, measuring liquidity, and high frequency trading.

Applications of our results include estimating the autocovariances and autocorrelations of the noise. There are a few papers which contain results on second order features of the noise. In particular, Zhang, Mykland, and Aït-Sahalia (2005) and Bandi and Russell (2008) proposed an estimator of the variance of the noise. When the noise is colored, the estimator no longer estimates the variance; see our discussions in Section 3.1.3. More generally, Hansen and Lunde (2006) considered a number of specifications for the dependence of the noise including white noise, finite dependence in calendar time, and a specific form of dependence in tick time, studied the impact of the dependence of the noise on volatility estimation, and obtained interesting empirical findings regarding the noise as well as the dependence between the price and the noise. In this paper, we study broader aspects of the microstructure noise. And our method works under much more general settings; in particular, we allow the noise to have rather general dependence structures and further to have diurnal features. Furthermore, we establish the corresponding CLTs, based on which one can, for example, test whether autocorrelations of particular lags vanish or not.

Empirically, when applying our estimators to tick-by-tick transaction data, we find that (1) microstructure noise tends to be positively autocorrelated; and (2) while the autocorrelations do appear to decay to zero as the lag increases, the decay can be as slow as polynomial. We believe that it is important and useful to incorporate these features into new models for market microstructure noise.

The paper is organized as follows. Section 2 introduces the setting and assumptions. Section 3 presents the consistency and asymptotic normality results of our proposed estimators for the (joint) moments of the noise. Section 4 demonstrates our results via simulations. Empirical studies are carried out in Section 5. Section 6 concludes. The proofs are given in the Appendix, with some technical details given in the Supplemental Material (Jacod, Li, and Zheng (2017)), which also contains some supplementary numerical results.

2. SETTING AND ASSUMPTIONS

We have three basic ingredients. The first one is the underlying process X , typically the latent efficient log-price of an asset, the second one is the observation scheme, and the third one is the noise. Below, we explain the respective assumptions in order.

2.1. Assumptions on the Latent Process X

The assumptions on the latent process X are rather general, and we allow for both unpredictable and predictable jumps. Predictable jumps are incorporated to account for individual company or macroeconomic announcements occurring at predictable (typically pre-announced) times. Basically, X is the sum of an Itô semimartingale X' , possibly discontinuous, plus the sum X'' of jumps occurring at these pre-announced times. That is, X is defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, and it takes the following form, where we use the Grigelionis representation of X' :

$$\begin{aligned} X &= X' + X'', \quad \text{where} \\ X'_t &= X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + (\delta 1_{\{|\delta| \leq 1\}}) \star (\underline{p} - \underline{q})_t + (\delta 1_{\{|\delta| > 1\}}) \star \underline{p}_t, \quad \text{and} \\ X''_t &= \sum_{i \geq 1} \Gamma_i 1_{\{0 < S_i \leq t\}}. \end{aligned} \quad (2.1)$$

Here, t stands for time, W is a standard Brownian motion, $\underline{p}$ is a Poisson random measure on $\mathbb{R}_+ \times E$, where $(E, \mathcal{E})$ is a Polish space, with a non-random intensity measure of the form $\underline{q}(dt, dz) = dt \otimes \lambda(dz)$ with λ a σ -finite measure on $(E, \mathcal{E})$, and $(\delta 1_{\{|\delta| \leq 1\}}) \star (\underline{p} - \underline{q})_t = \int_{[0, t] \times E} \delta(\omega, s, z) 1_{\{|\delta(\omega, s, z)| \leq 1\}} (\underline{p} - \underline{q})(\omega, ds, dz)$ and similarly for $(\delta 1_{\{|\delta| > 1\}}) \star \underline{p}_t$. Of course, we have $X'_0 = X_0$, and the formula for X' is the general form of an Itô semimartingale starting at 0. We assume the following (see Jacod and Shiryaev (2003, pp. 5 and 66), for the precise definitions of an optional process and a predictable function on $\Omega \times \mathbb{R}_+ \times E$):

ASSUMPTION (H): (i) (for X') *The process b is optional locally bounded, the process σ is adapted càdlàg, the function δ is predictable, and there is a localizing sequence (τ_n) of stopping times and, for each n , a deterministic nonnegative function J_n on E satisfying $\int J_n(z)^2 \lambda(dz) < \infty$ and such that $|\delta(\omega, t, z)| \wedge 1 \leq J_n(z)$ for all (ω, t, z) with $t \leq \tau_n(\omega)$.*

(ii) (for X'') *$(S_i)_{i \geq 1}$ is a strictly increasing sequence of stopping times tending to ∞ , and the process X'' is adapted, that is, each Γ_i is $\mathcal{F}_{S_i}$ -measurable.*

Note that (i) above is only slightly more restrictive than asking X' to be an Itô semimartingale. In virtually all the continuous-time econometrics literature, the price or log-price process is assumed to satisfy this assumption, and often much more; see, for example, Aït-Sahalia and Jacod (2014) for a comprehensive review. Our ability to estimate the

characteristics of the noise is not impaired by the presence of jumps, with finite or infinite activity, nor for that matter by the presence or absence of a Brownian component (no Brownian component means that σ vanishes identically).

Note also that, although not implied by (H), the S_i 's are typically predictable times or even non-random times.

2.2. Assumptions on the Observation Times

Next, we describe how observations take place. At stage n , that is, for a given frequency of observations, the successive observations occur at times $0 = T(n, 0) < T(n, 1) < \dots$, for a sequence $T(n, i)$ of (possibly random) finite times, and we set

$$N_t^n = \sum_{i \geq 1} 1_{\{T(n, i) \leq t\}}, \quad \Delta(n, i) = T(n, i) - T(n, i-1). \quad (2.2)$$

So the number of observations up to time t is $N_t^n + 1$, and $\Delta(n, i)$ is the i th inter-observation lag at stage n , and knowing the times $T(n, i)$ for $i \geq 0$ amounts to knowing the path of the counting process N^n . Of course, in the case of regularly spaced observations with lag Δ_n at stage n , we have $\Delta(n, i) = \Delta_n$ and $N_t^n = [t/\Delta_n]$.

The assumption is then as follows, where Δ_n is a sequence of non-random positive numbers going to 0 and, as usual, we write $\mathcal{F}_\infty = \bigvee_{t \geq 0} \mathcal{F}_t$:

ASSUMPTION (O): *On the space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, we have two semimartingales $\alpha = (\alpha_t)$ and $\bar{\alpha} = (\bar{\alpha}_t)$ which satisfy the same Assumption (H-i) as X' (with different coefficients than in (2.1)) and a localizing sequence (τ_m) of stopping times and positive constants $\kappa_{m,p}$ and κ such that:*

- (i) *If $t < \tau_m$, we have $1/\kappa_{m,1} \leq \alpha_{t-} \leq \kappa_{m,1}$ and $\bar{\alpha}_{t-} \leq \kappa_{m,1}$.*
- (ii) *If $(\mathcal{F}_t^n)$ is the smallest filtration containing $(\mathcal{F}_t)$ and with respect to which all $T(n, i)$ for $i = 1, 2, \dots$ are stopping times, for each i the variable $\Delta(n, i)$ is, conditionally on $\mathcal{F}_{T(n, i-1)}^n$, independent of $\mathcal{F}_\infty$.*
- (iii) *In restriction to the set $\{T(n, i-1) < \tau_m\}$, and for all $p > 0$, we have*

$$\begin{aligned} \left| \mathbb{E}(\Delta(n, i) \mid \mathcal{F}_{T(n, i-1)}^n) - \frac{\Delta_n}{\alpha_{T(n, i-1)}} \right| &\leq \kappa_{m,1} \Delta_n^{3/2+\kappa}, \\ \left| \mathbb{E}((\Delta(n, i) \alpha_{T(n, i-1)} - \Delta_n)^2 \mid \mathcal{F}_{T(n, i-1)}^n) - \Delta_n^2 \bar{\alpha}_{T(n, i-1)} \right| &\leq \kappa_{m,2} \Delta_n^{2+\kappa}, \\ \mathbb{E}(\Delta(n, i)^p \mid \mathcal{F}_{T(n, i-1)}^n) &\leq \kappa_{m,p} \Delta_n^p. \end{aligned} \quad (2.3)$$

Note that in practice, the observation times $T(n, i)$ are transaction times (or revision times if one uses quotes data) which are irregularly spaced and are naturally stopping times. The Δ_n can be thought of as an “average mesh size,” which characterizes how frequently the observations/transactions occur. In contrast with the times $T(n, i)$ and the lags $\Delta(n, i)$, it is not observable and is rather a mathematical abstraction, and it does not appear as such in the feasible statistics given below. For example, if we replace Δ_n by $a\Delta_n$ for some constant $a > 0$, Assumption (O) is still satisfied, upon replacing α_s by $a\alpha_s$ and with $\bar{\alpha}_s$ unchanged, and such a modification does not affect the feasible statistics below, nor the asymptotic variances in the feasible CLTs.

Assumption (O-ii) means that the time lag $\Delta(n, i)$ may depend on the process X (or on any other $(\mathcal{F}_t)$ -adapted process), and on the previous sampling times, up to $T(n, i-1)$,

but is independent of the evolution of X after that time. We will see that it implies that X is still a semimartingale relative to each filtration $(\mathcal{F}_t^n)$, with the same dynamics (2.1).

The process α_t is a sort of “density” of observations, as revealed by the following convergence in probability:

$$\Delta_n N_t^n \xrightarrow{\mathbb{P}} A_t := \int_0^t \alpha_s ds \quad (2.4)$$

(which is even uniform in t over each compact interval), as will be shown later. The second part of (2.3) can be interpreted as a regularity condition, which in particular entails a rate of convergence $1/\sqrt{\Delta_n}$ in (2.4), as shown later again.

Below, we give some simple examples for which Assumption (O) is satisfied.

EXAMPLE 2.1—Regular Sampling Scheme: We have $\Delta(n, i) \equiv \Delta_n$, and Assumption (O) is trivially satisfied with $\alpha_t = 1$ and $\bar{\alpha}_t = 0$.

EXAMPLE 2.2—Time-Changed Regular Sampling Scheme: We let α be as in Assumption (O) and set $T(n, i) = \inf\{t : A_t = i\Delta_n\}$, with A given by (2.4). We then have $N_t^n = [A_t/\Delta_n]$ and Assumption (O) holds with α_t as above and $\bar{\alpha}_t = 0$.

EXAMPLE 2.3—Modulated Poisson Sampling Scheme: We have a sequence of standard Poisson process $\bar{N}$ with rate 1, independent of $\mathcal{F}_\infty$, and we let α be as in Assumption (O) and associate A by (2.4). The observation times at stage n are defined, through the first part of (2.2), by $N_t^n = \bar{N}_{A_t/\Delta_n}$. In this case, N^n is, conditionally on $\mathcal{F}_\infty$, a non-homogeneous Poisson process with intensity α_t/Δ_n . Assumption (O) holds with α as above and $\bar{\alpha}_t = 1$.

EXAMPLE 2.4—Predictably-Modulated Random Walk Sampling Scheme: Let α_t be as in the previous example and $(\Phi_i^n : i, n \geq 1)$ be a double sequence of i.i.d. positive random variables, independent of $\mathcal{F}_\infty$ and with mean 1, variance a^2 , and finite fourth moment. At stage n , the observation times are defined recursively, starting with $T(n, 0) = 0$ and proceeding through $T(n, i) = T(n, i-1) + \frac{\Delta_n}{\alpha_{T(n, i-1)}} \Phi_i^n$. Then Assumption (O) holds with α_t as above and $\bar{\alpha}_t = a^2$.

One can mention that there are other natural sampling schemes which do not satisfy (O). For example, if $T(n, 0) = 0$ and $T(n, i+1) = \inf(t > T(n, i) : |X_t - X_{T(n, i)}| > a_n)$ for a sequence a_n decreasing to 0, the mesh of this scheme goes to 0 as soon as X has no interval of constancy, but the conditional independence in (ii) of Assumption (O) is *not* satisfied.

2.3. Assumptions on the Noise

Finally, at time $T(n, i)$ the process X is contaminated by some noise, meaning that we observe the variable

$$Y_i^n = X_{T(n, i)} + \varepsilon_i^n, \quad (2.5)$$

where the noise is ε_i^n .

For the sake of motivation about our forthcoming assumption on the noise, we (temporarily) suppose that the noise is independent of X , stationary, with mean zero and with a negative exponential covariance. This covers a whole range of “natural” situations, the two extreme ones being as follows:

(1) Conditionally on the observation times, the covariance between ε_i^n and ε_{i+j}^n is of the form $ae^{-d'(T(n,i+j)-T(n,i))}$: so the exponential covariance is in terms of calendar time and does not depend on the observation scheme.

(2) The covariance between ε_i^n and ε_{i+j}^n is $ae^{-d'j}$: so the covariance between two values of the noise depends only on how many observations (or, transactions) occurred in between the corresponding times.

And, of course, there are mid-term possibilities, like the covariance being $ae^{-d'ju_n}$ with a “scaling” sequence u_n going to 0 slower than $(T(n, i+1) - T(n, i))$.

In the first situation above, it is of course impossible to obtain consistent estimators for the characteristics of the noise, such as the autocovariance function (in the negative exponential case, as well as in a completely general case), unless the time horizon considered goes to infinity. Such a setting has been studied (see, e.g., [Ubukata and Oya \(2009\)](#)), and here we are interested in the case where the time horizon is fixed. In the second extremal situation, and in all intermediate cases, it is in principle possible to consistently estimate the characteristics of the noise, under appropriate assumptions, of course. The extreme case (2) above is obviously simpler than the intermediate cases, and should already provide useful insight, so below we focus on this second extremal case. Furthermore, our empirical studies suggest that it is reasonable to assume the autocorrelations as a function of tick time.

Before stating the precise assumption on the noise, and for completeness, we recall the ρ -mixing property of a stationary sequence $\chi = (\chi_i)_{i \in \mathbb{Z}}$ of variables, indexed by $\mathbb{Z}$: letting $\mathcal{G}_j = \sigma(\chi_i : i \leq j)$ and $\mathcal{G}^j = \sigma(\chi_i : i \geq j)$ be the pre- and post- σ -fields at time j , the ρ -mixing coefficients of χ for $k \geq 1$ are

$$\rho_k(\chi) = \sup\{|\mathbb{E}(UV)| : \mathbb{E}(U) = \mathbb{E}(V) = 0, \mathbb{E}(U^2) \leq 1, \mathbb{E}(V^2) \leq 1, \\ U \text{ is } \mathcal{G}_0\text{-measurable, } V \text{ is } \mathcal{G}^k\text{-measurable}\}, \quad (2.6)$$

and we say that χ is v -polynomially ρ -mixing if, for some $K > 0$, $\rho_k(\chi) \leq K/k^v$ for all $k \geq 1$, where v is a positive number. Then, the assumption on the noise is as follows:

ASSUMPTION (N): *The noise $(\varepsilon_i^n)_{i \geq 0}$ can be realized as*

$$\varepsilon_i^n = \gamma_{T(n,i)} \cdot \chi_i, \quad (2.7)$$

where γ is a nonnegative Itô semimartingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, which satisfies the same Assumption (H-i) as X' (with different coefficients than in (2.1)) and is not identically zero on any interval. Further, $(\chi_i)_{i \in \mathbb{Z}}$ is a stationary process, independent of the σ -field $\mathcal{F}_\infty = \bigvee_{t \geq 0} \mathcal{F}_t$, and with mean 0 and variance 1, with finite moments of all orders, and which is v -polynomially ρ -mixing for some $v > 0$.

Here, the process γ is included to accommodate possible diurnal features of the noise so that the size of the noise may change over time, while the autocorrelation remains roughly unchanged. The condition that γ is not identically zero on any interval guarantees that the noise is really present. In the empirical studies, we do find that the noise exhibits diurnal features (even though in a not-so-strong manner).

REMARK 2.5: This assumption could be weakened by asking finite moments up to a suitable order only: for example, if one is interested in estimating the autocovariance

function of the process χ , we only need finite moments up to order q , bigger than but arbitrarily close to 4.

The ρ -mixing condition could also be replaced by α -mixing or ϕ -mixing, or by any other condition implying ergodicity and a CLT for all functionals of the type $\sum_{i=1}^n f(\chi_i)$ when $\mathbb{E}(f(\chi_0)) = 0$ and $\mathbb{E}(|f(\chi_0)|^q) < \infty$ for all $q > 0$.

REMARK 2.6: The noise can be *dependent* on X , a form of dependency being induced by the presence of the process γ . However, our assumption implies that the noise and the returns of X are not correlated; this is a drawback of the model used here.

REMARK 2.7: It should be noted that the observations in our model have the form

$$Y_i^n = X_{T(n,i)} + \gamma_{T(n,i)} \chi_i^n, \quad (2.8)$$

where each sequence $(\chi_i^n : i \geq 0)$ might depend on n , but has a (joint) distribution *independent* of n . In other words, our model does *not* provide a definition of noise which is “consistent” with a change of observation times, unless of course the noise is white.

This inconsistency might be considered as a major drawback in the case of a colored noise. However, in practice, the frequency of observations is indeed fixed by the available data and does not really go to infinity. The interpretation of the asymptotic as $\Delta_n \rightarrow 0$ is thus that the frequency of our observations is “high enough” for considering that we are “almost” in the asymptotic regime. With this view, adding a noise which has a structure as specified above seems legitimate.¹

3. ESTIMATION OF THE MOMENTS OF THE NOISE

We are primarily interested in estimating the autocovariance of the noise. For a feasible CLT of the estimators of the autocovariance, we also need to estimate appropriate fourth moments, so below we develop a theory of estimation of all composite integer moments of the noise, which, apart from notational complexity, is not more difficult than for the autocovariance itself.

Toward this aim, we introduce some general notation. Let $\mathcal{J}$ be the set of all finite sequences of integers $\mathbf{j} = (j_1, j_2, \dots, j_q)$ (they are neither necessarily ordered, nor necessarily distinct, and $q \geq 1$), and $\mathcal{J}^+$ be the subset of all of $\mathbf{j} = (j_1, \dots, j_q)$ with $\min(j_\ell) = 0$ and $q \geq 2$. With each $\mathbf{j} \in \mathcal{J}$, we associate the integer composite moments of χ as

$$\mathbf{r}(\mathbf{j}) = \mathbf{r}((j_1, \dots, j_q)) = \mathbb{E} \left(\prod_{\ell=1}^q \chi_{j_\ell} \right). \quad (3.1)$$

We will further use the notation

$$\mathbf{j} = (j_1, \dots, j_q), \mathbf{j}' = (j'_1, \dots, j'_q) \mapsto \begin{cases} q(\mathbf{j}) = q, & \mu(\mathbf{j}) = \max(j_1, \dots, j_q), \\ \mathbf{j} \oplus \mathbf{j}' = (j_1, \dots, j_q, j'_1, \dots, j'_q), \\ \mathbf{j} + m = (m + j_1, \dots, m + j_q) & \text{if } m \in \mathbb{Z}. \end{cases} \quad (3.2)$$

¹In some cases, people might want to do the analysis with subsampled data such as the observations at times $T(n, ki)$ for $i = 0, 1, \dots$ for various values of $k = 2, 3, \dots$ (for instance, to check the “stability” of the method). In that case, one simply needs to replace the autocovariance $r(j) = \mathbb{E}(\chi_0 \chi_j)$ of the noise by the re-scaled autocovariance $r(jk)$.

We have $\mathbf{r}(\mathbf{j}) = 0$ when $q(\mathbf{j}) = 1$, and $\mathbf{r}(\mathbf{j}) = \mathbf{r}(\mathbf{j}_{+m})$ for all $m \in \mathbb{Z}$, so we restrict our attention to the estimation of $\mathbf{r}(\mathbf{j})$ when $\mathbf{j} \in \mathcal{J}^+$.

Of special interest is the autocovariance $r(j)$ (which, since $r(0) = 1$, is also the autocorrelation) of χ , that is,

$$r(j) = \mathbf{r}((0, j)). \quad (3.3)$$

Recall that the noise takes the form of $\varepsilon_i^n = \gamma_{T(n,i)} \cdot \chi_i$. In the special case when (γ_t) is constant, $r(j)$ is truly the autocorrelation of the noise. In general, since (γ_t) is a semimartingale and hence relatively stable within small time period, $r(j)$ still roughly measures the autocorrelation of the noise for small values of j .

Next, we discuss the size of the noise. Due to the convention $r(0) = 1$, knowing the $r(j)$'s tells us nothing about the actual size of the noise. The more relevant quantity is the following “average value” of the moments over the time interval of interest $[0, t]$, that is,

$$\mathbf{R}(\mathbf{j})_t = \frac{\mathbf{r}(\mathbf{j})}{A_t} \int_0^t \gamma_s^{q(\mathbf{j})} dA_s, \quad \text{and} \quad R(j)_t = \frac{r(j)}{A_t} \int_0^t \gamma_s^2 dA_s, \quad (3.4)$$

where A_t is defined in (2.4). Observe that $R(j)_t$ is indeed the limit of the average value of $\varepsilon_i^n \varepsilon_{i+j}^n$ for the “true noise” ε_i^n , over all $i = 0, 1, \dots, N_t^n$ (i.e., of the empirical covariance of the noise), as $n \rightarrow \infty$. In particular, $R(0)_t$ is the limit of the average value of $(\varepsilon_i^n)^2$, hence provides a natural measure of the size of the noise. We shall call $R(j)_t$ as the “scaled autocovariance” below. Note that $r(j) = R(j)_t / R(0)_t$.

3.1. Consistency Results

3.1.1. Convergences of Moment Estimates

For estimating $\mathbf{R}(\mathbf{j})$, we first choose a sequence $k_n \geq 2$ of integers which satisfies, with Δ_n as in Assumption (O) and some $\eta \in (0, 1/2)$:

$$k_n \rightarrow \infty, \quad k_n \Delta_n^\eta \rightarrow 0. \quad (3.5)$$

Then we set

$$\bar{X}_i^n = \frac{1}{k_n} \sum_{j=0}^{k_n-1} X_{T(n,i+j)}, \quad \bar{Y}_i^n = \frac{1}{k_n} \sum_{j=0}^{k_n-1} Y_{i+j}^n, \quad \bar{\chi}_i^n = \frac{1}{k_n} \sum_{j=0}^{k_n-1} \chi_{i+j}, \quad (3.6)$$

and for $\mathbf{j} = (j_1, \dots, j_q)$ and $\mu = \mu(\mathbf{j})$, consider the processes

$$U(\mathbf{j})_t^n = \sum_{i=0}^{N_t^n + 1 - \mu - 2qk_n} \hat{Y}(\mathbf{j})_i^n, \quad \text{where} \quad \hat{Y}(\mathbf{j})_i^n = \prod_{\ell=1}^q (Y_{i+j_\ell}^n - \bar{Y}_{i+\mu+(2\ell-1)k_n}^n). \quad (3.7)$$

Observe that the second condition in (3.5) ensures that in (3.6) we are taking *local* averages. The main intuition behind the definition in (3.7) is that $\bar{Y}_{i+\mu+(2\ell-1)k_n}^n \approx X_{T(n,i+j_\ell)}$, and so $Y_{i+j_\ell}^n - \bar{Y}_{i+\mu+(2\ell-1)k_n}^n \approx \varepsilon_{i+j_\ell}^n$. The index for $\bar{Y}^n$ above is chosen to ensure that the noise components in $Y_{i+j_\ell}^n$ and in the $\bar{Y}_{i+\mu+(2\ell-1)k_n}^n$'s are separated by at least k_n indices, implying that they are “independent enough.” The sum above, as everywhere else below, is set to be 0 when the upper limit is smaller than the lower limit, that is, $N_t^n < \mu + 2qk_n - 1$, but

for any $t > 0$ this is not the case when n is large enough. The upper limit of the sum above is such that $U(\mathbf{j})_t^n$ uses only data within the time interval $[0, t]$, and all these data.

The main consistency results are as follows.

THEOREM 3.1: *Under Assumptions (H), (O), (N), and (3.5), for all $t > 0$ and $\mathbf{j} \in \mathcal{J}^+$, we have the convergences in probability:*

$$\Delta_n U(\mathbf{j})_t^n \xrightarrow{\mathbb{P}} \mathbf{r}(\mathbf{j}) \int_0^t \gamma_s^{q(\mathbf{j})} dA_s \quad (3.8)$$

and

$$\frac{1}{N_t^n} U(\mathbf{j})_t^n \xrightarrow{\mathbb{P}} \mathbf{R}(\mathbf{j})_t. \quad (3.9)$$

REMARK 3.2: A seemingly more natural definition of $U(\mathbf{j})_t^n$ is the following:

$$U(\mathbf{j})_t^{n,\text{alt}} = \sum_{i=0}^{N_t^n+1-\mu-qk_n} \prod_{\ell=1}^q (Y_{i+j_\ell}^n - \bar{Y}_{i+j_\ell}^n), \quad (3.10)$$

namely, when taking differences, one just subtracts the averaged price at the same time point from the observed price. For such an alternative definition, by using the same arguments as for Theorem 3.1, we can prove that $\frac{1}{N_t^n} U(\mathbf{j})_t^{n,\text{alt}} \xrightarrow{\mathbb{P}} \mathbf{R}(\mathbf{j})_t$ under (3.5). However, for the associated CLT, one needs (3.5) and additionally, $k_n^{1 \wedge v} \Delta_n \rightarrow \infty$ and $v > 1$. These conditions are, however, all together contradictory, which is why we do not adopt $U(\mathbf{j})_t^{n,\text{alt}}$, because it does not enjoy a CLT within the asymptotic framework of this paper. It is, however, possible that $U(\mathbf{j})_t^{n,\text{alt}}$ is slightly better behaved in finite samples especially when $\mu(\mathbf{j})$ is large.

Since $\Delta_n N_t^n \xrightarrow{\mathbb{P}} A_t$, (3.9) is a trivial consequence of (3.8). Note that because Δ_n is a priori unknown, the convergence (3.8) is a purely mathematical result that does not provide an estimator for the right-hand side. In contrast, N_t^n and $U(\mathbf{j})_t^n$ are observed and (3.9) provides a feasible estimator for $\mathbf{R}(\mathbf{j})_t$, in particular,

$$\widehat{R}(j)_t^n := \frac{1}{N_t^n} U(0, j)_t^n \xrightarrow{\mathbb{P}} R(j)_t, \quad \text{for } j = 0, 1, \dots \quad (3.11)$$

Moreover, about the autocorrelations $r(j) = E(\chi_0 \chi_j)$, since $r(j) = R(j)_t / R(0)_t$ for all $t > 0$, we obtain that

$$\widehat{r}(j)_t^n := \frac{\widehat{R}(j)_t^n}{\widehat{R}(0)_t^n} = \frac{U(0, j)_t^n}{U(0, 0)_t^n} \xrightarrow{\mathbb{P}} r(j). \quad (3.12)$$

3.1.2. Convergences in Preparation for the CLTs

Next, we give some convergences in preparation for the CLTs associated with the previous convergences, under the additional assumption $v > 1$ in Assumption (N). The CLTs involve some limiting variances-covariances, based on the following quantities, where

$\mathbf{j}, \mathbf{j}' \in \mathcal{J}^+$:

$$\mathbf{s}(\mathbf{j}, \mathbf{j}') = \sum_{m \in \mathbb{Z}} (\mathbf{r}(\mathbf{j} \oplus \mathbf{j}'_{+m}) - \mathbf{r}(\mathbf{j})\mathbf{r}(\mathbf{j}')) \quad (3.13)$$

(this series is absolutely convergent when $v > 1$, as seen later). If $\mathbf{j}_l \in \mathcal{J}^+$ for $l = 1, \dots, d$, the $d \times d$ symmetric matrix $(\mathbf{s}(\mathbf{j}_l, \mathbf{j}_m))_{1 \leq l, m \leq d}$ is a covariance matrix, sometimes referred to as the long-run covariance matrix in econometrics.

More specifically, in order to have “feasible” CLTs, we need consistent estimators for the following quantities:

$$\begin{aligned} \frac{1}{A_t} \int_0^t \bar{\alpha}_s dA_s, \quad \frac{\mathbf{r}(\mathbf{j})}{A_t} \int_0^t \gamma_s^{q(\mathbf{j})} \bar{\alpha}_s dA_s, \\ \frac{\mathbf{r}(\mathbf{j})\mathbf{r}(\mathbf{j}')}{A_t} \int_0^t \gamma_s^{q(\mathbf{j})+q(\mathbf{j}')} \bar{\alpha}_s dA_s, \quad \frac{\mathbf{s}(\mathbf{j}, \mathbf{j}')}{A_t} \int_0^t \gamma_s^{q(\mathbf{j})+q(\mathbf{j}')} dA_s. \end{aligned}$$

Toward this end, we choose a sequence ϕ_n of positive numbers satisfying

$$\frac{\phi_n}{k_n \Delta_n} \rightarrow 0, \quad \frac{k_n^{3/4} \Delta_n}{\phi_n} \rightarrow 0. \quad (3.14)$$

For $\mathbf{j} = (j_1, \dots, j_d)$ and $\mathbf{j}' = (j'_1, \dots, j'_d)$ in $\mathcal{J}^+$, with $\mu = \mu(\mathbf{j})$ and $\mu'' = \mu + \mu(\mathbf{j}')$ and $q'' = q + q'$, and with the notation $\hat{Y}(\mathbf{j})_i^n$ of (3.7), we set

$$\begin{aligned} U_t^{1,n} &= \sum_{i=0}^{N_t^n - 2 - 2k_n} \hat{\Delta}_i^n, \\ \text{with } \hat{\Delta}_i^n &= \left(\frac{k_n \Delta(n, i+1+k_n) - T(n, i+2+2k_n) + T(n, i+2+k_n)}{(T(n, i+k_n) - T(n, i)) \vee \phi_n} \right)^2, \\ U_t^2(\mathbf{j})_t^n &= \sum_{i=0}^{N_t^n - 1 - \mu - (2q+2)k_n} \hat{\Delta}_i^n \hat{Y}(\mathbf{j})_{i+2+2k_n}^n, \\ U_t^3(\mathbf{j}, \mathbf{j}')_t^n &= \sum_{i=0}^{N_t^n - 1 - \mu'' - (2q''+3)k_n} \hat{\Delta}_i^n \hat{Y}(\mathbf{j})_{i+2+2k_n}^n \hat{Y}(\mathbf{j}')_{i+2+\mu+(2q+3)k_n}^n, \\ U_t^4(\mathbf{j}, \mathbf{j}')_t^n &= \sum_{i=0}^{N_t^n + 1 - \mu'' - (2q''+1)k_n} \hat{Y}(\mathbf{j})_i^n \hat{Y}(\mathbf{j}')_{i+\mu+(2q+1)k_n}^n, \end{aligned} \quad (3.15)$$

and, with k'_n another sequence of integers,

$$S(\mathbf{j}, \mathbf{j}')_t^n = \left(U(\mathbf{j} \oplus \mathbf{j}')_t^n + \sum_{m=1}^{k'_n} (U(\mathbf{j} \oplus \mathbf{j}'_{+m})_t^n + U(\mathbf{j}_{+m} \oplus \mathbf{j}')_t^n) \right) - (2k'_n + 1) U^4(\mathbf{j}, \mathbf{j}')_t^n. \quad (3.16)$$

The first part on the right-hand side above is a kind of heteroscedastic and autocorrelation consistent (HAC) estimator. The intuition is that $\mathbf{s}(\mathbf{j}, \mathbf{j}') \approx \sum_{|m| \leq k'_n} (\mathbf{r}(\mathbf{j} \oplus \mathbf{j}'_{+m}) - \mathbf{r}(\mathbf{j})\mathbf{r}(\mathbf{j}'))$ for a reasonably large (but finite) k'_n , and so we only need to estimate a finite sum.

THEOREM 3.3: Assume (H), (O), and (N), and let $\mathbf{j}, \mathbf{j}' \in \mathcal{J}^+$. If k_n and ϕ_n satisfy (3.5) and (3.14), we have

$$\Delta_n U_t^{1,n} \xrightarrow{\mathbb{P}} \int_0^t \bar{\alpha}_s dA_s, \quad (3.17)$$

$$\frac{1}{N_t^n} U_t^{1,n} \xrightarrow{\mathbb{P}} \frac{1}{A_t} \int_0^t \bar{\alpha}_s dA_s;$$

$$\Delta_n U^2(\mathbf{j})_t^n \xrightarrow{\mathbb{P}} \mathbf{r}(\mathbf{j}) \int_0^t \gamma_s^{q(\mathbf{j})} \bar{\alpha}_s dA_s, \quad (3.18)$$

$$\frac{1}{N_t^n} U^2(\mathbf{j})_t^n \xrightarrow{\mathbb{P}} \frac{\mathbf{r}(\mathbf{j})}{A_t} \int_0^t \gamma_s^{q(\mathbf{j})} \bar{\alpha}_s dA_s;$$

$$\Delta_n U^3(\mathbf{j}, \bar{\mathbf{j}})_t^n \xrightarrow{\mathbb{P}} \mathbf{r}(\mathbf{j}) \mathbf{r}(\bar{\mathbf{j}}) \int_0^t \gamma_s^{q(\mathbf{j})+q(\bar{\mathbf{j}})} \bar{\alpha}_s dA_s, \quad (3.19)$$

$$\frac{1}{N_t^n} U^3(\bar{\mathbf{j}}, \bar{\mathbf{j}})_t^n \xrightarrow{\mathbb{P}} \frac{\mathbf{r}(\mathbf{j}) \mathbf{r}(\bar{\mathbf{j}})}{A_t} \int_0^t \gamma_s^{q(\mathbf{j})+q(\bar{\mathbf{j}})} \bar{\alpha}_s dA_s;$$

$$\Delta_n U^4(\mathbf{j}, \mathbf{j}')_t^n \xrightarrow{\mathbb{P}} \mathbf{r}(\mathbf{j}) \mathbf{r}(\mathbf{j}') \int_0^t \gamma_s^{q(\mathbf{j})+q(\mathbf{j}')} dA_s, \quad (3.20)$$

$$\frac{1}{N_t^n} U^4(\mathbf{j}, \mathbf{j}')_t^n \xrightarrow{\mathbb{P}} \frac{\mathbf{r}(\mathbf{j}) \mathbf{r}(\mathbf{j}')}{A_t} \int_0^t \gamma_s^{q(\mathbf{j})+q(\mathbf{j}')} dA_s.$$

If, further, $v > 1$ in (N) and k'_n satisfies

$$k'_n \rightarrow \infty, \quad \frac{k'_n}{k_n^{v/2}} + k_n^3 k_n \Delta_n \rightarrow 0, \quad (3.21)$$

we also have

$$\Delta_n S(\mathbf{j}, \mathbf{j}')_t^n \xrightarrow{\mathbb{P}} \mathbf{s}(\mathbf{j}, \mathbf{j}') \int_0^t \gamma_s^{q(\mathbf{j})+q(\mathbf{j}')} dA_s, \quad (3.22)$$

$$\frac{1}{N_t^n} S(\mathbf{j}, \mathbf{j}')_t^n \xrightarrow{\mathbb{P}} \frac{\mathbf{s}(\mathbf{j}, \mathbf{j}')}{A_t} \int_0^t \gamma_s^{q(\mathbf{j})+q(\mathbf{j}')} dA_s.$$

3.1.3. An Alternative Method

The previous estimators require a tuning parameter k_n . It is worth noticing that we also have much simpler estimators for the quantity $R(0)_t - R(j)_t$. Namely, for any integer $j > 0$, we set

$$\begin{aligned} U'(j)_t^n &= \sum_{i=0}^{N_t^n-j} (Y_{i+j}^n - Y_i^n)^2 \\ &= \sum_{i=0}^{N_t^n-j} (X_{i+j}^n - X_i^n)^2 + 2 \sum_{i=0}^{N_t^n-j} (X_{i+j}^n - X_i^n)(\varepsilon_{i+j}^n - \varepsilon_i^n) + \sum_{i=0}^{N_t^n-j} (\varepsilon_{i+j}^n - \varepsilon_i^n)^2. \end{aligned} \quad (3.23)$$

One can then prove, under Assumptions (H), (O), and (N), that

$$\widehat{\Delta R(j)}_t^n := \frac{1}{2N_t^n} U'(j)_t^n \xrightarrow{\mathbb{P}} R(0)_t - R(j)_t = (1 - r(j)) \frac{1}{A_t} \int_0^t \gamma_s^2 dA_s, \quad (3.24)$$

and a CLT, analogous to the forthcoming one for $\widehat{R}(j)_t^n$, is also available, with indeed a simpler proof. Under the white noise setting, some authors have used this convergence with $j = 1$ to estimate the variance of the noise; see, for example, [Zhang, Mykland, and Aït-Sahalia \(2005\)](#) and [Bandi and Russell \(2008\)](#). These estimators enjoy a CLT analogous to the one for $U(\mathbf{j})_t^n$ below, and we even have joint CLTs for the $U(\mathbf{j})_t^n$'s and the $U'(j)_t^n$. Alternatively, observe that the first summand in (3.23) converges in probability to jQV_t , where $QV_t = \int_0^t \sigma_s^2 ds + \sum_{s \leq 1} (\Delta X_s)^2$ is the quadratic variation of X over the time interval $[0, t]$, whereas the second summand converges in probability to 0. Therefore, if $\widehat{QV}_t$ is a reasonable estimator of QV_t (see Section 3.4 for some details about how to find $\widehat{QV}_t$), we can do a small sample correction and obtain that

$$\widehat{\Delta R(j)}_t^{n,\text{adj}} := \widehat{\Delta R(j)}_t^n - j \frac{\widehat{QV}_t}{2N_t^n} \xrightarrow{\mathbb{P}} R(0)_t - R(j)_t. \quad (3.25)$$

It is likely that $\widehat{\Delta R(j)}_t^n$ (or $\widehat{\Delta R(j)}_t^{n,\text{adj}}$) is better behaved, for small samples, than $\widehat{R}(0)_t^n - \widehat{R}(j)_t^n = \frac{1}{N_t^n} (U(0, 0)_t^n - U(0, j)_t^n)$, for estimating $R(0)_t - R(j)_t$, mainly because the tuning parameter k_n does not show up. However, there seems to be no simple way to estimate $R(0)_t$, apart from using $\widehat{R}(0)_t^n$, and there k_n comes in again (we could also use $\widehat{\Delta R(j_n)}_t^n$ or $\widehat{\Delta R(j)}_t^{n,\text{adj}}$ instead, for some $j_n \rightarrow \infty$, but this again involves the choice of j_n). This is why we do not pursue this method here, but we will provide a number of additional comments in Section 5.1 below.

3.2. Central Limit Theorems

Under exactly the same assumptions as in Theorem 3.1, plus $v > 1$ in Assumption (N), one can derive a CLT for $\Delta_n U(\mathbf{j})_t^n$ with a centering term of

$$\mathbf{r}(k_n; \mathbf{j}) \int_0^t \gamma_s^{q(\mathbf{j})} dA_s, \quad \text{where } \mathbf{r}(k_n; \mathbf{j}) = \mathbb{E} \left(\prod_{\ell=1}^{q(\mathbf{j})} (\chi_{j_\ell} - \bar{\chi}_{\mu(\mathbf{j}) + (2\ell-1)k_n}^n) \right). \quad (3.26)$$

For the CLT with centering term $\mathbf{r}(\mathbf{j}) \int_0^t \gamma_s^{q(\mathbf{j})} dA_s$, we need the convergence $\mathbf{r}(k_n; \mathbf{j}) \rightarrow \mathbf{r}(\mathbf{j})$ to be faster than the rate of convergence in the CLT, that is, $\sqrt{\Delta_n}$. This will be the case if k_n goes fast enough to ∞ , and more precisely if, instead of (3.5), we have the following, stronger, property with v the mixing exponent:

$$k_n \Delta_n^{1/(2v+1)} \rightarrow \infty, \quad k_n \Delta_n^\eta \rightarrow 0 \quad \text{for some } \eta \in \left(\frac{1}{2v+1}, \frac{1}{2} \right). \quad (3.27)$$

Similarly to Theorem 3.1, we have two formulations for the CLT: the first one is theoretically interesting, but only the second one can really be used by the statistician. They

concern the following processes, for $\mathbf{j} \in \mathcal{J}^+$:

$$\begin{aligned} Z(\mathbf{j})_t^n &= \frac{1}{\sqrt{\Delta_n}} \left(\Delta_n U(\mathbf{j})_t^n - \mathbf{r}(\mathbf{j}) \int_0^t \gamma_s^{q(\mathbf{j})} dA_s \right), \\ \bar{Z}(\mathbf{j})_t^n &= \sqrt{N_t^n} \left(\frac{1}{N_t^n} U(\mathbf{j})_t^n - \mathbf{R}(\mathbf{j})_t \right). \end{aligned} \quad (3.28)$$

Before stating the result, let us recall the meaning of $\mathcal{F}_\infty$ -stable convergence in law, abbreviated as $U_n \xrightarrow{\mathcal{L}_s - \mathcal{F}_\infty} U$, for a sequence of E -valued variables U_n on $(\Omega, \mathcal{F}, \mathbb{P})$, with E a metric space: U is also an E -valued variable on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$ (i.e., $\tilde{\Omega} = \Omega \times \Omega'$ and $\tilde{\mathcal{F}} = \mathcal{F} \otimes \mathcal{F}'$ for some extra measurable space $(\Omega', \mathcal{F}')$ and $\tilde{\mathbb{P}}$ is a probability on $\tilde{\mathcal{F}}$ with $\tilde{\mathbb{P}}(A \times \Omega') = \mathbb{P}(A)$ if $A \in \mathcal{F}$), and for all continuous bounded functions f on E and bounded $\mathcal{F}_\infty$ -measurable variables Ψ , we have $\mathbb{E}(\Psi f(U_n)) \rightarrow \tilde{\mathbb{E}}(\Psi f(U))$. This implies the convergence in law of U_n , and also the joint convergence in law of (U_n, U'_n) to (U, U') for any other sequence U'_n on $(\Omega, \mathcal{F}, \mathbb{P})$ going in probability to an $\mathcal{F}_\infty$ -measurable limit U' .

THEOREM 3.4: Assume (H), (O), (N) with $v > 1$, and let k_n satisfy (3.27) and $t > 0$.

(a) We have $(Z(\mathbf{j})_t^n)_{\mathbf{j} \in \mathcal{J}^+} \xrightarrow{\mathcal{L}_s - \mathcal{F}_\infty} (Z(\mathbf{j})_t)_{\mathbf{j} \in \mathcal{J}^+}$, where the limit is defined on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$ and, conditionally on $\mathcal{F}$, is centered Gaussian with (conditional) covariances

$$\tilde{\mathbb{E}}(Z(\mathbf{j})_t Z(\mathbf{j}')_t | \mathcal{F}) = \mathbf{s}(\mathbf{j}, \mathbf{j}') \int_0^t \gamma_s^{q(\mathbf{j})+q(\mathbf{j}')} dA_s + \mathbf{r}(\mathbf{j}) \mathbf{r}(\mathbf{j}') \int_0^t \gamma_s^{q(\mathbf{j})+q(\mathbf{j}')} \bar{\alpha}_s dA_s. \quad (3.29)$$

(b) We have $(\bar{Z}(\mathbf{j})_t^n)_{\mathbf{j} \in \mathcal{J}^+} \xrightarrow{\mathcal{L}_s - \mathcal{F}_\infty} (\bar{Z}(\mathbf{j})_t)_{\mathbf{j} \in \mathcal{J}^+}$, where the limit is defined on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$ and, conditionally on $\mathcal{F}$, is centered Gaussian with (conditional) covariance

$$\begin{aligned} \mathbb{E}(\bar{Z}(\mathbf{j})_t \bar{Z}(\mathbf{j}')_t | \mathcal{F}) &= \bar{Z}(\mathbf{j}, \mathbf{j}')_t \\ &:= \frac{\mathbf{s}(\mathbf{j}, \mathbf{j}')}{A_t} \int_0^t \gamma_s^{q(\mathbf{j})+q(\mathbf{j}')} dA_s + \frac{\mathbf{r}(\mathbf{j}) \mathbf{r}(\mathbf{j}')}{A_t} \int_0^t \gamma_s^{q(\mathbf{j})+q(\mathbf{j}')} \bar{\alpha}_s dA_s \\ &\quad - \frac{\mathbf{r}(\mathbf{j}')}{A_t} \mathbf{R}(\mathbf{j})_t \int_0^t \gamma_s^{q(\mathbf{j}')} \bar{\alpha}_s dA_s - \frac{\mathbf{r}(\mathbf{j})}{A_t} \mathbf{R}(\mathbf{j}')_t \int_0^t \gamma_s^{q(\mathbf{j})} \bar{\alpha}_s dA_s \\ &\quad + \frac{1}{A_t} \mathbf{R}(\mathbf{j})_t \mathbf{R}(\mathbf{j}')_t \int_0^t \bar{\alpha}_s dA_s. \end{aligned} \quad (3.30)$$

In order to make this CLT feasible, one needs consistent estimators $\hat{Z}(\mathbf{j}, \mathbf{j}')_t^n$ for the conditional variances $\bar{Z}(\mathbf{j}, \mathbf{j}')_t$ of (3.30). In view of Theorems 3.1 and 3.3, this is easily achieved by setting

$$\begin{aligned} \hat{Z}(\mathbf{j}, \mathbf{j}')_t^n &= \frac{1}{N_t^n} \left(S(\mathbf{j}, \mathbf{j}')_t^n + U^3(\mathbf{j}, \mathbf{j}')_t^n - \frac{1}{N_t^n} U(\mathbf{j})_t^n U^2(\mathbf{j}')_t^n - \frac{1}{N_t^n} U(\mathbf{j}')_t^n U^2(\mathbf{j})_t^n \right. \\ &\quad \left. + \frac{1}{(N_t^n)^2} U(\mathbf{j})_t^n U(\mathbf{j}')_t^n U_t^{1,n} \right). \end{aligned} \quad (3.31)$$

This complicated expression greatly simplifies under stronger assumptions on the sampling scheme: if it is “close enough” to a regular scheme, we have $\bar{\alpha}_t \equiv 0$ in (2.3) (Example 2.2 is an instance of this, but more general schemes are, of course, of this type); if it is “close enough” to a modulated Poisson scheme, see Example 2.3, then we have $\bar{\alpha}_t \equiv 1$. In these two situations, we may use the following consistent estimators for $\bar{\mathcal{Z}}(\mathbf{j}, \mathbf{j})_t$:

$$\widehat{\mathcal{Z}}'(\mathbf{j}, \mathbf{j})_t^n = \begin{cases} \frac{1}{N_t^n} S(\mathbf{j}, \mathbf{j})_t^n & \text{if } \bar{\alpha}_t = 0, \\ \frac{1}{N_t^n} S(\mathbf{j}, \mathbf{j})_t^n + \frac{1}{N_t^n} \bar{U}(\mathbf{j}, \mathbf{j})_t^n - \frac{1}{(N_t^n)^2} U(\mathbf{j})_t^n U(\mathbf{j}')_t^n & \text{if } \bar{\alpha}_t \equiv 1. \end{cases} \quad (3.32)$$

The advantage of the above is that it does not necessitate the first three processes of (3.15), hence of the variables $\widehat{\Delta}_t^n$ which probably have a somewhat unstable behavior because of their denominators and necessitate the supplementary tuning parameters ϕ_n . But of course, (3.32) can only be used if we are (reasonably) sure about the additional assumptions $\bar{\alpha} \equiv 0$ or $\bar{\alpha} \equiv 1$.

At this stage, stating for simplicity the one-dimensional case only, by standard properties of the stable convergence in law and since $\bar{\mathcal{Z}}(\mathbf{j}, \mathbf{j})_t$ is $\mathcal{F}_\infty$ -measurable, we deduce from Theorems 3.1, 3.3, and 3.4 the following standardized version of the CLT.

COROLLARY 3.5: *Assume (H), (O), and (N), and let k_n , ϕ_n , and k'_n satisfy (3.27), (3.14), and (3.21). For any $t > 0$ and $\mathbf{j} \in \mathcal{J}$, we have the following $\mathcal{F}_\infty$ -stable convergence in law, with Φ an $\mathcal{N}(0, 1)$ -distributed variable defined on an extension of the space and independent of $\mathcal{F}$:*

$$\frac{\sqrt{N_t^n}}{\sqrt{\widehat{\mathcal{Z}}'(\mathbf{j}, \mathbf{j})_t^n}} \left(\frac{1}{N_t^n} U(\mathbf{j})_t^n - \mathbf{R}(\mathbf{j})_t \right) \xrightarrow{\mathcal{L}_s - \mathcal{F}_\infty} \Phi. \quad (3.33)$$

When, further, $\bar{\alpha}_t \equiv 0$ or $\bar{\alpha}_t \equiv 1$, so ϕ_n does not show, the same holds for the variables $\sqrt{N_t^n / \widehat{\mathcal{Z}}'(\mathbf{j}, \mathbf{j})_t^n} (\frac{1}{N_t^n} U(\mathbf{j})_t^n - \mathbf{R}(\mathbf{j})_t)$.

REMARK 3.6: At first sight, the choice of k_n in (3.27) requires the knowledge of v , or at least of the fact that v is bigger than some known value $v' > 1$: in this case, one may take $k_n \asymp 1/\Delta_n^{1/2v'}$ (and hence a smaller k_n for shorter memory dependence). This is unfortunate, since in general one does not a priori know the law of χ , and in particular whether it is stationary, or mixing, not to mention the number v for which it is v -polynomially ρ -mixing.

Fortunately, based on the convergences (3.24) and (3.25), which do not depend on any tuning parameters, we can choose k_n according to a heuristic approach that we will describe in detail in Section 5.1.2 without worrying about the above.

REMARK 3.7: The question of estimating v is important for the choice of the tuning parameter k_n , but deciding whether $v > 1$ or not is crucial for the validity of the CLT. Indeed, whereas the consistency results hold (except for (3.22)) as soon as $v > 0$, the assumption $v > 1$ is necessary for the CLT; otherwise, the numbers $\mathbf{s}(\mathbf{j}, \mathbf{j}')$ of (3.13) are in general not even well defined and (3.29), for example, makes no sense.

In the empirical section, we will show how to (consistently) estimate v , and the empirical results are somewhat mixed: it looks like, in practice, v ranges between 0.5 and 1.7. If, for

a particular set of observations, we have $v \leq 1$, then estimating $r(j)$, for instance, is still possible, but we can no longer build a confidence interval.

We must admit that we have no clear-cut solution to that problem, although methods based on bootstrap might help. While we can derive some convergence rate results for the case when $v \leq 1$, a CLT seems to be out of reach. One major complication is due to the slow decay of the (necessary) local-averaging term in our estimators. The γ process makes the problem even more complicated.

3.3. Estimation of the Autocorrelations of the Noise Component χ_i

Although the sequence (χ_i) is not the genuine noise, it is useful to estimate its autocorrelation function $r(j)$; see discussions below (3.3). This may be used to check whether $r(j) = 0$ for all $j \geq 1$ (so we basically have a white noise, or modulated white noise), and otherwise to assert its possible decay as j increases.

As discussed at the end of Section 3.1.1, $\widehat{r}(j)_t^n := \widehat{R}(j)_t^n / \widehat{R}(j)_t^n \xrightarrow{\mathbb{P}} r(j)$. The associated CLT is a straightforward consequence of Theorem 3.4, used with $\mathbf{j} = (0, j)$ and $\mathbf{j}' = (0, 0)$. Namely, since then $\mathbf{R}(\mathbf{j})_t = R(j)_t$ and $\mathbf{R}(\mathbf{j}')_t = R(0)_t$, under the assumptions of this theorem, the random variable $\sqrt{N_t^n}(\widehat{r}(j)_t^n - r(j))$ converges $\mathcal{F}_\infty$ -stably in law to a variable which is $\mathcal{F}_\infty$ -conditionally centered Gaussian with (conditional) variance

$$S(j)_t = \frac{R(0)_t^2 \overline{\mathcal{Z}}(\mathbf{j}, \mathbf{j})_t + R(j)_t^2 \overline{\mathcal{Z}}(\mathbf{j}', \mathbf{j}')_t - 2R(j)_t R(0)_t \overline{\mathcal{Z}}(\mathbf{j}, \mathbf{j}')_t}{R(0)_t^4}.$$

With the notation (3.31), consistent estimators for $S(j)_t$ are

$$\widehat{S}(j)_t^n = \frac{(U(\mathbf{j}')_t^n / N_t^n)^2 \cdot \widehat{\mathcal{Z}}(\mathbf{j}, \mathbf{j})_t^n + (U(\mathbf{j})_t^n / N_t^n)^2 \widehat{\mathcal{Z}}(\mathbf{j}', \mathbf{j}')_t^n - 2U(\mathbf{j})_t^n U(\mathbf{j}')_t^n / (N_t^n)^2 \cdot \widehat{\mathcal{Z}}(\mathbf{j}, \mathbf{j}')_t^n}{(U(\mathbf{j}')_t^n / N_t^n)^4}.$$

At this stage, the following result is obvious:

THEOREM 3.8: Assume (H), (O), (N), and let k_n and k'_n satisfy (3.27) and (3.21). For any $t > 0$ and $j \geq 1$, we have the following $\mathcal{F}_\infty$ -stable convergence in law, where Φ is as in Corollary 3.5:

$$\sqrt{\frac{N_t^n}{\widehat{S}(j)_t^n}} (\widehat{r}(j)_t^n - r(j)) \xrightarrow{\mathcal{L}} \Phi. \quad (3.34)$$

When, further, $\overline{\alpha}_t \equiv 0$ or $\overline{\alpha}_t \equiv 1$, the same holds for the variables $\sqrt{N_t^n / \widehat{S}'(\mathbf{j}, \mathbf{j})_t^n} (\widehat{r}(j)_t^n - r(j))$, where $\widehat{S}'(\mathbf{j})_t^n$ is given by the same formula as $\widehat{S}(j)_t^n$, with $\widehat{\mathcal{Z}}(\mathbf{j}, \mathbf{j})_t^n$ substituted with $\widehat{\mathcal{Z}}'(\mathbf{j}, \mathbf{j})_t^n$.

3.4. Removing the Finite Sample Bias

Since we do not observe the noise process, rather, we observe the noise process “contaminated” by the latent price, it is inevitable that our estimators are subject to some degree of (finite sample) bias. In particular, it is enlightening to see what happens to the estimators $\frac{1}{N_t^n} U(\mathbf{j})_t^n$ when there is no noise at all: they converge to 0, of course, but what is the actual rate?

To check this, it is convenient to restrict our attention to simple cases: below, we only consider autocovariance estimation, that is, $R(j)_t$ for some integer $j \geq 0$. Recalling the quadratic variation process $QV_t = \int_0^t \sigma_s^2 ds + \sum_{s \leq t} (\Delta X_s)^2$ of X , it is simple enough to show that, under (H) and (O) and when the noise is absent (i.e., $\gamma_t \equiv 0$ for all t),

$$\frac{1}{k_n} U((0, j))_t^n \xrightarrow{\mathbb{P}} \frac{3}{2} QV_t. \quad (3.35)$$

This leads us to propose, in general (when noise is present), the following bias correction. We consider a reasonable estimator $\widehat{QV}_t$ for the quadratic variation QV_t , as in (3.25), and then we take the estimator

$$\widehat{R}(j)_t^{n,\text{adj}} = \frac{1}{N_t^n} U((0, j))_t^n - \frac{3k_n}{2N_t^n} \widehat{QV}_t, \quad (3.36)$$

instead of the original $\widehat{R}(j)_t^n = \frac{1}{N_t^n} U((0, j))_t^n$ for estimating $R(j)_t$, and correspondingly,

$$\widehat{r}(j)_t^{n,\text{adj}} = \frac{\widehat{R}(j)_t^{n,\text{adj}}}{\widehat{R}(0)_t^{n,\text{adj}}} \quad (3.37)$$

for estimating $r(j)$. The ratio k_n/N_t^n is negligible in front of $1/\sqrt{\Delta_n}$, so the bias correction has no asymptotic effect, but it may be significant for small samples or when $R(j)_t/r(j)$ is small, for example, when j is large.

Let us end this section with some comments about the construction of an estimator $\widehat{QV}_t$ for QV_t . Recall that we want a “reasonable” approximation of QV_t , but not necessarily a consistent sequence $\widehat{QV}_t^n$ of estimators, since in any case the correction term in (3.36) is asymptotically negligible. There are two possibilities:

1. We can drastically subsample, by looking, for example, at observations at every 5 minutes, and then take for $\widehat{QV}_t$ the “realized variance” (i.e., the sum of all squared returns) at this frequency: typically, the noise level is much smaller than the average return at this frequency, so $\widehat{QV}_t$ is a reasonable approximation of QV_t . Of course, this does not give us a sequence of consistent estimators because, if we increase the subsampling frequency, the noise becomes prevalent, but it seems that subsampling at 5 minutes provides a good enough approximation for QV_t ; see [Liu, Patton, and Sheppard \(2015\)](#).

2. We can use estimators based on complete tick-by-tick data. When the efficient price process X is continuous, in [Jacod, Li, and Zheng \(2015\)](#) we showed that the dependence in the noise induces a bias in the usual volatility estimators like the pre-averaging method, and a bias correction is necessary if one wants to apply the method to complete tick-by-tick data. Making use of the estimators of autocovariances of noise that we propose in this paper, we are able to correct for the bias and develop a volatility estimator that works in the presence of (i) dependent noise which can have diurnal features and be dependent on the latent price process, and (ii) irregular observation times which can be endogenous. In the same paper, when the process jumps, we also gave consistent estimators for the integrated volatility by using the *truncated* pre-averaged values (and a CLT holds in that case as well, under appropriate conditions on the jumps). If we use the same estimators *without truncation*, we get consistent estimators for QV even when X jumps (and a CLT would still be available in this case, although we did not prove it in that paper).

4. SIMULATIONS

Throughout this section, we focus on the time interval $[0, 1]$, which represents one trading day.

4.1. When Assumption (O) Is Satisfied

We consider the following design: X is an Ornstein–Uhlenbeck process with jumps

$$dX_t = -\rho(X_t - \mu) dt + \sigma dW_t + dJ_t, \quad t \in [0, 1], \quad (4.1)$$

where W is a standard Brownian motion, and J is a compound Poisson process independent of W as follows:

$$J_t = \sum_{i=0}^{N_t} D_i,$$

where (N_t) is a Poisson process with rate λ , and D_i 's are i.i.d. symmetric mixed normals: $D_i = B_i \cdot Z'_i$, where B_i takes values 1 and -1 with equal probability 0.5, and $Z'_i \sim N(\mu', \sigma'^2)$.

The observation times $T(n, i)$ follow an *inhomogeneous* Poisson process with rate $n\alpha_t$ with $\alpha_t = (1 + \cos(2\pi t)/2)$, which is a special case of Example 2.3 with $\Delta_n = 1/n$ and α_t as above, and we thus have $\bar{\alpha}_t = 1$. Observe that the rate function is U-shaped, and is used to mimic the empirical feature that there are more transactions in the early morning and late afternoon than in the middle of the day.

As to the noise process, we find in the empirical studies in Section 5 that the autocorrelations of the noise decay slowly. For this reason, we choose χ to be a moving-average (MA) series which approximates a fractionally differenced process. More specifically, for an exponent $d \in (-0.5, 0.5)$ and for a large cutoff value M , χ admits the following representation:

$$\begin{aligned} \chi_i &= Z_i'' + \sum_{j=1}^M \psi_j Z_{i-j}'', \quad \text{where} \\ \psi_j &= \frac{d(1+d) \cdots (j-1+d)}{j!}, \quad \text{and} \quad Z_i'' \sim_{\text{i.i.d.}} N(0, \sigma_0^2). \end{aligned} \quad (4.2)$$

If one lets $M = \infty$, then one obtains a fractionally differenced process whose autocorrelation function decays polynomially at the rate of $2d - 1$; see Tsay (2002, pp. 72–73). For finite but large M , the autocorrelation function first decays slowly, and after lag M , the autocorrelation vanishes. We have an additional process γ , which we assume to be an Ornstein–Uhlenbeck-type process featuring a U-shaped pattern as follows:

$$d\gamma_t = -\rho_\gamma(\gamma_t - \mu_t) dt + \sigma_\gamma dW_t, \quad (4.3)$$

where μ_t is a deterministic process which is U-shaped, and W is the *same* Brownian motion that is used in (4.1). The motivation for specifying γ to be U-shaped comes from the empirical studies where a (vague) U-shaped pattern is observed when examining the size of noise within each day; see Section 5.2.1 below. The noise is $\varepsilon_i^n = \gamma_{T(n,i)} \chi_i$. Note that, because of the dependence between X and γ , the noise ε_i^n is *dependent* on the X process. The observations are $Y_i^n = X_{T(n,i)} + \varepsilon_i^n = X_{T(n,i)} + \gamma_{T(n,i)} \chi_i$.

The parameters in (4.1), (4.2), and (4.3) are taken as follows:

$$\begin{aligned} \rho &= 0.5, & \mu &= 1.6, & \sigma &= 0.01, & \lambda &= 3, & \mu' &= \sigma/10, & \sigma' &= \sigma/30, \\ d &= 0.3, & M &= 100, & \sigma_0 &= 0.0005, \\ \rho_\gamma &= 10, & \mu_t &= 1 + 0.1 \cos(2\pi t), & \text{and} & \sigma_\gamma &= 0.1. \end{aligned} \quad (4.4)$$

Under these specifications, the price ($\exp(X_t)$) is roughly around $\exp(\mu) = \exp(1.6) \approx \5 ; the volatility is about 0.01, which is typical in practice; and the (average) standard deviation of the noise (ε_i^n) is around $1 \times \sqrt{1 + \sum_{j=1}^M \psi_j^2} \times \sigma_0 \approx 5.6 \times 10^{-4}$, which is similar to what we find in the empirical studies below. We further take $n = 23,400$.

In the estimation, the tuning parameter k_n is chosen to be 6 based on the heuristic criterion in Section 5.1.2, which we illustrate with real data. To save space, in the following we only report the results about autocorrelations; those about autocovariances are given in the Supplemental Material (Jacod, Li, and Zheng (2017, Appendix B)). Figure 1 compares the estimates of autocorrelations based on Theorem 3.8 with the infeasible estimates based on the noise process and the theoretical values. The estimates are for one simulated path. More specifically, we compare the estimates of autocorrelations based on Theorem 3.8 ($\hat{r}(j)_1^n$ as in (3.12), see red dashed curve) with the infeasible estimates based on the noise process (autocorrelations estimated from (ε_i^n) , see blue dotted curve), and the theoretical values $r(j)$ given by the black solid curve. The theoretical values $r(j)$ are, with $\psi_0 = 1$ and $\psi_j, j = 1, 2, \dots$ as in (4.2),

$$r(j) = \frac{\sum_{i=0}^{M-j} \psi_i \psi_{i+j}}{\sum_{i=0}^M \psi_i^2}, \quad \text{for } j = 0, 1, \dots, M. \quad (4.5)$$

Figure 1 demonstrates that our estimates are comparable to the infeasible estimates based on the noise process, and both are close to the theoretical values.

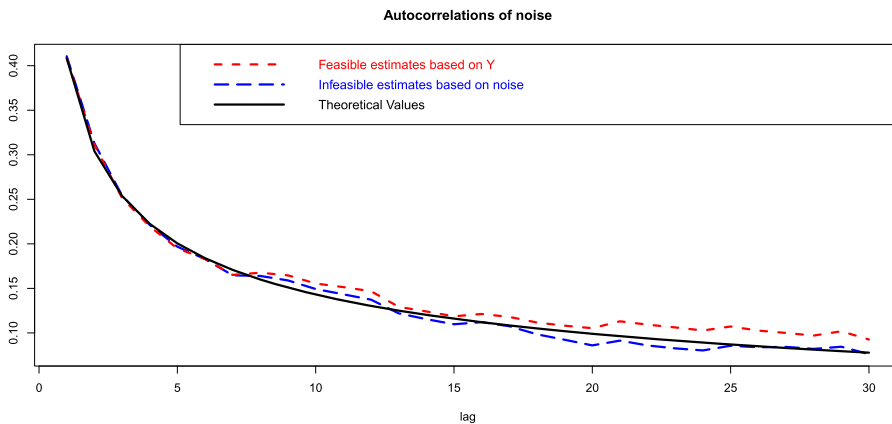


FIGURE 1.—Estimates of autocorrelations. The feasible estimates based on the observed prices are compared with the infeasible estimates based on unobservable noise process, and with the theoretical values.

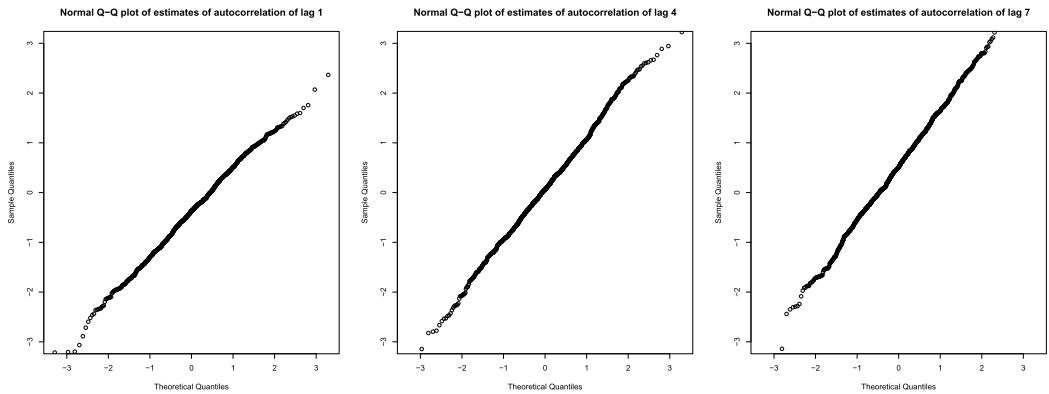


FIGURE 2.—Normal QQ-plots of (4.6) for lags 1, 4, and 7, based on 1000 replications.

Next, we examine the CLT. We plot the normal quantile-quantile plots of

$$\sqrt{\frac{N_1^n}{\hat{S}(j)_1^n}}(\hat{r}(j)_1^n - r(j)) \quad (4.6)$$

as in the last part of Theorem 3.8 for $j = 1, 4$, and 7 in Figure 2, based on 1000 replications. The plots support the normality established in the theorem.

4.2. When There Is Rounding

Rounding is an inevitable element in financial prices. While rounding is not incorporated in the current framework and the related theoretical properties are beyond the scope of this paper, the simulation studies do show that our method still performs well in the presence of rounding. Below, we report a brief summary of our estimates in a setting that involves rounding and either colored or white noise. Some supplementary discussions are given in the Supplemental Material (Appendix B).

Specifically, we consider the following specification for the observed price process:

$$\tilde{S}_i^n = [\exp(X_{T(n,i)} + \varepsilon_i^n)/0.01] \times 0.01. \quad (4.7)$$

In other words, the observations are the contaminated prices further rounded to cents.² The latent process X is specified by equation (4.1) as before. As to the noise ε_i^n , we consider two cases: (i) the colored case when $\varepsilon_i^n = \gamma_{T(n,i)}\chi_i$ with γ and χ specified by equations (4.2)–(4.4); and (ii) the white case when χ_i are i.i.d. $N(0, \sigma_0^2)$.

We then apply our estimators to the observed prices. The estimation results for the autocorrelations are given in Figure 3.

Figure 3 shows that, for both cases, our estimates are close to the infeasible estimates based on the noise, suggesting that our method works well even in the presence of rounding. Furthermore, the lower panel in Figure 3 for the white noise case suggests that rounding alone hardly induces dependence.

²This specification had been considered in Li and Mykland (2007) and Jacod et al. (2009) with ε_i^n being i.i.d. Gaussian or following a particular specification, respectively.

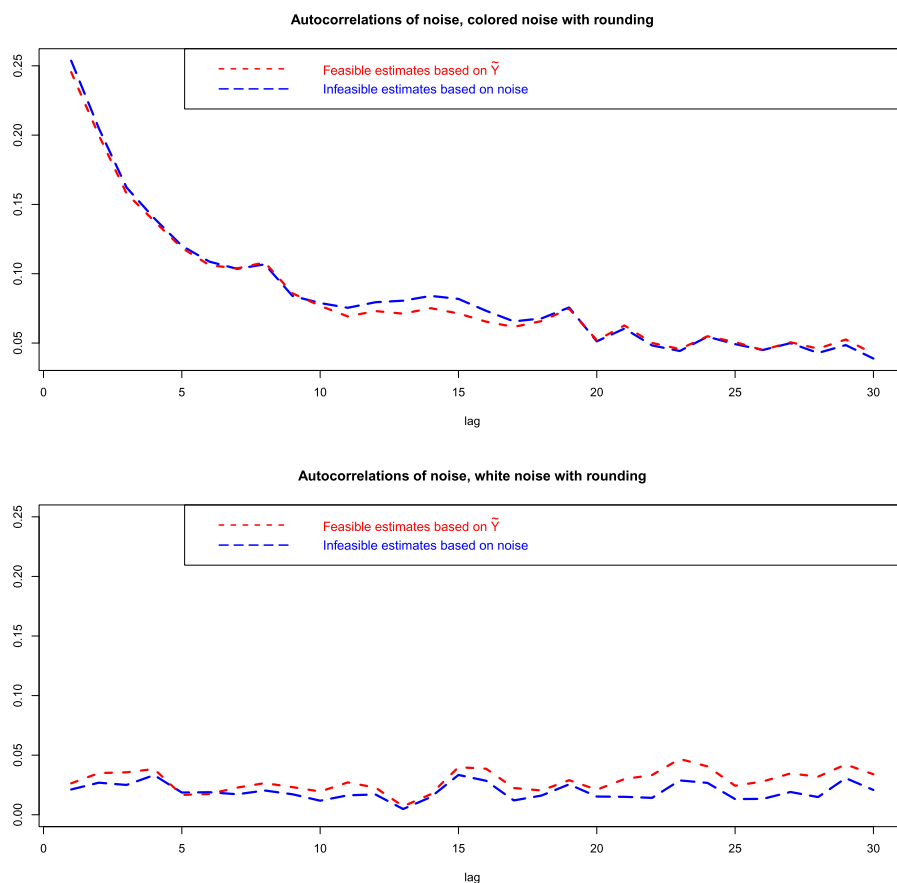


FIGURE 3.—Estimates of autocorrelations with rounded observations. Upper panel is for the case when the noise is colored, and lower panel for the white noise case. The feasible estimates based on the observed prices are compared with the infeasible estimates based on the noise.

5. EMPIRICAL STUDIES

We now apply our estimators to analyze the statistical properties of microstructure noise for the stock Citigroup Inc. (NYSE: C). The analysis is based on the tick-by-tick trade data (during regular trading hours) of Citigroup Inc. in January 2011. The average observation frequency for Citigroup Inc. during this period is about 246,000 per trading day.³ In the Supplemental Material (Appendix B), we report parallel analysis results for

³We include all transactions in our analysis; in particular, those transactions with the same price as the previous ones are also included. This is different than, for example, Aït-Sahalia, Mykland, and Zhang (2011) where only transactions that lead to a price change were included in the analysis (see Table 1 therein). The reasons are as follows. First, our motivation is to understand the statistical properties of noise at the tick-by-tick level, hence it is natural to include all transactions in our analysis. Second, we believe that even transactions that do not lead to a price change contain information about the noise. This is best seen in the simulation studies when there is rounding (Section 4.2). There, both when the noise is dependent or white, the proportion of flat trading, that is, the proportion of zero intraday returns, is higher than 70%. With such high proportions of flat trading, our method can still identify the true size as well as the dependence structure of the noise. For

the same stock on May 6, 2010, the day on which a flash crash occurred, and for another stock, Intel Corporation (NASDAQ:INTC), during January 2011.

5.1. Choosing the Tuning Parameter k_n

Our estimators of both autocovariances and autocorrelations involve the tuning parameter k_n . In this subsection, we describe a heuristic approach for choosing k_n . We illustrate the procedure for January 3, 2011, the first trading day in 2011. We start with the “robust” estimates of the differences of autocovariances.

5.1.1. Examining the Differences of Autocovariances

We apply the alternative method in Section 3.1.3 to estimate the differences of autocovariances of noise. More specifically, we compute $\widehat{\Delta R}(j)_t^n := \frac{1}{2N_t^n} U'(j)_t^n$ for $t = 1$ and $j = 0, 1, \dots$, which, according to (3.24), converge to $R(0)_1 - R(j)_1 = (1 - r(j)) \cdot \int_0^1 \gamma_s^2 dA_s / A_1$ for any fixed j . To account for the finite sample bias, we also compute $\widehat{\Delta R}(j)_1^{\text{adj}} = \widehat{\Delta R}(j)_1 - j \cdot \widehat{QV}_1 / (2N_1^n)$ with $\widehat{QV}_1$ being the recently proposed volatility estimator in Jacod, Li, and Zheng (2015) based on all tick-by-tick data and without truncations, so it is consistent for QV_1 even when the process X jumps (see Section 3.4 for comments about this). The results are given in Figure 4.

Several comments about Figure 4 are in order.

(i) Finite sample bias correction is indeed necessary. When the observation frequency is high, $\widehat{\Delta R}(j)_1$ is approximately $(R(0)_1 - R(j)_1) + j \times QV_1 / (2N_1^n)$, hence it increases linearly in j for large j 's, and this explains the linear increasing pattern of the black curve in the first plot.⁴

(ii) On the other hand, the corrected $\widehat{\Delta R}(j)_1^{n,\text{adj}}$ does stabilize as the lag j increases. In light of the convergence (3.25), this suggests that indeed the autocovariance $R(j)$ (and hence the autocorrelation $r(j)$) approaches 0 as the lag j increases to ∞ . Such a reasoning also suggests that the level that the curve approaches should be $R(0)_1 = \int_0^1 \gamma_s^2 dA_s / A_1$.

(iii) Furthermore, there is a clear increasing trend in $\widehat{\Delta R}(j)_1^{n,\text{adj}}$ for small to moderate j 's (up to 100 or so, roughly within 10 seconds window). Combining with (ii), we see that

$$(1 - r(j)) \cdot \frac{\int_0^1 \gamma_s^2 dA_s}{A_1} \quad \text{increases to} \quad \frac{\int_0^1 \gamma_s^2 dA_s}{A_1}, \quad \text{as } j \rightarrow \infty,$$

and therefore we reach the following conclusion:

$$r(j)\text{'s are positive and decay to zero as } j \rightarrow \infty. \quad (5.1)$$

Citi in Jan 2011, the proportion of flat trading is also about 70%. The results in Section 4.2 suggest that our method is applicable to complete tick-by-tick data.

⁴An interesting observation is that such a curve suggests that we can estimate QV_1 by simply estimating the slope of the curve (maybe restricted to large j 's). In this way, we are combining all different sampling frequencies. This is certainly related to the two-scales estimator (Aït-Sahalia, Mykland, and Zhang (2005)) and the multi-scales estimator (Zhang (2006)), and also Li, Zhang, and Li (2015).

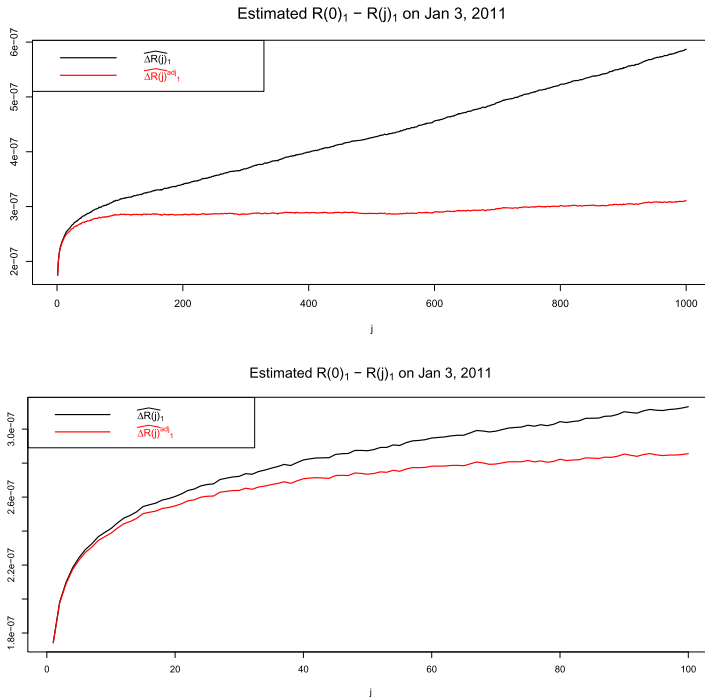


FIGURE 4.—Estimates of $R(0)_1 - R(j)_1$ for Citigroup (C) stock on January 3, 2011. Upper panel: for $j = 1, 2, \dots$ up to 1000; lower panel: for $j = 1, 2, \dots$ up to 100.

It is worth emphasizing that there is *no* tuning parameter involved in drawing Figure 4 (except in estimating QV_1 , for which there are various well-established estimators; in addition, in Jacod, Li, and Zheng (2015) we developed a new volatility estimator that accommodates both dependent noise and irregular observation times), hence the conclusions above are robust.

REMARK 5.1: Note that the conclusion (5.1) is about the autocorrelations of noise. In earlier studies, for example, in Aït-Sahalia, Mykland, and Zhang (2011), it was found that observed *returns* may exhibit *negative* autocorrelations.⁵ This is *not* in contradiction with our finding (5.1), and actually, (5.1) provides a more informative explanation to the negative autocorrelations of observed returns. To see this, suppose for simplicity that the noise is simply $\varepsilon_i^n = \chi_i$. Since the noise dominates the return at high frequency, and recalling that $\Delta_i^n Y = Y_i^n - Y_{i-1}^n$ and $\Delta_i^n \varepsilon = \varepsilon_i^n - \varepsilon_{i-1}^n$, we have for any $j \geq 1$:

$$\text{Cov}(\Delta_{i+j}^n Y, \Delta_i^n Y) \approx \text{Cov}(\Delta_{i+j}^n \varepsilon, \Delta_i^n \varepsilon) = 2r(j) - r(j+1) - r(j-1).$$

A closer look at Figure 4 reveals that the differences $r(j) - r(j+1)$ are decreasing in j , hence $2r(j) - r(j+1) - r(j-1)$ would be negative, and so would be $\text{Cov}(\Delta_{i+j}^n Y, \Delta_i^n Y)$.⁶

⁵The results in Aït-Sahalia, Mykland, and Zhang (2011) are based on those transactions that lead to a price change; in other words, zero intraday returns are removed from the analysis.

⁶For the stocks Intel (INTC) and Microsoft (MSFT), the autocorrelations of their observed returns during April 2004 exhibit alternating signs; see the middle panel of Figure 4 in Aït-Sahalia, Mykland, and Zhang (2011). The results are based on only those transactions that lead to a price change, and zero intraday returns

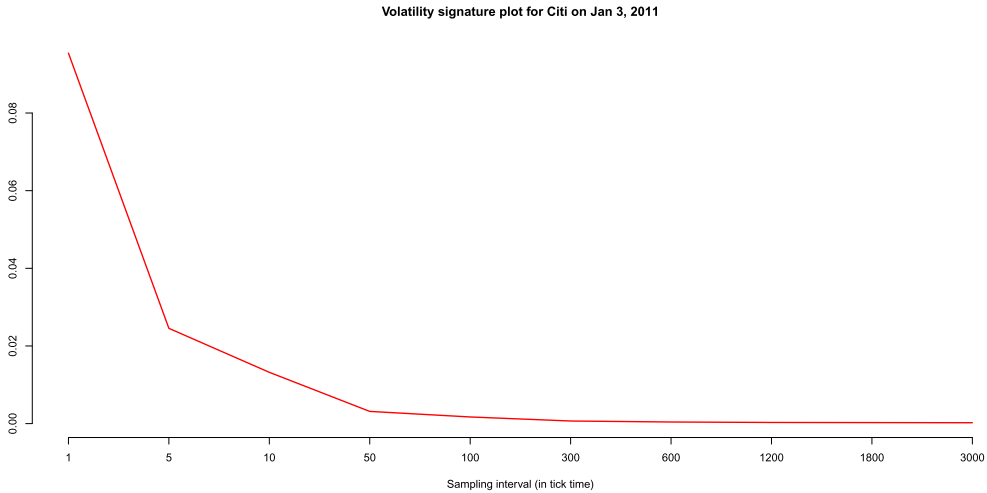


FIGURE 5.—Volatility signature plot for Citigroup (C) stock on January 3, 2011.

REMARK 5.2: Figure 4 is closely related to, but different than, the usual volatility signature plot (Andersen, Bollerslev, Diebold, and Labys (2000b)). In the usual volatility signature plot, one plots the realized volatility based on subsampled data against the sampling interval. In other words, if the sampling is conducted according to the tick time, then in the usual volatility signature plot, instead of plotting $U'(j)_1^n/(2N_1^n)$ against j as in Figure 4, one plots $U'(j)_1^n/j$ against j (we divide $U'(j)_1^n$ by j because it is a sum of j copies of realized volatility based on data sampled every j observations). Figure 5 is the volatility signature plot for Citigroup (C) stock on January 3, 2011. Not surprisingly, the realized volatility blows up as one samples more frequently. The fact that the trends in Figures 4 and 5 are opposite to each other is due to the difference in the two quantities that we are plotting; in particular, in Figure 4 we normalize $U'(j)_1^n$ by N_1^n to account for the accumulation of noise, which estimates $(1 - r(j)) \cdot \int_0^1 \gamma_s^2 dA_s/A_1$ and possesses an opposite trend.

5.1.2. Choosing k_n

We now explain our heuristic approach for choosing k_n . The approach relies on the “robust” convergence (3.25), which estimates the difference of autocovariances. On the other hand, by (3.11) we can consistently estimate the autocovariances. Therefore, combining the two convergences (3.11) and (3.25), we have

$$\widehat{R}(0)_t^n - \widehat{R}(j)_t^n - \widehat{\Delta R}(j)^{n, \text{adj}} \xrightarrow{\mathbb{P}} 0.$$

are not included in the analysis. Such a strange feature led the authors to propose a dependent model for the noise; see equation (33) therein. The authors then used the model to fit the data and obtained a *negative* autocorrelation among the noise. We examined the same stocks, Intel (INTC) and Microsoft (MSFT), during January 2011. When we include all the transactions in our analysis, we find that the autocorrelations of observed returns do not exhibit alternating signs; instead, are all negative and decay to zero rather quickly. When further applying our method to study the noise for these two stocks, we observe positive autocorrelations just like for Citigroup (C).

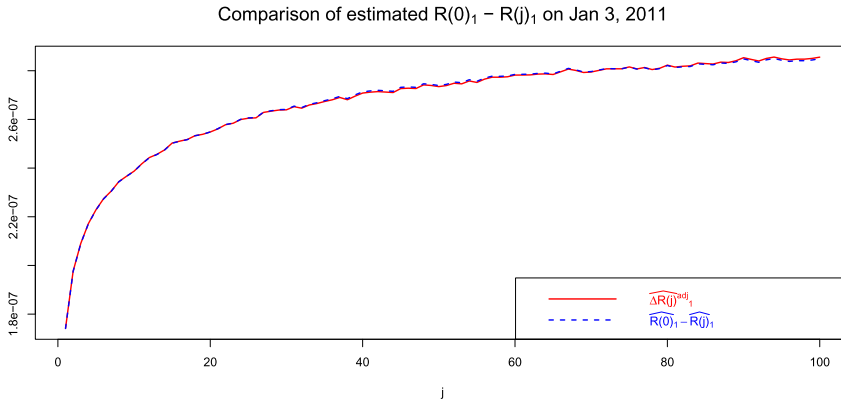


FIGURE 6.—Comparison of the two estimates of $R(0)_1 - R(j)_1$, one based on $\widehat{\Delta R(j)}_1^{n,adj}$, the other based on $(\widehat{R}(0)_1^n - \widehat{R}(j)_1^n)$, for Citigroup (C) stock, on January 3, 2011.

Consequently, a suitable value of the tuning parameter k_n should satisfy the following:

HEURISTIC CRITERION: *The two sequences $(\widehat{R}(0)_1^n - \widehat{R}(j)_1^n)_{j=1,2,\dots}$ and $(\widehat{\Delta R(j)}_1^{n,adj})_{j=1,2,\dots}$ are close to each other.*

Checking this criterion gives a simple way of choosing k_n .

Yet another application in checking this criterion is to test for the uncorrelatedness between the noise and efficient returns. Recall that, as we discussed in Remark 2.6, our model does imply that the noise and efficient returns are uncorrelated. To check for such uncorrelatedness, observe that in the convergence (3.25), the correlation between the noise and efficient returns is not accounted for. Therefore, if the *heuristic criterion* can be satisfied, then it is reasonable to conclude that the correlation between the noise and efficient returns is not significant.^{7,8}

We now apply the heuristic criterion to choose k_n . Figure 6 gives the comparison between $\widehat{\Delta R(j)}_1^{n,adj}$ and $(\widehat{R}(0)_1^n - \widehat{R}(j)_1^n)$, the latter computed by using $k_n = 40$. Observe how close the two curves are to each other. It is worth emphasizing again that the shape of $\widehat{\Delta R(j)}_1^{n,adj}$ does not depend on the tuning parameter k_n , and therefore such a comparison provides a strong support of the robustness of our estimates of $R(0)_1 - R(j)_1$ and, consequently, of $R(j)_1$.

Finally, as we discussed above, the fact that the heuristic criterion can be satisfied so well suggests that there is no clear violation of the uncorrelatedness between the noise and efficient returns implied by Assumption (N).

⁷Observe that if the correlation $\text{Corr}(X_{i+1}^n - X_i^n, \varepsilon_{i+1}^n - \varepsilon_i^n) \equiv \rho \neq 0$, then the cross term $\sum_{i=0}^{N^n-j} (X_{i+j}^n - X_i^n)(\varepsilon_{i+j}^n - \varepsilon_i^n)$ will be $O_p(j/\sqrt{\Delta_n})$, hence unlikely to be of smaller order than $\sum_{i=0}^{N^n-j} (X_{i+j}^n - X_i^n)^2$ which is $O_p(j)$.

⁸In Hansen and Lunde (2006), the authors applied a similar convergence to (3.24) to different sampling frequencies, drew the volatility signature plot, and checked for the correlation between the noise and efficient returns (see Section 2.2 therein for detailed discussions). Alternatively, one can first estimate the efficient prices and microstructure noise, using, for example, the method in Hansen and Lunde (2006) or that in Li, Xie, and Zheng (2016), compute the covariance/correlation between the efficient returns and the noise, and then conduct a test.

5.2. Diurnal Features in the Noise

Motivated by related empirical findings like volatility clustering, we examine whether there are diurnal features in the noise. To do so, we divide the whole trading day into a total of thirteen half-hour time intervals, namely, 9:30–10:00, 10:00–10:30, and so on, and within each period, estimate the autocovariance and autocorrelation functions of the noise. The number of observations within each half-hour interval ranges from about 6000 to 92,000. Since we plan to look at large lags, we adopt the estimators with finite sample corrections that we discussed in Section 3.4. More specifically, for the estimation of (scaled) autocovariances of the noise, we use $\widehat{R}(j)_t^{n,\text{adj}}$ in (3.36) with $\widehat{QV}_t$ as in Section 5.1, and for estimating $r(j)$ we use $\widehat{r}(j)_t^{n,\text{adj}}$ in (3.37). The tuning parameter k_n is chosen in the same way as in the previous subsection, using observations within each half-hour time interval. It turned out that for the majority of half-hour intervals, with $k_n = 40$, the heuristic criterion was well satisfied; while for the remaining, slightly different values of k_n worked.

5.2.1. Diurnal Features in the Size

We first check whether the size of the noise admits any diurnal features, by looking at the standard deviation of noise, namely,

$$\sqrt{R(0)_t} = \sqrt{\frac{\int_0^t \gamma_s^2 dA_s}{A_t}}$$

for $t = 1/13$, to be estimated based on the observations within each of the thirteen half-hour time intervals. The estimation results are shown in Figure 7. To save space, we only plot the results for the first five trading days in January 2011 and an average curve representing the standard deviations of noise during different half-hour intervals averaged over the 20 trading days in January 2011.

Figure 7 indicates that the size of noise is around 0.0006. This is similar to the simulation studies: in the case without rounding in Section 4.1, the standard deviation is about $\sqrt{3.5 \times 10^{-7}} \approx 0.0006$, and in the case with rounding in Section 4.2, the standard deviation is about $\sqrt{6 \times 10^{-7}} \approx 0.00077$. Compared with the estimates for the same stock in Hansen and Lunde (2006) for the years of 2000–2004, the size of 0.0006 is smaller even after adjusting for the price difference.⁹ One possible reason for such a comparison is that the market has become more liquid. In 2004, the daily transaction number at the NYSE for Citigroup (C) is about 4800 (see Table 1 in Hansen and Lunde (2006)), while for the period under study, the daily transaction number at the same exchange is about 8600, about 80% larger. Such a change echoes the findings, in particular, Fact III in Hansen and Lunde (2006) which states that “The noise is smaller than one might think.” Of course, even with such a small size of noise, the bias in the realized volatility would still be significant; in particular, it will be of similar size to the underlying volatility as soon as the sampling frequency is higher than one observation per minute.

On the other hand, notice that the price level is about \$5, and the typical bid-ask spread as well as the tick size is one cent, hence half a spread/tick size corresponds to around

⁹The stock price of Citigroup (C) is around \$50 in 2004, and around \$5 in 2011, about one tenth of the earlier price due to splitting. Presumably, the size of noise would also shrink by about 10 times. Table 2 in Hansen and Lunde (2006) reports the estimated variance of noise to be about 0.01×10^{-2} , which corresponds to a standard deviation of $\sqrt{0.01 \times 10^{-2}}/10 = 0.001$ after adjusting for the price level in such a naive way.

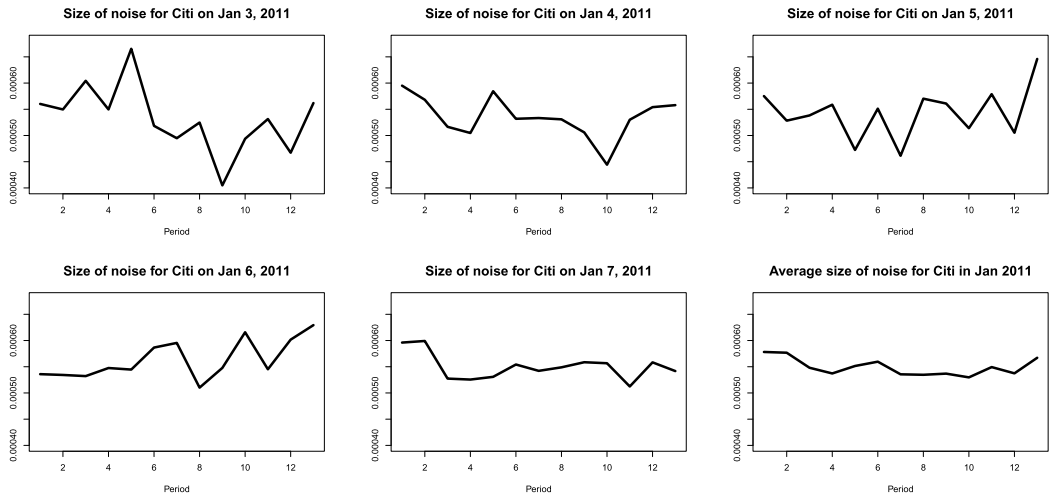


FIGURE 7.—Estimated sizes of noise for Citigroup (C) stock, during 13 half-hour intervals and for different trading days. The first five curves plot the estimates for the first five trading days in January 2011. The lower-right curve plots the sizes of noise during different half-hour intervals averaged over the 20 trading days in January 2011.

$\log(5.005/5) \approx 0.001$, which is larger than 0.0006, the estimated size of noise. At first sight, this comparison may look worrisome as one may wonder whether the dependence estimated afterwards would be purely due to rounding. However, our simulation study in Section 4.2 ensures that it will NOT be the case: our method works well even when rounding is involved, and even when rounding appears to dominate the noise.

About the diurnal patterns of the size of noise, the plots in Figure 7 do not show a clear pattern. For some days like Jan 4 and 5, 2011, it seems that the size of noise is larger at the beginning and towards the end of the trading hours and smaller in the middle, which resembles the behavior of volatility, as shown in Figure 8.

Plotting the estimated size of noise against volatility during different half-hour intervals yields Figure 9.

The left panel in Figure 9 suggests that when the volatility is unusually high, the size of noise also tends to be unusually high. If we remove the two outliers on its far right, we obtain the plot on the right in Figure 9, which no longer shows a clear pattern. The correlations in two plots are 0.13 and -0.19 , respectively, both of which are statistically significant at the 5% level. Using the alternative estimate of noise size, namely, (3.25) with j a large enough number, yields similar conclusions.

5.2.2. Diurnal Features in the Dependence Structure

We next check whether the dependence changes over the day. Since the autocovariances and autocorrelations are related to each other by a factor of the variance of noise, which we have examined above (by looking at its square root), for the dependence we only look at the autocorrelations estimated by $\hat{r}(j)_t^{n, \text{adj}}$ as in (3.37). The estimation results are shown in Figure 10.

We see from Figure 10 that the autocorrelations are relatively stable across different half-hour intervals especially after averaging, which presumably reduces the estimation errors. This suggests that the dependence structure of the noise may not change much within each trading day.

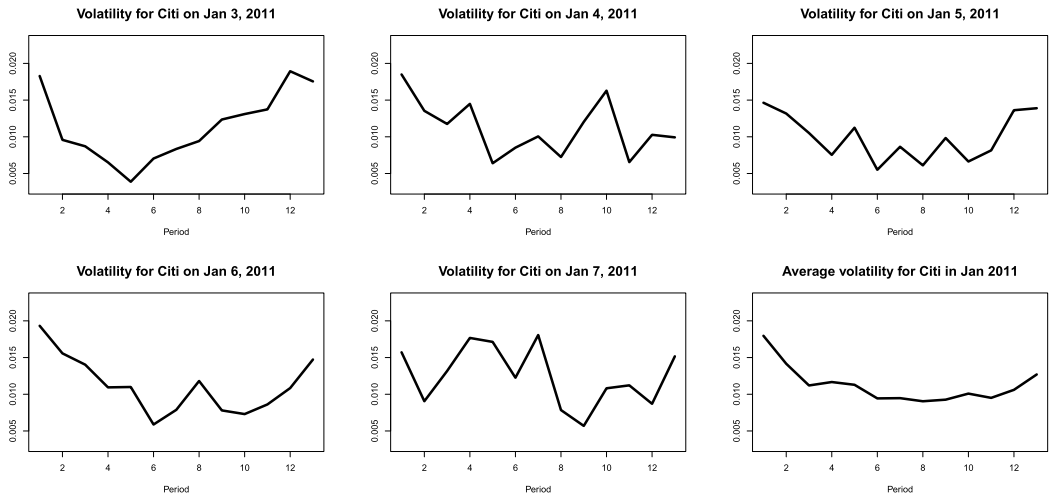


FIGURE 8.—Estimated volatilities for Citigroup (C) stock, during 13 half-hour intervals and for different trading days. The first five curves plot the estimates for the first five trading days in January 2011. The lower-right curve plots the estimated volatilities during different half-hour intervals averaged over the 20 trading days in January 2011.

Furthermore, recall that we assume the autocorrelations *as a function of tick time*. Since the observation times (= transaction times) are irregular, the above finding that the (estimated) autocorrelations seem to be quite stable across different half-hour intervals supports our assumption, and suggests that it is reasonable to assume the autocorrelations as a function of tick time.

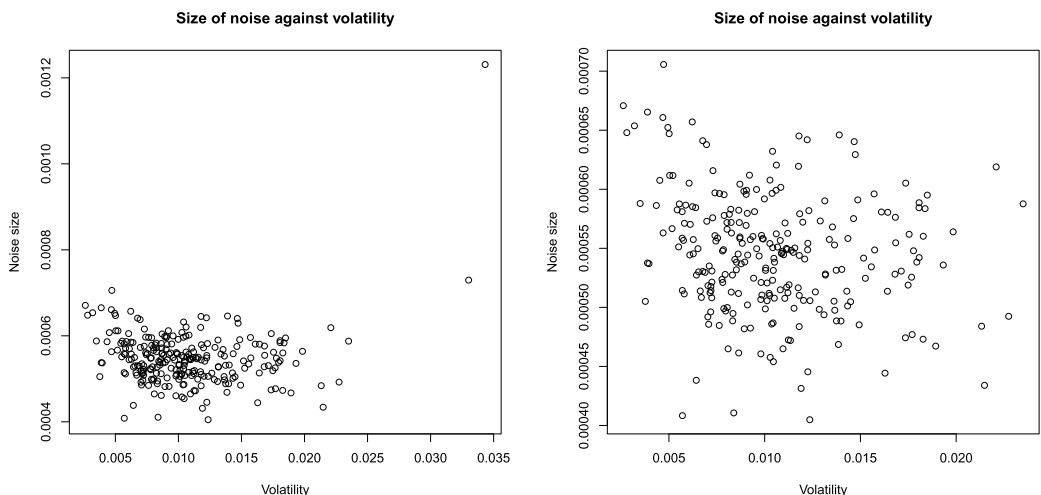


FIGURE 9.—Scatterplots of the estimated sizes of noise against volatilities during different half-hour intervals for Citigroup (C) stock, in January 2011. The right panel is a zoomed-in version of the left panel with the two outliers on its far right removed.

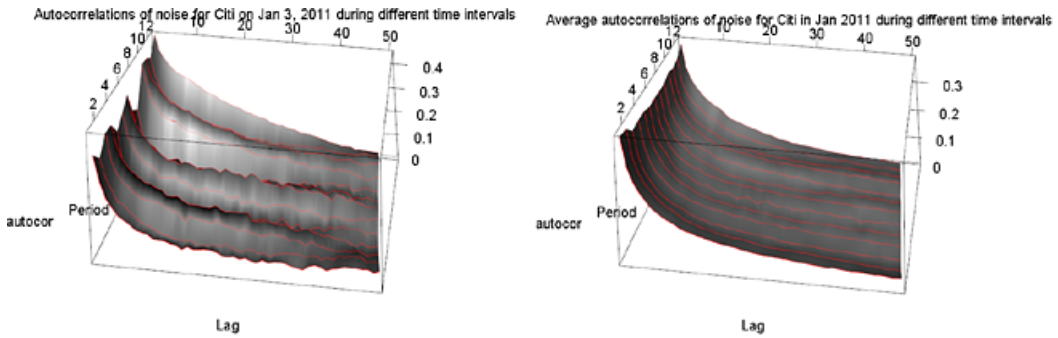


FIGURE 10.—Estimated autocorrelations of noise up to lag 50 for Citigroup (C) stock, during different half-hour intervals. Each red curve represents the estimates during one half-hour interval. Left: autocorrelations on January 3, 2011; right: autocorrelations averaged over the 20 trading days in January 2011.

5.3. Estimating the Autocorrelations of Noise

Having seen that the autocorrelations are relatively stable within each trading day, we next use the whole-day data for estimation. For each of the 20 trading days, we estimate the autocorrelations of noise by using (3.37) (with $t = 1$), and plot the estimates in Figure 11.

We see from Figure 11 that the noise is positively autocorrelated (up to a large lag) for all the days under study. Again, we reach the same conclusion as (5.1). Of course, Figure 11 tells us far more than the qualitative statement (5.1). It shows the magnitude of the autocorrelations as well as how they decay.

5.4. Tests and Confidence Intervals

Based on Theorem 3.8, we can further test whether the autocorrelations are equal to zero. More specifically, for any $j \geq 1$, under the null hypothesis $H_0 : r(j)_1 = 0$, by Theo-

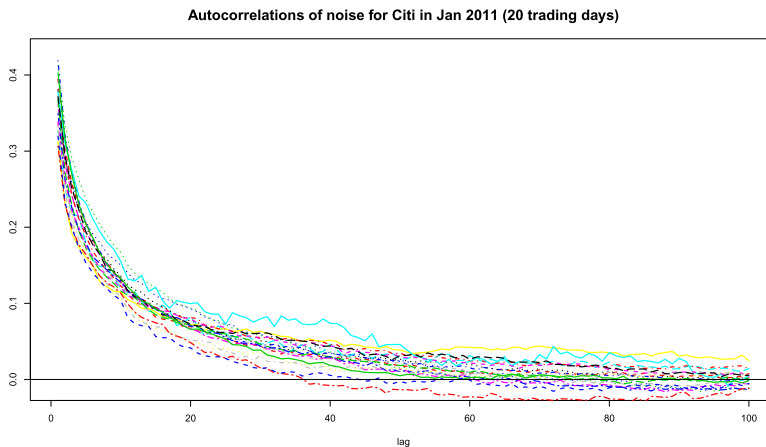


FIGURE 11.—Estimated autocorrelations of noise for Citigroup (C) stock in January 2011. Each curve is for one trading day, and plots the estimated autocorrelations of lags up to 100.

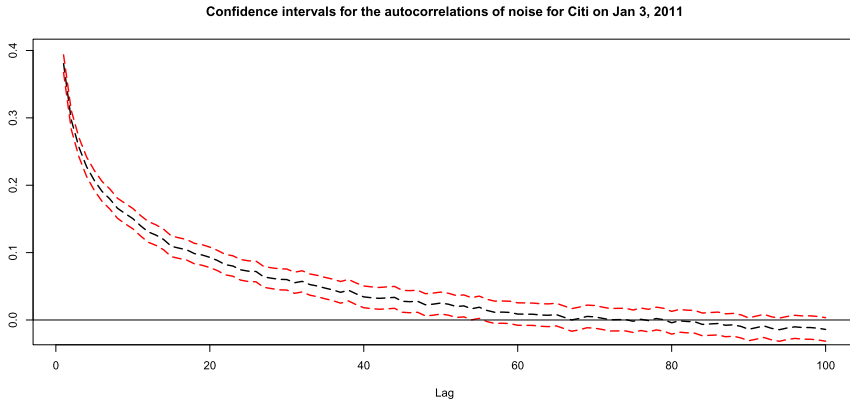


FIGURE 12.—Confidence intervals for the autocorrelations of noise up to lag 100 for Citigroup (C) stock on January 3, 2011.

rem 3.8 applied with $\mathbf{j} = (0, j)$ and that $k_n/\sqrt{N_1^n} = o_p(1)$, we have

$$T_n := \sqrt{\frac{N_t^n}{\widehat{S}(j)_t^n}} \widehat{r}(j)_t^{n,\text{adj}} \xrightarrow{\mathcal{L}} \Phi,$$

where Φ is an $\mathcal{N}(0, 1)$ -distributed random variable. So we can compute the p -value for testing $H_0 : r(j)_1 = 0$ as $P(|\Phi| > |t_n|)$ for any realization t_n of T_n . The tuning parameter k_n is taken to be 40, just the same as before, and the k'_n in $\widehat{\mathcal{Z}}(\mathbf{j}, \mathbf{j})_1^n$ is taken to be 15. For lags up to 20 and for the 20 trading days under study, the p -values turn out to be all small, mostly much smaller than 5%. Furthermore, since the estimated autocorrelations are positive, the results also imply that if one conducts a one-sided test $H_0 : r(j) \leq 0$, then one rejects these hypotheses at the 5% significance level. We hence conclude that the autocorrelations are statistically significantly positive, for lags up to 20 and for the 20 trading days under study.

Theorem 3.8 also allows us to build confidence intervals for the autocorrelations. More specifically, by Theorem 3.8, a 95% confidence interval for the autocorrelations $r(j)$ is given by

$$\left[\widehat{r}(j)_1^{n,\text{adj}} \pm 1.96 \cdot \sqrt{\frac{\widehat{S}(j)_1^n}{N_1^n}} \right], \quad j = 1, 2, \dots \quad (5.2)$$

Applying these formulas to the data on January 3, 2011 yields the confidence intervals shown in Figure 12 for the autocorrelations of noise. We can then conclude that on January 3, 2011, the autocorrelation of the noise of any lag up to 15 is greater than 0.15 or so, with 95% confidence; on the other hand, for lags higher than 50 or so, the autocorrelations are statistically insignificant.

5.5. Decay Rate in the Autocorrelations of the Noise

Finally, if indeed the autocorrelations of the noise decay polynomially with rate ν in the sense that $\lim_{j \rightarrow \infty} j^\nu |r(j)| \rightarrow \alpha$ for some $\alpha > 0$ (or slightly more generally, $0 <$

$\liminf_{j \rightarrow \infty} j^v |r(j)| \leq \limsup_{j \rightarrow \infty} j^v |r(j)| < \infty$), then we can consistently estimate the decay rate v as follows. First, the decay assumption implies that

$$\lim_{j \rightarrow \infty} \frac{\log(|r(j)|)}{\log j} = v.$$

By Theorem 3.4, we have

$$\widehat{r}(j)_t^n = r(j) + \frac{1}{\sqrt{N_t^n}} \widehat{\varepsilon}_t^n(j),$$

where $\widehat{\varepsilon}_t^n(j)$ converges $\mathcal{F}_\infty$ -stably in law to a variable which is $\mathcal{F}_\infty$ -conditionally centered Gaussian with (conditional) variance $\mathcal{S}(j)_t$. By noticing that $\sup_j \mathcal{S}(j)_t = O_p(1)$, and

$$\frac{\log(|\widehat{r}(j)_t^n|)}{\log j} = \frac{\log(|r(j)|)}{\log j} + \frac{\log\left(1 + \widehat{\varepsilon}_t^n(j)/(\sqrt{N_t^n}|r(j)|)\right)}{\log j},$$

we immediately have $\frac{\log(|\widehat{r}(j)_t^n|)}{\log j} \xrightarrow{\mathbb{P}} v$ so long as $j \rightarrow \infty$ and $j^v/\sqrt{N_t^n} \rightarrow 0$.

In practice, we can either use $\frac{\log(|\widehat{r}(j)_t^n|)}{\log j}$ with a large j to estimate v , or possibly more efficiently, conduct a linear regression between $\log(|\widehat{r}(j)_t^n|)$ and $\log j$ for j 's in a wide range and estimate v by the slope in the regression. Below, we adopt the second approach. Plotting the logarithms of the estimated autocorrelations against the logarithms of the lags gives the curves shown in Figure 13. The plots do show a roughly linear pattern (except that towards the right tail, the curves fluctuate drastically, likely due to the big estimation errors in the *log scale* when the true autocorrelations are almost zero). The linear pattern suggests a linear relationship between the log-autocorrelation and log-lag. Conducting linear regressions between them reveals that autocorrelations decay polynomially at rate between 0.5 and 1.7 (note that the consistency result Theorem 3.1 only requires

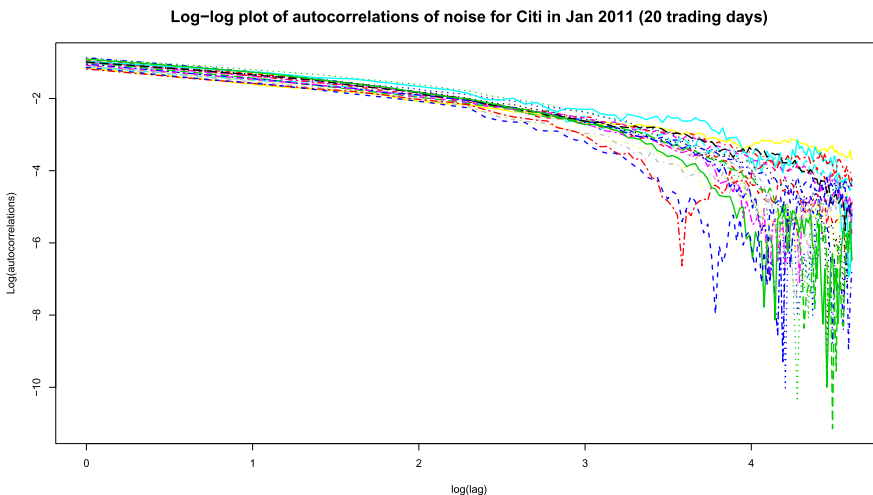


FIGURE 13.—Log-log plots of estimated autocorrelations of noise up to lag 100 for Citigroup (C) stock in January 2011. Each curve is for one trading day, and plots the logarithms of estimated autocorrelations of lags 1 through 100 against the logarithms of lags.

that $v > 0$). Such a slow decay rate indicates a possible long range dependence of the noise. What is the reason behind such a slow decay? Can it be due to rounding? The studies in Section 4.2 suggest that rounding cannot be the main cause. While we do not quite understand this phenomenon at the moment, we believe that it is an important factor to be taken into account in various applications.

6. CONCLUSION AND DISCUSSIONS

In this paper, we study the estimation of the (joint) moments, in particular, the autocovariances/autocorrelations of the microstructure noise, based on high frequency data. The estimation is conducted under a relatively general nonparametric setting which allows the underlying price process to have jumps, the observation times to be irregularly spaced, and the noise to be dependent on the price process and to have diurnal features. We establish consistency as well as central limit theorems for our proposed estimators. Simulation studies demonstrate that our estimators perform well in the presence of jumps, irregular observation times, and even rounding. Extensive empirical studies show that for the stocks tested, the microstructure noises are not white, but are actually (moderately) positively correlated. Some diurnal features of the noise are also observed.

When the noises have general autocorrelations, existing theory based on i.i.d. noises or noises of other simple specific forms has to be modified. We show in a subsequent paper (Jacod, Li, and Zheng (2015)) that when the microstructure noise can be dependent, the autocorrelations of the noise lead to a bias in existing volatility estimators, and there we introduce a new estimator which corrects for such a bias.

Besides volatility estimation, our results are also valuable to other areas such as measuring liquidity and high frequency trading. In terms of measuring liquidity, our method allows one to estimate the size of microstructure noise, not only about the second moment, but also about all the higher order moments. Regarding high frequency trading, the mere fact that microstructure noise is (positively) autocorrelated means that it is possible to make forecasts about the microstructure noise. More specifically, since we can estimate both the size of noise and the decay rate of the autocorrelations, a time series model for microstructure noise can be fitted or calibrated, which in turn can be used to make forecasts about noise, and consequently to help with designing better trading strategies.

To sum up, microstructure noise exhibits rich structure which has impacts on and potential applications in various areas, and in this paper we provide an approach to investigate this rich structure.

APPENDIX A: PROOFS

In this appendix, we indicate the main stream of the proofs. All details can be found in the Supplemental Material, Appendix C.

A.1. The Sampling Scheme and the Localization Procedure

In all the sequel, the constant K varies from line to line, but never depends on n and on the various indices $i, j, \dots$. If it depends on auxiliary parameters $p, q, \dots$, we write it as $K_{p,q,\dots}$.

We start with some consequences of Assumption (O). First, recalling the enlarged filtration $(\mathcal{F}_t^n)$ defined in that assumption, we have the following:

LEMMA A.1: *Under (O), any $(\mathcal{F}_t)$ -martingale is also an $(\mathcal{F}_t^n)$ -martingale.*

In particular, the process X is still an $(\mathcal{F}_t^n)$ -semimartingale and (2.1) for X' holds with the same W and $\underline{p}$, still a Brownian motion and a Poisson measure relative to $(\mathcal{F}_t^n)$, and α and γ are also $(\mathcal{F}_t^n)$ -semimartingales.

Next, we describe how far in a sense the sampling scheme is from a regular scheme. We have a sequence u_n of positive integers and a d -dimensional càdlàg $(\mathcal{F}_t)$ -adapted process $V = (V^j)_{1 \leq j \leq d}$, and we associate the following d -dimensional processes (with the convention $\sum_{i=0}^m = 0$ if $m \leq 0$):

$$H_t^n = H(V)_t^n = \frac{1}{\sqrt{\Delta_n}} \sum_{i=0}^{N_t^n - u_n} V_{T(n,i)} (\alpha_{T(n,i)} \Delta(n, i+1) - \Delta_n). \quad (\text{A.1})$$

LEMMA A.2: *Assume $u_n = 0$, or $\Delta(n, i) \leq K \Delta_n^\rho$ for some $\rho \in (1/2, 1)$ and $u_n \Delta_n^{\rho-1/2} \rightarrow 0$.*

(a) *We have the following, which implies (2.4):*

$$H_t^m := \Delta_n \sum_{i=0}^{N_t^n - u_n} V_{T(n,i)} \xrightarrow{\text{u.c.p.}} \int_0^t V_s dA_s. \quad (\text{A.2})$$

(b) *The processes H^n converge $\mathcal{F}_\infty$ -stably in law to a limiting process $H = (H^{(j)})_{1 \leq j \leq d}$ defined on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$, and which, conditionally on $\mathcal{F}$, is a centered continuous Gaussian martingale with (conditional) covariances*

$$\mathbb{E}(H_t^{(j)} H_t^{(m)} | \mathcal{F}) = \int_0^t V_s^j V_s^m \alpha_s \bar{\alpha}_s ds = \int_0^t V_s^j V_s^m \bar{\alpha}_s dA_s. \quad (\text{A.3})$$

Next, we turn to the so-called “localization procedure,” which allows us to assume that the characteristics of the processes involved (such as $X, \alpha, \bar{\alpha}, \gamma$) are bounded. Unlike in the case of regular sampling, this procedure is not totally trivial here. We need first to state an assumption on an arbitrary Itô semimartingale V on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, which can thus be written as X' in (2.1) with some coefficients b^V, σ^V, δ^V :

ASSUMPTION (K): *The processes V, b^V, σ^V are bounded, and $|\delta^V(\omega, t, z)| \leq J(z)$ for some bounded function J on E satisfying $\int J(z)^2 \lambda(dz) < \infty$.*

The strengthened set of assumptions goes as follows. It depends on some number, a priori arbitrary in $(0, 1)$. However, we will need this number to be related to the sequence k_n which satisfies (3.5) or (3.27), according to the case and we choose some $\rho \in (1/2 + \eta, 1)$, where η is as in (3.5) or (3.27). In particular, we have for some $\varepsilon > 0$:

$$k_n \Delta_n^{\rho-1/2-\varepsilon} \rightarrow 0. \quad (\text{A.4})$$

ASSUMPTION (SHON): *We have (H), (N), (O), with further:*

- (i) *The processes $X', \alpha, \bar{\alpha}$, and γ satisfy (K), and the process $1/\alpha$ is bounded.*
- (ii) *The variables Γ_i in (H-ii) are bounded and $S_{I+1} \equiv \infty$ for a (non-random) integer I .*
- (iii) *We have $\Delta(n, i) \leq K \Delta_n^\rho$ for all n, i and some constant K .*

Then, we have the following:

LEMMA A.3: *If the results of Section 3 hold under (SHON), they also hold under (H), (O), and (N).*

We always assume (SHON) below, mostly without special mention. This implies identically, for some constant C , and by (2.4) for the last property:

$$\Delta(n, i) \leq C\Delta_n^p, \quad A_t \leq Ct, \quad \mathbb{P}(\Omega_t^n) \rightarrow 1 \quad \text{if } \Omega_t^n = \{\Delta_n N_t^n \leq (1 + Ct)\}. \quad (\text{A.5})$$

Note that, in view of classical estimates and of Lemma A.1, for any Itô semimartingale V satisfying (K) and all $p \geq 2$ and finite $(\mathcal{F}_t^n)$ -stopping times $S \leq T$, we have

$$\begin{aligned} |\mathbb{E}(V_T - V_S | \mathcal{F}_S^n)| &\leq K\mathbb{E}(T - S | \mathcal{F}_S^n), \\ \mathbb{E}(|V_T - V_S|^p | \mathcal{F}_S^n) &\leq K_p\mathbb{E}(T - S | \mathcal{F}_S^n). \end{aligned} \quad (\text{A.6})$$

By conditioning first on $\mathcal{F}_{T(n,i-1)}^n \vee \sigma(\Delta(n, i))$, and upon using (A.5), plus Hölder's inequality for the second estimate, we also get

$$\begin{aligned} \left| \mathbb{E} \left(\int_{T(n,i-1)}^{T(n,i)} (V_s - V_{T(n,i-1)}) ds \mid \mathcal{F}_{T(n,i-1)}^n \right) \right| &\leq K\mathbb{E}(\Delta(n, i)^2 | \mathcal{F}_{T(n,i-1)}^n) \leq K\Delta_n^2, \\ \mathbb{E} \left(\left| \int_{T(n,i-1)}^{T(n,i)} (V_s - V_{T(n,i-1)}) ds \right|^p \mid \mathcal{F}_{T(n,i-1)}^n \right) &\leq K_p\mathbb{E}(\Delta(n, i)^{p+1} | \mathcal{F}_{T(n,i-1)}^n) \leq K_p\Delta_n^{p+1}. \end{aligned} \quad (\text{A.7})$$

A.2. Some Facts About Stationary Sequences

Whether the noise and the underlying process X are a priori defined on the same space or not is irrelevant for the results. However, for the proofs it is convenient to suppose that X , α , $\bar{\alpha}$, γ , and the observation times $T(n, i)$ are defined (and satisfy the relevant assumptions) on a space $(\Omega^{(0)}, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}^{(0)})$, whereas the sequence $(\chi_i)_{i \in \mathbb{Z}}$ is defined on another space $(\Omega^{(1)}, \mathcal{G}, (\mathcal{G}_i)_{i \in \mathbb{Z}}, \mathbb{P}^{(1)})$, with $\mathcal{G}_i = \sigma(\chi_k : k \leq i)$ and $\mathcal{G}^i = \sigma(\chi_k : k \geq i)$ and $\mathcal{G} = \bigvee_i \mathcal{G}_i$, and we set

$$\Omega = \Omega^{(0)} \times \Omega^{(1)}, \quad \mathcal{F} = \mathcal{F}^{(0)} \otimes \mathcal{G}, \quad \mathbb{P} = \mathbb{P}^{(0)} \otimes \mathbb{P}^{(1)}.$$

As usual, any variable or process or σ -field on $\Omega^{(j)}$ for $j = 1, 2$ is also considered, with the same notation, as defined on the product Ω (this is in particular the case of the filtration $(\mathcal{F}_t^n)$, a priori defined on Ω^0 but which can be considered as a filtration on Ω).

The space $(\Omega^{(1)}, \mathcal{G}, (\mathcal{G}_i)_{i \in \mathbb{Z}}, \mathbb{P}^{(1)})$ is naturally endowed with a measure-preserving and invertible transformation θ such that $\chi_{i+j} = \chi_i \circ \theta^j$ for all $i, j \in \mathbb{Z}$, and θ is ergodic by the ρ -mixing property. Let us also recall a consequence of the definition of the mixing coefficients $\rho_j(\chi)$, and of the product structure of the space $(\Omega, \mathcal{F}, \mathbb{P})$. We set $\mathcal{H}_i = \mathcal{F}^0 \otimes \mathcal{G}_i$. If ξ is a centered square-integrable variable on $(\Omega^{(1)}, \mathcal{G}, (\mathcal{G}_i)_{i \in \mathbb{Z}}, \mathbb{P}^{(1)})$, which is measurable with respect to $\mathcal{G}^0 = \sigma(\chi_i : i \geq 0)$, when $j \geq 1$ we have

$$\mathbb{E}(|\mathbb{E}(\xi \circ \theta^{i+j} | \mathcal{H}_i)|^2 | \mathcal{F}^0) \leq \rho_j(\chi)^2 \mathbb{E}(\xi^2) \leq \frac{K\mathbb{E}(\xi^2)}{j^{2v}}. \quad (\text{A.8})$$

We also have the following: if ξ is $\mathcal{G}_i$ -measurable and ξ' is $\mathcal{G}^{i+j}$ -measurable, both square-integrable, by (2.6) applied to $U = (\xi \circ \theta^{-i} - E(\xi))/\sqrt{\text{var}(\xi)}$ and $U' = (\xi' \circ \theta^{-i} -$

$E(\xi')/\sqrt{\text{var}(\xi')}$, plus $\rho_k(\chi) \leq K/k^v$ and $\mathbb{E}(U) = \mathbb{E}(U') = 0$ and the stationarity, we have

$$\begin{aligned} |\mathbb{E}(\xi\xi' | \mathcal{F}^0)| &= |\mathbb{E}(\xi)\mathbb{E}(\xi') + \mathbb{E}(UU')\sqrt{\text{var}(\xi)\text{var}(\xi')}| \\ &\leq |\mathbb{E}(\xi)\mathbb{E}(\xi')| + \frac{K}{j^v}\sqrt{\mathbb{E}(\xi^2)\mathbb{E}(\xi'^2)}. \end{aligned} \quad (\text{A.9})$$

In the rest of this subsection, we consider a sequence $\xi^n = (\xi^{n,j})_{1 \leq j \leq d}$ of d -dimensional variables on the space $(\Omega^{(1)}, \mathcal{G}, \mathbb{P}^{(1)})$, satisfying the following, where w_n and w are non-negative integers with $w_n \geq w \vee 1$:

$$\begin{aligned} \xi^n &\xrightarrow{\mathbb{P}} \xi, \quad \mathbb{E}(\xi^n) = 0, \quad \sup_{n \in \mathbb{N}} \mathbb{E}(\|\xi^n\|^p) < \infty \quad \text{for all } p > 0, \\ \xi^n &\text{ is measurable with respect to the } \sigma\text{-field } \mathcal{G}^0 \cap \mathcal{G}_{w_n} = \sigma(\chi_i : 0 \leq i \leq w_n), \\ \xi &\text{ is measurable with respect to the } \sigma\text{-field } \mathcal{G}^0 \cap \mathcal{G}_w = \sigma(\chi_i : 0 \leq i \leq w). \end{aligned} \quad (\text{A.10})$$

Note that $\mathbb{E}(\xi) = 0$, whereas $w_n \rightarrow w$ is *not* assumed. We write $\xi_i^n = \xi^n \circ \theta^i$ and $\xi_i = \xi \circ \theta^i$. Then, for $i \geq 1$, (A.8) and (A.9) yield

$$\begin{aligned} \mathbb{E}(\|\mathbb{E}(\xi_{l+i}^n | \mathcal{H}_l)\|^2 | \mathcal{F}^0) &= \mathbb{E}(\|\mathbb{E}(\xi_{l+i}^n | \mathcal{G}_l)\|^2) \leq K\mathbb{E}(\|\xi^n\|^2)/i^{2v} \leq K/i^{2v}, \\ |\mathbb{E}(\xi_{l+i}^{n,j} \xi_{l+i}^{n,k} | \mathcal{F}^0)| &\leq K\mathbb{E}(\|\xi^n\|^2)/((i - w_n) \vee 1)^v \leq K/((i - w_n) \vee 1)^v, \end{aligned} \quad (\text{A.11})$$

and the same for ξ_{l+i} , with w_n replaced by w in the second inequality. As soon as $v > 1$, we deduce that the following series converge and define a covariance matrix:

$$a^{jk} = \mathbb{E}(\xi_0^j \xi_0^k) + \sum_{i=1}^{\infty} (\mathbb{E}(\xi_0^j \xi_i^k) + \mathbb{E}(\xi_i^j \xi_0^k)). \quad (\text{A.12})$$

In the simple situation where $w = 0$ (so ξ is a function of χ_0), a trivial multi-dimensional extension of Corollary VIII.3.106 of Jacod and Shiryaev (2003) yields a Central Limit Theorem which says that, when $v > 1$,

$$\frac{1}{\sqrt{v_n}} \sum_{i=0}^{\lfloor tv_n \rfloor} \xi_i \xrightarrow{\mathcal{L}} B_t, \quad (\text{A.13})$$

for any sequence $v_n \rightarrow \infty$, where $B = (B_t)$ is a d -dimensional Brownian motion with covariance $\mathbb{E}(B_1^j B_1^k) = a^{jk}$.

We need to extend this result into two directions: when ξ^n depends on n , subject to (A.10), with $w_n \geq 1$; and when, further, each ξ_i^n is affected by a random weight which is the value at time $T(n, i)$ of a d -dimensional bounded càdlàg process $V = (V^j)_{1 \leq j \leq d}$ on $(\Omega^{(0)}, \mathcal{F}^0, (\mathcal{F}_t), \mathbb{P}^{(0)})$, whose components satisfy (A.6). This leads us to consider the following processes, where u_n is a sequence of integers:

$$G^n = (G^{n,j})_{1 \leq j \leq d}, \quad \text{where } G_t^{n,j} = \sqrt{\Delta_n} \sum_{i=0}^{N_t^n - u_n} V_{T(n,i)}^j \xi_i^{n,j}. \quad (\text{A.14})$$

We also need a joint convergence for G^n and the processes $H^n = H(V)^n$ of (A.1).

THEOREM A.4: Assume (A.10) and $v > 1$ in (N), and also $w_n \mathbb{E}(\|\xi^n - \xi\|^2) \rightarrow 0$ and $u_n \Delta_n^{\rho-1/2} \rightarrow 0$. For any $t \geq 0$, the $(2d)$ -dimensional variables (G_t^n, H_t^n) converge $\mathcal{F}_\infty$ -stably in law to a variable (G_t, H_t) with components $G_t^{(j)}$ and $H_t^{(j)}$, defined on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$, and which, conditionally on $\mathcal{F}$, is a centered Gaussian martingale with (conditional) covariances

$$\begin{aligned}\tilde{\mathbb{E}}(G_t^{(j)} G_t^{(m)} | \mathcal{F}) &= a^{jm} \int_0^t V_s^j V_s^m dA_s, \\ \tilde{\mathbb{E}}(H_t^{(j)} H_t^{(m)} | \mathcal{F}) &= \int_0^t V_s^j V_s^m \bar{\alpha}_s dA_s, \\ \tilde{\mathbb{E}}(G_t^{(j)} H_t^{(m)} | \mathcal{F}) &= 0.\end{aligned}\tag{A.15}$$

We will also need bounds for processes of the same type as G^n , under slightly different circumstances, and whether $v > 1$ or not. We still consider the setting (A.10), with $d = 2$, and w_n^j is the smallest integer such that $\xi^{n,j}$ is $\mathcal{G}^0 \cap \mathcal{G}_{w_n^j}$ -measurable (so $w_n = \max(w_n^1, w_n^2)$). We also consider a one-dimensional bounded $\mathcal{F}^0$ -measurable process V and a sequence of integers $u_n \geq 0$. Then we set (note that $G^{n,1} = \frac{1}{\sqrt{\Delta_n}} G^{n,1}$ with the notation (A.14))

$$\begin{aligned}G_t^{n,1} &= \sum_{i=0}^{N_t^n - u_n} V_{T(n,i)} \xi_i^{n,1}, & G_t^{n,2} &= \sum_{i=0}^{N_t^n - u_n} V_{T(n,i)} \xi_{i+w_n^1+k_n}^{n,2}, \\ G_t^{n,m} &= \sum_{i=0}^{N_t^n - u_n} V_{T(n,i)} \xi_i^{n,1} \xi_{i+w_n^1+k_n}^{n,2}.\end{aligned}\tag{A.16}$$

LEMMA A.5: In the previous setting we have, for $j = 1, 2$,

$$\begin{aligned}\mathbb{E}(|G_t^{n,j}|^2 | \mathcal{F}^0) &\leq K(w_n^j N_t^n + (N_t^n)^{2-v} (1 + \log(1 + N_t^n 1_{\{v=1\}}))), \\ \mathbb{E}(|G_t^{n,m}|^2 | \mathcal{F}^0) &\leq K((k_n + w_n^1 + w_n^2) N_t^n + k_n^{-2v} (N_t^n)^2 + (N_t^n)^{2-v} (1 + \log(1 + N_t^n 1_{\{v=1\}}))).\end{aligned}$$

A.3. Further Auxiliary Results

In this subsection, we gather a few useful results of a technical character.

The first of these results is a non-standard criterion for a triangular array of variables to be asymptotically negligible. Again, the reason for introducing this criterion is the fact that our sampling times are non-regularly spaced. Below, for each $n \geq 1$, we have a sequence $(\delta_i^n)_{i \geq 0}$ of real variables, a discrete-time filtration $(\tilde{\mathcal{H}}_i^n)_{i \geq 0}$, and an integer $w_n \geq 1$.

LEMMA A.6: In the above setting, if $\mathbb{E}(\mathbb{E}(\delta_i^n | \tilde{\mathcal{H}}_i^n)) \leq a_n$ and $\mathbb{E}(|\delta_i^n|^2) \leq a'_n$ and if each δ_i^n is $\tilde{\mathcal{H}}_{i+w_n}^n$ -measurable, the variables $B_i^n = \sum_{l=0}^i \delta_l^n$ satisfy

$$\mathbb{E}\left(\sup_{l \leq k} |B_l^n|\right) \leq (k+1)a_n + K\sqrt{ka'_n w_n}.\tag{A.17}$$

Next, with any process V and sequence u_n of nonnegative integers, we associate the following processes:

$$J(V)_t^n = \frac{1}{\sqrt{\Delta_n}} \left(\int_0^t V_s dA_s - \sum_{i=0}^{N_t^n - u_n} V_{T(n,i)} \alpha_{T(n,i)} \Delta(n, i+1) \right). \quad (\text{A.18})$$

The following is a well-known result when observation times are regularly spaced (so $T(n, N_s^n) = [s/\Delta_n]$ and $A_t = t$), but in our setting it needs a proof.

LEMMA A.7: *If V is bounded and satisfies (A.6) and $u_n \Delta_n^{\rho-1/2} \rightarrow 0$, then $J(V)_t^n \xrightarrow{\text{u.c.p.}} 0$.*

Our third auxiliary result mainly compares $\mathbf{r}(k_n, \mathbf{j})$ defined in (3.26) with $\mathbf{r}(\mathbf{j})$ in (3.1).

LEMMA A.8: *For $k \geq 1$, let $f_v(k) = k^{-(v \wedge 1)}$ if $v \neq 1$ and $f_1(k) = \frac{\log(1+k)}{k}$. We have $\mathbb{E}((\bar{\chi}_j^n)^2) \leq K f_v(k_n)$ and, if $\mathbf{j} = (j_1, \dots, j_q) \in \mathcal{J}$ and $\mu = \mu(\mathbf{j})$ and $q \geq 1$,*

$$|\mathbf{r}(k_n; \mathbf{j}) - \mathbf{r}(\mathbf{j})| \leq K_j \frac{\sqrt{f_v(k_n)}}{k_n^v}, \quad (\text{A.19})$$

$$\mathbb{E} \left(\left| \prod_{m=1}^q (\chi_{j_m} - \bar{\chi}_{\mu+(2m-1)k_n}^n) - \prod_{m=1}^q \chi_{j_m} \right|^2 \right) \leq K_{p,j,\varepsilon} f_v(k_n).$$

We will need bounds on the differences between $U(\mathbf{j})_t^n$ or $U^4(\mathbf{j}, \mathbf{j}')_t^n$, and their counterparts when there is only noise and the process γ_t is suitably frozen, which are defined as follows. Let $\mathbf{j} = (j_1, \dots, j_q)$ and $\mathbf{j}' = (j'_1, \dots, j'_{q'})$ be in $\mathcal{J}^+$, and set $\mu = \mu(\mathbf{j})$ and $\mu' = \mu(\mathbf{j}')$, and also $q'' = q + q'$ and $\mu'' = \mu + \mu'$. Then, with $\hat{\chi}(\mathbf{j})_i = \prod_{\ell=1}^q (\chi_{i+j_\ell} - \bar{\chi}_{i+\mu+(2\ell-1)k_n}^n)$, we set

$$U'(\mathbf{j})_t^n = \sum_{i=0}^{N_t^n - u_n} \gamma_{T(n,i)}^q \hat{\chi}(\mathbf{j})_i^n, \quad U^4(\mathbf{j}, \mathbf{j}')_t^n = \sum_{i=0}^{N_t^n - u_n} \gamma_{T(n,i)}^{q''} \hat{\chi}(\mathbf{j})_i^n \hat{\chi}(\mathbf{j}')_{i+\mu+(2q+1)k_n}^n. \quad (\text{A.20})$$

LEMMA A.9: *Assume (3.5) and $X'' \equiv 0$ in (2.1). Let $r > 1$ and $q, q' \geq 0$ and $t > 0$. There are constants $K_{r,q,t}, K_{r,q,q',t}$ such that, for any $\mathbf{j}, \mathbf{j}' \in \mathcal{J}^+$ with $q(\mathbf{j}) = q$ and $q(\mathbf{j}') = q'$ and $\mu(\mathbf{j}) + \mu(\mathbf{j}') \leq k_n$, and any sequence of integers u_n with $0 \leq u_n \leq \mu'' + (2q'' + 1)k_n$, we have*

$$\mathbb{E}(|U(\mathbf{j})_t^n - U'(\mathbf{j})_t^n| 1_{\Omega_t^n}) \leq K_{r,q,t} k_n^{1/r} \Delta_n^{\rho/r-1}, \quad (\text{A.21})$$

$$\mathbb{E}(|U^4(\mathbf{j}, \mathbf{j}')_t^n - U^4(\mathbf{j}, \mathbf{j}')_t^n| 1_{\Omega_t^n}) \leq K_{r,q,q',t} k_n^{1/r} \Delta_n^{\rho/r-1}.$$

Our last two auxiliary results are related with the variables $\hat{\Delta}_i^n$ showing in (3.15).

LEMMA A.10: *Let $\phi'_n = \Delta_n^\kappa + \frac{1}{\sqrt{k_n}} + \frac{k_n \Delta_n^2}{\phi_n^2}$. If V is a bounded $(\mathcal{F}_t)$ -adapted process, we have for $j \geq 1$:*

$$\mathbb{E} \left(\sup_{k \leq j} \left| \sum_{i=1}^k V_{T(n,i)} (\hat{\Delta}_i^n - \bar{\alpha}_{T(n,i)}) \right| \right) \leq K(j \phi'_n + \sqrt{j k_n}). \quad (\text{A.22})$$

LEMMA A.11: *Under (3.5) and (3.14) for k_n and ϕ_n , we have (3.17), (3.18), and (3.19).*

A.4. Proof of the Main Results When $X'' \equiv 0$

From now on, we give the full proofs. The proof of the consistency results is simple enough, and for the CLT the idea consists in reducing the problem to an application of Theorem A.4. We first describe how to perform this reduction, in the setting of Theorem 3.4.

For later convenience, we add to the set $\mathcal{J}$ the “empty sequence” $\mathbf{e}$, with which we associate $\mu(\mathbf{e}) = 1$ and $q(\mathbf{e}) = 0$. With the natural convention that an empty product equals 1, this leads to $\mathbf{r}(\mathbf{e}) = \mathbf{r}(k_n; \mathbf{e}) = 1$ and to $U(\mathbf{e})_t^n = N_t^n$ and (3.28) becomes $Z(\mathbf{e})_t^n = \frac{1}{\sqrt{\Delta_n}}(\Delta_n N_t^n - A_t)$. All previous results where some $\mathbf{j}$ appears are true (usually trivially so) for $\mathbf{j} = \mathbf{e}$.

For $m = 1, \dots, d-1$, we take $\mathbf{j}_m = (j_{m,1}, \dots, j_{m,q(\mathbf{j}_m)})$ in $\mathcal{J}^+$, and we add the empty sequence $\mathbf{j}_d = \mathbf{e}$. We associate integers $u_n(m)$ and u_n , and d -dimensional variables ξ^n and ξ , and d -dimensional processes V , as follows (below, $m \in \{1, \dots, d\}$):

$$\begin{aligned} u_n(m) &= 2q(\mathbf{j}_m)k_n + \mu(\mathbf{j}_m) - 1, & u_n &= \max_{1 \leq m \leq d} u_n(m), & V_t^m &= \gamma_t^{q(\mathbf{j}_m)}, \\ \xi^{n,m} &= \prod_{\ell=1}^{q(\mathbf{j}_m)} (\chi_{j_{m,\ell}} - \bar{\chi}_{\mu(\mathbf{j}_m) + (2\ell-1)k_n}^n) - \mathbf{r}(k_n; \mathbf{j}_m), \\ \xi^m &= \prod_{\ell=1}^{q(\mathbf{j}_m)} \chi_{j_{m,\ell}} - \mathbf{r}(\mathbf{j}_m). \end{aligned} \quad (\text{A.23})$$

Note that $u_n(d) = 0$ and $V_t^d = 1$ and $\xi^{n,d} = \xi^d = 0$. By Lemma A.8, the variables ξ^n and ξ satisfy (A.10) with $w_n = u_n$ and $w = \max_{1 \leq m \leq d} \mu(\mathbf{j}_m)$. Note also that $u_n \Delta_n \rightarrow 0$ by (3.5), and $w_n \mathbb{E}(\|\xi^n - \xi\|^2) \rightarrow 0$ by another application of Lemma A.8, when $v > 1$.

With G^n , H^n , $J(V^m)^n$, and $U'(\mathbf{j})_t^n$ given by (A.14), (A.1), (A.18), and (A.20), all with u_n as above, a simple calculation shows us that for $1 \leq m \leq d$, the processes $Z(\mathbf{j}_m)^n$ of (3.28) satisfy

$$\begin{aligned} Z(\mathbf{j}_m)_t^n &= G_t^{n,m} - \mathbf{r}(k_n; \mathbf{j}_m) H_t^{n,m} + \sum_{j=1}^3 B(m, j)_t^n \quad \text{where} \\ B(m, 1)_t^n &= -\mathbf{r}(k_n; \mathbf{j}_m) J(V^m)_t^n, & B(m, 2)_t^n &= \sqrt{\Delta_n} (U(\mathbf{j}_m)_t^n - U'(\mathbf{j}_m)_t^n), \\ B(m, 3)_t^n &= \frac{\mathbf{r}(k_n; \mathbf{j}_m) - \mathbf{r}(\mathbf{j}_m)}{\sqrt{\Delta_n}} \int_0^t V_s^m dA_s. \end{aligned} \quad (\text{A.24})$$

By Lemmas A.7, A.8, and A.9, plus $u_n \leq Kk_n$ and (A.4), we obtain, for any $r > 1$ and when $X'' \equiv 0$ (below, K depends on $\mathbf{j}_m, t, r$):

$$\begin{aligned} B(m, 1)_t^n &\xrightarrow{\mathbb{P}} 0, & \mathbb{E}(|B(m, 2)_t| 1_{\Omega_t^n}) &\leq K k_n^{1/r} \Delta_n^{\rho/r-1/2}, \\ |B(m, 3)_t^n| &\leq K \frac{\sqrt{f_v(k_n)}}{k_n^v \sqrt{\Delta_n}}. \end{aligned} \quad (\text{A.25})$$

PROOF OF THEOREM 3.1 WHEN $X'' \equiv 0$: We consider $d = 2$ above, and $\mathbf{j}_1 = \mathbf{j}$. We have $\sqrt{\Delta_n} B(1, j)_t^n \xrightarrow{\mathbb{P}} 0$ for $j = 1, 2, 3$ by (A.25). Lemma A.2 and $\mathbf{r}(k_n; \mathbf{j}) \rightarrow \mathbf{r}(\mathbf{j})$ yield

$\sqrt{\Delta_n} \mathbf{r}(k_n; \mathbf{j}) H_t^n \xrightarrow{\mathbb{P}} 0$. With the notation (A.16), we have $\sqrt{\Delta_n} G^{n,1} = \Delta_n G^{m,1}$, and $w_n^1 \leq K k_n$. Since $N_t^n \leq K_t / \Delta_n$ on Ω_t^n , we deduce from Lemma A.5 (assuming without loss of generality here that $v < 1$) that

$$\mathbb{E}((\sqrt{\Delta_n} G_t^{n,1})^2 1_{\Omega_t^n}) \leq K_t (k_n \Delta_n + \Delta_n^v) \rightarrow 0.$$

Since $\mathbb{P}(\Omega_t^n) \rightarrow 1$, we deduce $\sqrt{\Delta_n} G_t^{n,1} \xrightarrow{\mathbb{P}} 0$. This proves (3.8), from which (3.9) follows because of (2.4). Q.E.D.

PROOF OF THEOREM 3.3 WHEN $X'' \equiv 0$: By Lemma A.11, it remains to prove (3.20) and (3.22), and in each of these two cases the second part is implied by the first one and (2.4).

(1) Let $\mathbf{j}, \mathbf{j}' \in \mathcal{J}^+$, and set $\mu = \mu(\mathbf{j})$, $\mu' = \mu(\mathbf{j}')$, $\mu'' = \mu + \mu'$, and $q'' = q + q'$. Set $C(\mathbf{j}, \mathbf{j}')_t^n = \Delta_n U^4(\mathbf{j}, \mathbf{j}')_t^n - \mathbf{r}(\mathbf{j}) \mathbf{r}(\mathbf{j}') \int_0^t \gamma_s^{q''} dA_s$.

We use the notation (A.10) with $d = 3$ and $\mathbf{j}_1 = \mathbf{j}$ and $\mathbf{j}_2 = \mathbf{j}'$, and set $V = \gamma^{q''} \bar{\alpha}$. We use the processes $J(V)^n$ of (A.18), $H^n = H^n(V)$ of (A.1), $U^4(\mathbf{j}, \mathbf{j}')^n$ of (A.20), $G^{m,1}$, $G^{m,2}$, $G^{m,n}$ of (A.16), all with $u_n = \mu'' - 1 + (2q'' + 3)k_n$; note that for (A.16) we have $w_n^j = u_n(j)$. Similarly with (A.24), we have

$$\begin{aligned} C(\mathbf{j}, \mathbf{j}')_t^n &= \sum_{l=1}^5 A(l)_t^n \quad \text{where} \\ A(1)_t^n &= \Delta_n (U^4(\mathbf{j}, \mathbf{j}')_t^n - U^4(\mathbf{j}, \mathbf{j}')_t^n), \\ A(2)_t^n &= (\mathbf{r}(k_n; \mathbf{j}) \mathbf{r}(k_n; \mathbf{j}') - \mathbf{r}(\mathbf{j}) \mathbf{r}(\mathbf{j}')) \int_0^t V_s dA_s, \\ A(3)_t^n &= -\sqrt{\Delta_n} \mathbf{r}(k_n; \mathbf{j}) \mathbf{r}(k_n; \mathbf{j}') J(V)_t^n, \\ A(4)_t^n &= -\sqrt{\Delta_n} \mathbf{r}(k_n; \mathbf{j}) \mathbf{r}(k_n; \mathbf{j}') H_t^n, \\ A(5)_t^n &= \Delta_n (G_t^{m,n} + \mathbf{r}(k_n; \mathbf{j}) G_t^{m,2} + \mathbf{r}(k_n; \mathbf{j}') G_t^{m,1}). \end{aligned}$$

Since $|\mathbf{r}(k_n; \mathbf{j})| + |\mathbf{r}(k_n; \mathbf{j}')| \leq K$ and $v_n^j \leq K(q'' + \mu'')k_n$, we deduce from Lemmas A.5 (again we can assume $v \neq 1$), A.8, and A.9, and since under (A.4) one may find $r > 1$ such that $(k_n \Delta_n^\rho)^{1/r} \leq K \sqrt{k_n \Delta_n}$ (recall $\rho > 1/2$), that we have

$$|C(\mathbf{j}, \mathbf{j}')_t^n| \leq K_{q''} L_t^n + Y(\mathbf{j}, \mathbf{j}')_t^n, \quad \text{where } L_t^n = \sqrt{\Delta_n} (|H_t^n| + |J(V)_t^n|) \quad \text{and} \quad (A.26)$$

$$\mathbb{E}(Y(\mathbf{j}, \mathbf{j}')_t^n 1_{\Omega_t^n}) \leq a_n(t, q'', \mu'') := K_{r,t} (k_n^{-(1 \wedge v)} + \sqrt{k_n \Delta_n (q'' + \mu'')} + \Delta_n^{-v/2}).$$

Exactly the same argument shows that $\Delta_n U(\mathbf{j})_t^n - \mathbf{r}(\mathbf{j}) \int_0^t \gamma_s^q dA_s = \sqrt{\Delta_n} Z(\mathbf{j})_t^n$ satisfies, with the same L_t^n :

$$\sqrt{\Delta_n} |Z(\mathbf{j})_t^n| \leq K_q L_t^n + Y(\mathbf{j})_t^n, \quad \text{where } \mathbb{E}(Y(\mathbf{j})_t^n 1_{\Omega_t^n}) \leq a_n(t, q, \mu). \quad (A.27)$$

(2) We have $L_t^n \xrightarrow{\mathbb{P}} 0$ by Lemmas A.2 and A.7, and $a_n(r, t, q'', \mu'') \rightarrow 0$ as $n \rightarrow 0$ under (3.5), hence the first part of (3.20) follows from $\mathbb{P}(\Omega_t^n) \rightarrow 1$.

For (3.22), we assume $v > 1$ and (3.21). Set $b_n = \sum_{m \in \mathbb{Z}: |m| > k'_n} |\mathbf{r}(\mathbf{j} \oplus \mathbf{j}'_{+m}) - \mathbf{r}(\mathbf{j})\mathbf{r}(\mathbf{j}')|$. The definitions of $\mathbf{s}(\mathbf{j}, \mathbf{j}')$ and $S(\mathbf{j}, \mathbf{j}')^n$, plus $\alpha_t, \gamma_t \leq K$, yield

$$\begin{aligned} & \left| \Delta_n S(\mathbf{j}, \mathbf{j}')^n_t - \mathbf{s}(\mathbf{j}, \mathbf{j}') \int_0^t \gamma_s^{q''} dA_s \right| \\ & \leq K t b_n + \sqrt{\Delta_n} \left(|Z_t^{n, \mathbf{j} \oplus \mathbf{j}'}| + \sum_{m=1}^{k'_n} (|Z_t^{n, \mathbf{j} \oplus \mathbf{j}'_{+m}}| + |Z_t^{n, \mathbf{j}_{+m} \oplus \mathbf{j}'}|) \right) + (2k'_n + 1) |C(\mathbf{j}, \mathbf{j}')^n_t|. \end{aligned}$$

Observe that $q(\mathbf{j} \oplus \mathbf{j}'_{+m}) = q''$ and $\mu(\mathbf{j} \oplus \mathbf{j}'_{+m}) \leq \mu'' + m$, and the same for $\mathbf{j}_{+m} \oplus \mathbf{j}'$. Then, omitting the dependency on q, q', q'' of the constants, (A.26) and (A.27) yield

$$\begin{aligned} & \left| \Delta_n S(\mathbf{j}, \mathbf{j}')^n_t - \mathbf{s}(\mathbf{j}, \mathbf{j}') \int_0^t \gamma_s^{q''} dA_s \right| \leq K t b_n + K k'_n L_t^n + \tilde{Y}_t^n \quad \text{on } \Omega_t^n, \text{ where} \\ & \mathbb{E}(\tilde{Y}_t^n 1_{\Omega_t^n}) \leq K k'_n a_n(t, q'', \mu'' + k'_n). \end{aligned}$$

We have $|\mathbf{r}(\mathbf{j} \oplus \mathbf{j}'_{+m}) - \mathbf{r}(\mathbf{j})\mathbf{r}(\mathbf{j}')| \leq K(|m| - \mu'')^{-v}$ when $|m| > \mu''$, hence $b_n \rightarrow 0$ because $k'_n \rightarrow \infty$ and $v > 1$. We have $k'_n L_t^n \xrightarrow{\mathbb{P}} 0$ by Lemmas A.2 and A.7 as soon as $k'_n \sqrt{\Delta_n} \rightarrow 0$, so the first part of (3.22) will hold if, further, $k'_n a_n(t, q'', \mu'' + k'_n) \rightarrow 0$. These conditions are implied by (3.21), and the proof of the first part of (3.22) is complete. *Q.E.D.*

PROOF OF THEOREM 3.4 WHEN $X'' \equiv 0$: We assume $v > 1$ and (3.27). Since $\mathbb{R}^{\mathcal{J}^+}$ is endowed with the product topology, it is enough to prove the two claims for any finite family $(\mathbf{j}_1, \dots, \mathbf{j}_{d-1})$ of elements of $\mathcal{J}^+$, so we can place ourselves in the setting of (A.23) and in particular we add the empty sequence $\mathbf{j}_d = \mathbf{e}$ and we have the associated d -dimensional processes with components $G^{n,m}$ and $H^{n,m}$ (note that $G^{n,d} = 0$).

We have $B(m)_t^n \xrightarrow{\mathbb{P}} 0$ for $m = 1, 2, 3$ by (A.25) and (A.4) and $\mathbb{P}(\Omega_t^n) \rightarrow 1$ (take r close enough to 1 for the case $m = 2$). Then (A.24) and Theorem A.4 yield the $\mathcal{F}_\infty$ -stable convergence

$$(Z(\mathbf{j}_m)_t^n)_{1 \leq m \leq d} \xrightarrow{\mathcal{L}_S - \mathcal{F}_\infty} (Z(\mathbf{j}_m)_t := G_t^{(m)} + \mathbf{r}(\mathbf{j}_m) H_t^{(m)})_{1 \leq m \leq d}, \quad (\text{A.28})$$

where the $2d$ -dimensional variable $(G_t^{(m)}, H_t^{(m)})$ is, conditionally on $\mathcal{F}$, Gaussian centered with covariance given by (A.15). Note here that the covariance of (A.12) is in fact, with the notation (3.13), $a^{lm} = \mathbf{s}(\mathbf{j}_l, \mathbf{j}_m)$. Restricting (A.28) to the $d - 1$ first components, we readily deduce (a), with the limiting variable having the covariance given by (3.29).

Next, recalling $Z(\mathbf{e})_t^n = \frac{1}{\sqrt{\Delta_n}} (\Delta_n N_t^n - A_t)$, a simple calculation shows us that, for $m \leq d - 1$,

$$\bar{Z}(\mathbf{j}_m)_t^n = \frac{\sqrt{\Delta_n N_t^n}}{A_t + \sqrt{\Delta_n} Z(\mathbf{e})_t^n} \left(Z(\mathbf{j}_m)_t^n - \frac{\mathbf{r}(\mathbf{j}_m) Z(\mathbf{e})_t^n}{A_t} \int_0^t \gamma_s^{q(\mathbf{j}_m)} dA_s \right).$$

Since $\Delta_n N_t^n \xrightarrow{\mathbb{P}} A_t$, (A.28) implies the stable convergence in (b), with the limit

$$\bar{Z}(\mathbf{j})_t = \frac{1}{\sqrt{A_t}} (Z(\mathbf{j})_t - \mathbf{R}(\mathbf{j})_t Z(\beta)_t^n),$$

whose conditional variance is given by (3.30).

Q.E.D.

A.5. Proof of the Main Results When X'' Is Not Vanishing

We are left to proving Theorems 3.1, 3.3, and 3.4 when the process X'' is present. For simplicity, we focus on the behavior of $U(\mathbf{j})_t^n$, hence on Theorems 3.1 and 3.4, showing how one can reduce the case $X'' \neq 0$ to the case $X'' \equiv 0$, but the same would apply to Theorem 3.3 as well.

Toward this aim, and recalling (2.5), we set

$$Y_i^m = X_0 + X'_{T(n,i)} + \varepsilon_i^n = Y_i^n - X''_{T(n,i)}$$

(that would be the observations if X'' were identically vanishing), and we denote by $V(\mathbf{j})_t^n$ the processes defined by (3.7), upon replacing Y_i^n by Y_i^m . From what precedes, the processes $V(\mathbf{j})_t^n$ satisfy Theorems 3.1 and 3.4, so in order to obtain the results for $U(\mathbf{j})_t^n$, it clearly suffices to prove that

$$\sqrt{\Delta_n}(U(\mathbf{j})_t^n - V(\mathbf{j})_t^n) \xrightarrow{\mathbb{P}} 0. \quad (\text{A.29})$$

We write ζ_i^n for the i th summand in the definition (3.7) of $U(\mathbf{j})_t^n$, and ζ_i^m for the i th summand in the definition of $V(\mathbf{j})_t^n$. Set $M_n = \sup(|\varepsilon_i^n| : 0 \leq i \leq N_t^n)$ and recall that X and X' are bounded, so if $q = q(\mathbf{j})$, we clearly have

$$|\zeta_i^n|, |\zeta_i^m| \leq M'_n := K(1 + M_n^q).$$

On the other hand, we have $X_{t+s} - X_t = X'_{t+s} - X'_t$ as long as $[t, t+s] \cap D = \emptyset$, with the random set $D = \bigcup_{m=1} \{S_m\}$. Since ζ_i^n and ζ_i^m make use of the observations within the time interval $[T(n, i), T(n, i + w_n)]$ only with $w_n = \mu(\mathbf{j}) + (q+1)k_n$, we deduce

$$\begin{aligned} [T(n, i), T(n, i + w_n)] \cap D = \emptyset &\Rightarrow \zeta_i^n = \zeta_i^m, \\ [T(n, i), T(n, i + w_n)] \cap D \neq \emptyset &\Rightarrow |\zeta_i^n - \zeta_i^m| \leq 2M'_n, \end{aligned}$$

which yields

$$\sqrt{\Delta_n}|U(\mathbf{j})_t^n - V(\mathbf{j})_t^n| \leq 2M'_n \bar{N}_n \sqrt{\Delta_n}, \quad (\text{A.30})$$

where $\bar{N}_n$ is the number of integers i between 0 and N_t^n for which $[T(n, i), T(n, i + w_n)] \cap D \neq \emptyset$, so we are left to proving that $M'_n \bar{N}_n \sqrt{\Delta_n} \xrightarrow{\mathbb{P}} 0$.

For this, we observe that there are either w_n or $w_n + 1$ values of i such that S_m belongs to the interval $[(T(n, i), T(n, i + w_n))]$, whereas the number of finite S_m 's is at most I by (SHON-ii), hence $\bar{N}_n \leq I(w_n + 1) \leq Kk_n$. Recalling the definitions of M'_n and M_n , the boundedness of γ , the finiteness of all moments of χ_i , the independence of $\mathcal{F}^0$ and $\mathcal{G}$, and the property $N_t^n \leq (1 + Ct)/\Delta_n$ on Ω_t^n , we see that $\mathbb{E}((M'_n)^p 1_{\Omega_t^n}) \leq K_{p,t}/\Delta_n$ for any $p > 0$. Therefore, for all $a, p > 0$, we have by Markov's inequality:

$$\mathbb{P}(M'_n \bar{N}_n \sqrt{\Delta_n} > a) \leq K_{p,t} \frac{\Delta_n^{p/2-1} k_n^p}{a^p} + \mathbb{P}((\Omega_t^n)^c).$$

Since $k_n \Delta_n^\eta \rightarrow 0$ for some $\eta < 1/2$ by (3.5) or (3.27), whereas $\mathbb{P}((\Omega_t^n)^c) \rightarrow 0$, by taking above p large enough, we deduce that $\mathbb{P}(M'_n \bar{N}_n \sqrt{\Delta_n} > a) \rightarrow 0$ for all $a > 0$. Then (A.30) implies (A.29), and the proof is complete.

REFERENCES

- AÏT-SAHALIA, Y., AND J. JACOD (2014): *High-Frequency Financial Econometrics*. Princeton: Princeton University Press. [1135]
- 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. [1153]
- (2011): "Ultra High Frequency Volatility Estimation With Dependent Microstructure Noise," *Journal of Econometrics*, 160, 190–203. [1134,1152,1154]
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2000a): "Great Realizations," *Risk*, 13, 105–108. [1133]
- (2000b): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55. [1155]
- BANDI, F. M., AND J. R. RUSSELL (2008): "Microstructure Noise, Realized Variance, and Optimal Sampling," *Review of Economic Studies*, 75, 339–369. [1134,1144]
- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing Realized Kernels to Measure ex-post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76, 1481–1536. [1134]
- BLACK, F. (1986): "Noise," *The Journal of Finance*, 41 (3), 529–543. [1133]
- GLOSTEN, L. R., AND L. E. HARRIS (1988): "Estimating the Components of the Bid/Ask Spread," *Journal of Financial Economics*, 21, 123–142. [1133]
- HANSEN, P. R., AND A. LUNDE (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 127–161. [1134,1156,1157]
- HARRIS, L. (1990a): "Estimation of Stock Variance and Serial Covariance From Discrete Observations," *Journal of Financial and Quantitative Analysis*, 25, 291–306. [1133]
- (1990b): "Stock Price Clustering and Discreteness," *Review of Financial Studies*, 4, 389–415. [1133]
- HASBROUCK, J. (2007): *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford. [1133]
- JACOD, J., AND A. SHIRYAEV (2003): *Limit Theorems for Stochastic Processes* (Second Ed.). Berlin: Springer-Verlag. [1135,1166]
- JACOD, J., Y. LI, P. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and Their Applications*, 119, 2249–2276. [1134, 1151]
- JACOD, J., Y. LI, AND X. ZHENG (2015): "Estimating the Integrated Volatility With Tick Observations," Working Paper, Available at SSRN, <https://ssrn.com/abstract=2659615>. [1148,1153,1154,1163]
- (2017): "Supplement to 'Statistical Properties of Microstructure Noise'," *Econometrica Supplemental Material*, 85, <http://dx.doi.org/10.3982/ECTA13085>. [1135,1150]
- LI, Y., AND P. A. MYKLAND (2007): "Are Volatility Estimators Robust With Respect to Modeling Assumptions?" *Bernoulli*, 13, 601–622. [1151]
- LI, Y., S. XIE, AND X. ZHENG (2016): "Efficient Estimation of Integrated Volatility Incorporating Trading Information," *Journal of Econometrics*, 195, 33–50. [1156]
- LI, Y., Z. ZHANG, AND Y. LI (2015): "A Unified Approach to Volatility Estimation in the Presence of Both Rounding and Random Market Microstructure Noise," Working Paper, Available at SSRN, <https://ssrn.com/abstract=2707177>. [1153]
- LIU, L., A. J. PATTON, AND K. SHEPPARD (2015): "Does Anything Beat 5-Minute RV? A Comparison of Realized Measures Across Multiple Asset Classes," *Journal of Econometrics*, 187, 293–311. [1148]
- MADHAVAN, A. (2000): "Market Microstructure: A Survey," *Journal of Financial Markets*, 3, 205–258. [1133]
- O'HARA, M. (1995): *Market Microstructure Theory*. Oxford: Blackwell. [1133]
- PODOLSKIJ, M., AND M. VETTER (2009): "Estimation of Volatility Functionals in the Simultaneous Presence of Microstructure Noise and Jumps," *Bernoulli*, 15, 634–658. [1134]
- ROLL, R. (1984): "A Simple Implicit Measure of the Effective Bid–Ask Spread in an Efficient Market," *Journal of Finance*, 39, 1127–1139. [1133]
- STOLL, H. (2003): *Market Microstructure*. Handbook of Economics and Finance. Elsevier Science B.V. [1133]
- TSAY, R. (2002): *Analysis of Financial Time Series*. Hoboken, NJ: John Wiley & Sons, Inc. [1149]
- UBUKATA, M., AND K. OYA (2009): "Estimation and Testing for Dependence in Market Microstructure Noise," *Journal of Financial Econometrics*, 7, 106–151. [1134,1138]
- XIU, D. (2010): "Quasi-Maximum Likelihood Estimation of Volatility With High Frequency Data," *Journal of Econometrics*, 159, 235–250. [1134]
- ZHANG, L. (2006): "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12, 1019–1043. [1134,1153]

ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): “A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data,” *Journal of the American Statistical Association*, 100, 1394–1411. [1134,1144]

Co-editor Lars Peter Hansen handled this manuscript.

Manuscript received 29 December, 2014; final version accepted 13 February, 2017; available online 2 March, 2017.