8-2021

Glivenko-Cantelli Theorems for integrated functionals of stochastic processes

Jia LI
Singapore Management University, jiali@smu.edu.sg

Congshan ZHANG
Duke University

Yunxiao LIU
University of North Carolina at Chapel Hill

GLIVENKO-CANTELLI THEOREMS FOR INTEGRATED FUNCTIONALS OF STOCHASTIC PROCESSES

By Jia Li and Congshan Zhang and Yunxiao Liu

Duke University and University of North Carolina at Chapel Hill

We prove a Glivenko–Cantelli theorem for integrated functionals of latent continuous-time stochastic processes. Based on a bracketing condition via random brackets, the theorem establishes the uniform convergence of a sequence of empirical occupation measures towards the occupation measure induced by underlying processes over large classes of test functions, including indicator functions, bounded monotone functions, Lipschitz-in-parameter functions, and Hölder classes as special cases. The general Glivenko–Cantelli theorem is then applied in more concrete high-frequency statistical settings to establish uniform convergence results for general integrated functionals of the volatility of efficient price and local moments of microstructure noise.

1. Introduction. There is now a large and burgeoning literature on the estimation of integrated functionals of stochastic processes based on high-frequency data. The functional of interest takes the form

$$\mathbb{F}g \equiv \int_0^1 g(V_s)ds \tag{1.1}$$

for some latent continuous-time process V and some test function $g(\cdot)$. The most prominent example is integrated volatility functionals, where V is the stochastic volatility (or, more generally, spot covariance matrix) of a semimartingale process. Special cases include the integrated variance-covariance matrix ([8], [9], [11], [5], [12]), integrated betas ([10], [35], [36], [40], [27]), correlation/leverage effects ([25]), idiosyncratic variance ([29]), volatility Laplace transforms ([41]), volatility occupation times ([26]), and eigenvalues ([4]). In a more complicated setting in which the underlying semimartingale is contaminated with the so-called microstructure noise, the latent process V may also include the other processes such as the stochastic variance of microstructure noise. In applications, the integrated functional $\mathbb{F}g$ can be used to directly measure risk or to construct criterion functions

AMS 2000 subject classifications: 60F17, 60G44, 60G57, 62G05

Keywords and phrases: Glivenko–Cantelli, high-frequency data, spot volatility, microstructure noise, occupation measure.

for extremum estimation and specification test ([30]), and it may also appear as the asymptotic variance of another estimator that needs to be estimated for conducting feasible inference.

The estimation of integrated functionals can be viewed equivalently based on the notion of occupation measure ([26]). Indeed, equation (1.1) also defines $\mathbb{F}$ as the *occupation measure* induced by the process V: With $g(\cdot) = 1_A(\cdot)$ for some subset A, $\mathbb{F}g$ reports the amount of time which the process V spends in A. The occupation measure is thus a natural "realized" analogue of the (conventional) probability distribution P, with $P(A)$ denoting the probability that V_t falls in A. In conventional statistical settings, a law of large numbers may be written as $P_n g \xrightarrow{\mathbb{P}} Pg$, where P_n denotes the empirical measure ([42]). In the same fashion, we can construct an *empirical occupation measure* $\mathbb{F}_n$ such that

$$\mathbb{F}_n g \xrightarrow{\mathbb{P}} \mathbb{F}g. \tag{1.2}$$

As such, a convergence result for an integrated functional for some $g(\cdot)$ function can be understood as the pointwise convergence of $\mathbb{F}_n$ towards $\mathbb{F}$ at the "point" $g(\cdot)$.

A theoretical question naturally arises from this perspective: Does the pointwise convergence (1.2) holds uniformly for g ranging over a general class $\mathcal{G}$ of test functions (e.g., all bounded monotone functions)? A similar question has been well studied by various Glivenko–Cantelli theorems in the classical empirical process literature, but it remains to be largely open for occupation measures. The general lesson from the classical theory is that the uniform convergence holds if the class $\mathcal{G}$ has "restricted complexity," which can be more precisely stated in terms of covering or bracketing numbers.

Set against this background, we aim to establish a general Glivenko–Cantelli theorem for the occupation measure, as a first effort for bridging the gap between the two literatures on high-frequency data and empirical processes. Specifically, we consider a "plug-in" type estimator $\mathbb{F}_n$ given by

$$\mathbb{F}_n g \equiv \int_0^1 g(\widehat{V}_s)ds, \tag{1.3}$$

where $\widehat{V}$ is a preliminary nonparametric estimator of the latent process V. We show (see Theorem 1) that $\mathbb{F}_n g$ converges in probability to $\mathbb{F}g$ uniformly over a class $\mathcal{G}$ under two high-level conditions: (i) a bracketing condition stating that $\mathcal{G}$ can be covered by a finite number of brackets with any given size; and (ii) $\mathbb{F}_n f \xrightarrow{\mathbb{P}} \mathbb{F}f$ for each bracket function f.

Our result resembles the classical Glivenko–Cantelli theorem for empirical measures based on bracketing, but there are three important distinctions, which make our analysis notably different from prior work (cf. [42]). First, unlike the (nonrandom) probability measure P, the occupation measure $\mathbb{F}$ is itself a random quantity. As a result, the aforementioned bracketing condition cannot always be verified using (only) deterministic bracket functions as in the classical setting. We address this issue by introducing *random brackets*, accompanied by a novel notion of bracketing condition, which can be verified for broad classes of test functions (e.g., bounded monotone functions, Hölder class, and Lipschitz-in-parameter class) in the present setting.

Second, in order to accommodate the use of random brackets, we establish (see Theorem 2) a general pointwise convergence result $\mathbb{F}_n f \xrightarrow{\mathbb{P}} \mathbb{F} f$, allowing the test function f to be random. This more general pointwise convergence result is needed for applying our Glivenko–Cantelli theorem despite the fact that each test function $g \in \mathcal{G}$ is deterministic.

Third, even for deterministic test functions, the pointwise convergence of $\mathbb{F}_n$ is more complicated than that of the classical empirical measure: While the latter follows simply from a law of large numbers for sample averages, the former concerns integrated functionals of the nonparametric estimator $\widehat{V}$ (recall (1.3)). Of course, when V is the spot variance-covariance matrix process, the pointwise convergence of $\mathbb{F}_n$ is now relatively well understood in the literature on integrated volatility functionals ([23], [24], [26], [30], [27]). However, these papers focus exclusively on volatility estimation in the basic setting without microstructure noise. In the presence of noise, V involves not only the volatility but also local moments of the observation noise, and requires a more deliberate method of estimation. To the best of our knowledge, the (pointwise) estimation theory for the integrated functional with a general test function $g(\cdot)$ has not been considered in the literature on noisy high-frequency data.

To accommodate noisy data, we further establish a convergence result for integrated functionals with general test functions, allowing for the presence of (a certain type of) microstructure noise. Specifically, we adopt the pre-averaging approach ([19], [38], [21]) to construct a nonparametric spot estimator $\widehat{V}$ that contains a noise-robust estimator for the efficient price's spot volatility and an estimator of the (time-varying and stochastic) spot variance of microstructure noise, and we show that the associated empirical occupation measure $\mathbb{F}_n$ satisfies the pointwise convergence $\mathbb{F}_n f \xrightarrow{\mathbb{P}} \mathbb{F} f$ for general (possibly random and discontinuous) test function f. We then "upgrade" this pointwise convergence to a uniform version using our general Glivenko–Cantelli theorem; see Section 3.2 for details. As an intermediate step of this analysis, we develop a uniform approximation result for the pre-averaging nonparametric spot estimators; this appears to be new to the noise literature, and may be useful in other types of analysis as well.

The present paper is closely related to two strands of the high-frequency econometrics/statistics literature. The first is the literature on the estimation of general integrated volatility functionals. Theorem 9.3.2 of [23] provides a pointwise convergence result for continuous test functions with polynomial growth. The polynomial-growth condition is restrictive as it rules out many commonly used test functions (e.g., beta, correlation, and idiosyncratic variance). [27] relaxes this restriction based on a uniform approximation argument. Some uniform convergence results have appeared in the literature, but only for very narrow classes of test functions. For example, [26] studies the estimation of volatility occupation time, for which the test functions are indicator functions; [30] and [29] study the uniform estimation of asymptotic conditional variance-covariance functions of a mixed Gaussian process (indexed by a scalar) in order to conduct Bierens-type nonparametric specification tests. In contrast, the uniformity of our Glivenko–Cantelli theorem holds for substantially broader classes of test functions. Moreover, we also study the estimation in the presence of microstructure noise, which is not considered in the aforementioned prior work.

Meanwhile, the extant literature on noisy high-frequency data has mostly focused on the estimation of integrated variance or covariance matrix ([48], [46], [31], [3], [47], [6], [7], [19], [22], [38], [16], [18], [21], [13], [44]), which is a special case of $\mathbb{F}g$ with g being the identity function (i.e., $g(x) = x$). The asymptotic variance of these estimators are also integrated functionals, typically taking form as integrated integer polynomials of the spot volatility of the efficient price and the spot variance of the noise. To the best of our knowledge, the present paper appears to be the first to establish the pointwise estimation of general integrated functionals in the noisy setting, while our main contribution, namely the Glivenko–Cantelli theorems, clearly set our theory further apart from the extant literature on microstructure noise.

Finally, this paper is clearly inspired by the classical empirical process literature (see, e.g., [42]), but it concerns a very different non-stationary non-ergodic infill asymptotic setting involving latent processes. In addition, an interesting novelty of our theory is the use of random brackets, which stem naturally from the stochastic nature of the occupation measure. Our Glivenko–Cantelli theorem is a necessary first step for a more complete development of empirical process theory (e.g., Donsker theorems) in the high-frequency setting.

The rest of the paper is organized as follows. Section 2 presents our generic Glivenko–Cantelli theorem. Section 3 specializes the general theorem in more concrete settings with and without microstructure noise, and provides an empirical application. Section 4 concludes. All proofs are given in Section 5.

Notation: We use $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{\mathbb{P}^*}$ to denote convergence in probability and convergence in outer probability, respectively. For two real-valued sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$, we write $a_n \asymp b_n$ if there exists some constant $C \geq 1$, such that $a_n/C \leq b_n \leq Ca_n$. For any real number a, $\lfloor a \rfloor$ (resp. $\lceil a \rceil$) denotes the greatest (resp. smallest) integer smaller (resp. larger) than a. All limits are for $n \to \infty$.

2. A Glivenko–Cantelli theorem for integrated functionals.

This section presents our Glivenko–Cantelli theorem for empirical occupation measures. After introducing the setting in Section 2.1, we present our generic Glivenko–Cantelli theorem in Section 2.2. This result is based on a bracketing condition via random brackets, and a pointwise convergence condition for the brackets. The latter condition is studied in detail in Section 2.3.

2.1. *Empirical occupation measure.*

Consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, on which a generic continuous-time stochastic process V_t is defined over a *fixed* time interval normalized to be $[0, 1]$. The process V_t is càdlàg, adapted, and takes values in an Euclidean space $\mathcal{V}$. We are interested in situations where V_t is not directly observable, but needs to be (nonparametrically) estimated. In applications, V_t typically plays the role of the spot variance of a financial asset, or the spot covariance matrix of multiple assets. In more general settings, V_t may include the local moment processes of other quantities such as the observational error on asset price (i.e., microstructure noise), trading duration, volume, bid-ask spread, etc.

Much attention of the high-frequency econometrics and statistics literature has been devoted to estimating integrated functionals of the form

$$\int_0^1 g\left(V_s\right) ds \tag{2.1}$$

for some measurable function $g : \mathcal{V} \mapsto \mathbb{R}$. The most prominent example is the case when V_t is the spot variance of an asset price process, and the resulting integrated volatility functionals are of direct interest as measures of financial risks. In addition, these functionals are more generally relevant for statistical inference, in that the estimators' asymptotic variances are also integrated functionals. The latter type of functionals often have more complicated forms, especially for estimators constructed using noisy high-frequency data.

The study of integrated functionals in fact concerns the occupation measure. Specifically, we recall (see [17] and [26]) that the occupation measure

$\mathbb{F}$ induced by V_t over the $[0,1]$ time interval is defined as

$$\mathbb{F}(A) \equiv \int_0^1 1_{\{V_s \in A\}} ds,$$

for any Borel subset $A \subseteq \mathcal{V}$, which records the amount of time spent by the V process in the set A. The occupation measure is thus evidently the "realized" analogue of the probability distribution of V_t. With an appeal to basic integration theory, we can equivalently view the occupation measure $\mathbb{F}$ as a linear functional that acts on measurable test functions, that is,

$$\mathbb{F}g \equiv \int_{\mathcal{V}} g(x)\mathbb{F}(dx) = \int_0^1 g(V_s)ds. \tag{2.2}$$

To construct the empirical occupation measure, we suppose that a nonparametric estimator $\widehat{V}_t$ of the V_t process is available, satisfying the following assumption.

Assumption 1. *$\widehat{V}_t \xrightarrow{\mathbb{P}} V_t$ for Lebesgue almost every $t \in (0,1)$.*

Assumption 1 requires $\widehat{V}_t$ to be a consistent "spot" estimator of V_t for almost every $t \in (0,1)$. Although this condition is high-level in nature, it is relatively mild. When V_t is the spot covariance of a multivariate semimartingale, the spot estimation theory is well known under very general conditions in the case without microstructure noise; see Theorem 9.3.2 of [23] for a general result. Section 3.2 presents similar results in a setting with noise.

Equipped with this spot estimator, we can naturally define the corresponding empirical occupation measure $\mathbb{F}_n$ as follows:

$$\mathbb{F}_n g \equiv \int_0^1 g(\widehat{V}_s)ds. \tag{2.3}$$

A relatively basic problem concerns the pointwise consistency of $\mathbb{F}_n$, that is, for a fixed g function,

$$\mathbb{F}_n g \xrightarrow{\mathbb{P}} \mathbb{F}g.$$

Going one step further, our main theoretical goal in this paper is to address the uniform convergence problem in a general setting. Specifically, we aim to establish a Glivenko–Cantelli theorem of the form

$$\sup_{g \in \mathcal{G}} |\mathbb{F}_n g - \mathbb{F}g| \xrightarrow{\mathbb{P}^*} 0,$$

where the supremum is taken over a large class $\mathcal{G}$ of test functions. As in [42], we state the convergence under the outer probability $\mathbb{P}^*$ to circumvent measurability issues.

Before diving into the formal results, it is instructive to illustrate the usefulness of the Glivenko–Cantelli theorem in three types of applied scenarios.

Example (Risk Measurement). The integrated functional $\mathbb{F}g$ may serve as a measure of risk in financial applications. One case in point is the volatility occupation time studied by [26]. In that case, V is the stochastic volatility of an asset, and $\mathcal{G}$ consists of test functions of the form $g(x) = 1_{\{x \leq c\}}$ indexed by the constant $c > 0$. The volatility occupation time complements the popular integrated volatility measure $\int_0^1 V_s ds$ by providing additional "distributional" information regarding the stochastic volatility process. In a more general multivariate case, we may take V to be the spot beta process of an asset with respect to the market portfolio, which is a conventional measure of systematic risk and is formally defined as the ratio between their spot covariance and the market portfolio's spot variance. The beta occupation time is then informative about the firm's time-varying systematic risk, which we will further illustrate in the empirical application (Section 3.3). In the same vein, one can study a firm's idiosyncratic risk, asymmetric information, and liquidity by setting V to be the corresponding process that can be formed as transformations of assets' spot covariance matrix and instantaneous moments of microstructure noise. □

Example (M-Estimation). Establishing the consistency of an M-estimator often relies on the uniform convergence of the criterion function. In this scenario, $\mathcal{G}$ typically collects the criterion function indexed by a finite-dimensional parameter θ. To fix ideas, consider a nonlinear model $V_{1,t} = h(V_{2,t}; \theta)$, and set $V_t = (V_{1,t}, V_{2,t})$. For instance, $V_{1,t}$ may be the price of an option contract, $V_{2,t}$ may be a vector consisting of the state variables in the option pricing model such as the price of the underlying asset and its stochastic volatility, and $h(\cdot)$ specifies the pricing formula parameterized by θ, which in turn governs the risk-neutral dynamics of the state processes. In this case, V_t is not directly observable for two reasons: The stochastic volatility is latent, and the "efficient" prices of the option contract and the underlying asset may be contaminated by microstructure noise (which is particularly relevant for options). The estimation of θ can be carried out by minimizing the sample version of $\int_0^1 (V_{1,s} - h(V_{2,s}; \theta))^2 ds$, which corresponds to $\mathbb{F}g$ with $g(x_1, x_2) = (x_1 - h(x_2; \theta))^2$ and, correspondingly, the family $\mathcal{G}$ collects all such test functions indexed by θ over the parameter space. More generally, the θ parameter may be a function itself, giving rise to a nonparametric M-estimation problem. □

Example (Feasible Functional Inference). The sampling variability of a functional estimator is often captured by a limiting Gaussian process with unknown (conditional) covariance function, typically taking a much more complicated form than the original functional estimator of interest. The associated feasible inference requires the uniform consistent estimation of the asymptotic covariance function. One case in point is the nonparametric specification test of [29] for testing the constancy of smooth transformations of the spot covariance matrix (e.g., volatility level, beta, idiosyncratic variance, correlation, eigenvalue, variance beta, etc.), which is based on a functional central limit theorem for a continuum of integrated volatility functionals. The Glivenko–Cantelli theorem can be used to design the estimator for the asymptotic covariance function and provides a generic tool for justifying the estimator's theoretical validity. □

2.2. *A Glivenko–Cantelli theorem based on random brackets.* Our theory is inspired by the classical empirical process theory based on bracketing (see Theorem 2.4.1 in [42]). In order to explain clearly the similarity and the difference between these results, it is instructive to briefly review the essentials of the classical theory first. Consider an i.i.d. sequence of random variables $(Z_i)_{1\leq i\leq n}$ and let P and P_n denote the corresponding probability measure and empirical measure, respectively. For each P-integrable measurable function g, we denote

$$Pg \equiv \int g\left(z\right) P\left(dz\right), \quad P_n g \equiv \frac{1}{n}\sum_{i=1}^{n} g\left(Z_i\right).$$

The pointwise convergence $P_n g \xrightarrow{\mathbb{P}} Pg$ for a given g function follows directly from a conventional law of large numbers. The corresponding Glivenko–Cantelli theorem has the form

$$\sup_{g\in\mathcal{G}} \left|P_n g - Pg\right| \xrightarrow{\mathbb{P}^*} 0. \tag{2.4}$$

To obtain this uniform convergence, one approach is to restrict the complexity of $\mathcal{G}$ using a notion of bracketing. Specifically, for a constant $\varepsilon > 0$, a pair of functions $(l(\cdot), u(\cdot))$ is said to form an ε-bracket if $l(\cdot) \leq u(\cdot)$ and $Pu - Pl \leq \varepsilon$. If for any $\varepsilon > 0$, there exists a finite collection of ε-brackets given by $\{(l_j, u_j) : 1 \leq j \leq N_\varepsilon\}$ such that each $g \in \mathcal{G}$ satisfies $l_j(\cdot) \leq g(\cdot) \leq u_j(\cdot)$ for some j, then the uniform convergence in (2.4) holds.

Our theory is also based on a bracketing condition—which is described in Assumption 2—but with an important difference. Unlike the (non-random) probability measure P in the conventional setting, the occupation measure $\mathbb{F}$ considered here is a random measure. Consequently, we need to consider random bracket functions in certain applications as illustrated below. In the sequel, a function f is called a random function on $\mathcal{V}$ if it is a (jointly) measurable function defined on the product space $\Omega \times \mathcal{V}$. We then define $\mathbb{F}f$ as $\mathbb{F}f(\omega) \equiv \int f(\omega, x)\, \mathbb{F}_\omega(dx)$, $\omega \in \Omega$, where $\mathbb{F}_\omega$ is the realization of $\mathbb{F}$ on the sample path ω.

ASSUMPTION 2. *For each constant $\varepsilon > 0$, there exist a finite constant N_ε and a set of (possibly random) bracket functions $\{(l_j, u_j) : 1 \leq j \leq N_\varepsilon\}$, such that for each $g \in \mathcal{G}$ and $\omega \in \Omega$, we can find constants $j, k \in \{1, \ldots, N_\varepsilon\}$ that may depend on ω and satisfy $l_j(\omega, \cdot) \leq g(\cdot) \leq u_k(\omega, \cdot)$ and*

$$\int \left(u_k(\omega, x) - l_j(\omega, x)\right) \mathbb{F}_\omega(dx) \leq \varepsilon.$$

Assumption 2 is more flexible than the aforementioned conventional bracketing condition in two related ways. One is to allow the bracket functions to be random. The other is that the brackets for each test function g is allowed to vary across different sample paths (i.e., j and k may depend on ω). In particular, the lower and upper bracket functions l_j and u_k are allowed to be "mixed and matched" to cover the test functions. Such flexibility is particularly useful in our proof of the uniform convergence over bounded monotone functions (see Theorem 3 below).

We are now ready to state our generic Glivenko–Cantelli theorem for empirical occupation measures.

THEOREM 1. *Suppose that (i) Assumption 2 holds, and (ii) $\mathbb{F}_n f \xrightarrow{\mathbb{P}} \mathbb{F}f$ for each bracket function $f \in \{l_j, u_j : 1 \leq j \leq N_\epsilon\}$. Then, we have*

$$\sup_{g \in \mathcal{G}} |\mathbb{F}_n g - \mathbb{F}g| \xrightarrow{\mathbb{P}^*} 0.$$

Theorem 1 relies on two sufficient conditions: the bracketing condition and the pointwise convergence for bracket functions. In the remaining part of this subsection, we illustrate how to verify the bracketing condition for general classes of test functions. Meanwhile, the pointwise convergence for the bracket function introduces an interesting complication. That is, although the original test functions (i.e., g) are deterministic, we nonetheless need to consider random test functions (i.e., f) as well. To our knowledge, a general theory for $\mathbb{F}_n f \xrightarrow{\mathbb{P}} \mathbb{F}_n f$ with f being random is not available in the literature. For this reason, we provide such a result in Section 2.3 below. In Section 3, we further specialize Theorem 1 in more concrete settings under primitive conditions.

Turning to examples for Assumption 2, we start with the class of indicator functions of the form $g(\cdot) = 1_{\{\cdot \leq x\}}$. When V_t is scalar-valued, $\mathbb{F}g = \int_0^1 1_{\{V_s \leq x\}} ds$ reports the time spent by the process V below a certain level x. This is the occupation time studied by [26], which we denote by $\mathbb{F}(x)$ for simplicity. These test functions satisfy Assumption 2 as illustrated by Example 1 below.

Example 1 (Indicator Functions). Consider a class of indicator functions $\mathcal{G}_I = \left\{1_{(-\infty,x]}(\cdot) : x \in \mathbb{R}\right\}$. To construct the brackets for a given $\varepsilon > 0$, let x_j be the $(j\varepsilon \wedge 1)$-quantile of $\mathbb{F}(\cdot)$. Observe that $\mathbb{F}(x_j) - \mathbb{F}(x_{j-1}) \leq \varepsilon$, for $j = 1, \ldots, N_\varepsilon$, where $N_\varepsilon = \lceil 1/\varepsilon \rceil$ is finite. Then Assumption 2 can be verified by setting the bracket functions as $l_1(\cdot) = 0$ and $l_{j+1}(\cdot) = u_j(\cdot) = 1_{(-\infty,x_j]}(\cdot)$ for $j \geq 1$. □

Example 1 is of particular interest because it directly resembles the original Glivenko–Cantelli problem regarding cumulative distribution functions, as the occupation time $\mathbb{F}(x)$ is exactly the "realized" analogue of the distribution function $\mathbb{P}(V_t \leq x)$. This example also demonstrates how to verify the bracketing condition by adapting classical results from the empirical process literature. Meanwhile, we also see that random brackets naturally arise from this context because quantiles of the random measure $\mathbb{F}$ are also random, resulting in random brackets. Indicator functions are special cases of the class of bounded monotone functions, which also satisfy Assumption 2 as we now illustrate.

Example 2 (Bounded Monotone Functions). Consider a collection of bounded monotone functions

$$\mathcal{G}_M \equiv \{\text{monotone functions } g : \mathcal{V} \mapsto [0,1]\},$$

where we have normalized the range of g to be $[0,1]$. Obviously, $\mathcal{G}_I \subset \mathcal{G}_M$. In the classical setting, Theorem 2.7.5 in [42] shows that, for every $\varepsilon > 0$, the number of ε-brackets needed to cover $\mathcal{G}_M$ is bounded by $\exp\left(K\varepsilon^{-1}\right)$ for some universal constant K and is thus finite. Although this result cannot be directly used to verify Assumption 2, the underlying method can be adapted to construct random brackets in the current setting. The exact construction is somewhat technical (especially when the occupation time

$\mathbb{F}(x)$ is not strictly increasing), and we provide the details in the proof of Theorem 3 below. □

Note that monotone functions in the above examples do not need to be smooth, as the indicator functions are in fact discontinuous. Large classes of smooth functions also satisfy the bracketing condition, as demonstrated in the following two examples.

EXAMPLE 3 (LIPSCHITZ-IN-PARAMETER CLASS). Let $\Theta \subseteq \mathbb{R}^q$ be a compact parameter space. We consider functions of the form $g : \mathcal{V} \times \Theta \mapsto \mathbb{R}$, and a collection $\mathcal{G}_L = \{g(\cdot, \theta) : \theta \in \Theta\}$. We further assume that $g(x, \theta)$ is continuous in x for each $\theta \in \Theta$ and, for any compact set $\mathcal{K} \subseteq \mathcal{V}$, there exists a finite constant $M_{\mathcal{K}} > 0$, such that for all $\theta, \theta' \in \Theta$,

$$\sup_{x \in \mathcal{K}} \left| g(x, \theta) - g(x, \theta') \right| \leq M_{\mathcal{K}} \left\| \theta - \theta' \right\|. \tag{2.5}$$

Since Θ is bounded, for each $\varepsilon > 0$ we can select a finite number of points $\theta_1, \ldots, \theta_{N_\varepsilon}$ from Θ such that each $\theta \in \Theta$ satisfies $\|\theta - \theta_j\| \leq \varepsilon/(2M_{\mathcal{K}})$ for some j. Bracket functions can then be constructed as

$$l_j(x) = g(x, \theta_j) - \varepsilon/2, \quad u_j(x) = g(x, \theta_j) + \varepsilon/2.$$

These functions are deterministic and satisfy $\mathbb{F}u_j - \mathbb{F}l_j \leq \varepsilon$. If V_t takes values in a compact set $\mathcal{K}$, it is also easy to see that $\mathcal{G}_L$ can be covered by these brackets because of the Lipschitz condition. □

EXAMPLE 4 (HÖLDER CLASS). Let $\mathcal{G}_H$ be a class of real-valued functions on $\mathcal{V}$ satisfying the following condition: for any compact subset $\mathcal{K}$ of $\mathcal{V}$, there exist constants $\alpha_{\mathcal{K}} \in (0, 1]$ and $M_{\mathcal{K}} > 0$, such that

$$\sup_{x, y \in \mathcal{K}, x \neq y} \frac{|g(x) - g(y)|}{\|x - y\|^{\alpha_{\mathcal{K}}}} \leq M_{\mathcal{K}}.$$

That is, the functions in $\mathcal{G}$ are uniformly $\alpha_{\mathcal{K}}$-Hölder continuous on each $\mathcal{K}$. If V_t takes value in some compact set $\mathcal{K}$, then we can apply Corollary 2.7.2 of [42] to show that for every $\varepsilon > 0$, there exists deterministic bracket functions $\{(l_j, u_j) : 1 \leq j \leq N_\varepsilon\}$ covering $\mathcal{G}_H$, with $\sup_{x \in \mathcal{K}}(u_j(x) - l_j(x)) \leq \varepsilon$ and $N_\varepsilon \leq \exp\left(K\varepsilon^{-d/\alpha_{\mathcal{K}}}\right)$ for some constant K. It follows that $\mathbb{F}u_j - \mathbb{F}l_j \leq \varepsilon$ on all sample paths. Assumption 2 is then verified. □

A few remarks on Example 3 and Example 4 are in order. First, Example 3 concerns functions that are smoothly (in the Lipschitz sense) parametrized by a parameter θ. The Lipschitz condition (2.5) is easy to verify because $\mathcal{K}$ and Θ are compact. Example 4 concerns a nonparametric family of test functions, which is not parametrized by any finite-dimensional parameter. The latter example only requires the functions to be Hölder continuous in the state variable. This is a mild requirement in that the Hölder index can be arbitrarily small.

Second, for illustration purpose, we have assumed in these examples that the process V_t takes value in a compact subset $\mathcal{K}$. This condition is seemingly restrictive: For example, a stochastic volatility process typically takes values in $(0, \infty)$ rather than in any fixed compact set. That said, we actually only need this condition to hold locally in time when deriving limit theorems, thanks to the standard localization procedure (see Section 4.4.1 of [23]). Theorems 4 and 6, below, provide concrete uniform convergence results under the latter more general condition.

2.3. *Pointwise convergence for random test functions.* Under the bracketing condition, Theorem 1 reduces the Glivenko–Cantelli problem into a pointwise convergence for bracket functions. The examples in Section 2.2 show that the bracket functions may be random and/or discontinuous. For this reason, we establish in this subsection a pointwise convergence result

$$\mathbb{F}_n f \xrightarrow{\mathbb{P}} \mathbb{F} f$$

while allowing the test function f to be random and discontinuous. That said, we do not restrict f to be a bracket function described in Section 2.2, so this pointwise convergence result is of its own independent theoretical interest. We impose the following condition on the test function f.

Assumption 3. *(i) For Lebesgue almost every $t \in [0,1]$, f is almost surely continuous at V_t, that is, $\mathbb{P}(\{\omega : f(\omega, \cdot)$ is continuous at $V_t(\omega)\}) = 1$.*

(ii) For a deterministic function $F : \mathcal{V} \mapsto \mathbb{R}_+$, $|f(\omega, x)| \leq F(x)$ all $\omega \in \Omega$ and $x \in \mathcal{V}$. Moreover, $\sup_{x \in \mathcal{K}} F(x) < \infty$ for any compact subset $\mathcal{K} \subseteq \mathcal{V}$.

Condition (i) of Assumption 3 imposes a continuity requirement on f, while allowing for a form of discontinuity. Consider Example 1 for illustration, where $f(\cdot) = 1_{(-\infty, x_j]}(\cdot)$ and x_j is a quantile of the occupation time $\mathbb{F}(x)$. Recall that $\mathbb{F}$, x_j, and $f(\cdot)$ are all random. Under the assumption that the occupation time $x \mapsto \mathbb{F}(x)$ is continuous almost surely, we have

$$\int_0^1 \mathbb{P}\left(\{\omega : f(\omega,\cdot) \text{ is discontinuous at } V_s(\omega)\}\right) ds$$
$$= \int_0^1 \mathbb{P}\left(V_s = x_j\right) ds = \mathbb{E}\left[\int_0^1 1_{\{V_s = x_j\}} ds\right] = \mathbb{E}\left[\mathbb{F}(x_j) - \mathbb{F}(x_j -)\right] = 0,$$

which implies condition (i).

Assumption 3(ii) introduces a deterministic envelope F for the random function f. If f is a bounded function (such as the bracket functions in Examples 1 and 2), F can be simply taken as a constant. The envelope becomes more relevant when f is not uniformly bounded. In this case, we require F to be bounded on compact sets. This requirement is mild because we do not need F to be bounded over the entire state space. Specifically, in Examples 3 and 4, each bracket function f can be chosen to be a continuous deterministic function, and we can simply set $F = |f|$. In this case, $\sup_{x\in\mathcal{K}} F(x) < \infty$ is ensured by the continuity of F because $\mathcal{K}$ is compact.

In order to accommodate general test functions (particularly without the polynomial growth restriction used by [23] and [24]), we impose the following local compactness condition, which permits the use of spatial localization as in [30] and [27].

Assumption 4. *There exist a sequence of stopping times $(T_m)_{m\geq 1}$ increasing to infinity and a sequence of compact subsets $\mathcal{K}_m \subseteq \mathcal{V}$ such that for each $m \geq 1$, $V_t, \widehat{V}_t \in \mathcal{K}_m$ for all $t \leq T_m \wedge T$ with probability approaching one.*

By localization, Assumption 4 allows us to assume that V_t and $\widehat{V}_t$ take values in a compact set $\mathcal{K}$ without loss of generality when deriving limit theorems. It has two requirements. The first concerns the pathwise regularity of the V_t process, namely, it is locally compactly valued. This condition is mild and easy to verify. For example, if V_t is the spot volatility process taking values in $\mathcal{V} = (0,\infty)$, this condition is satisfied if both V_t and $1/V_t$ are locally bounded. In this case, we can find a localizing sequence of stopping times $(T_m)_{m\geq 1}$ such that $V_t \in [2/m, m]$ for $t \leq T_m$.

The second requirement is high-level in nature, that is, the spot estimator $\widehat{V}_t \in \mathcal{K}_m$ for all $t \leq T_m \wedge T$, with probability approaching one. A sufficient condition is the following uniform convergence:

$$\sup_{t\in[0,T]} \|\widehat{V}_t - V_t\| = o_p(1), \tag{2.6}$$

which then implies that $\widehat{V}_t$ falls in a small enlargement of the range of V_t. Specifically, if $V_t \in [2/m, m]$ as in the aforementioned example, then we can set $\mathcal{K}_m = [1/m, m+1]$, which verifies Assumption 4. It should be noted that the uniform approximation in (2.6) typically does not hold when the process V_t contains jumps; however, for the purpose of verifying Assumption 4, (2.6) is not necessary, either. In more general settings, we often can still show that

$$\sup_{t\in[0,T]} \|\widehat{V}_t - \overline{V}_t\| = o_p(1), \tag{2.7}$$

where $\overline{V}_t$ is defined as a moving average of the V_t process, which is also (locally) compactly valued. The above argument can then be adapted straightforwardly to verify Assumption 4. The results in Section 3 provide more details.

We are now ready to state the pointwise convergence for general random test functions.

THEOREM 2. *Under Assumptions 1, 3, and 4,* $\mathbb{F}_n f \xrightarrow{\mathbb{P}} \mathbb{F} f$.

Theorem 2 is related to several known results on the pointwise convergence of integrated volatility functionals. For example, Theorem 9.4.1 of [23] and Theorem 3 of [27] establish results for continuous deterministic test functions, and Lemma 1 of [26] concerns bounded deterministic test functions that may be discontinuous. Theorem 2 here is more general in that it can be applied not only to those test functions, but also allows them to be random, which is needed for verifying the high-level condition of our Glivenko–Cantelli theorem.

3. Applications.

In this section, we apply the general theory developed in Section 2 to a range of statistical applications on high-frequency data. Section 3.1 collects results for the uniform estimation of integrated volatility functionals. Section 3.2 provides further extensions by incorporating microstructure noise. Section 3.3 presents an empirical application.

3.1. *Glivenko–Cantelli theorems for integrated volatility functionals .*

The estimation of spot covariance process and the related integrated volatility functionals are relatively well understood in the literature. We first demonstrate in this familiar context how to use our general Glivenko–Cantelli theorem (Theorem 1) to establish new uniform convergence results over general classes of test functions. We do so under primitive conditions commonly seen in the high-frequency econometrics and statistics literature.

Consider a d-dimensional Itô semimartingale X_t of the form

$$
\begin{aligned}
X_t &= x_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s \\
&\quad + \int_0^t \int_{\mathbb{R}} \delta(s,x) 1_{\{\|\delta(s,x)\| \le 1\}} (\mu - \nu)(ds, dx) \\
&\quad + \int_0^t \int_{\mathbb{R}} \delta(s,x) 1_{\{\|\delta(s,x)\| > 1\}} \mu(ds, dx),
\end{aligned} \tag{3.1}
$$

where the drift b_t takes values in $\mathbb{R}^d$, the spot volatility matrix σ_t takes values in $\mathbb{R}^{d \otimes d}$, W is a d-dimensional standard Brownian motion, $\delta : \Omega \times \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}^d$ is a predictable function, μ is a Poisson random measure on $\Omega \times \mathbb{R}$, with its compensator $\nu(ds, dx) = ds \otimes \lambda(dx)$ for some σ-finite measure $\lambda(\cdot)$. In practical applications, X_t models the (log) price process of d financial assets. The process of interest in this context is the spot covariance process defined as $c_t = \sigma_t \sigma_t^\top$, which takes values in an open set $\mathcal{C}$ (e.g., the collection of positive definite matrices). For ease of discussion, all processes are assumed to be càdlàg and adapted. We further impose a few (mild) regularity conditions.

Assumption 5. *Let $r \in [0,2)$ be a constant. The process X is an Itô semimartingale given by (3.1). The processes b_t and σ_t are locally bounded. There exists a sequence of stopping times $(T_m)_{m \ge 1}$ increasing to infinity such that for each $m \ge 1$, all $\omega \in \Omega$, and $t \le T_m(\omega)$:*

(i) $\|\delta(\omega, t, z)\|^r \le D_m(z)$, where $D_m(\cdot)$ is a sequence of nonnegative bounded λ-integrable functions on $\mathbb{R}$;

(ii) $c_t \in \mathcal{K}_m$, where $\mathcal{K}_m$ is a sequence of compact convex subsets of $\mathcal{C}$.

We consider an infill asymptotic framework in which X is sampled at discrete times $i\Delta_n$, $i = 0, 1, \ldots, n$ within a fixed time span normalized to be $[0,1]$, where $\Delta_n = 1/n$ goes to zero asymptotically. The spot covariance process can be estimated using a truncated local realized covariance estimator. Specifically, we choose an integer sequence k_n of local windows and a sequence u_n of truncation thresholds that satisfy (with r described in Assumption 5):

$$
k_n \asymp n^\gamma, \quad u_n \asymp n^{-\varpi}, \quad \gamma \in \left(\frac{r}{2}, 1\right), \quad \varpi \in \left[\frac{1-\gamma}{2-r}, \frac{1}{2}\right). \tag{3.2}
$$

For each $i \in \{1, \ldots, n - k_n\}$, the spot covariance estimator for each $t \in$

$((i-1)\Delta_n, i\Delta_n]$ is given by

$$\hat{c}_t \equiv \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \left(\Delta_{i+j}^n X\right) \left(\Delta_{i+j}^n X\right)^\top 1_{\{\|\Delta_{i+j}^n X\| \le u_n\}}, \tag{3.3}$$

and we set $\hat{c}_t \equiv \hat{c}_{(n-k_n)\Delta_n}$ for $(n-k_n)\Delta_n \le t \le 1$.

Asymptotic properties of the spot estimator $\hat{c}_t$ are known from prior literature under Assumption 5 and (3.2). Theorem 9.3.2 of [23] shows that

$$\hat{c}_t \xrightarrow{\mathbb{P}} c_t, \quad t \in (0,1), \tag{3.4}$$

which directly implies Assumption 1. In addition, [30] (see Lemma A2 in that paper's supplemental appendix) shows that

$$\sup_{t \in [0,1]} \|\hat{c}_t - \bar{c}_t\| = o_p(1), \tag{3.5}$$

where $\bar{c}_t$ is defined as a local average of c_t given by

$$\bar{c}_t \equiv \frac{1}{k_n \Delta_n} \int_{i\Delta_n}^{i\Delta_n + k_n \Delta_n} c_s ds, \quad t \in ((i-1)\Delta_n, i\Delta_n], \quad i \in \{1, \ldots, n-k_n\},$$

and $\bar{c}_t \equiv \bar{c}_{(n-k_n)\Delta_n}$ for $(n-k_n)\Delta_n \le t \le 1$. We can then use the uniform approximation (3.5) to verify Assumption 4 (taking $V_t = c_t$), following the discussion in Section 2.3.

Equipped with these results, we are ready to specialize Theorem 1 under more primitive conditions regarding the function class $\mathcal{G}$ and the underlying processes, starting with the bounded monotone class $\mathcal{G}_M$ described in Example 2. Since the input of these monotone functions is one-dimensional, we consider $V_t = h(c_t)$ for some continuous function $h : \mathcal{C} \mapsto \mathbb{R}$. In the univariate setting, $h(\cdot)$ can transform the spot variance into spot volatility, log volatility, quarticity, volatility Laplace transform, etc. In the multivariate setting, $h(\cdot)$ may transform the spot covariance matrix into beta, correlation, idiosyncratic variance, eigenvalue, etc.

THEOREM 3. *Let $V_t = h(c_t)$ for some continuous function $h : \mathcal{C} \mapsto \mathbb{R}$. Suppose that the following conditions hold: (i) Assumption 5; (ii) k_n and u_n satisfy (3.2); (iii) the occupation time $\mathbb{F}(x) = \int_0^1 1_{\{V_s \le x\}} ds$ is continuous in x almost surely. Then,*

$$\sup_{g \in \mathcal{G}_M} |\mathbb{F}_n g - \mathbb{F} g| \xrightarrow{\mathbb{P}^*} 0.$$

Theorem 3 is proved by verifying the high-level assumptions in Theorem 1. In the proof, we verify Assumption 2 by properly constructing random brackets. To further establish the pointwise convergence for these bracket functions, we verify the conditions in Theorem 2 under the primitive conditions imposed by Theorem 3. Specifically, conditions (i) and (ii) lead to (3.4) and (3.5) as explained above, and condition (iii) implies Assumption 3(i). Note that a sufficient condition of the continuity of the occupation time $\mathbb{F}(x)$ is the existence of occupation density (meaning that $\mathbb{F}(x)$ is differentiable in x). The latter existence results for Markov and Gaussian processes are reviewed in [17] and [34]. [28] provides similar results for jump-diffusion models that are commonly used in economics and finance (see Lemma 2.1 in that paper).

Based on a "more elementary" proof without bracketing, [26] establishes (see Theorem 1 there) the uniform convergence of occupation time estimators for univariate volatility, which corresponds to indicator test functions (i.e., $\mathcal{G}_I$), and hence, is a special case of Theorem 3. Our result is far more general as it shows that the convergence holds uniformly for all bounded monotone functions.

We next consider the uniform convergence over the Lipschitz-in-parameter and the Hölder classes of functions described in Examples 3 and 4, respectively, with $V_t = c_t$. Let $\mathcal{G}_L$ and $\mathcal{G}_H$ be defined as in those examples. As mentioned before, we do not need to assume V_t to be compactly valued; instead, we only need the localized version of this condition given by Assumption 5(ii).

Theorem 4. *Suppose that the following conditions hold: (i) Assumption 5; (ii) k_n and u_n satisfy (3.2). Then, for $\mathcal{G} = \mathcal{G}_L$ or $\mathcal{G}_H$,*

$$\sup_{g\in\mathcal{G}} |\mathbb{F}_n g - \mathbb{F} g| \xrightarrow{\mathbb{P}^*} 0.$$

Theorem 4 is also proved by verifying the high-level assumptions in Theorem 1. The resulting Glivenko–Cantelli theorem is very general, as it establishes uniformity for both parametric families ($\mathcal{G}_L$) and Hölder-continuous nonparametric families ($\mathcal{G}_H$). These classes cover most—if not all—smooth test functions considered in the literature on integrated volatility functionals. Theorem 4 reveals that the convergence is in fact uniform across all these test functions, which is theoretically interesting.

3.2. *Glivenko–Cantelli theorems in the noisy setting.* In this subsection, we further extend the results in Section 3.1 to a setting in which the semimartingale X is contaminated with microstructure noise. In financial applications, the noise is often attributed to various market frictions and has long been recognized in the literature ([15], [32], [39]). The related statistics literature has mainly focused on the estimation of integrated variance and covariance matrix using various approaches including the multi-scale estimator ([48], [46], [3], [47]), the realized kernel ([6], [7]), the pre-averaging method ([19], [38], [2], [21]), the local method of moments ([13]), and the likelihood method ([44]). By "localizing" these integrated estimators, several papers also propose spot estimators of volatility ([49], [33], [45], [14]). In a recent paper, [20] focuses on the noise and provides estimators for integrated moments of noise.

To the best of our knowledge, however, the estimation of integrated functionals for general test functions in the noisy setting remains largely to be an open question, and the related Glivenko–Cantelli problem has not been studied at all. In this subsection we establish such results for integrated functionals of the volatility of efficient price and/or the spot variance of microstructure noise, by applying our general Glivenko-Cantelli theorem (Theorem 1).

We consider a similar setting as [22] for modeling noisy high-frequency data. Suppose the latent efficient price process X (recall equation (3.1)) is observed with error. That is, we observe

$$Y_{i\Delta_n} = X_{i\Delta_n} + \epsilon_i, \quad i = 0, \dots, \lfloor 1/\Delta_n \rfloor, \tag{3.6}$$

where ϵ_i denotes the d-dimensional noise, for which we impose the following assumption.

ASSUMPTION 6. *The variables $(\epsilon_i)_{i\geq 0}$ are $\mathcal{F}$-conditionally independent with zero mean. Moreover,*

(i) for every $q > 0$, $\mathbb{E}\left[\|\epsilon_i\|^q|\mathcal{F}\right] = M_{q,i\Delta_n}$ for some locally bounded process $(M_{q,t})_{t\geq 0}$;

(ii) $\mathbb{E}[\epsilon_i\epsilon_i^\top|\mathcal{F}] = v_{i\Delta_n}$ for some càdlàg, $\mathcal{F}_t$-adapted process $(v_t)_{t\geq 0}$ taking values in an open set $\mathcal{C}$, and there exist a sequence of stopping times $(T_m)_{m\geq 1}$ increasing to infinity and a sequence $(\mathcal{K}_m)_{m\geq 1}$ of compact convex subsets of $\mathcal{C}$ such that $v_t \in \mathcal{K}_m$ for each $m \geq 1$ and $t \leq T_m$.

A few remarks are in order. Condition (i) of Assumption 6 is a mild regularity condition. We assume local boundedness of conditional moments at all orders only for ease of exposition. Condition (ii) mainly requires that the spot covariance of the noise vector is locally compactly valued. This assumption is akin to Assumption 5(ii) on the spot covariance of efficient price. Like the latter, we also allow the noise volatility process to be stochastic with jumps. This is empirically relevant and leads to some interesting theoretical complications in our analysis.

We acknowledge that this setting is not completely general. While we allow for unconditional dependence among the noises (as the v and M_q processes can be highly persistent), we rule out their $\mathcal{F}$-conditional dependence which has been studied by [20]. In addition, we have implicitly assumed in equation (3.6) that the data are regularly sampled, and hence, ruled out random sampling. That being said, our main goal is not to estimate the integrated covariance matrix under even more general settings, but rather to illustrate how to establish new Glivenko–Cantelli theorems for general integrated functionals in the noisy setting, which extends the noise literature in a new dimension. Our results may be extended to allow for more general forms of noise as studied by [20], which would be an interesting topic for future research.

The latent processes of interest are the efficient price's spot covariance process c_t and the noise spot covariance v_t. Similar to the development in the no-noise setting considered in Section 3.1, the key to verifying the high-level conditions of Theorem 1 is to establish a uniform approximation of spot estimators for moving averages of the corresponding underlying processes (recall (3.5)). But this type of uniform approximation is unavailable in the extant literature on noisy high-frequency data. For this reason, we first establish such a result, which is of independent theoretical interest. We consider spot estimators using the pre-averaging method. We make this choice only for concreteness, while noting that similar results may be established using the other approaches mentioned above as well.

To implement the pre-averaging method, we consider a weight function $w : \mathbb{R} \mapsto \mathbb{R}$ that is continuous, piecewise continuously differentiable with Lipschitz-continuous derivative, and further satisfies $w(s) = 0$ for $s \notin (0, 1)$ and $\int_0^1 w(s)^2 ds > 0$. For an integer sequence h_n of pre-averaging windows, we set

$$\phi_n \equiv \frac{1}{h_n} \sum_{j \geq 0} w\left(\frac{j}{h_n}\right)^2, \quad \bar{\phi}_n \equiv h_n \sum_{j \geq 0} \left(w\left(\frac{j}{h_n}\right) - w\left(\frac{j-1}{h_n}\right)\right)^2.$$

The pre-averaging estimators are constructed using the following pre-averaged quantities

$$\begin{aligned}\widetilde{Y}_i^n &\equiv \sum_{j=1}^{h_n-1} w\left(\frac{j}{h_n}\right)\Delta_{i+j}^n Y,\\ \widehat{Y}_i^n &\equiv \sum_{j=1}^{h_n}\left(w\left(\frac{j}{h_n}\right)-w\left(\frac{j-1}{h_n}\right)\right)^2\left(\Delta_{i+j}^n Y\right)\left(\Delta_{i+j}^n Y\right)^\top.\end{aligned} \tag{3.7}$$

The spot estimators $\hat{c}_t, \hat{v}_t$ are defined as follows. We choose a sequence k_n of integers satisfying $k_n/h_n \to \infty$, and divide the sample span into $\lfloor n/k_n \rfloor$ non-overlapping blocks. For each block $i \in \{0, \ldots, \lfloor n/k_n \rfloor - 1\}$, we form a spot estimator for the noise spot covariance

$$\hat{v}_i^n \equiv \frac{h_n}{2\bar{\phi}_n\left(k_n - h_n + 1\right)} \sum_{j=0}^{k_n-h_n} \widehat{Y}_{ik_n+j}^n, \tag{3.8}$$

and the spot covariance of the efficient price

$$\begin{aligned}\hat{c}_i^n \equiv{}& \frac{1}{\left(k_n - h_n + 1\right) h_n \phi_n \Delta_n} \sum_{j=0}^{k_n-h_n} \left(\widetilde{Y}_{ik_n+j}^n\right)\left(\widetilde{Y}_{ik_n+j}^n\right)^\top 1_{\{\|\widetilde{Y}_{ik_n+j}^n\| \leq u_n\}}\\ &-\frac{\bar{\phi}_n}{h_n^2 \Delta_n \phi_n}\hat{v}_i^n,\end{aligned} \tag{3.9}$$

where the second term corrects the bias induced by the microstructure noise. Finally, we set for each $i \in \{1, \ldots, \lfloor n/k_n \rfloor - 1\}$ and $t \in \left(\left(i-1\right) k_n \Delta_n, i k_n \Delta_n\right]$

$$\hat{c}_t \equiv \hat{c}_i^n, \quad \hat{v}_t \equiv \hat{v}_i^n,$$

and $\hat{c}_t \equiv \hat{c}_{(\lfloor n/k_n \rfloor - 1)k_n\Delta_n}$ and $\hat{v}_t \equiv \hat{v}_{(\lfloor n/k_n \rfloor - 1)k_n\Delta_n}$ for $\left(\lfloor n/k_n \rfloor - 1\right) k_n \Delta_n \leq t \leq 1$. The pointwise and uniform approximation properties of these spot estimators are given by the following proposition.

Proposition 1. *Suppose that (i) Assumptions 5 and 6 hold; (ii) $k_n \asymp n^\gamma$, $h_n \asymp n^{1/2}$, and $u_n \asymp (h_n \Delta_n)^\varpi$, where*

$$\gamma \in \left(\frac{6+r}{8}, 1\right), \quad \varpi \in \left[\frac{4(1-\gamma)}{2-r}, \frac{1}{2}\right).$$

Then, (a)

$$\max_{0\le i\le \lfloor n/k_n\rfloor -1}\left\|\hat{c}_i^n-\frac{1}{k_n\Delta_n}\int_{ik_n\Delta_n}^{(i+1)k_n\Delta_n}c_s ds\right\|=o_p(1),$$

$$\max_{0\le i\le \lfloor n/k_n\rfloor -1}\Bigg\|\hat{v}_i^n-\frac{h_n}{2\bar{\phi}_n(k_n-h_n+1)}\sum_{j=0}^{k_n-h_n}\sum_{l=1}^{h_n}\left(w\left(\frac{l}{h_n}\right)-w\left(\frac{l-1}{h_n}\right)\right)^2 \times\left(v_{(ik_n+j+l)\Delta_n}+v_{(ik_n+j+l-1)\Delta_n}\right)\Bigg\|=o_p(1).$$

(b) $\hat{c}_t\xrightarrow{\mathbb{P}}c_t$ *and* $\hat{v}_t\xrightarrow{\mathbb{P}}v_t$ *for* $t\in[0,1)$.

Compared with extant results on spot estimation in the noisy setting, the key novelty of Proposition 1 is the uniform approximation described in part (a), which implies part (b) as a corollary because of the right continuity of the processes c and v. This result is analogous to the uniform approximation (3.5) in the no-noise setting, and describes what the spot estimators are "directly" approximating. We remind the reader that this type of uniform approximation is used to show that the spot estimators uniformly fall in a small enlargement around the compact set in which the underlying processes take value (up to a localizing stopping time), which in turn allows us to accommodate general test functions without requiring them to have polynomial growth (cf. [23]); see [27] for a more detailed discussion on this "spacial localization" technique. Although the approximating variable for $\hat{v}_i^n$ takes a somewhat complicated form as a "double moving average," it is enough for our purpose mentioned above. We stress that this uniform approximation result holds even if the underlying c and v processes are discontinuous. In this situation, the seemingly natural uniform approximation $\sup_{t\in[0,1]}(\|\hat{c}_t-c_t\|+\|\hat{v}_t-v_t\|)=o_p(1)$ would generally fail.

Equipped with Proposition 1, we can readily establish Glivenko–Cantelli theorems in the noisy setting in parallel to Theorems 3 and 4, as described in the following theorems.

THEOREM 5. *Let* $V_t=h(c_t,v_t)$ *for some continuous function* $h:\mathcal{C}\times\mathcal{C}\mapsto\mathbb{R}$, *and let* $\widehat{V}_t=h(\hat{c}_t,\hat{v}_t)$. *Suppose that (i) the conditions in Proposition 1 hold; (ii) the occupation time* $\mathbb{F}(x)=\int_0^1 1_{\{V_s\le x\}}ds$ *is continuous in* x *almost surely. Then,*

$$\sup_{g\in\mathcal{G}_M}|\mathbb{F}_n g-\mathbb{F}g|\xrightarrow{\mathbb{P}^*}0.$$

THEOREM 6. *Let $V_t = (c_t, v_t)$ and $\widehat{V}_t = (\hat{c}_t, \hat{v}_t)$. Suppose that the conditions in Proposition 1 hold. Then, for $\mathcal{G} = \mathcal{G}_L$ or $\mathcal{G}_H$,*

$$\sup_{g \in \mathcal{G}} |\mathbb{F}_n g - \mathbb{F} g| \xrightarrow{\mathbb{P}^*} 0.$$

Theorems 5 and 6 provide Glivenko–Cantelli results for bounded monotone functions, Lipschitz-in-parameter, and Hölder classes in the noisy setting. Of course, a special case of these results is the pointwise estimation of integrated functionals for any specific test function $g(\cdot)$, which is often needed for estimating asymptotic variances of estimators constructed based on noisy high-frequency data (also see [37]). But our results are not restricted to this specific purpose, and they may serve as a starting point for the study of general integrated functionals and an associated empirical process theory in the noisy setting for future research.

3.3. *Empirical illustration.* We illustrate the usefulness of the proposed theory with an empirical application pertaining to the time-varying systematic risk of individual firms. The recent COVID-19 pandemic has made an enormous negative impact on the global financial markets, and the effect also varies substantially across industries. On one hand, airline companies have suffered from arguably the most severe loss due to consumers' safety concerns and governments' travel bans, with their solvency relying critically on government bailouts. On the other hand, technology companies appear to be much less influenced, some of which (e.g., Amazon.com) have even witnessed significant gains in their equity values during the ongoing crisis.

Set against this background, we study the market betas of four individual "stocks" in the US equity market using a recent sample from January 2, 2019 to June 30, 2020 (recall that the market beta of a firm measures the sensitivity of the firm's stock price with respect to changes in the price of the market portfolio). The first three companies are Amazon.com, Facebook, and Boeing. The last one is an equally weighted portfolio of the three major US airline companies: American Airlines, Delta Air Lines, and United Airlines. For ease of discussion, we refer to the latter portfolio as "Airlines." The market portfolio is proxied by the S&P 500 ETF. We obtain high-frequency transaction data during regular trading hours from the TAQ database, and sample the data sparsely at the 1-minute frequency so as to mitigate the effect of microstructure noise. Empirically, we are interested in how the betas of these companies have evolved during the COVID-19 pandemic. We thus divide the full sample into four periods. The first spans the entire year of 2019, and is used as the pre-COVID benchmark. The six months in 2020 are then equally divided into three 2-month periods. Note that although the COVID-19 virus was first reported in China on December 31, 2019, it did not appear to be a major concern in the US during the first two months of 2020 (the first COVID death in the US was reported on February 29, 2020). The number of confirmed cases then substantially increased across the country and caused an enormous turmoil in the stock market.

Our numerical analysis proceeds more precisely as follows. We suppose that the asset prices are modeled by (3.1). Denoting the log price of the market portfolio by $X_{1,t}$ and the log price of a firm by $X_{2,t}$, the firm's beta is given by $\beta_t = h_\beta(c_t) \equiv c_{12,t}/c_{11,t}$, which is exactly the "local" regression coefficient obtained by regressing the firm's instantaneous return on that of the market portfolio without intercept. A large (resp. small) beta indicates that the firm bears high (resp. low) systematic risk. We summarize the distributional feature of the firm's beta process on a time interval $[0, T]$ using its occupation time, defined as $T^{-1}\int_0^T 1_{\{\beta_s \le x\}} ds$, which measures the proportion of time spent by the beta process below a certain level $x \in \mathbb{R}$. We estimate the spot beta using $\hat{\beta}_t = h_\beta(\hat{c}_t)$, with $\hat{c}_t$ defined in (3.3). We set the block size $k_n = 30$ and, following [26], set the truncation threshold adaptively as $u_n = 5\sqrt{BV}\Delta_n^{0.49}$, where BV is the bipower estimator (see [11]) of integrated volatility computed separately for each trading day and each asset.

Figure 1 plots the estimated beta occupation time for each stock in each of the four subsamples. Looking at the estimates for Amazon on the top-left panel, we see that in 2019 the occupation measure of the e-commerce firm's beta concentrates near (indeed moderately above) unity, and we get essentially the same estimate in the first two months of 2020. However, we find a remarkable left-shift of the occupation time for the later two subsamples during the pandemic, suggesting that Amazon's market beta has become much lower and, in particular, is below unity most of the time. It is interesting to note that the estimates for the March-April period and the May-June period are virtually the same, despite that they were estimated separately without any restriction. This distributional evidence for Amazon's reduced sensitivity with respect to the market portfolio suggests that the Amazon stock may play a useful role for diversifying the market-level risk during the crisis. The estimates for Facebook, shown on the top-right panel of Figure 1, are similar to those of Amazon in the pre-COVID subsamples (i.e., 2019 and 2020 Jan-Feb), as well as the early pandemic period of March and April. However, in the more recent May-June period, the estimate appears to have largely returned to its pre-COVID level, which shows an interesting contrast to Amazon.

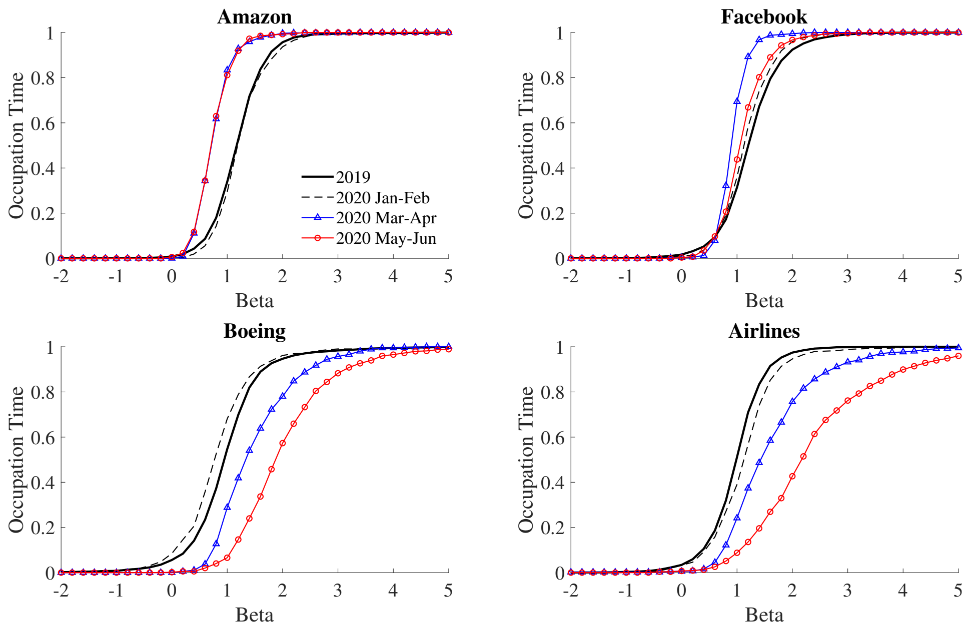

FIG 1. *Occupation Times of Market Betas around the COVID-19 Pandemic.*

The estimates of Boeing and Airlines, shown on the bottom panels of Figure 1, are drastically different from the technology companies. Indeed, their occupation time estimates have evidently shifted to the right during the early pandemic period (March and April), and even more so in the more recent period (May and June). More specifically, we see that, during the pandemic, not only do their occupation measures center at much higher levels of beta, they also become much more dispersed. These findings suggest that firms in the airline industry load on an elevated level of systematic risk, and their risk loadings also bear an additional "layer" of high uncertainty.

Overall, this empirical example demonstrates how the proposed uniform estimation theory may be applied in practice. Admittedly, a limitation of the theory developed here is that it only provides consistent functional estimation. Formal hypothesis testing would further require empirical process-type functional central limit theorems (i.e., Donsker theorems) for the family of integrated functionals. Existing results on this issue are rather limited (see [30] and [29]), and the empirical-process theory for either nonsmooth test functions (e.g., the occupation time) or noisy data remains to be challenging open questions. These may be interesting topics for future research.

4. Conclusion. We prove a general Glivenko-Cantelli theorem for integrated functionals of latent stochastic processes, which is based on a bracketing condition with (possibly) random brackets. The theorem is shown to be broadly applicable for general classes of test functions, and various underlying stochastic processes commonly studied in high-frequency econometrics and statistics including particularly the volatility of efficient price and spot variance of microstructure noise.

5. Proofs.

5.1. *Proofs for Section 2.* In this subsection, we prove Theorem 1 and Theorem 2.

PROOF OF THEOREM 1. Let $\varepsilon > 0$ and consider brackets $(l_j, u_j)_{1\leq j\leq N_\varepsilon}$ given by Assumption 2. Then, for each $g \in \mathcal{G}$ and $\omega \in \Omega$, we can find $j(\omega)$ and $k(\omega)$ in $\{1, \dots, N_\varepsilon\}$, such that

$$l_{j(\omega)}(\omega,\cdot) \leq g(\cdot) \leq u_{k(\omega)}(\omega,\cdot)$$

and

$$\int \left(u_{k(\omega)}(\omega,x) - l_{j(\omega)}(\omega,x)\right)\mathbb{F}_\omega(dx) \leq \varepsilon,$$

where the integers $j(\omega)$ and $k(\omega)$ may depend on ω. These conditions further imply that

$$\int \left(u_{k(\omega)}(\omega,x) - g(x)\right)\mathbb{F}_\omega(dx) \leq \varepsilon, \quad \int \left(g(x) - l_{j(\omega)}(\omega,x)\right)\mathbb{F}_\omega(dx) \leq \varepsilon.$$

We then observe

$$\begin{aligned}
&\int g(x)\,\mathbb{F}_{n,\omega}(dx) - \int g(x)\,\mathbb{F}_\omega(dx) \\
&= \int g(x)\,\mathbb{F}_{n,\omega}(dx) - \int u_{k(\omega)}(\omega,x)\,\mathbb{F}_\omega(dx) \\
&\quad + \int u_{k(\omega)}(\omega,x)\,\mathbb{F}_\omega(dx) - \int g(x)\,\mathbb{F}_\omega(dx) \\
&\leq \int u_{k(\omega)}(\omega,x)\,\mathbb{F}_{n,\omega}(dx) - \int u_{k(\omega)}(\omega,x)\,\mathbb{F}_\omega(dx) + \varepsilon.
\end{aligned}$$

Since $k(\omega) \in \{1, \dots, N_\varepsilon\}$, we further have

$$\mathbb{F}_n g - \mathbb{F} g \leq \max_{1\leq j\leq N_\varepsilon}\left(\mathbb{F}_n u_j - \mathbb{F} u_j\right) + \varepsilon.$$

Hence,

$$\sup_{g\in\mathcal{G}} (\mathbb{F}_n - \mathbb{F})\, g \le \max_{1\le j\le N_\varepsilon} (\mathbb{F}_n - \mathbb{F})\, u_j + \varepsilon. \tag{5.1}$$

Similarly, we can deduce that

$$\sup_{g\in\mathcal{G}} (\mathbb{F} - \mathbb{F}_n)\, g \le - \min_{1\le j\le N_\varepsilon} (\mathbb{F}_n - \mathbb{F})\, l_j + \varepsilon. \tag{5.2}$$

From (5.1) and (5.2), we see that the following holds almost surely

$$\sup_{g\in\mathcal{G}} |(\mathbb{F}_n - \mathbb{F})\, g| \le \max_{1\le j\le N_\varepsilon} |(\mathbb{F}_n - \mathbb{F})\, u_j| + \max_{1\le j\le N_\varepsilon} |(\mathbb{F}_n - \mathbb{F})\, l_j| + \varepsilon.$$

Therefore,

$$\begin{aligned}
&\mathbb{P}^*\left(\sup_{g\in\mathcal{G}} |(\mathbb{F}_n - \mathbb{F})\, g| > 3\varepsilon\right)\\
&\quad\le \mathbb{P}\left(\max_{1\le j\le N_\varepsilon} |(\mathbb{F}_n - \mathbb{F})\, u_j| + \max_{1\le j\le N_\varepsilon} |(\mathbb{F}_n - \mathbb{F})\, l_j| > 2\varepsilon\right)\\
&\quad\le \mathbb{P}\left(\max_{1\le j\le N_\varepsilon} |(\mathbb{F}_n - \mathbb{F})\, u_j| > \varepsilon\right) + \mathbb{P}\left(\max_{1\le j\le N_\varepsilon} |(\mathbb{F}_n - \mathbb{F})\, l_j| > \varepsilon\right)\\
&\quad\to 0,
\end{aligned}$$

where the convergence follows from condition (ii) of Theorem 1. This finishes the proof. □

Proof of Theorem 2. With an appeal to the standard localization technique (see, e.g., Section 4.4.1 of [23]), we can assume that for some compact subset $\mathcal{K} \subseteq \mathcal{V}$,

$$\Omega_n \equiv \left\{V_t, \widehat{V}_t \in \mathcal{K} \text{ for all } t \in [0,T]\right\}$$

satisfies $\mathbb{P}(\Omega_n) \to 1$.

Fix some $s \in (0,1)$ such that

$$\widehat{V}_s \xrightarrow{\mathbb{P}} V_s \quad \text{and} \quad \mathbb{P}\left(\{\omega : f(\omega,\cdot) \text{ is continuous at } V_s(\omega)\}\right) = 1.$$

By Assumption 1 and Assumption 3(i), this holds for Lebesgue almost every $s \in (0,1)$. By the subsequence characterization, for each subsequence $\mathbb{N}_1 \subseteq \mathbb{N}$, there exists a further subsequence $\mathbb{N}_2 \subseteq \mathbb{N}_1$ such that for an event $\Omega_1 \subseteq \Omega$ satisfying $\mathbb{P}(\Omega_1) = 1$, we have

$$\widehat{V}_s(\omega) \to V_s(\omega),$$

as $n \to \infty$ along $\mathbb{N}_2$ for all $\omega \in \Omega_1$. Let $\Omega_2 \equiv \{\omega : f(\omega, \cdot)$ is continuous at $V_s(\omega)\}$. Note that $\mathbb{P}(\Omega_1 \cap \Omega_2) = 1$. For each $\omega \in \Omega_1 \cap \Omega_2$, the mapping $x \mapsto f(\omega, x)$ is continuous at $V_s(\omega)$, and hence,

$$f(\omega, \widehat{V}_s(\omega)) \to f(\omega, V_s(\omega))$$

as $n \to \infty$ along $\mathbb{N}_2$. With an appeal to the subsequence characterization of convergence in probability again, we see that

$$f(\widehat{V}_s) \xrightarrow{\mathbb{P}} f(V_s), \quad \text{as} \quad n \to \infty. \tag{5.3}$$

By Assumption 3(ii), $\sup_{x\in\mathcal{K}} |f(x)| \le \sup_{x\in\mathcal{K}} F(x) < \infty$. By the bounded convergence theorem, (5.3) further implies

$$\mathbb{E}\left[\left|f(\widehat{V}_s) - f(V_s)\right| 1_{\Omega_n}\right] \to 0.$$

Since this holds for Lebesgue almost every $s \in [0,1]$, we can use the bounded convergence theorem again to deduce that

$$\mathbb{E}\left[\left|\int_0^1 f(\widehat{V}_s)ds - \int_0^1 f(V_s)ds\right| 1_{\Omega_n}\right] \le \int_0^1 \mathbb{E}\left[\left|f(\widehat{V}_s) - f(V_s)\right| 1_{\Omega_n}\right] ds \to 0.$$

Since $\mathbb{P}(\Omega_n) \to 1$, the assertion $\mathbb{F}_n f \xrightarrow{\mathbb{P}} \mathbb{F} f$ readily follows from the above convergence. □

5.2. *Proofs for Section 3.1.* In this subsection, we prove Theorem 3 and Theorem 4 in Section 3.1.

Proof of Theorem 3. The proof is done by verifying the high-level conditions of Theorem 1.

We start with recalling a known result. Let $\tilde{\mathcal{G}}_M$ denote the collection of monotone functions $\tilde{g} : [0,1] \mapsto [0,1]$. By Theorem 2.7.5 of [42], for each $\varepsilon > 0$, there exist piecewise constant deterministic bracket functions $\{(\tilde{l}_j, \tilde{u}_j) : 1 \le j \le N_\varepsilon\}$ such that each $\tilde{g} \in \tilde{\mathcal{G}}_M$ satisfies $\tilde{l}_j \le \tilde{g} \le \tilde{u}_j$ for some j, and $\int_0^1 (\tilde{u}_j(q) - \tilde{l}_j(q))dq \le \varepsilon/2$. Consequently,

$$\int_0^1 (\tilde{u}_j(q) - \tilde{g}(q))dq \le \varepsilon/2, \quad \int_0^1 (\tilde{g}(q) - \tilde{l}_j(q))dq \le \varepsilon/2.$$

We now proceed to construct the random brackets. For each $\omega \in \Omega$, we denote

$$\mathbb{F}_\omega(x) \equiv \int_0^1 1_{\{V_s(\omega) \le x\}} ds,$$

and denote its left-continuous and right-continuous inverse respectively by

$$\mathbb{F}_{\omega}^{-1}(q) \equiv \inf\{x : \mathbb{F}_{\omega}(x) \geq q\}, \quad \overline{\mathbb{F}}_{\omega}^{-1}(q) \equiv \sup\{x : \mathbb{F}_{\omega}(x) \leq q\}.$$

Below, we verify that Assumption 2 is satisfied with the following bracket functions:

$$l_j(\omega, x) \equiv \tilde{l}_j(\mathbb{F}_{\omega}(x)), \quad u_j(\omega, x) \equiv \tilde{u}_j(\mathbb{F}_{\omega}(x)).$$

For each $g \in \mathcal{G}_M$ and $\omega \in \Omega$, the (deterministic) function $g \circ \mathbb{F}_{\omega}^{-1}$ belongs to $\tilde{\mathcal{G}}_M$. Hence, for some $j(\omega) \in \{1, \ldots, N_{\varepsilon}\}$, we have

$$\tilde{l}_{j(\omega)}(\cdot) \leq g \circ \mathbb{F}_{\omega}^{-1}(\cdot),$$

and

$$\int_0^1 \left(g \circ \mathbb{F}_{\omega}^{-1}(q) - \tilde{l}_{j(\omega)}(q)\right) dq \leq \varepsilon/2.$$

These estimates further imply (by Lemma 21.1 of [43])

$$l_{j(\omega)}(\omega, \cdot) = \tilde{l}_{j(\omega)} \circ \mathbb{F}_{\omega}(\cdot) \leq g(\cdot), \tag{5.4}$$

and

$$\begin{aligned} &\int \left(g(x) - l_{j(\omega)}(\omega, x)\right) \mathbb{F}_{\omega}(dx) \\ &= \int_0^1 \left(g \circ \mathbb{F}_{\omega}^{-1}(q) - \tilde{l}_{j(\omega)}(q)\right) dq \leq \varepsilon/2. \end{aligned} \tag{5.5}$$

By repeating a similar argument for the function $g \circ \overline{\mathbb{F}}_{\omega}^{-1} \in \tilde{\mathcal{G}}_M$, we can find some $k(\omega) \in \{1, \ldots, N_{\varepsilon}\}$, such that

$$g(\cdot) \leq u_{k(\omega)}(\omega, \cdot), \quad \int \left(u_{k(\omega)}(\omega, x) - g(x)\right) \mathbb{F}_{\omega}(dx) \leq \varepsilon/2. \tag{5.6}$$

(We note that $k(\omega)$ may be different from $j(\omega)$ because they are associated with two different functions in $\tilde{\mathcal{G}}_M$, namely, $g \circ \mathbb{F}_{\omega}^{-1}$ and $g \circ \overline{\mathbb{F}}_{\omega}^{-1}$, and this distinction is relevant because the occupation time may not be strictly increasing.) The requirements of Assumption 2 readily follow from (5.4), (5.5), and (5.6). Condition (i) of Theorem 1 is now verified.

It remains to verify condition (ii) of Theorem 1. We do this by applying Theorem 2. Under conditions (i) and (ii) of Theorem 3, we have (3.4) and (3.5) by Theorem 9.3.2 of [23] and Lemma 2 of [27], that is,

$$\hat{c}_t \xrightarrow{\mathbb{P}} c_t, \quad \sup_{t \in [0,1]} \|\hat{c}_t - \bar{c}_t\| = o_p(1). \tag{5.7}$$

Under Assumption 5, $c_t \in \mathcal{K}_m$ for all $t \leq T_m$. By the convexity of $\mathcal{K}_m$, we also have $\bar{c}_t \in \mathcal{K}_m$. By the uniform approximation in (5.7), we see that $\hat{c}_t$ falls in some enlargement $\mathcal{K}'_m$ of $\mathcal{K}_m$ for all $t \in [0, T]$ with probability approaching 1. From these results and the continuity of $h(\cdot)$, Assumptions 1 and 4 readily follow.

The remaining task is to verify Assumption 3 for each of the bracket functions l_j and u_j defined above. We provide the details for the lower bracket l_j. Note that l_j takes values in $[0, 1]$, so we can set the envelope function $F(x) = 1$ identically. To verify Assumption 3(i), we recall that $\tilde{l}_j$ is piecewise constant and $\mathbb{F}(\cdot)$ is almost surely continuous. Let $x_0 \in (0, 1)$ be any point of the finitely many discontinuity points of $\tilde{l}_j$, it suffices to show that, for Lebesgue a.e. $t \in [0, 1]$,

$$\mathbb{P}(\mathbb{F}(V_t) = x_0) = 0.$$

This holds true because

$$\int_0^1 \mathbb{P}(\mathbb{F}(V_s) = x_0)\, ds = \mathbb{E}\left[\int_0^1 1_{\{\mathbb{F}(V_s)=x_0\}} ds\right] = 0.$$

We can then apply Theorem 2 to show that $\mathbb{F}_n l_j \xrightarrow{\mathbb{P}} \mathbb{F} l_j$ for each j. The proof for $\mathbb{F}_n u_j \xrightarrow{\mathbb{P}} \mathbb{F} u_j$ is similar. □

Proof of Theorem 4. The proof is done by verifying the high-level conditions of Theorem 1. As explained in Examples 3 and 4, we can verify Assumption 2 with deterministic bracket functions which are also continuous. Let f be a generic bracket function in these examples, it remains to verify that $\mathbb{F}_n f \xrightarrow{\mathbb{P}} \mathbb{F} f$ by applying Theorem 2. To this end, we note that Assumptions 1 and 4 can be verified as in the proof Theorem 3, and Assumption 3 holds with $F = |f|$ because f is continuous. □

5.3. *Proofs for Section 3.2.* We first prove Proposition 1. It is easy to see that, upon using a polarization argument, there is no loss of generality to consider the univariate case with $d = 1$. Moreover, by a classical localization argument (see Section 4.4.1 of [23]), we can also strengthen Assumptions 5 and 6 as the following condition without loss of generality.

Assumption 7. *We have Assumptions 5 and 6 with $T_1 = \infty$. Moreover, the processes b and M_q are uniformly bounded, and c, v take values in a compact convex set $\mathcal{K} \subset \mathcal{C}$. For some bounded λ-integrable function $D(\cdot)$ on $\mathbb{R}$, $\|\delta(\omega, t, z)\|^r \leq D(z)$.*

Throughout the proofs, we use K to denote a generic constant that may change from line to line; we sometimes write K_p in order to emphasize its dependence on some parameter p. We use $\|\cdot\|_p$ to denote the L_p norm and write $\mathbb{E}_{\mathcal{F}}[\,\cdot\,]$ in place of $\mathbb{E}[\,\cdot\,|\mathcal{F}]$. For notational simplicity, we denote $w_j^n \equiv w\left(j/h_n\right)$ and $\bar{w}_j^n \equiv w\left(j/h_n\right) - w\left((j-1)/h_n\right)$. In addition, let X^c be the continuous part (i.e., drift plus diffusion) of X, and then set $Y_{i\Delta_n}^c = X_{i\Delta_n}^c + \epsilon_i$. We define the pre-averaged quantities $\widetilde{Y}_j^{c,n}$, $\widetilde{X}_i^{c,n}$, $\widetilde{\epsilon}_i^n$ in the same way as in (3.7), with Y replaced by Y^c, X^c, ϵ, respectively. Finally, we set

$$\hat{c}_i^{n\prime} \equiv \frac{1}{(k_n - h_n + 1)h_n\Delta_n\phi_n} \sum_{j=0}^{k_n-h_n} (\widetilde{Y}_{ik_n+j}^{c,n})^2 - \frac{\bar{\phi}_n}{h_n^2\Delta_n\phi_n}\hat{v}_i^n.$$

The proof of Proposition 1 is divided into three lemmas. Lemma 1 and Lemma 2 establish uniform approximations for $\hat{v}_i^n$ and $\hat{c}_i^{n\prime}$, respectively. Lemma 3 shows the uniform asymptotic negligibility of $\hat{c}_i^n - \hat{c}_i^{n\prime}$.

Lemma 1. *Suppose that (i) Assumption 7 holds, and (ii) $h_n \to \infty$ and $k_n \asymp n^\gamma$ for some $\gamma \in (0,1)$. Then,*

$$\max_{0\le i\le \lfloor n/k_n\rfloor -1} \left| \hat{v}_i^n - \frac{h_n}{2\bar{\phi}_n(k_n-h_n+1)} \sum_{j=0}^{k_n-h_n}\sum_{l=1}^{h_n} (\bar{w}_l^n)^2 \times \left(v_{(ik_n+j+l)\Delta_n} + v_{(ik_n+j+l-1)\Delta_n}\right)\right| = o_p(1).$$

Proof. We set $\eta_{i,j}^n \equiv \sum_{l=1}^{h_n} (\bar{w}_l^n)^2 \left((\epsilon_{ik_n+j+l} - \epsilon_{ik_n+j+l-1})^2 - (v_{(ik_n+j+l)\Delta_n} + v_{(ik_n+j+l-1)\Delta_n})\right)$. Our proof replies on the following decomposition:

$$\hat{v}_i^n - \frac{h_n}{2\bar{\phi}_n(k_n-h_n+1)} \sum_{j=0}^{k_n-h_n}\sum_{l=1}^{h_n}(\bar{w}_l^n)^2 \left(v_{(ik_n+j+l)\Delta_n} + v_{(ik_n+j+l-1)\Delta_n}\right) = U_{1,i}^n + U_{2,i}^n + U_{3,i}^n,$$

where

$$U_{1,i}^n \equiv \frac{h_n}{2\bar{\phi}_n(k_n-h_n+1)} \sum_{j=0}^{k_n-h_n} \eta_{i,j}^n,$$

$$U_{2,i}^n \equiv \frac{h_n}{2\bar{\phi}_n(k_n-h_n+1)} \sum_{j=0}^{k_n-h_n}\sum_{l=1}^{h_n} (\bar{w}_l^n)^2 \left(\Delta_{ik_n+j+l}^n X\right)^2,$$

$$U_{3,i}^n \equiv \frac{h_n}{\bar{\phi}_n(k_n-h_n+1)} \sum_{j=0}^{k_n-h_n}\sum_{l=1}^{h_n} (\bar{w}_l^n)^2 \Delta_{ik_n+j+l}^n X \left(\epsilon_{ik_n+j+l} - \epsilon_{ik_n+j+l-1}\right).$$

We now consider these terms in turn, starting with $U^n_{1,i}$. Note that for any $p \geq 1$,

$$\mathbb{E}_{\mathcal{F}}\left[|\eta^n_{i,j}|^p\right] \leq K_p h_n^{-3p/2},$$

which follows from the fact that $|\bar{w}^n_l| \leq K/h_n$, the conditional independence of $(\epsilon_i)_{i\geq 0}$, the Burkholder–Davis–Gundy inequality and Hölder's inequality. We further observe that, conditional on $\mathcal{F}$, $\eta^n_{i,j}$ has zero mean and $\eta^n_{i,j}$ is independent of $\eta^n_{i,k}$ for $|j-k| > h_n$. From here, we deduce

$$\mathbb{E}_{\mathcal{F}}\left[\left|\frac{h_n}{2\bar{\phi}_n(k_n-h_n+1)}\sum_{j=0}^{k_n-h_n}\eta^n_{i,j}\right|^p\right] \leq K_p k_n^{-p/2},$$

which further implies

$$\left\|U^n_{1,i}\right\|_p \leq K_p k_n^{-1/2}.$$

With an appeal to the maximal inequality, we deduce

$$\left\|\max_{0\leq i\leq \lfloor n/k_n\rfloor -1}\left|U^n_{1,i}\right|\right\|_p \leq K_p n^{1/p} k_n^{-1/2-1/p}.$$

Since $k_n \asymp n^\gamma$, picking $p > 2(1-\gamma)/\gamma$ yields

$$\max_{0\leq i\leq \lfloor n/k_n\rfloor -1}\left|U^n_{1,i}\right| = o_p(1). \tag{5.8}$$

Next, we note that

$$\mathbb{E}\left[\left|U^n_{2,i}\right|\right] = \frac{h_n}{2\bar{\phi}_n(k_n-h_n+1)}\sum_{j=0}^{k_n-h_n}\sum_{l=1}^{h_n}(\bar{w}^n_l)^2\,\mathbb{E}\left[\left(\Delta^n_{ik_n+j+l}X\right)^2\right] \leq K\Delta_n.$$

Hence, $\mathbb{E}[\max_{0\leq i\leq\lfloor n/k_n\rfloor-1}|U^n_{2,i}|] \leq K k_n^{-1}$, which further implies

$$\max_{0\leq i\leq \lfloor n/k_n\rfloor -1}\left|U^n_{2,i}\right| = o_p(1). \tag{5.9}$$

Finally, we consider $U^n_{3,i}$. Note that, conditional on $\mathcal{F}$, the variables

$$\Delta^n_{ik_n+j+l}X(\epsilon_{ik_n+j+l}-\epsilon_{ik_n+j+l-1}), \quad 1\leq l\leq h_n,$$

are independent with zero mean. Hence,

$$\begin{aligned}
&\mathbb{E}_{\mathcal{F}}\left[\left(\sum_{l=1}^{h_n}(\bar{w}^n_l)^2\,\Delta^n_{ik_n+j+l}X\left(\epsilon_{ik_n+j+l}-\epsilon_{ik_n+j+l-1}\right)\right)^2\right]\\
&\quad = \sum_{l=1}^{h_n}(\bar{w}^n_l)^4\left(\Delta^n_{ik_n+j+l}X\right)^2\mathbb{E}_{\mathcal{F}}\left[\left(\epsilon_{ik_n+j+l}-\epsilon_{ik_n+j+l-1}\right)^2\right]\\
&\quad \leq K h_n^{-4}\sum_{l=1}^{h_n}\left(\Delta^n_{ik_n+j+l}X\right)^2.
\end{aligned}$$

Note that $\mathbb{E}[(\Delta_i^n X)^2] \leq K\Delta_n$. Hence,

$$\left\| \sum_{l=1}^{h_n} (\bar{w}_l^n)^2 \, \Delta_{ik_n+j+l}^n X \left(\epsilon_{ik_n+j+l} - \epsilon_{ik_n+j+l-1} \right) \right\|_2 \leq K h_n^{-3/2} \Delta_n^{1/2}.$$

This estimate further implies $\|U_{3,i}^n\|_2 \leq K h_n^{-1/2} \Delta_n^{1/2}$. By maximal inequality,

$$\left\| \max_{0 \leq i \leq \lfloor n/k_n \rfloor - 1} \left| U_{3,i}^n \right| \right\|_2 \leq K k_n^{-1/2} h_n^{-1/2} = o(1). \tag{5.10}$$

The assertion of the lemma then follows from (5.8), (5.9), and (5.10). □

LEMMA 2. *Suppose that (i) Assumption 7 holds, and (ii)* $h_n \asymp n^{1/2}$, $k_n \asymp n^{\gamma}$, *with* $1/2 < \gamma < 1$. *Then,*

$$\max_{0 \leq i \leq \lfloor n/k_n \rfloor - 1} \left| \hat{c}_i^{n\prime} - \frac{1}{k_n \Delta_n} \int_{ik_n\Delta_n}^{(i+1)k_n\Delta_n} c_s ds \right| = o_p(1).$$

PROOF. Step 1. We outline the proof in this step. Throughout the proof, we use $o_{pu}(1)$ to denote a generic sequence of random variables that is $o_p(1)$ uniformly for $0 \leq i \leq \lfloor n/k_n \rfloor - 1$. For ease of notation, we denote

$$A_n \equiv \frac{1}{(k_n - h_n + 1) h_n \Delta_n \phi_n}, \quad B_n \equiv \frac{\bar{\phi}_n}{h_n^2 \Delta_n \phi_n}.$$

We can then rewrite $\hat{c}_i^{n\prime} = A_n \sum_{j=0}^{k_n - h_n} (\widetilde{Y}_{ik_n+j}^{c,n})^2 - B_n \hat{v}_i^n$. Since $\widetilde{Y}_i^{c,n} = \widetilde{X}_i^{c,n} + \widetilde{\epsilon}_i^n$ by definition, we can decompose

$$\begin{aligned} \hat{c}_i^{n\prime} &= A_n \sum_{j=0}^{k_n - h_n} (\widetilde{X}_{ik_n+j}^{c,n})^2 + A_n \sum_{j=0}^{k_n - h_n} (\widetilde{\epsilon}_{ik_n+j}^n)^2 - B_n \hat{v}_i^n \\ &\quad + 2A_n \sum_{j=0}^{k_n - h_n} \widetilde{X}_{ik_n+j}^{c,n} \widetilde{\epsilon}_{ik_n+j}^n. \end{aligned}$$

The assertion of the lemma then follows from

$$\begin{cases} A_n \sum_{j=0}^{k_n - h_n} (\widetilde{X}_{ik_n+j}^{c,n})^2 - \frac{1}{k_n \Delta_n} \int_{ik_n\Delta_n}^{(i+1)k_n\Delta_n} c_s ds = o_{pu}(1), \\ A_n \sum_{j=0}^{k_n - h_n} (\widetilde{\epsilon}_{ik_n+j}^n)^2 - B_n \hat{v}_i^n = o_{pu}(1), \\ A_n \sum_{j=0}^{k_n - h_n} \widetilde{X}_{ik_n+j}^{c,n} \widetilde{\epsilon}_{ik_n+j}^n = o_{pu}(1). \end{cases} \tag{5.11}$$

These claims will be proved in steps 2, 3, and 4, respectively.

Step 2. In this step, we prove the first claim in (5.11). For each $i \geq 0$, we define a function $w_{n,i}(\cdot)$ as follows: $w_{n,i}(s) \equiv w(j/h_n)$, for $s \in [(i+j-1)\Delta_n, (i+j)\Delta_n)$ and $j \geq 1$. We can then rewrite

$$\widetilde{X}_i^{c,n} = \int_{i\Delta_n}^{(i+h_n-1)\Delta_n} w_{n,i}(s) dX_s^c.$$

By Itô's formula,

$$\begin{aligned}
&\left(\widetilde{X}_i^{c,n}\right)^2 \\
&= \left(\int_{i\Delta_n}^{(i+h_n-1)\Delta_n} w_{n,i}(s) dX_s^c\right)^2 \\
&= 2\int_{i\Delta_n}^{(i+h_n-1)\Delta_n} \left(\int_{i\Delta_n}^{s} w_{n,i}(u) dX_u^c\right) w_{n,i}(s)(b_s ds + \sigma_s dW_s) \\
&\quad + \int_{i\Delta_n}^{(i+h_n-1)\Delta_n} w_{n,i}(s)^2 \sigma_s^2 ds.
\end{aligned} \tag{5.12}$$

For ease of notation, let $\Gamma_i^n \equiv \int_{i\Delta_n}^{(i+h_n-1)\Delta_n} w_{n,i}(s)^2\sigma_s^2 ds$. By (5.12), we can decompose

$$(\widetilde{X}_i^{c,n})^2 - \Gamma_i^n = \xi_{1,i}^{c,n} + \xi_{2,i}^{c,n}, \tag{5.13}$$

where

$$\begin{aligned}
\xi_{1,i}^{c,n} &\equiv 2\int_{i\Delta_n}^{(i+h_n-1)\Delta_n} \left(\int_{i\Delta_n}^{s} w_{n,i}(u) dX_u^c\right) w_{n,i}(s)\sigma_s dW_s, \\
&\quad +2\int_{i\Delta_n}^{(i+h_n-1)\Delta_n} \left(\int_{i\Delta_n}^{s} w_{n,i}(u)\sigma_u dW_u\right) w_{n,i}(s) b_s ds, \\
\xi_{2,i}^{c,n} &\equiv 2\int_{i\Delta_n}^{(i+h_n-1)\Delta_n} \left(\int_{i\Delta_n}^{s} w_{n,i}(u) b_u du\right) w_{n,i}(s) b_s ds.
\end{aligned}$$

Note that $w_{n,i}(\cdot)$ is bounded. By standard estimates for continuous Itô semimartingales, we have for any $p \geq 1$,

$$\mathbb{E}\left[\left|\xi_{1,i}^{c,n}\right|^p\right] \leq K_p (h_n\Delta_n)^p, \quad \mathbb{E}\left[\left|\xi_{2,i}^{c,n}\right|^p\right] \leq K_p (h_n\Delta_n)^{2p}. \tag{5.14}$$

We observe that $\xi_{1,i}^{c,n}$ is $\mathcal{F}_{(i+h_n-1)\Delta_n}$-measurable and has zero $\mathcal{F}_{i\Delta_n}$-conditional mean. Hence, for each i, $(\xi_{1,i+l(h_n-1)}^{c,n}, \mathcal{F}_{(i+(l+1)(h_n-1))\Delta_n})_{l\geq 0}$ forms a martingale difference array, and accordingly, we can then decompose $\sum_{j=0}^{k_n-h_n} \xi_{1,ik_n+j}^{c,n}$

as the sum of $h_n - 1$ terms, each of which is a sum of martingale differences with the aforementioned form. Then, by the Burkholder–Davis–Gundy inequality and Hölder's inequality, we deduce from (5.14) that

$$\left\| A_n \sum_{j=0}^{k_n - h_n} \xi_{1,ik_n+j}^{c,n} \right\|_p \leq K_p h_n^{1/2} k_n^{-1/2}.$$

By a maximal inequality under L_p, we further have

$$\left\| \max_{0 \leq i \leq \lfloor n/k_n \rfloor - 1} \left| A_n \sum_{j=0}^{k_n - h_n} \xi_{1,ik_n+j}^{c,n} \right| \right\|_p \leq K_p n^{1/p} h_n^{1/2} k_n^{-1/2-1/p} \leq K_p n^{\frac{1-\gamma}{p} - \frac{2\gamma-1}{4}}.$$

Since $\gamma \in (1/2, 1)$, we can pick $p > 4(1-\gamma)/(2\gamma - 1)$, so that the majorant side of the above estimates goes to zero as $n \to \infty$. Hence,

$$A_n \sum_{j=0}^{k_n - h_n} \xi_{1,ik_n+j}^{c,n} = o_{pu}(1). \tag{5.15}$$

In addition, we note that (5.14) implies that $\|\xi_{2,i}^{c,n}\|_p \leq K_p \left(h_n \Delta_n\right)^2$, and hence, by the triangle inequality,

$$\left\| A_n \sum_{j=0}^{k_n - h_n} \xi_{2,ik_n+j}^{c,n} \right\|_p \leq K_p h_n \Delta_n.$$

With an appeal to a maximal inequality, we deduce

$$\left\| \max_{0 \leq i \leq \lfloor n/k_n \rfloor - 1} \left| A_n \sum_{j=0}^{k_n - h_n} \xi_{2,ik_n+j}^{c,n} \right| \right\|_p \leq K_p n^{1/p-1} h_n k_n^{-1/p} \leq K_p n^{\frac{1-\gamma}{p} - \frac{1}{2}}.$$

Since $\gamma < 1$, we can pick $p > 2\left(1-\gamma\right)$ so that the above bound vanishes to zero as $n \to \infty$. Hence,

$$A_n \sum_{j=0}^{k_n - h_n} \xi_{2,ik_n+j}^{c,n} = o_{pu}(1). \tag{5.16}$$

Combining (5.13), (5.15) and (5.16), we deduce

$$A_n \sum_{j=0}^{k_n - h_n} (\widetilde{X}_{ik_n+j}^{c,n})^2 - A_n \sum_{j=0}^{k_n - h_n} \Gamma_{ik_n+j}^n = o_{pu}\left(1\right). \tag{5.17}$$

Let $C_t \equiv \int_0^t c_s ds$. We can rewrite $\Gamma_{ik_n+j}^n = \sum_{l=1}^{h_n} (w_l^n)^2 \Delta_{ik_n+j+l}^n C$. Hence,

$$
\begin{aligned}
&A_n \sum_{j=0}^{k_n-h_n} \Gamma_{ik_n+j}^n \\
&\quad = A_n \sum_{j=0}^{k_n-h_n} \sum_{l=1}^{h_n} (w_l^n)^2 \Delta_{ik_n+j+l}^n C \\
&\quad = A_n \sum_{u=1}^{k_n} \left(\sum_{l=1 \vee (u-k_n+h_n)}^{h_n \wedge u} (w_l^n)^2 \right) \Delta_{ik_n+u}^n C \\
&\quad = \frac{1}{(k_n-h_n+1)\Delta_n} \sum_{u=1}^{k_n} \left(\frac{\sum_{l=1\vee(u-k_n+h_n)}^{h_n\wedge u} (w_l^n)^2}{\sum_{l=1}^{h_n} (w_l^n)^2} \right) \Delta_{ik_n+u}^n C \\
&\quad = \frac{1}{k_n\Delta_n} \sum_{u=1}^{k_n} \Delta_{ik_n+u}^n C + o_{pu}(1),
\end{aligned}
$$

where the second equality follows from a change of variable, the third equality is by the definition of A_n and $h_n\phi_n = \sum_{l=1}^{h_n}(w_l^n)^2$, and the last line follows from the fact that $k_n/h_n \to \infty$ and the process c is bounded. We can write the above equivalently as

$$
A_n \sum_{j=0}^{k_n-h_n} \Gamma_{ik_n+j}^n - \frac{1}{k_n\Delta_n} \int_{ik_n\Delta_n}^{(i+1)k_n\Delta_n} c_s ds = o_{pu}(1). \tag{5.18}
$$

The first claim of (5.11) then readily follows from (5.17) and (5.18).

Step 3. We can rewrite $\widetilde{\epsilon}_j^n = -\sum_{l=0}^{h_n} \bar{w}_{l+1}^n \epsilon_{j+l}$. Since the variables $(\epsilon_i)_{i\geq 0}$ are $\mathcal{F}$-conditionally independent with bounded moments and $|\bar{w}_l^n| \leq K h_n^{-1}$,

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{F}}[(\widetilde{\epsilon}_j^n)^2] = \sum_{l=0}^{h_n} \left(\bar{w}_{l+1}^n\right)^2 v_{(j+l)\Delta_n}, \\
&\mathbb{E}_{\mathcal{F}}\left[\left|\widetilde{\epsilon}_j^n\right|^p\right] \leq K_p h_n^{-p/2}, \quad \text{for any } p > 0.
\end{aligned} \tag{5.19}
$$

For notational simplicity, we set $\Lambda_j^n \equiv (\widetilde{\epsilon}_j^n)^2 - \mathbb{E}_{\mathcal{F}}[(\widetilde{\epsilon}_j^n)^2]$. From (5.19), we can further deduce

$$
\mathbb{E}_{\mathcal{F}}\left[\left|\Lambda_j^n\right|^p\right] \leq K_p h_n^{-p}. \tag{5.20}
$$

Note that, conditional on $\mathcal{F}$, the series $(\Lambda_j^n)_{j\geq 1}$ has zero mean with h_n-dependence. By the triangle inequality, the Burkholder–Davis–Gundy inequality, and Hölder's inequality, we see that (5.20) further implies (recall

$h_n \asymp n^{1/2}$)

$$\mathbb{E}_{\mathcal{F}}\left[\left|A_n \sum_{j=0}^{k_n-h_n} \Lambda^n_{ik_n+j}\right|^p\right] \leq K_p A_n^p k_n^{p/2} h_n^{-p/2} \leq K_p n^{p/4} k_n^{-p/2}.$$

Taking an unconditional expectation yields

$$\left\|A_n \sum_{j=0}^{k_n-h_n} \Lambda^n_{ik_n+j}\right\|_p \leq K_p n^{1/4} k_n^{-1/2}.$$

By using a maximal inequality,

$$\left\|\max_{0\leq i\leq \lfloor n/k_n\rfloor -1}\left|A_n \sum_{j=0}^{k_n-h_n} \Lambda^n_{ik_n+j}\right|\right\|_p \leq K_p n^{\frac{1-\gamma}{p}-\frac{2\gamma-1}{4}}.$$

The majorant side is $o(1)$ if we pick $p > 4(1-\gamma)/(2\gamma-1)$. We thus have $A_n \sum_{j=0}^{k_n-h_n} \Lambda^n_{ik_n+j} = o_{pu}(1)$, that is,

$$A_n \sum_{j=0}^{k_n-h_n} (\widetilde{\epsilon}^n_{ik_n+j})^2 - A_n \sum_{j=0}^{k_n-h_n} \sum_{l=0}^{h_n} \left(\bar{w}^n_{l+1}\right)^2 v_{(ik_n+j+l)\Delta_n} = o_{pu}\left(1\right). \tag{5.21}$$

By Lemma 1 and the fact that $B_n = O_p(1)$,

$$B_n \hat{v}^n_i - \frac{A_n}{2} \sum_{j=0}^{k_n-h_n} \sum_{l=1}^{h_n} (\bar{w}^n_l)^2 \left(v_{(ik_n+j+l)\Delta_n} + v_{(ik_n+j+l-1)\Delta_n}\right) = o_{pu}(1). \tag{5.22}$$

In addition,

$$\begin{aligned}
&\frac{A_n}{2} \sum_{j=0}^{k_n-h_n} \sum_{l=1}^{h_n} (\bar{w}^n_l)^2 \left(v_{(ik_n+j+l)\Delta_n} + v_{(ik_n+j+l-1)\Delta_n}\right) \\
&\quad = \frac{A_n}{2} \sum_{j=0}^{k_n-h_n} \sum_{l=0}^{h_n} \left((\bar{w}^n_l)^2 + \left(\bar{w}^n_{l+1}\right)^2\right) v_{(ik_n+j+l)\Delta_n} \\
&\quad = A_n \sum_{j=0}^{k_n-h_n} \sum_{l=0}^{h_n} \left(\bar{w}^n_{l+1}\right)^2 v_{(ik_n+j+l)\Delta_n} + o_{pu}\left(1\right),
\end{aligned} \tag{5.23}$$

where the first equality is by direct calculation, and the last line follows from $A_n \leq Knk_n^{-1}h_n^{-1}$, $|(\bar{w}^n_{l+1})^2 - (\bar{w}^n_l)^2| \leq Kh_n^{-3} \leq Kn^{-1}h_n^{-1}$, and the boundedness of v. The second claim of (5.11) then readily follows from (5.21), (5.22), and (5.23).

Step 4. In this step, we show the third claim in (5.11). Note that, conditional on $\mathcal{F}$, the series $(\widetilde{X}^{c,n}_{ik_n+j}\widetilde{\epsilon}^n_{ik_n+j})_{j\geq 0}$ has zero mean and is h_n-dependent. For $q = 0, \ldots, h_n - 1$, we set

$$\mathcal{I}^n_{i,q} \equiv \{q, q + h_n, q + 2h_n, \ldots\} \cap \{0, \ldots, k_n - h_n\}.$$

We can then decompose

$$\sum_{j=0}^{k_n-h_n} \widetilde{X}^{c,n}_{ik_n+j}\widetilde{\epsilon}^n_{ik_n+j} = \sum_{q=0}^{h_n-1} \sum_{j\in\mathcal{I}^n_{i,q}} \widetilde{X}^{c,n}_{ik_n+j}\widetilde{\epsilon}^n_{ik_n+j},$$

where each $\sum_{j\in\mathcal{I}^n_{i,q}} \widetilde{X}^{c,n}_{ik_n+j}\widetilde{\epsilon}^n_{ik_n+j}$ summation only involves $\mathcal{F}$-conditionally independent variables. Recall from (5.19) that $\mathbb{E}_{\mathcal{F}}[|\widetilde{\epsilon}^n_{ik_n+j}|^p] \leq K_p h_n^{-p/2}$. For each q, we can then apply the Burkholder–Davis–Gundy inequality and Hölder's inequality to deduce that

$$\begin{aligned}\mathbb{E}_{\mathcal{F}}\left[\left|\sum_{j\in\mathcal{I}^n_{i,q}} \widetilde{X}^{c,n}_{ik_n+j}\widetilde{\epsilon}^n_{ik_n+j}\right|^p\right] &\leq K_p h_n^{-p/2}\mathbb{E}_{\mathcal{F}}\left[\left|\sum_{j\in\mathcal{I}^n_{i,q}}\left(\widetilde{X}^{c,n}_{ik_n+j}\right)^2\right|^{p/2}\right]\\ &\leq K_p h_n^{-p/2}\left(\frac{k_n}{h_n}\right)^{p/2-1}\sum_{j\in\mathcal{I}^n_{i,q}}\left|\widetilde{X}^{c,n}_{ik_n+j}\right|^p.\end{aligned}$$

By a standard estimate for continuous Itô semimartingale, $\mathbb{E}[|\widetilde{X}^{c,n}_{ik_n+j}|^p] \leq K_p\,(h_n\Delta_n)^{p/2}$. Hence,

$$\left\|\sum_{j\in\mathcal{I}^n_{i,q}} \widetilde{X}^{c,n}_{ik_n+j}\widetilde{\epsilon}^n_{ik_n+j}\right\|_p \leq K_p h_n^{-1/2}k_n^{1/2}\Delta_n^{1/2}.$$

By the triangle inequality, we further deduce

$$\begin{aligned}\left\|A_n\sum_{j=0}^{k_n-h_n} \widetilde{X}^{c,n}_{ik_n+j}\widetilde{\epsilon}^n_{ik_n+j}\right\|_p &\leq A_n\sum_{q=0}^{h_n-1}\left\|\sum_{j\in\mathcal{I}^n_{i,q}} \widetilde{X}^{c,n}_{ik_n+j}\widetilde{\epsilon}^n_{ik_n+j}\right\|_p\\ &\leq K_p h_n^{1/2}k_n^{-1/2}.\end{aligned}$$

By using a maximal inequality,

$$\left\|\max_{0\leq i\leq\lfloor n/k_n\rfloor-1}\left|A_n\sum_{j=0}^{k_n-h_n} \widetilde{X}^{c,n}_{ik_n+j}\widetilde{\epsilon}^n_{ik_n+j}\right|\right\|_p \leq K_p n^{\frac{1-\gamma}{p}-\frac{2\gamma-1}{4}}.$$

Taking any fixed $p > 4(1-\gamma)/(2\gamma-1)$ implies that the majorant side of the above estimate is $o(1)$. This further implies the third claim of (5.11).□

LEMMA 3. *Suppose that (i) Assumption 7 holds, and (ii) $k_n \asymp n^{\gamma}$, $h_n \asymp n^{1/2}$, and $u_n \asymp (h_n\Delta_n)^{\varpi}$, where*

$$\gamma \in \left(\frac{6+r}{8}, 1\right), \quad \varpi \in \left[\frac{4(1-\gamma)}{2-r}, \frac{1}{2}\right).$$

Then,

$$\max_{0\le i\le \lfloor n/k_n\rfloor -1}\left|\frac{\sum_{j=0}^{k_n-h_n}\left((\widetilde{Y}^n_{ik_n+j})^2 1_{\{|\widetilde{Y}^n_{ik_n+j}|\le u_n\}} - (\widetilde{Y}^{c,n}_{ik_n+j})^2\right)}{(k_n-h_n+1)h_n\Delta_n\phi_n}\right| = o_p(1).$$

PROOF. Recall the definition of $w_{n,i}(s)$ in the proof of Lemma 2. We can write $\widetilde{Y}^n_i - \widetilde{Y}^{c,n}_i$ as

$$\begin{aligned}\widetilde{Y}^n_i - \widetilde{Y}^{c,n}_i &= \int_{i\Delta_n}^{(i+h_n)\Delta_n}\int_{\mathbb{R}} \delta(s,z)1_{\{|\delta(s,z)|>1\}} w_{n,i}(s)\mu(ds,dz)\\ &\quad + \int_{i\Delta_n}^{(i+h_n)\Delta_n}\int_{\mathbb{R}} \delta(s,z)1_{\{|\delta(s,z)|\le 1\}} w_{n,i}(s)\,(\mu-\nu)\,(ds,dz).\end{aligned}$$

By Corollary 2.1.9 in [23],

$$\begin{aligned}\mathbb{E}\left[\left(\widetilde{Y}^n_i - \widetilde{Y}^{c,n}_i\right)^2 \wedge u_n^2\right] &\le K\,(h_n\Delta_n)^{2\varpi}\,\mathbb{E}\left[\left(\frac{\left|\widetilde{Y}^n_i - \widetilde{Y}^{c,n}_i\right|}{(h_n\Delta_n)^{\varpi}}\wedge 1\right)^2\right]\\ &\le K n^{-\frac{1}{2}-\frac{(2-r)\varpi}{2}} a_n, \end{aligned} \tag{5.24}$$

for some deterministic sequence $a_n = o(1)$. Note that $\mathbb{E}[|\widetilde{Y}^{c,n}_i|^2] \le Kn^{-1/2}$. Therefore, by the Cauchy–Schwarz inequality and the above estimate,

$$\mathbb{E}\left[|\widetilde{Y}^{c,n}_i|\left(\left(\widetilde{Y}^n_i - \widetilde{Y}^{c,n}_i\right)\wedge u_n\right)\right] \le Kn^{-\frac{1}{2}-\frac{(2-r)\varpi}{4}} a_n^{1/2}. \tag{5.25}$$

In addition, we observe that

$$\begin{aligned}\mathbb{E}\left[|\widetilde{Y}^{c,n}_i|^2 1_{\{|\widetilde{Y}^{c,n}_i|>u_n/2\}}\right] &\le K_p\frac{\mathbb{E}\left[|\widetilde{Y}^{c,n}_i|^{p+2}\right]}{u_n^p}\\ &\le K_p n^{-\frac{1}{2}-\frac{p(1/2-\varpi)}{2}}, \end{aligned} \tag{5.26}$$

where the first inequality is by Markov's inequality, and the second inequality follows from $\mathbb{E}[|\widetilde{Y}_i^{c,n}|^{p+2}] \leq K_p n^{-(p+2)/4}$.

Recall the following elementary inequality (see, e.g., equation (47) of [1]): for $x, y \in \mathbb{R}$ and $u \in (0,1)$,

$$\begin{aligned}&\left| |x+y|^2 1_{\{|x+y|\leq u\}} - |x|^2 \right| \\ &\quad \leq K\left(|x|^2 1_{\{|x|>u/2\}} + y^2 \wedge u^2 + |x| \left(|y| \wedge u\right) \right).\end{aligned}$$

Combining this inequality with (5.24), (5.25), and (5.26), and further fixing some $p > (2-r)\varpi/(1-2\varpi)$, we deduce

$$\mathbb{E}\left[\left| (\widetilde{Y}_i^n)^2 1_{\{|\widetilde{Y}_i^n|\leq u_n\}} - (\widetilde{Y}_i^{c,n})^2 \right| \right] \leq K n^{-\frac{1}{2}-\frac{(2-r)\varpi}{4}} a_n',$$

where $a_n' = a_n^{1/2} + n^{-p(1/4-\varpi/2)+(2-r)\varpi/4} = o(1)$. Hence,

$$\begin{aligned}&\mathbb{E}\left[\max_{0\leq i\leq \lfloor n/k_n\rfloor -1} \left| \frac{\sum_{j=0}^{k_n-h_n} \left((\widetilde{Y}_{ik_n+j}^n)^2 1_{\{|\widetilde{Y}_{ik_n+j}^n|\leq u_n\}} - (\widetilde{Y}_{ik_n+j}^{c,n})^2 \right)}{(k_n-h_n+1)h_n\Delta_n\phi_n} \right| \right] \\ &\leq K k_n^{-1} h_n^{-1} n \sum_{i=0}^{\lfloor n/k_n\rfloor -1} \sum_{j=0}^{k_n-h_n} \mathbb{E}\left[(\widetilde{Y}_{ik_n+j}^n)^2 1_{\{|\widetilde{Y}_{ik_n+j}^n|\leq u_n\}} - (\widetilde{Y}_{ik_n+j}^{c,n})^2 \right] \\ &\leq K n^{1-\gamma-\frac{(2-r)\varpi}{4}} a_n'.\end{aligned}$$

Under the maintained assumption on ϖ and γ, the majorant side of the above display can be further bounded by $Ka_n' = o(1)$. The assertion of the lemma then readily follows. □

Proof of Proposition 1. Part (a) of Proposition 1 follows directly from Lemmas 1, 2, and 3. Part (b) is implied by part (a) because of the the right continuity of v and c. □

Proof of Theorems 5 and 6. The proof of Theorem 5 (resp. Theorem 6) is the same as that of Theorem 3 (resp. Theorem 4), except that we use Proposition 1 in place of equation (5.7). □

Acknowledgements. We would like to thank Tim Bollerslev and George Tauchen for helpful discussions.

References.

[1] Aït-Sahalia, Y. and J. Jacod (2009). Testing for jumps in a discretely observed process. *Annals of Statistics 37*, 184–222.

[2] Aït-Sahalia, Y., J. Jacod, and J. Li (2012). Testing for jumps in noisy high frequency data. *Journal of Econometrics 168*(2), 207–222.

[3] Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics 160*(1), 160–175.

[4] Aït-Sahalia, Y. and D. Xiu (2019). Principal component analysis of high-frequency data. *Journal of the American Statistical Association 114*(525), 287–303.

[5] Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2006). A central limit theorem for realised power and bipower variations of continuous semimartingales. In *From Stochastic Calculus to Mathematical Finance*, pp. 33–68. Springer.

[6] Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica 76*(6), 1481–1536.

[7] Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics 162*(2), 149–169.

[8] Barndorff-Nielsen, O. E. and N. Shephard (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(2), 253–280.

[9] Barndorff-Nielsen, O. E. and N. Shephard (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics 17*(5), 457–477.

[10] Barndorff-Nielsen, O. E. and N. Shephard (2004a). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica 72*(3), 885–925.

[11] Barndorff-Nielsen, O. E. and N. Shephard (2004b). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 2*(1), 1–37.

[12] Barndorff-Nielsen, O. E., N. Shephard, and M. Winkel (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Processes and Their Applications 116*(5), 796–806.

[13] Bibinger, M., N. Hautsch, P. Malec, and M. Reiß (2014). Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *Annals of Statistics 42*(4), 1312–1346.

[14] Bibinger, M., N. Hautsch, P. Malec, and M. Reiss (2019). Estimating the spot covariation of asset prices–statistical theory and empirical evidence. *Journal of Business & Economic Statistics 37*(3), 419–435.

[15] Black, F. (1986). Noise. *Journal of Finance 41*(3), 528–543.

[16] Christensen, K., M. Podolskij, and M. Vetter (2013). On covariation estimation for multivariate continuous itô semimartingales with noise in non-synchronous observation schemes. *Journal of Multivariate Analysis 120*, 59–84.

[17] Geman, D. and J. Horowitz (1980). Occupation Densities. *Annals of Probability 8*, 1–67.

[18] Hautsch, N. and M. Podolskij (2013). Preaveraging-based estimation of quadratic variation in the presence of noise and jumps: Theory, implementation, and empirical evidence. *Journal of Business & Economic Statistics 31*(2), 165–183.

[19] Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications 119*(7), 2249–2276.
[20] Jacod, J., Y. Li, and X. Zheng (2017). Statistical properties of microstructure noise. *Econometrica 85*(4), 1133–1174.
[21] Jacod, J., Y. Li, and X. Zheng (2019). Estimating the integrated volatility with tick observations. *Journal of Econometrics 208*(1), 80–100.
[22] Jacod, J., M. Podolskij, and M. Vetter (2010). Limit theorems for moving averages of discretized processes plus noise. *Annals of Statistics 38*(3), 1478–1545.
[23] Jacod, J. and P. Protter (2012). *Discretization of Processes*. Springer.
[24] Jacod, J. and M. Rosenbaum (2013). Quarticity and other functionals of volatility: Efficient estimation. *Annals of Statistics 41*, 1462–1484.
[25] Kalnina, I. and D. Xiu (2016). Nonparametric estimation of the leverage effect: A trade-off between robustness and efficiency. *Journal of the American Statistical Association*.
[26] Li, J., V. Todorov, and G. Tauchen (2013). Volatility occupation times. *Annals of Statistics 41*(4), 1865–1891.
[27] Li, J., V. Todorov, and G. Tauchen (2016a). Adaptive estimation of continuous-time regression models using high-frequency data. *Journal of Econometrics*.
[28] Li, J., V. Todorov, and G. Tauchen (2016b). Estimating the volatility occupation time via regularized laplace inversion. *Econometric Theory 32*(5), 12531288.
[29] Li, J., V. Todorov, and G. Tauchen (2016c). Inference theory on volatility functional dependencies. *Journal of Econometrics 193*, 17–34.
[30] Li, J. and D. Xiu (2016). Generalized method of integrated moments for high-frequency data. *Econometrica 84*(4), 1613–1633.
[31] Li, Y. and P. A. Mykland (2007). Are volatility estimators robust with respect to modeling assumptions? *Bernoulli 13*(3), 601–622.
[32] Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets 3*(3), 205–258.
[33] Mancini, C., V. Mattiussi, and R. Renò (2015). Spot volatility estimation using delta sequences. *Finance and Stochastics 19*(2), 261–293.
[34] Marcus, M. B. and J. Rosen (2006). *Markov Processes, Gaussian Processes, and Local Times*. Cambridge University Press.
[35] Mykland, P. A. and L. Zhang (2006). ANOVA for diffusions and Ito processes. *Annals of Statistics 34*(4), 1931–1963.
[36] Mykland, P. A. and L. Zhang (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica 77*(5), 1403–1445.
[37] Mykland, P. A. and L. Zhang (2017). Assessment of uncertainty in high frequency data: The observed asymptotic variance. *Econometrica 85*(1), 197–231.
[38] Podolskij, M. and M. Vetter (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli 15*(3), 634–658.
[39] Stoll, H. (2003). *Market microstructure, Handbook of Economics and Finance*. Elsevier Science B.V.
[40] Todorov, V. and T. Bollerslev (2010). Jumps and betas: A new framework for disentangling and estimating systematic risks. *Journal of Econometrics 157*(2), 220–235.
[41] Todorov, V. and G. Tauchen (2012). The realized Laplace transform of volatility. *Economtrica 80*, 1105–1127.
[42] van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag.
[43] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

[44] Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics 159*(1), 235–250.

[45] Zhang, C., J. Li, and T. Bollerslev (2018). Occupation density estimation for noisy high-frequency data. Technical report.

[46] Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli 12*(6), 1019–1043.

[47] Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics 160*(1), 33–47.

[48] 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*(472), 1394–1411.

[49] Zu, Y. and H. P. Boswijk (2014). Estimating spot volatility with high-frequency financial data. *Journal of Econometrics 181*(2), 117–135.

Department of Economics
Duke University
Durham, NC 27708-0097
E-mail: jl410@duke.edu

Department of Economics
Duke University
Durham, NC 27708-0097
E-mail: congshan.zhang@duke.edu

Department of Statistics and Operations Research
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
E-mail: yunxiao@live.unc.edu