VOLATILITY COUPLING

By Jean Jacod and Jia Li and Zhipeng Liao

Sorbonne Université, Duke University and UCLA

This paper provides a strong approximation, or coupling, theory for spot volatility estimators formed using high-frequency data. We show that the t-statistic process associated with the nonparametric spot volatility estimator can be strongly approximated by a growing-dimensional vector of independent variables defined as functions of Brownian increments. We use this coupling theory to study the uniform inference for the volatility process in an infill asymptotic setting. Specifically, we propose uniform confidence bands for spot volatility, beta, idiosyncratic variance processes, and their nonlinear transforms. The theory is also applied to address an open question concerning the inference of monotone nonsmooth integrated volatility functionals such as the occupation time and its quantiles.

1. Introduction. During the past two decades, a large literature in time-series analysis has been devoted to estimating the volatility (i.e., diffusion coefficient) of continuous-time Itô semimartingale processes using high-frequency data ([1], [15]). The inference theory is well understood in two scenarios. One concerns the nonparametric estimation of the stochastic volatility at a fixed time point (see, e.g., [13] and [11]), that is, the "spot" volatility. The other pertains to a semiparametric setting for estimating integrated volatility functionals (see, e.g., [2], [3], [4], [26], and [28]). A more recent literature shows that these two problems are tightly related in that the nonparametric spot volatility estimator can be used to construct semiparametrically efficient estimators of general (smooth) integrated volatility functionals; see, for example, [18], [16], [27], [23], and [21].

An open question, however, is how to make *uniform* nonparametric inference for the spot volatility process in a general nonstationary setting. The difficulty for making uniform inference is well understood: The estimation errors of the spot estimators at any two distinct time points are asymptotically independent, and hence, the collection of these estimators cannot admit a functional central limit theorem in the sense of weak convergence or stable convergence in law (see p. 394 in [15] for a detailed discussion). In other words, the uniform inference for the entire volatility process is a non-Donsker

AMS 2000 subject classifications: 60F15, 60G44, 62G20

Keywords and phrases: coupling, high-frequency data, occupation measure, quantiles, semimartingale, uniform inference.

problem. In an earlier contribution, [12] proposes a uniform confidence band for the volatility process based on an extreme-value approximation for Gaussian processes in a setting with stationary volatility, following the classical approach of [6] and [17]. Meanwhile, non-Donsker problems also arise from other statistical contexts, such as the recent literature on the uniform inference in nonparametric and/or high-dimensional settings (see, e.g., [10], [5], [29], and [19]). The key insight from the latter literature is that, in the absence of functional central limit theorems, one can still strongly approximate the nonparametric functional estimator using variables with known finite-sample distributions (typically Gaussian with consistently estimable covariance). The classical device for obtaining the strong approximation is through Yurinskii's coupling.

Set against this background, our contribution in this paper is to establish a uniform inference theory for the spot volatility process, through the development of a new coupling result for the spot volatility estimators. Specifically, we show that the t-statistic process associated with the spot volatility estimator can be strongly approximated by a growing-dimensional vector of independent normalized chi-squared variables. The sup-t statistic can then be strongly approximated by the maximum of these approximating variables. The distribution of the latter is known in finite sample, which permits the feasible computation of critical values and the corresponding uniform confidence bands, without relying on further Gaussian or extreme-value approximations employed in prior work. We also allow for important empirical features such as essentially unrestricted nonstationarity in volatility, and jumps in the price and volatility processes, which are not considered in [12]. Moreover, we extend the theory to a multivariate setting for constructing uniform confidence bands for the beta and idiosyncratic variance processes, which is a new result in the literature.

As a by-product, we show that the uniform confidence band for the spot volatility process can be further used to construct (joint) confidence sets for monotone functionals of the volatility process. One example is the volatility occupation time ([14], [20]), defined as the time spent by the volatility process below a specific level, which is the "realized" analogue of the cumulative distribution function. The inference for this seemingly basic quantity, however, is an open question in the literature to date because it corresponds to integrated volatility functionals with nonsmooth test functions. In contrast, central limit theorems for integrated volatility functionals are available in the current literature for test functions that are three-time continuously differentiable as shown by [16] among others. Moreover, volatility quantiles inverted from the occupation time are also monotone functionals of the volatility process, for which our theory provides feasible inference. Our coupling result may shed light on a broader range of high-frequency inference problems involving growing dimensions, uniform inference, and nonsmooth test functions, which may be studied in future research.

Below, we present the theory in Section 2 and a simulation study in Section 3. All proofs are contained in Section 4. The following notation will be used. We denote by $\mathcal{M}_d$ the set of all $d \times d$ matrices and by $\mathcal{M}_d^+$ the subset of all symmetric positive semi-definite elements of $\mathcal{M}_d$. The d-dimensional identity matrix is denoted by I_d. The notation $\|x\|$ is the Euclidean norm if $x \in \mathbb{R}^d$, and the operator norm if $x \in \mathcal{M}_d$. For a matrix A, we use A_{jk}, $A^\top$, $\operatorname{tr}(A)$ to denote its (j,k) element, transpose, and trace, respectively. For a function $f : \mathcal{M}_d \mapsto \mathbb{R}$, we denote $\partial_{jk} f(A) = \partial f(A) / \partial A_{jk}$ and collect all partial derivatives using $\partial f(\cdot) = [\partial_{jk} f(\cdot)]_{1 \le j,k \le d}$. For two real sequences a_n and b_n, we write $a_n \asymp b_n$ if $a_n / C \le b_n \le C a_n$ for some finite constant $C \ge 1$.

2. Theory. Section 2.1 describes the setting. Section 2.2 shows our main result for coupling the spot covariance estimator. Further extensions regarding the uniform inference for the stochastic beta and idiosyncratic variance processes are provided in Section 2.3. In Section 2.4, we apply the uniform inference theory to construct confidence sets for monotone functionals of the volatility process.

2.1. *The setting.* Suppose that the vector process X is a d-dimensional Itô semimartingale defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \ge 0}, \mathbb{P})$ that can be written as

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \tag{2.1}$$

where the drift process b and the stochastic volatility matrix process σ are optional, taking values in $\mathbb{R}^d$ and $\mathcal{M}_d$, respectively, W is a d-dimensional standard Brownian motion, and J is a pure jump process driven by a homogeneous Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$. In particular, $J_t = \sum_{s \le t} \Delta X_s$, where ΔX_s is the jump size of X at time s. In financial applications, X plays the role of a vector of asset prices, which are routinely modeled as semimartingales.

The process X is observed at discrete times $i\Delta_n$ for $i = 0, 1, \ldots, n$ within the time interval $[0, T]$ with $\Delta_n = T/n$. We consider an infill asymptotic setting with the sample span T fixed and the sampling interval $\Delta_n \to 0$ asymptotically. The ith increment of X is denoted by

$$\Delta_i^n X \equiv X_{i\Delta_n} - X_{(i-1)\Delta_n}, \quad i \in \{1, \ldots, n\}.$$

We are interested in the uniform inference concerning the spot covariance process $c_t \equiv \sigma_t \sigma_t^\top$. To construct the spot estimator, we divide the sample into m_n nonoverlapping blocks. Specifically, we partition $\{1, \ldots, n\} = \cup_{j=1}^{m_n} \mathcal{I}_{n,j}$, with $\mathcal{I}_{n,j}$ collecting the indices of $k_{n,j}$ consecutive increments in the jth block. Correspondingly, we partition the sample span $[0, T] = \cup_{j=1}^{m_n} \mathcal{T}_{n,j}$, where with $t(n,j) \equiv (\min \mathcal{I}_{n,j} - 1) \Delta_n$,

$$\mathcal{T}_{n,j} \equiv \begin{cases} [t(n,j), t(n,j+1)) & \text{if } 1 \leq j < m_n, \\ [t(n,m_n), T] & \text{if } j = m_n. \end{cases}$$

Allowing the local window size $k_{n,j}$ to vary across blocks is convenient for applications, without introducing additional technical difficulty.

To obtain jump-robust volatility estimates, we adopt a standard truncation technique ([24]) using a sequence of truncation threshold u_n satisfying $u_n \asymp \Delta_n^{\varpi}$ for some $\varpi \in (0, 1/2)$. The spot covariance estimator for the jth block is then defined as

$$\hat{c}_{n,j} \equiv \frac{1}{k_{n,j}\Delta_n} \sum_{i \in \mathcal{I}_{n,j}} \Delta_i^n X \Delta_i^n X^\top 1_{\{\|\Delta_i^n X\| \leq u_n\}}. \tag{2.2}$$

The collection of blockwise estimators $(\hat{c}_{n,j})_{1 \leq j \leq m_n}$ serve as the functional estimator for the process $(c_t)_{t \in [0,T]}$. To reflect this idea more clearly in our notation, we identify $(\hat{c}_{n,j})_{1 \leq j \leq m_n}$ with a t-indexed functional estimator $(\hat{c}_{n,t})_{t \in [0,T]}$ by setting

$$\hat{c}_{n,t} \equiv \hat{c}_{n,j}, \quad \text{for} \quad t \in \mathcal{T}_{n,j} \text{ and } j \in \{1, \ldots, m_n\}.$$

2.2. *Uniform strong approximations for spot covariance estimators.* In this subsection, we present our key coupling result for the spot covariance estimator, and describe how to construct uniform confidence band for the stochastic volatility process. We need a few regularity conditions.

Assumption 1. *Suppose that X has the form (2.1) and there exists a sequence $(T_m)_{m \geq 1}$ of stopping times increasing to infinity such that the following conditions hold for each $m \geq 1$:*

(i) for some constant $r \in (0, 1/2)$, $\|b_t\| + \|\sigma_t\| + \int (\|x\|^r \wedge 1) F_t(dx) \leq K_m$, for all $t \in [0, T_m]$ and some constant K_m, where F_t denotes the spot Lévy measure of J;

(ii) for each $p > 0$, $\mathbb{E}[\sup_{t,s \in \mathcal{T}_{n,j}} \|\sigma_{t \wedge T_m} - \sigma_{s \wedge T_m}\|^p] \leq K_{m,p} |t - s|^{p/2}$ for all $1 \leq j \leq m_n$ and some constant $K_{m,p}$.

Assumption 1 imposes some regularity conditions on the underlying processes, which allow for essentially unrestricted nonstationarity and persistence in the price and volatility dynamics, as well as important empirical features such as intraday volatility seasonality and leverage effect (e.g., negatively correlated price and volatility shocks). Specifically, condition (i) imposes local boundedness on various processes plus a degree of jump activity less than $1/2$ for J, and condition (ii) imposes some smoothness on the stochastic volatility process σ_t. These conditions are stronger than those needed for conducting pointwise inference for spot volatility (cf. Chapter 13.3 in [15]), which is our "cost" to pay for making uniform inference. The main restriction is condition (ii), which in particular requires the paths of the volatility process to be Hölder continuous with any index strictly smaller than $1/2$ on each subinterval $\mathcal{T}_{n,j}$. This condition holds if the volatility process behaves as a continuous Itô semimartingale or long-memory process within each subinterval, and it is allowed to jump on the boundary time points between the $\mathcal{T}_{n,j}$ subintervals. The (piecewise) continuity condition is necessary for making uniform inference on the volatility process because, otherwise, the volatility process cannot even be uniformly consistently estimated (see Remark 1 of [21] and additional discussions therein). In economic and financial applications, this setting accommodates an important type of jumps triggered by macroeconomic announcements, for which we can use announcement times to divide the nearby estimation windows. Consistently estimable jump times (see, e.g., Proposition 1 of [22]) can be used for the same purpose.

In applications, we are often interested in certain nonlinear transformations of the spot covariance matrix, say $f(c_t)$, for some smooth function $f : \mathcal{M}_d^+ \mapsto \mathbb{R}$. In the univariate case, one may be interested in the log volatility by taking $f(c) = \log(c)$. An important financial example in the multivariate cases is "beta," which corresponds to $f(c) = c_{12}/c_{11}$, so that $f(c_t)$ is the spot beta of asset 2 with respect to asset 1. We impose the following condition on the $f(\cdot)$ function.

ASSUMPTION 2. *There exists a sequence $(T_m)_{m\geq 1}$ of stopping times increasing to infinity and a sequence $(\mathcal{K}_m)_{m\geq 1}$ of convex compact subsets of $\mathcal{M}_d^+$ such that c_t takes values in $\mathcal{K}_m$ for all $t \in [0, T_m]$ and f is twice continuously differentiable on $\mathcal{K}_m^\eta \equiv \{x \in \mathcal{M}_d : \inf_{y\in\mathcal{K}_m} \|x - y\| < \eta\}$ for some $\eta > 0$.*

Assumption 2 is easy to verify. For example, if $f(\cdot) = \log(\cdot)$, this condition can be verified provided that the c_t process is locally bounded from above and away from zero, with $\mathcal{K}_m = [1/m, m]$. We also note that the stopping times $(T_m)_{m\geq 1}$ can be taken as the same ones in Assumption 1 without loss of generality, because the (component-wise) minimum of two localizing sequences of stopping times is also a localizing sequence.

We are now ready to state our key result for the strong approximation of the functional estimator of the spot covariance process.

THEOREM 1. *Suppose that (i) Assumption 1 holds and (ii)* $k_{n,j} \asymp \Delta_n^{-\rho}$ *uniformly for all* $j \in \{1, \ldots m_n\}$ *and* $u_n \asymp \Delta_n^{\varpi}$ *such that* $\rho \in (r, 1/2)$ *and* $\varpi \in ((1-\rho/2)/(2-r), 1/2)$. *The following statements hold for some constant* $\epsilon > 0$: *(a) With* $U_{n,j} \equiv k_{n,j}^{-1/2} \sum_{i \in \mathcal{I}_{n,j}} \left(\Delta_i^n W \Delta_i^n W^\top / \Delta_n - I_d\right)$ *for each* $1 \leq j \leq m_n$, *we have*

$$\max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \left\| k_{n,j}^{1/2} \left(\hat{c}_{n,t} - c_t\right) - \sigma_{t(n,j)} U_{n,j} \sigma_{t(n,j)}^\top \right\| = o_p\left(\Delta_n^\epsilon\right); \tag{2.3}$$

(b) If Assumption 2 holds in addition, we further have

$$\max_{1 \leq j \leq m_n} \sup_{t \in \mathcal{T}_{n,j}} \Big| k_{n,j}^{1/2} \left(f\left(\hat{c}_{n,t}\right) - f\left(c_t\right)\right) - \operatorname{tr}\left[\partial f\left(c_{t(n,j)}\right) \sigma_{t(n,j)} U_{n,j} \sigma_{t(n,j)}^\top\right] \Big| = o_p\left(\Delta_n^\epsilon\right).$$

Part (a) of Theorem 1 shows that, uniformly over all estimation blocks, the normalized estimation errors (i.e., $k_{n,j}^{1/2}\left(\hat{c}_{n,t} - c_t\right)$) can be approximated by $(\sigma_{t(n,j)} U_{n,j} \sigma_{t(n,j)}^\top)_{1 \leq j \leq m_n}$. Part (b) provides a similar result for $f\left(\hat{c}_{n,t}\right)$. Theorem 1 is the foundation of all our uniform inference procedures discussed below, for which the knowledge that the strong approximation occurs at some polynomial rate (captured by $o_p(\Delta_n^\epsilon)$) is enough.

We note that the coupling error in Theorem 1 stems from three sources: the drift component of X, the time-variation in volatility, and the "residual" error from the jump truncation. [12] proposes an alternative strong approximation with Gaussian coupling variables (see Proposition 2 there). In general, both coupling results hold only approximately. However, in the baseline model with $b_t = J_t = 0$ identically and volatility being block-wise constant, our "approximation" in (2.3) is actually *exact*, whereas [12]'s Gaussian coupling still carries a non-zero approximation error. Hence, in this "weak" sense, the former is more precise than the latter.

The uniform inference across all estimation blocks involves some additional complication. While the $\mathcal{F}_{t(n,j)}$-conditional distribution of the variable $\sigma_{t(n,j)} U_{n,j} \sigma_{t(n,j)}^\top$ is known, we generally cannot characterize the joint distribution of $(\sigma_{t(n,j)} U_{n,j} \sigma_{t(n,j)}^\top)_{1 \leq j \leq m_n}$, because the $\sigma_{t(n,k)}$ variables are heterogeneous in our nonstationary setting and may be dependent on $U_{n,j}$ for $j < k$ when volatility loads on past Brownian price shocks. We address this issue by uniformly pivotalizing the spot estimators. This can be generically achieved in the univariate setting (i.e., $d = 1$) as described in

Theorem 2 below. The situation becomes more intricate in the multivariate case, which is deferred to Section 2.3.

THEOREM 2. *Suppose that the conditions of Theorem 1 hold with $d = 1$, and the process $\partial f(c_t)\, c_t$ is locally bounded away from zero. Then, for some $\epsilon > 0$,*

$$\max_{1 \le j \le m_n} \sup_{t \in \mathcal{T}_{n,j}} |S_{n,t} - U_{n,j}| = o_p(\Delta_n^{\epsilon}), \tag{2.4}$$

where the t-statistic process $S_{n,t}$ is defined by

$$S_{n,t} \equiv \frac{k_{n,j}^{1/2}\,(f(\hat{c}_{n,t}) - f(c_t))}{\partial f(\hat{c}_{n,t})\,\hat{c}_{n,t}}, \quad \text{for} \quad t \in \mathcal{T}_{n,j} \text{ and } j \in \{1, \ldots, m_n\},$$

under the convention that $S_{n,t} = 0$ when $\partial f(\hat{c}_{n,t})\,\hat{c}_{n,t} = 0$.

Theorem 2 establishes the uniform strong approximation of the t-statistic process $S_{n,t}$ associated with the estimation of the $f(c_t)$ process. The $S_{n,t}$ process is asymptotically uniformly pivotal in the sense that the joint distribution of the coupling variables $(U_{n,j})_{1 \le j \le m_n}$ is free of unknown nuisance. Indeed, they are mutually independent and each scaled variable $k_{n,j}^{1/2} U_{n,j} = \sum_{i \in \mathcal{I}_{n,j}} ((\Delta_i^n W)^2 / \Delta_n - 1)$ has a centered chi-squared distribution with degree of freedom $k_{n,j}$.

The strong approximation in (2.4) for the t-statistic process can be used to construct uniform inference for the $f(c_t)$ process. Specifically, we consider the sup-t statistic given by

$$S_n^* \equiv \sup_{t \in [0,T]} \left|S_{n,t}\right|.$$

Theorem 2 implies that

$$S_n^* - \max_{1 \le j \le m_n} |U_{n,j}| = o_p(\Delta_n^{\epsilon}).$$

This approximation naturally suggests that quantiles of $\max_{1 \le i \le m_n} |U_{n,j}|$, which are known in finite samples, can be used as critical values for the sup-t statistic.

This intuition, however, needs to be formalized theoretically with care. The relevant technicality is that $\max_{1 \le i \le m_n} |U_{n,j}|$ is the maximum of a *growing* number of variables, and its quantiles are generally divergent as $n \to \infty$. This is unlike the conventional (low-dimensional) statistical setting in which critical values defined as quantiles of the limiting distribution are finite. This issue is addressed in our proofs by using an anti-concentration inequality for divergent random sequences. Theorem 3, below, describes the uniform confidence band for the $f(c_t)$ process, and justifies its asymptotic validity.

THEOREM 3. *Suppose that the conditions in Theorem 2 hold. For any* $\alpha \in (0, 1/2)$, *with* κ_n^* *being the* $1-\alpha$ *quantile of* $\max_{1\leq j\leq m_n} |U_{n,j}|$, *we have* $\mathbb{P}(S_n^* \leq \kappa_n^*) \to 1-\alpha$. *Consequently, the confidence band defined by* $B_{n,t} \equiv [B_{n,t}^-, B_{n,t}^+]$,

$$B_{n,t}^{\pm} \equiv f(\hat{c}_{n,t}) \pm k_{n,j}^{-1/2} \kappa_n^* \partial f(\hat{c}_{n,t})\, \hat{c}_{n,t}, \quad \textit{for} \quad t \in \mathcal{T}_{n,j} \textit{ and } j \in \{1, \ldots, m_n\},$$

satisfies $\mathbb{P}(f(c_t) \in B_{n,t} \textit{ for all } t \in [0,T]) \to 1-\alpha$.

Theorem 3 shows that $(B_{n,t})_{t\in[0,T]}$ forms an asymptotically valid uniform confidence band with nominal level $1-\alpha$. A particularly interesting example is $f(c) = \log(c)$, which corresponds to $\partial f(c)\, c = 1$. In this case, there is no need to estimate the standard error and the uniform confidence band is given by $[\log(\hat{c}_{n,t}) - k_{n,j}^{-1/2}\kappa_n^*, \log(\hat{c}_{n,t}) + k_{n,j}^{-1/2}\kappa_n^*]$ when $t \in \mathcal{T}_{n,j}$, which is equally sized over time. We recommend using this confidence band in practice.

We close this subsection with some remarks on related work. The growing-dimensional strong approximation has been used in recent work on high-dimensional and/or uniform nonparametric inference; see, for example, [10], [8], [5], [29], and [19]. To our knowledge, the present paper is the first one that uses this type of technique in the nonstationary nonergodic high-frequency setting. It is also interesting to note that the aforementioned papers are all based on strong Gaussian approximation (e.g., implied by Yurinskii's coupling), whereas our coupling variables, $(U_{n,j})_{1\leq j\leq m_n}$, are normalized chi-squared. The known finite-sample distribution of these variables allows us to compute critical values without resorting to the additional Gaussian approximation. The uniform inference for spot volatility has been studied in [12]. Following the insight of [6] and [17], [12] proposes a uniform confidence band based on an extreme-value theory (without pivotalizing the estimator). In contrast, we do not rely on the extreme-value theory to approximate the distribution of the sup-t statistic. In addition, [12]'s theory rules out jumps in X and the volatility process, and requires the volatility process to be stationary (see Theorem 2 there). Stationarity would rule out a basic model with time-varying deterministic volatility, or more generally, the well-known U-shaped intraday pattern in volatility. Our theory is established in a more general setting without these restrictions.

2.3. *Uniform inference for spot beta and idiosyncratic variance.* In this subsection, we discuss how to feasibly conduct uniform inference for the $f(c_t)$ process in the multivariate case (i.e., $d > 1$). As discussed in the previous subsection, we need to uniformly pivotalize the estimator $f(\hat{c}_{n,t})$, which can be easily done in the univariate case as shown in Theorem 2. In the multivariate case, we need the following condition on the transform $f(\cdot)$.

Assumption 3. *(i) There exists $1 \leq k^* \leq l^* \leq d$ such that, for all $1 \leq k \leq l \leq d$,*

$$\sum_{i,j=1}^{d} \partial_{ij} f(c_t)(\sigma_{ik,t}\sigma_{jl,t} + \sigma_{il,t}\sigma_{jk,t}) = 0 \quad \textit{when} \quad (k,l) \neq (k^*, l^*). \tag{2.5}$$

(ii) The process $F(c_t) \equiv \sum_{i,j,k,l=1}^{d} \partial_{ij} f(c_t)\,\sigma_{ik,t}\sigma_{jl,t}$ is locally bounded away from zero.

Assumption 3 ensures that the approximating variable described in Theorem 1 loads on a single source of randomness, meaning that we can rewrite it as

$$\mathrm{tr}[\partial f\left(c_{t(n,j)}\right)\sigma_{t(n,j)}U_{n,j}\sigma_{t(n,j)}^{\top}] = F\left(c_{t(n,j)}\right)U_{n,j}^{*}, \tag{2.6}$$

where $U_{n,j}^{*}$ is the (k^*, l^*) element of $U_{n,j}$. Importantly, the coupling variable only depends on the scalar-valued random variable $U_{n,j}^{*}$ instead of the $U_{n,j}$ matrix itself. It is then clear that uniform pivotalization can be attained via the t-statistic defined by

$$S_{n,t} \equiv \frac{k_{n,j}^{1/2}\left(f(\hat{c}_{n,t}) - f(c_t)\right)}{F(\hat{c}_{n,t})}, \quad \text{for} \quad t \in \mathcal{T}_{n,j} \text{ and } j \in \{1, \ldots, m_n\}. \tag{2.7}$$

The seemingly peculiar restriction (2.5) is in fact valid for several important transformations. We illustrate this concretely with the following examples.

Example 1 (Beta and Idiosyncratic Variance). Consider a bivariate setting with the diffusive shocks represented by

$$dX_t^c = \sigma_t dW_t = \begin{pmatrix} v_t^{1/2} & 0 \\ \beta_t v_t^{1/2} & \varsigma_t^{1/2} \end{pmatrix} \begin{pmatrix} dW_{1,t} \\ dW_{2,t} \end{pmatrix}, \tag{2.8}$$

where X^c denotes the continuous martingale part of X. In a financial context, denoting $X = (M, Y)$, we can interpret v_t as the spot variance of the market portfolio M, β_t as the beta of asset price Y with respect to the market, and ς_t as the idiosyncratic variance of Y. The matrix-valued process σ_t is informationally equivalent to $(v_t, \beta_t, \varsigma_t)$. The latter scalar-valued processes are respectively associated with the following transformations:

$$f_v(c_t) = v_t = c_{11,t}, \quad f_\beta(c_t) = \beta_t = \frac{c_{12,t}}{c_{11,t}}, \quad f_\varsigma(c_t) = \varsigma_t = c_{22,t} - \frac{c_{12,t}^2}{c_{11,t}},$$

and the corresponding $F(\cdot)$ functions in Assumption 3 are given by $F_v(c_t) \equiv v_t$, $F_\beta(c_t) \equiv \sqrt{\varsigma_t / v_t}$, and $F_\varsigma(c_t) \equiv \varsigma_t$. By elementary calculation, it is easy to see that f_v, f_β, and f_ς satisfy Assumption 3(i) for (k^*, l^*) being $(1,1)$, $(1,2)$, and $(2,2)$, and the normalized coupling variable (recall (2.6)) can be written explicitly as

$$U_{n,j}^* = \begin{cases} k_{n,j}^{-1/2} \sum_{i \in \mathcal{I}_{n,j}} \left(\frac{(\Delta_i^n W_1)^2}{\Delta_n} - 1 \right) & \text{when } f = f_v, \\ k_{n,j}^{-1/2} \sum_{i \in \mathcal{I}_{n,j}} \frac{(\Delta_i^n W_1)(\Delta_i^n W_2)}{\Delta_n} & \text{when } f = f_\beta, \\ k_{n,j}^{-1/2} \sum_{i \in \mathcal{I}_{n,j}} \left(\frac{(\Delta_i^n W_2)^2}{\Delta_n} - 1 \right) & \text{when } f = f_\varsigma. \end{cases} \tag{2.9}$$

□

The setting of Example 1 can be viewed equivalently as a continuous-time regression:

$$dY_t^c = \beta_t dM_t^c + d\epsilon_t, \tag{2.10}$$

with $dM_t^c = v_t^{1/2} dW_{1,t}$ and $d\epsilon_t = \varsigma_t^{1/2} dW_{2,t}$. In particular, the independence between W_1 and W_2 ensures the orthogonality in the martingale sense between M^c and ϵ, which in turn permits the identification of β_t.

The univariate regression may be extended to the case with multiple regressors. To fix ideas, consider an additional regressor process Z in the following model:

$$dY_t^c = \beta_{1,t} dM_t^c + \beta_{2,t} dZ_t^c + d\epsilon_t, \tag{2.11}$$

under the orthogonality condition that $\langle M^c, \epsilon \rangle = 0$ and $\langle Z^c, \epsilon \rangle = 0$, where $\langle \cdot, \cdot \rangle$ denotes the angle bracket of local martingales. Obviously, if M^c is also orthogonal to Z^c (i.e., Z is "market neutral"), we can conduct analysis for $\beta_{1,t}$ as in the univariate setting (2.10) by rewriting the model as $dY_t^c = \beta_{1,t} dM_t^c + d\tilde{\epsilon}_t$ with $d\tilde{\epsilon}_t = \beta_{2,t} dZ_t^c + d\epsilon_t$.

The situation becomes more complicated when Z^c is not orthogonal to M^c. In this case, we consider the following decomposition of Z^c:

$$dZ_t^c = \gamma_t dM_t^c + d\tilde{Z}_t, \tag{2.12}$$

such that M^c and $\tilde{Z}$ are orthogonal in the martingale sense. Plugging (2.12) into (2.11), we can rewrite the latter as

$$dY_t^c = \bar{\beta}_{1,t} dM_t^c + \beta_{2,t} d\tilde{Z}_t + d\epsilon_t, \quad \text{where} \quad \bar{\beta}_{1,t} = \beta_{1,t} + \beta_{2,t}\gamma_t.$$

In this alternative model, the regressor processes M^c and $\tilde{Z}$ are orthogonal, with $\bar{\beta}_{1,t}$ capturing the "total" effect of the dM_t^c shock on dY_t^c. The betas can be identified as

$$\bar{\beta}_{1,t} = \frac{d\langle Y^c, M^c\rangle_t}{d\langle M^c, M^c\rangle_t}, \quad \beta_{2,t} = \frac{d\langle Y^c, \tilde{Z}\rangle_t}{d\langle \tilde{Z}, \tilde{Z}\rangle_t}.$$

As discussed in the previous paragraph, the analysis for $\bar{\beta}_{1,t}$ is the same as the univariate regression in Example 1. The nontrivial part is the analysis for $\beta_{2,t}$, because the residual process $\tilde{Z}$ is not directly observed. That being said, we can still identify $\beta_{2,t}$ from the spot covariance matrix of $X = (M, Z, Y)$. Indeed, we have $\gamma_t = c_{12,t}/c_{11,t}$,

$$\begin{aligned} d\langle Y^c, \tilde{Z}\rangle_t &= d\left\langle Y^c, Z^c\right\rangle_t - \gamma_t d\left\langle Y^c, M^c\right\rangle_t = \left(c_{23,t} - \frac{c_{12,t}}{c_{11,t}} c_{13,t}\right) dt, \\ d\langle \tilde{Z}, \tilde{Z}\rangle_t &= \left(c_{22,t} - \frac{c_{12,t}^2}{c_{11,t}}\right) dt, \end{aligned}$$

yielding

$$\beta_{2,t} = f_{\beta_2}(c_t) \equiv \frac{c_{11,t}c_{23,t} - c_{12,t}c_{13,t}}{c_{11,t}c_{22,t} - c_{12,t}^2}.$$

By direct calculation, we can verify that the function $f_{\beta_2}(\cdot)$ satisfies Assumption 3 with $(k^*, l^*) = (2, 3)$, corresponding to $F_{\beta_2}(c_t) \equiv \sqrt{\varsigma_t/\tilde{\varsigma}_t}$, where

$$\tilde{\varsigma}_t \equiv c_{22,t} - \frac{c_{12,t}^2}{c_{11,t}}, \quad \varsigma_t \equiv c_{33,t} - \frac{c_{13,t}^2}{c_{11,t}} - \beta_{2,t}^2 \tilde{\varsigma}_t,$$

which are the spot variance processes of $\tilde{Z}$ and ϵ, respectively.

Example 2 (Permanence). Assumption 3 satisfies a permanence property with respect to smooth transformations in the following sense. Suppose that

$g : \mathcal{M}_d^+ \mapsto \mathbb{R}$ satisfies Assumption 3 with $G(c_t) \equiv \sum_{i,j,k,l=1}^{d} \partial_{ij} g(c_t) \sigma_{ik,t}\sigma_{jl,t}$. Consider a continuously differentiable function h and define $f = h \circ g$. Since $\partial_{ij} f(c_t) = \partial h(g(c_t)) \partial_{ij} g(c_t)$, it is easy to see that f satisfies condition (i) as well. Condition (ii) can be verified if the process $F(c_t) = \partial h(g(c_t)) G(c_t)$ is locally bounded away from zero. □

Theorem 4, below, extends the results in Theorems 2 and 3 to a multivariate case under Assumption 3. Recall that the sup-t statistic is denoted by $S_n^* = \sup_{t\in[0,T]} |S_{n,t}|$.

THEOREM 4. *Under the conditions of Theorem 1 and Assumption 3, the following statements hold: (a) We have* $S_n^* - \sup_{1\le j\le m_n} |U_{n,j}^*| = o_p(\Delta_n^{\epsilon})$ *for some* $\epsilon > 0$, *where* $U_{n,j}^*$ *is the* (k^*, l^*) *element of* $U_{n,j}$*; (b) For any* $\alpha \in (0, 1/2)$, *with* κ_n^* *being the* $1-\alpha$ *quantile of* $\max_{1\le j\le m_n} |U_{n,j}^*|$, *we have* $\mathbb{P}(S_n^* \le \kappa_n^*) \to 1-\alpha$. *Consequently, the confidence band defined by* $B_{n,t} \equiv [B_{n,t}^-, B_{n,t}^+]$,

$$B_{n,t}^{\pm} \equiv f(\hat{c}_{n,t}) \pm k_{n,j}^{-1/2} \kappa_n^* F(\hat{c}_{n,t}), \quad \textit{for} \quad t \in \mathcal{T}_{n,j} \textit{ and } j \in \{1, \ldots, m_n\},$$

satisfies $\mathbb{P}(f(c_t) \in B_{n,t} \textit{ for all } t \in [0,T]) \to 1-\alpha$.

To guide application, we note that the critical value κ_n^* is known in finite samples and can be computed via Monte Carlo simulation. We illustrate this point with the running example.

EXAMPLE 1 (CONTINUED). We can simulate independent standard normal variables $\mathcal{N}_{1,i}$ and $\mathcal{N}_{2,i}$ and construct an identical copy (in distribution) of $U_{n,j}^*$ as

$$\widetilde{U}_{n,j}^* \equiv \begin{cases} k_{n,j}^{-1/2} \sum_{i\in\mathcal{I}_{n,j}} \left(\mathcal{N}_{1,i}^2 - 1\right) & \text{when } f = f_v, \\ k_{n,j}^{-1/2} \sum_{i\in\mathcal{I}_{n,j}} \mathcal{N}_{1,i}\mathcal{N}_{2,i} & \text{when } f = f_\beta, \\ k_{n,j}^{-1/2} \sum_{i\in\mathcal{I}_{n,j}} \left(\mathcal{N}_{2,i}^2 - 1\right) & \text{when } f = f_\varsigma. \end{cases}$$

The critical value κ_n^* can then be computed as the $1-\alpha$ quantile of the variable $\max_{1\le j\le m_n} |\widetilde{U}_{n,j}^*|$ from a large number of simulations. □

2.4. *Inference for monotone functionals of volatility.* The uniform confidence band studied above may be used to conduct other types of inference. An interesting application concerns monotone functionals. Let $\mathcal{H}$ denote the space of real-valued Borel measurable functions on $[0,T]$. We call a functional $\mathbb{F} : \mathcal{H} \mapsto \mathbb{R}$ increasing if $f(\cdot) \leq g(\cdot)$ implies $\mathbb{F}(f) \leq \mathbb{F}(g)$, and use $\mathbb{M}$ to collect all increasing functionals. By monotonicity,

$$\begin{aligned}
&\liminf_{n\to\infty} \mathbb{P}\left(\mathbb{F}\left(B_n^-\right) \leq \mathbb{F}\left(f\left(c\right)\right) \leq \mathbb{F}\left(B_n^+\right) \text{ for all } \mathbb{F} \in \mathbb{M}\right) \\
&\quad \geq \lim_{n\to\infty} \mathbb{P}\left(B_{n,t}^- \leq f\left(c_t\right) \leq B_{n,t}^+ \text{ for all } t \in [0,T]\right) = 1-\alpha,
\end{aligned} \tag{2.13}$$

where the convergence follows from Theorem 4. That is, $[\mathbb{F}\left(B_n^-\right), \mathbb{F}\left(B_n^+\right)]$ forms a confidence interval for the functional $\mathbb{F}\left(f\left(c\right)\right)$ with asymptotic coverage rate at least $1-\alpha$.

This simple implication addresses an open question regarding the high-frequency inference for occupation times ([20]); see [14] and [25] for additional background. Recall that the occupation time for the $f\left(c_t\right)$ process is defined as follows:

$$\mathbb{F}_x\left(f\left(c\right)\right) = \int_0^T 1_{\{f(c_s)\leq x\}} ds, \quad \text{for} \quad x \in \mathbb{R},$$

which measures the amount of time when the $f\left(c_t\right)$ process is below a certain value x over the time interval $[0,T]$. Obviously, the occupation time is the "realized" analogue of the cumulative distribution function $x \mapsto \mathbb{P}\left(f\left(c_t\right) \leq x\right)$. Correspondingly, we can define the occupational quantile of the $f\left(c_t\right)$ process as the functional inverse of the occupation time, that is,

$$\mathbb{Q}_q\left(f\left(c\right)\right) = \inf\{x \in \mathbb{R} : \mathbb{F}_x\left(f\left(c\right)\right) \geq qT\} \quad \text{for} \quad q \in (0,1).$$

The occupation time and quantile functionals capture the full "distributional" feature of the underlying process.

[20] first study the volatility occupation time by establishing consistency and rate of convergence results. The related inference problem, however, remains to be an open question. Many recent papers study the inference for integrated volatility functionals of the form $\int_0^T g\left(c_s\right) ds$ for some smooth (more precisely, three-time continuously differentiable) function $g\left(\cdot\right)$; see, for example, [18], [16], [23], and [21]. However, those theories are not applicable here because the occupation time corresponds to a nonsmooth test function of the form $1_{\{f(\cdot)\leq x\}}$. Using our uniform inference theory, we instead exploit the fact that $\mathbb{F}_x$ and $\mathbb{Q}_q$ are monotone functionals. In fact, $-\mathbb{F}_x$ and $\mathbb{Q}_q$ belong to $\mathbb{M}$ for all $x \in \mathbb{R}$ and $q \in (0,1)$, so we can conduct feasible inference on the basis of (2.13). This result is summarized in the following corollary.

Corollary 1. *Under the conditions of Theorem 4, we have*

$$\liminf_{n\to\infty} \mathbb{P}\left(\Omega_{1,n} \cap \Omega_{2,n}\right) \geq 1-\alpha,$$

where

$$\begin{aligned}\Omega_{1,n} &\equiv \{\mathbb{F}_x\left(B_n^+\right) \leq \mathbb{F}_x\left(f\left(c\right)\right) \leq \mathbb{F}_x\left(B_n^-\right) \text{ for all } x \in \mathbb{R}\},\\ \Omega_{2,n} &\equiv \{\mathbb{Q}_q(B_n^-) \leq \mathbb{Q}_q\left(f\left(c\right)\right) \leq \mathbb{Q}_q(B_n^+) \text{ for all } q \in (0,1)\}.\end{aligned}$$

This corollary shows that $[\mathbb{F}_x\left(B_n^+\right), \mathbb{F}_x\left(B_n^-\right)]_{x\in\mathbb{R}}$ provides a uniform confidence band for the occupation time function $x \mapsto \mathbb{F}_x(f\left(c\right))$. In addition, $[\mathbb{Q}_q(B_n^-), \mathbb{Q}_q(B_n^+)]_{q\in(0,1)}$ provides a uniform confidence band for the occupation quantile function $q \mapsto \mathbb{Q}_q(f\left(c\right))$. These confidence bands are valid simultaneously, but they may be conservative. Specifically in the context of Example 1, this result can be used to make inference about the occupation times and occupation quantiles of spot variance, beta, and idiosyncratic variance.

3. Simulations. In this section, we examine the finite-sample performance of the proposed inference method in a Monte Carlo experiment. Since the pointwise inference of spot volatility has been well studied in the literature, we focus on the uniform inference for the volatility process. Below, we fix the sample span to be $T = 1$ day.

We simulate the log price process from $dX_t = \sigma_t dW_t$. To simulate the volatility process, we follow [7] and generate two volatility factors, $V_{1,t}$ and $V_{2,t}$, from the following model:

$$\begin{aligned}dV_{1,t} &= 0.0128(0.4068 - V_{1,t})dt + 0.0954\sqrt{V_{1,t}}\left(\rho dW_t + \sqrt{1-\rho^2}dB_{1,t}\right),\\ dV_{2,t} &= 0.6930(0.4068 - V_{2,t})dt + 0.7023\sqrt{V_{2,t}}\left(\rho dW_t + \sqrt{1-\rho^2}dB_{2,t}\right),\end{aligned}$$

where B_1 and B_2 are independent standard Brownian motions that are also independent of W. The $\rho = -0.7$ parameter captures the well-documented negative correlation between price and volatility shocks (i.e., the "leverage" effect). The V_1 volatility factor is highly persistent with a half-life of 2.5 months, while the V_2 volatility factor is quickly mean-reverting with a half-life of only one day. We simulate the continuous-time processes using an Euler scheme on a 1-second mesh, and the observed returns actually used in the calculations are sampled at $\Delta_n = 1$ minute intervals. There are 390 returns within the trading day.

We then consider two models for the volatility process, with $\sigma_t^2 = 2V_{1,t}$ and $\sigma_t^2 = V_{1,t} + V_{2,t}$, which will be referred to as the one-factor and two-factor models, respectively. Under the one-factor model, the sample path of the "slow" volatility factor is relatively smooth, and hence, results in a relatively small nonparametric estimation bias. The two-factor model, on the other hand, presents a nontrivial challenge to spot volatility estimation, in that the volatility process can vary considerably within a day.

We implement the uniform confidence band described in Theorem 3 with $f(\cdot) = \log(\cdot)$, as recommended before. Recall that the uniform critical value (CV), κ_n^*, is defined as the $1-\alpha$ quantile of $\max_{1\leq j\leq m_n}|U_{n,j}|$. For example, when $k_n = 30$, there are $m_n = 13$ estimation blocks within a day, and $\kappa_n^* \approx 3.97$ (resp. 4.47) when $\alpha = 0.1$ (resp. 0.05). To highlight the difference between the uniform CV and the pointwise CV, we also examine the uniform coverage property of a confidence band based on the pointwise CV, say $\tilde{\kappa}_n^*$, which is defined as the $1-\alpha$ quantile of $|U_{n,j}|$; recall that the distribution of $U_{n,j}$ does not depend on j. Not surprisingly, $\tilde{\kappa}_n^*$ is generally much smaller than κ_n^*. For example, when $k_n = 30$, $\tilde{\kappa}_n^* \approx 2.27$ (resp. 2.72) when $\alpha = 0.1$ (resp. 0.05). Therefore, we expect the pointwise band based on $\tilde{\kappa}_n^*$ to have severe under-coverage. Finally, we implement the uniform confidence band proposed by [12], which employs a CV based on extreme-value theory.

In each setting, we consider two confidence levels: 90% and 95%. We also consider a range of block sizes $k_n \in \{10, 15, 26, 30\}$. These block sizes are chosen so that the 390 observations in each day can be divided into equally sized blocks. Note that the largest block size is three times larger than the smallest one, thus providing a meaningful robustness check regarding the "bandwidth" parameter k_n.

Table 1 reports the finite-sample coverage rates of the three versions of confidence bands (i.e., Uniform, Pointwise, and Extreme-Value), calculated based on 10,000 Monte Carlo replications. The top panel presents the results for the one-factor model. From the first two columns, we see that the coverage rates of the proposed uniform confidence bands are generally close to the nominal confidence levels. The size control is particularly good in the "undersmoothing" case with relatively small k_n, which mitigates the nonparametric estimation bias stemming from time-varying volatility. In sharp contrast, the pointwise confidence band suffers from severe size distortion, as shown on columns 3 and 4 of the table. This finding is expected and highlights concretely the distinction between the uniform and pointwise confidence bands. Looking at the last two columns of the table, we see that the extreme-value-based confidence band of [12] also bears nontrivial size distortion. This suggests that the extreme-value asymptotic approximation has not "kicked in." This may also reflect the fact that the assumption of stationary volatility underlying [12]'s theory does not capture well the "realized heterogeneity" in the volatility path within a relatively short sample.

Finally, we turn to the results with the two-factor volatility dynamics shown on the bottom panel of Table 1. The performance of all methods de-

TABLE 1
Finite-sample Coverage Rates of Confidence Bands

	Uniform		Pointwise		Extreme-Value	
k_n	90%	95%	90%	95%	90%	95%
Panel A: One-factor Volatility Model						
10	89.1	94.9	0.2	2.9	19.1	25.1
15	89.7	95.7	1.7	12.2	33.1	40.6
26	86.8	94.1	9.9	30.6	49.7	58.4
30	86.6	93.7	14.0	36.0	54.5	62.8
Panel B: Two-factor Volatility Model						
10	87.5	94.2	0.0	1.9	25.8	31.2
15	86.2	93.6	0.4	5.6	34.8	41.1
26	78.1	89.7	2.2	11.7	43.8	50.5
30	73.8	87.7	2.4	12.6	43.6	50.5

Note: The table reports the uniform coverage rates (in percentages) of three versions of confidence bands for the spot variance process $(c_t)_{t\in[0,1]}$. The first version is based on the uniform critical value (CV) described in Theorem 3 for $f(\cdot) = \log(\cdot)$. The second version instead uses the pointwise CV. The third version is the confidence band proposed by [12], which is based on an extreme-value approximation. We report results for two confidence levels, 90% and 95%, and various block sizes, $k_n \in \{10, 15, 26, 30\}$. The coverage rates are computed based on 10,000 Monte Carlo replications.

teriorates under this more challenging data generating process, as it features much rougher volatility paths. That being said, we still see clearly that the proposed uniform confidence band outperforms the pointwise band and the extreme-value-based band, with adequate performance in the undersmoothing case.

4. Proofs. Throughout the proofs, we use K to denote a positive constant that may change from line to line, and write K_p to emphasize its dependence on some parameter p. For $p \geq 1$, we use $\|\cdot\|_p$ to denote the L_p-norm of a random variable. In addition, by a standard localization procedure, we can strengthen Assumptions 1 and 2 by assuming that they hold with $T_1 = \infty$ without loss of generality; see Section 4.4.1 in [15] for details on localization.

PROOF OF THEOREM 1. (a) Fix any constant

$$\epsilon \in (0, (1/2-\rho) \wedge ((2-r)\,\varpi + \rho/2 - 1)),$$

which is possible given condition (ii) of the theorem. Denote $X' = X - J$. For each $j \in \{1, \ldots, m_n\}$, we set $\hat{c}'_{n,j} \equiv k_{n,j}^{-1}\Delta_n^{-1}\sum_{i\in\mathcal{I}_{n,j}} \Delta_i^n X' \Delta_i^n X'^{\top}$. Our proof relies on the following decomposition:

$$k_{n,j}^{1/2}\left(\hat{c}_{n,j}-c_t\right)-\sigma_{t(n,j)}U_{n,j}\sigma_{t(n,j)}^{\top}=k_{n,j}^{1/2}(c_{t(n,j)}-c_t)+R_{n,j}+R'_{n,j}, \tag{4.1}$$

where

$$\begin{aligned}
R_{n,j} &\equiv k_{n,j}^{1/2}\left(\hat{c}_{n,j}-\hat{c}'_{n,j}\right), \\
R'_{n,j} &\equiv k_{n,j}^{1/2}\left(\hat{c}'_{n,j}-\sigma_{t(n,j)}\left(\frac{1}{k_{n,j}\Delta_n}\sum_{i\in\mathcal{I}_{n,j}}\Delta_i^n W\Delta_i^n W^{\top}\right)\sigma_{t(n,j)}^{\top}\right).
\end{aligned}$$

By Assumption 1 and a maximal inequality, we have for any $p\geq 1$,

$$\left\|\max_{1\leq j\leq m_n}\sup_{t\in\mathcal{T}_{n,j}}\left\|k_{n,j}^{1/2}(c_{t(n,j)}-c_t)\right\|\right\|_p\leq K_p m_n^{1/p}\Delta_n^{1/2-\rho}.$$

With $p>(1-\rho)/(1/2-\rho-\epsilon)$, the right-hand side of the above estimate is $o\left(\Delta_n^{\epsilon}\right)$. Hence,

$$\max_{1\leq j\leq m_n}\sup_{t\in\mathcal{T}_{n,j}}\left\|k_{n,j}^{1/2}(c_{t(n,j)}-c_t)\right\|=o_p\left(\Delta_n^{\epsilon}\right). \tag{4.2}$$

We rewrite $R_{n,j}=\sum_{i\in\mathcal{I}_{n,j}}\zeta_{n,i}$ where for each $i\in\mathcal{I}_{n,j}$,

$$\zeta_{n,i}\equiv\frac{1}{k_{n,j}^{1/2}\Delta_n}\left(\Delta_i^n X\Delta_i^n X^{\top}1_{\{\|\Delta_i^n X\|\leq u_n\}}-\Delta_i^n X'\Delta_i^n X'^{\top}\right).$$

By some known estimates (see p. 1476 in [16]),

$$\mathbb{E}\left[\left\|\Delta_i^n X\Delta_i^n X^{\top}1_{\{\|\Delta_i^n X\|\leq u_n\}}-\Delta_i^n X'\Delta_i^n X'^{\top}\right\|\right]\leq K\Delta_n^{(2-r)\varpi+1},$$

which further implies $\mathbb{E}\left[\|\zeta_{n,i}\|\right]\leq K\Delta_n^{(2-r)\varpi+(\rho/2)}$. Note that $\|R_{n,j}\|\leq\sum_{i=1}^{[T/\Delta_n]}\|\zeta_{n,i}\|$ uniformly in j. Hence,

$$\mathbb{E}\left[\max_{1\leq i\leq m_n}\|R_{n,j}\|\right]\leq K\Delta_n^{(2-r)\varpi+(\rho/2)-1},$$

yielding

$$\max_{1\leq i\leq m_n}\|R_{n,j}\|=O_p\left(\Delta_n^{(2-r)\varpi+(\rho/2)-1}\right)=o_p\left(\Delta_n^{\epsilon}\right). \tag{4.3}$$

Finally, we consider $R'_{n,j}$. Define for each $j \in \{1, \ldots, m_n\}$ and each $i \in \mathcal{I}_{n,j}$,

$$\xi_{n,i} \equiv \Delta_i^n X' - \sigma_{t(n,j)} \Delta_i^n W = \int_{(i-1)\Delta_n}^{i\Delta_n} b_s ds + \int_{(i-1)\Delta_n}^{i\Delta_n} \left(\sigma_s - \sigma_{t(n,j)}\right) dW_s.$$

We can then rewrite

$$R'_{n,j} = \frac{1}{k_{n,j}^{1/2} \Delta_n} \sum_{i \in \mathcal{I}_{n,j}} \left(\xi_{n,i} \xi_{n,i}^\top + \xi_{n,i} \left(\sigma_{t(n,j)} \Delta_i^n W\right)^\top + \left(\sigma_{t(n,j)} \Delta_i^n W\right) \xi_{n,i}^\top \right).$$

By the Burkholder–Davis–Gundy inequality, Hölder's inequality, and Assumption 1, we have for any $p \geq 2$,

$$\begin{aligned}
&\mathbb{E}\left[\left\| \int_{(i-1)\Delta_n}^{i\Delta_n} \left(\sigma_s - \sigma_{t(n,j)}\right) dW_s \right\|^p \right] \\
&\quad \leq K_p \mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \left\| \sigma_s - \sigma_{t(n,j)} \right\|^2 ds \right)^{p/2} \right] \\
&\quad \leq K_p \Delta_n^{p/2-1} \mathbb{E}\left[\int_{(i-1)\Delta_n}^{i\Delta_n} \left\| \sigma_s - \sigma_{t(n,j)} \right\|^p ds \right] \\
&\quad \leq K_p k_{n,j}^{p/2} \Delta_n^p \leq K_p \Delta_n^{p(1-\rho/2)}.
\end{aligned}$$

It is then easy to see that

$$\|\xi_{n,i}\|_p \leq \left\| \int_{(i-1)\Delta_n}^{i\Delta_n} b_s ds \right\|_p + \left\| \int_{(i-1)\Delta_n}^{i\Delta_n} \left(\sigma_s - \sigma_{t(n,j)}\right) dW_s \right\|_p \leq K_p \Delta_n^{1-\rho/2}.$$

This estimate further implies that

$$\left\| \xi_{n,i} \xi_{n,i}^\top \right\|_p \leq K_p \Delta_n^{2-\rho}, \quad \left\| \xi_{n,i} \left(\sigma_{t(n,j)} \Delta_i^n W\right)^\top \right\|_p \leq K_p \Delta_n^{1-\rho/2} \Delta_n^{1/2},$$

and hence, $\|R'_{n,j}\|_p \leq K_p \Delta_n^{1/2-\rho}$. By a maximal inequality for $p > (1-\rho)/(1/2-\rho-\epsilon)$,

$$\max_{1 \leq j \leq m_n} \left\| R'_{n,j} \right\| \leq K_p m_n^{1/p} \Delta_n^{1/2-\rho} = o_p(\Delta_n^\epsilon). \tag{4.4}$$

The assertion of part (a) then follows from (4.1), (4.2), (4.3), and (4.4).

(b) Note that $\max_{1\leq j\leq m_n}\|\sigma_{t(n,j)}U_{n,j}\sigma_{t(n,j)}^\top\|\leq K\max_{1\leq j\leq m_n}\|U_{n,j}\|$. It is also easy to see that for any $p\geq 1$, $\max_{1\leq j\leq m_n}\|U_{n,j}\|_p\leq K_p$. Therefore, by a maximal inequality,

$$\mathbb{E}\left[\max_{1\leq j\leq m_n}\left\|\sigma_{t(n,j)}U_{n,j}\sigma_{t(n,j)}^\top\right\|\right]\leq K_p m_n^{1/p}. \tag{4.5}$$

Note that $\epsilon<\rho/2$, so we can find a (small) constant $\iota>0$ such that $\rho/2-2\iota>\epsilon$. By taking p sufficiently large, we deduce from (4.5) that

$$\max_{1\leq j\leq m_n}\left\|\sigma_{t(n,j)}U_{n,j}\sigma_{t(n,j)}^\top\right\|=O_p(\Delta_n^{-\iota}).$$

From this estimate and the result in part (a), we further deduce that

$$\max_{1\leq j\leq m_n}\sup_{t\in\mathcal{T}_{n,j}}\|\hat{c}_{n,j}-c_t\|=O_p(\Delta_n^{\rho/2-\iota})=o_p(1). \tag{4.6}$$

In particular, we see that $\{\hat{c}_{n,j}:1\leq j\leq m_n\}\subseteq\mathcal{K}_1^\eta$ for any fixed constant $\eta>0$ with probability approaching 1. Since f is twice continuously differentiable on the bounded set $\mathcal{K}_1^\eta$, we can use a mean-value expansion to deduce that uniformly for $1\leq j\leq m_n$ and $t\in\mathcal{T}_{n,j}$,

$$\begin{aligned}&\left|k_{n,j}^{1/2}\left(f\left(\hat{c}_{n,t}\right)-f\left(c_t\right)\right)-\operatorname{tr}[\partial f\left(c_{t(n,j)}\right)k_{n,j}^{1/2}(\hat{c}_{n,j}-c_t)]\right|\\&\quad\leq K\Delta_n^{-\rho/2}\left\|\hat{c}_{n,j}-c_t\right\|^2=O_p(\Delta_n^{\rho/2-2\iota})=o_p(\Delta_n^{\epsilon}),\end{aligned} \tag{4.7}$$

where the rate statements follow from (4.6) and the fact that $\rho/2-2\iota>\epsilon$. Since $\|\partial f\left(c_t\right)\|$ is bounded, the assertion of part (b) readily follows from (4.7) and part (a). □

Proof of Theorem 2. Since $\partial f\left(c_t\right)c_t$ is locally bounded away from zero, with an appeal to localization, we can assume that $\partial f\left(c_t\right)c_t\geq C$ for some constant $C>0$ without loss of generality. In addition, (4.6) implies that, for some constant $\iota>0$ satisfying $\rho/2-2\iota>\epsilon$,

$$\max_{1\leq j\leq m_n}\left\|\hat{c}_{n,j}-c_{t(n,j)}\right\|=O_p(\Delta_n^{\rho/2-\iota})=o_p(1).$$

From here, it is easy to see that uniformly for $1\leq j\leq m_n$,

$$\left(\partial f\left(\hat{c}_{n,j}\right)\hat{c}_{n,j}\right)^{-1}=O_p(1),\qquad\left|\frac{\partial f\left(c_{t(n,j)}\right)c_{t(n,j)}}{\partial f\left(\hat{c}_{n,j}\right)\hat{c}_{n,j}}-1\right|=O_p(\Delta_n^{\rho/2-\iota}). \tag{4.8}$$

We then finish the proof by observing

$$\begin{aligned}
&\max_{1\leq j\leq m_n}\sup_{t\in\mathcal{T}_{n,j}}\left|\frac{k_{n,j}^{1/2}\left(f\left(\hat{c}_{n,j}\right)-f\left(c_t\right)\right)}{\partial f\left(\hat{c}_{n,j}\right)\hat{c}_{n,j}}-U_{n,j}\right| \\
&\leq \max_{1\leq j\leq m_n}\frac{1}{\left|\partial f\left(\hat{c}_{n,j}\right)\hat{c}_{n,j}\right|} \\
&\quad\cdot\max_{1\leq j\leq m_n}\sup_{t\in\mathcal{T}_{n,j}}\left|k_{n,j}^{1/2}\left(f\left(\hat{c}_{n,j}\right)-f\left(c_t\right)\right)-\partial f\left(c_{t(n,j)}\right)c_{t(n,j)}U_{n,j}\right| \\
&\quad+\max_{1\leq j\leq m_n}\left|\frac{\partial f\left(c_{t(n,j)}\right)c_{t(n,j)}}{\partial f\left(\hat{c}_{n,j}\right)\hat{c}_{n,j}}-1\right|\cdot\max_{1\leq j\leq m_n}\left|U_{n,j}\right| \\
&= o_p\left(\Delta_n^{\epsilon}\right)+O_p(\Delta_n^{\rho/2-2\iota})=o_p\left(\Delta_n^{\epsilon}\right),
\end{aligned}$$

where the first inequality is by the triangle inequality, and the subsequent rate statement follows from Theorem 1, (4.8), $\max_{1\leq j\leq m_n}|U_{n,j}|=O_p(\Delta_n^{-\iota})$, and $\rho/2-2\iota>\epsilon$. □

Proof of Theorem 3. Step 1. In this step, we construct another coupling variable for $\max_{1\leq j\leq m_n}|U_{n,j}|$. For ease of notation, we denote $L_n=\log(\Delta_n^{-1})$. Let $\bar{k}_n=\max_{1\leq j\leq m_n}k_{n,j}$ and define an array of random variables $(\widetilde{X}_{i,j})_{1\leq i\leq\bar{k}_n,1\leq j\leq 2m_n}$ as follows: for $1\leq j\leq m_n$, we set

$$(\widetilde{X}_{i,j})_{1\leq i\leq k_{n,j}}=\left(k_{n,j}^{-1/2}\left(\frac{(\Delta_l^n W)^2}{\Delta_n}-1\right)\right)_{l\in\mathcal{I}_{n,j}},\tag{4.9}$$

and $\widetilde{X}_{i,j}=0$ if $k_{n,j}<i\leq\bar{k}_n$; we then set $\widetilde{X}_{i,j}=-\widetilde{X}_{i,j-m_n}$ for $m_n+1\leq j\leq 2m_n$ and $1\leq i\leq\bar{k}_n$. With these definitions, we can rewrite

$$\max_{1\leq j\leq m_n}|U_{n,j}|=\max_{1\leq j\leq 2m_n}\sum_{i=1}^{\bar{k}_n}\widetilde{X}_{i,j}.\tag{4.10}$$

We note that the $(2m_n)$-dimensional vectors $(\widetilde{X}_{i,j})_{1\leq j\leq 2m_n}$ are independent across different i's. Let $(\widetilde{Y}_{i,j})_{1\leq i\leq\bar{k}_n,1\leq j\leq 2m_n}$ be a generic array of centered Gaussian random vectors, such that $\mathbb{E}[\widetilde{Y}_{i,j}\widetilde{Y}_{i',j'}]=\mathbb{E}[\widetilde{X}_{i,j}\widetilde{X}_{i',j'}]$ for all $1\leq i,i'\leq\bar{k}_n$ and $1\leq j,j'\leq 2m_n$.

To construct a coupling variable for $\max_{1\leq j\leq m_n}|U_{n,j}|$, we first need to derive bounds for the following sequences:

$$
\begin{aligned}
B_{1,n} &\equiv \mathbb{E}\left[\max_{1\le j,k\le 2m_n}\left|\sum_{i=1}^{\bar{k}_n}\left(\widetilde{X}_{i,j}\widetilde{X}_{i,k}-\mathbb{E}\left[\widetilde{X}_{i,j}\widetilde{X}_{i,k}\right]\right)\right|\right],\\
B_{2,n} &\equiv \mathbb{E}\left[\max_{1\le j\le 2m_n}\sum_{i=1}^{\bar{k}_n}\left|\widetilde{X}_{i,j}\right|^3\right],\\
B_{3,n} &\equiv \sum_{i=1}^{\bar{k}_n}\mathbb{E}\left[\max_{1\le j\le 2m_n}\left|\widetilde{X}_{i,j}\right|^3\cdot 1_{\left\{\max_{1\le j\le 2m_n}\left|\widetilde{X}_{i,j}\right|>L_n^{-2}\log^{-1}\left(2m_n\vee\bar{k}_n\right)\right\}}\right].
\end{aligned}
$$

We start with $B_{1,n}$. Note that for any $p \ge 1$,

$$
\left\|\widetilde{X}_{i,j}\widetilde{X}_{i,k}-\mathbb{E}\left[\widetilde{X}_{i,j}\widetilde{X}_{i,k}\right]\right\|_p \le K_p\Delta_n^{\rho}. \tag{4.11}
$$

Since $(\widetilde{X}_{i,j}\widetilde{X}_{i,k}-\mathbb{E}[\widetilde{X}_{i,j}\widetilde{X}_{i,k}])_{1\le i\le\bar{k}_n}$ is a sequence of independent variables, by the Burkholder–Davis–Gundy inequality, Hölder's inequality, and (4.11), we have

$$
\left\|\sum_{i=1}^{\bar{k}_n}\left(\widetilde{X}_{i,j}\widetilde{X}_{i,k}-\mathbb{E}\left[\widetilde{X}_{i,j}\widetilde{X}_{i,k}\right]\right)\right\|_p \le K_p\Delta_n^{\rho/2}.
$$

Then, by using a maximal inequality, $B_{1,n} \le K_p m_n^{2/p}\Delta_n^{\rho/2}$. By taking p sufficiently large, we deduce that for any fixed constant $\iota > 0$,

$$
B_{1,n} = o\left(\Delta_n^{\rho/2-\iota}\right).
$$

Next, we consider $B_{2,n}$. It is easy to see that for any $p \ge 1$, $\|\sum_{i=1}^{\bar{k}_n}|\widetilde{X}_{i,j}|^3\|_p \le K_p\Delta_n^{\rho/2}$. By another use of maximal inequality, we then deduce that

$$
B_{2,n} = o\left(\Delta_n^{\rho/2-\iota}\right).
$$

Similarly,

$$
B_{3,n} \le \sum_{i=1}^{\bar{k}_n}\mathbb{E}\left[\max_{1\le j\le 2m_n}\left|\widetilde{X}_{i,j}\right|^3\right] = o\left(\Delta_n^{\rho/2-\iota}\right).
$$

By Corollary 4.1 of [9], there exists a sequence of random variables $(\widetilde{Z}_n)_{n\ge1}$

such that $\widetilde{Z}_n$ has the same distribution as $\max_{1\leq j\leq 2m_n}\sum_{i=1}^{\bar{k}_n}\widetilde{Y}_{i,j}$ and

$$\mathbb{P}\left(\left|\max_{1\leq j\leq 2m_n}\sum_{i=1}^{\bar{k}_n}\widetilde{X}_{i,j}-\widetilde{Z}_n\right|>16L_n^{-2}\right)$$
$$\leq L_n^4\left\{B_{1,n}+L_n^2\left(B_{2,n}+B_{3,n}\right)\right\}\log\left(2m_n\vee\bar{k}_n\right)+\frac{\log\left(\bar{k}_n\right)}{\bar{k}_n}.$$

Since $B_{1,n}$, $B_{2,n}$, and $B_{3,n}$ converge to zero at polynomial rates, it is easy to see that the majorant side of the above inequality is $o(1)$. In view of (4.10), we deduce

$$\max_{1\leq j\leq m_n}|U_{n,j}|-\widetilde{Z}_n=o_p\left(L_n^{-1}\right). \tag{4.12}$$

Step 2. We prove the assertion of the theorem in this step. Without loss of generality one can assume $|\partial f\left(c_t\right)c_t|\geq C$ for some constant $C>0$, and hence, in view of (4.6) the probability that $S_{n,t}$ takes the "dummy" value 0 for at least one $t\in[0,T]$ goes to zero as $n\to\infty$, yielding $S_n^*-\widetilde{Z}_n=o_p\left(L_n^{-1}\right)$ by Theorem 2 and (4.12). Therefore, there exists a positive real sequence $\delta_n=o(L_n^{-1})$, such that with probability approaching 1,

$$\widetilde{Z}_n-\delta_n\leq S_n^*\leq\widetilde{Z}_n+\delta_n. \tag{4.13}$$

Let $\tilde{\kappa}_n(q)$ denote the q-quantile of $\widetilde{Z}_n$. Recall that κ_n^* is the $1-\alpha$ quantile of $\max_{1\leq j\leq m_n}|U_{n,j}|$. By (4.12) and Lemma A.1 of [5], there exists a positive real sequence $v_n=o(1)$ such that with probability approaching 1,

$$\tilde{\kappa}_n(1-\alpha-v_n)-\delta_n\leq\kappa_n^*\leq\tilde{\kappa}_n(1-\alpha+v_n)+\delta_n, \tag{4.14}$$

where the δ_n sequence can be taken as the same one in (4.13) without loss of generality. Therefore,

$$\begin{aligned}\mathbb{P}\left(S_n^*\leq\kappa_n^*\right) &\leq \mathbb{P}\left(\widetilde{Z}_n\leq\kappa_n^*+\delta_n\right)+o(1)\\ &\leq \mathbb{P}\left(\widetilde{Z}_n\leq\tilde{\kappa}_n\left(1-\alpha+v_n\right)+2\delta_n\right)+o(1)\\ &\leq \mathbb{P}\left(\widetilde{Z}_n\leq\tilde{\kappa}_n\left(1-\alpha+v_n\right)\right)+o(1)\\ &= 1-\alpha+o(1),\end{aligned} \tag{4.15}$$

where the first inequality is by (4.13), the second inequality is by (4.14), the third inequality is by the anti-concentration inequality for the maximum of

Gaussian variables (see, e.g., Lemma 5.3 of [5]), and the last line is by the definition of $\tilde{\kappa}_n(\cdot)$. By a similar argument, we can also show that

$$\mathbb{P}\left(S_n^* \leq \kappa_n^*\right) \geq 1 - \alpha + o(1). \tag{4.16}$$

The assertions of the theorem then readily follow from (4.15) and (4.16). □

PROOF OF THEOREM 4. In view of equation (2.6), part (a) can be proved in the same way as Theorem 2 with the function $c \mapsto \partial f(c)\, c$ replaced by $F(\cdot)$. To prove part (b), we modify (4.9) as

$$(\widetilde{X}_{i,j})_{1\leq i\leq k_{n,j}} = \left(k_{n,j}^{-1/2}\left(\frac{\Delta_i^n W_{k^*}\Delta_i^n W_{l^*}}{\Delta_n} - \mathbb{E}\left[\frac{\Delta_i^n W_{k^*}\Delta_i^n W_{l^*}}{\Delta_n}\right]\right)\right)_{i\in\mathcal{I}_{n,j}},$$

and then follow the same proof as in Theorem 3. □

Acknowledgements. This research is conducted while Li is a visiting professor at Yale University and the Cowles foundation.

References.

[1] Aït-Sahalia, Y. and J. Jacod (2014). *High-Frequency Financial Econometrics*. Princeton University Press.
[2] Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*, 885–905.
[3] Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), pp. 579–625.
[4] Barndorff-Nielsen, O. E. and N. Shephard (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica 72*(3), 885 – 925.
[5] Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics 186*(2), 345 – 366. High Dimensional Problems in Econometrics.
[6] Bickel, P. J. and M. Rosenblatt (1973, 11). On some global measures of the deviations of density function estimates. *Annals of Statistics 1*(6), 1071–1095.
[7] Bollerslev, T. and V. Todorov (2011). Estimation of jump tails. *Econometrica 79*(6), 1727–1783.
[8] Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics 41*(6), 2786–2819.
[9] Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics 42*(4), 1564–1597.
[10] Chernozhukov, V., S. Lee, and A. M. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica 81*(2), 667–737.
[11] Comte, F. and E. Renault (1998). Long memory in continuous time stochastic volatility models. *Mathematical Finance 8*(4), 291–323.

[12] Fan, J. and Y. Wang (2008). Spot volatility estimation for high-frequency data. *Statistics and Its Interface 1*, 279–288.

[13] Foster, D. P. and D. B. Nelson (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica 64*(1), 139–174.

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

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

[16] Jacod, J. and M. Rosenbaum (2013). Quarticity and Other Functionals of Volatility: Efficient Estimation. *The Annals of Statistics 41*(3), 1462–1484.

[17] Johnston, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *Journal of Multivariate Analysis 12*(3), 402 – 414.

[18] Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory 26*, 60–93.

[19] Li, J. and Z. Liao (2020). Uniform nonparametric inference for time series. *Journal of Econometrics 219*, 38–51.

[20] Li, J., V. Todorov, and G. Tauchen (2013). Volatility occupation times. *The Annals of Statistics 41*(4), 1865–1891.

[21] Li, J., V. Todorov, and G. Tauchen (2017a). Adaptive estimation of continuous-time regression models using high-frequency data. *Journal of Econometrics 200*, 36–47.

[22] Li, J., V. Todorov, and G. Tauchen (2017b). Jump regressions. *Econometrica 85*, 173–195.

[23] Li, J. and D. Xiu (2016). Generalized method of integrated moments for high-frequency data. *Econometrica 84*(4), 1613–1633.

[24] Mancini, C. (2001). Disentangling the jumps of the diffusion in a Geometric jumping Brownian Motion. *Giornale dell'Istituto Italiano degli Attuari LXIV*, 19–47.

[25] Marcus, M. B. and J. Rosen (2006). *Markov Processes, Gaussian Processes, and Local Times*. Cambridge University Press.

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

[27] Renault, E., C. Sarisoy, and B. J. Werker (2016). Efficient estimation of integrated volatility and related processes. *Econometric Theory*.

[28] Todorov, V. and G. Tauchen (2012). The Realized Laplace Transform of Volatility. *Economtrica 80*, 1105–1127.

[29] Zhang, D. and W. B. Wu (2017). Gaussian approximation for high dimensional time series. *The Annals of Statistics 45*(5), 1895–1919.

Institut de Mathématiques de Jussieu
Sorbonne Université
Paris 75252
E-mail: jean.jacod@gmail.com

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

Department of Economics
UCLA
Los Angeles, CA 90095
E-mail: zhipeng.liao@econ.ucla.edu