

## EFFICIENT ESTIMATION OF INTEGRATED VOLATILITY FUNCTIONALS VIA MULTISCALE JACKKNIFE

BY JIA LI<sup>1</sup>, YUNXIAO LIU AND DACHENG XIU<sup>2</sup>

*Duke University, University of North Carolina at Chapel Hill and  
University of Chicago*

We propose semiparametrically efficient estimators for general integrated volatility functionals of multivariate semimartingale processes. A plug-in method that uses nonparametric estimates of spot volatilities is known to induce high-order biases that need to be corrected to obey a central limit theorem. Such bias terms arise from boundary effects, the diffusive and jump movements of stochastic volatility and the sampling error from the nonparametric spot volatility estimation. We propose a novel jackknife method for bias correction. The jackknife estimator is simply formed as a linear combination of a few uncorrected estimators associated with different local window sizes used in the estimation of spot volatility. We show theoretically that our estimator is asymptotically mixed Gaussian, semiparametrically efficient, and more robust to the choice of local windows. To facilitate the practical use, we introduce a simulation-based estimator of the asymptotic variance, so that our inference is derivative-free, and hence is convenient to implement.

**1. Introduction.** This paper concerns the efficient estimation of integrated volatility functionals of the form  $\int_0^T g(c_s) ds$ , where  $c$  is the spot covariance matrix process of a  $d$ -dimensional Itô semimartingale process  $X$ ,  $g(\cdot)$  is a smooth function and  $T$  is the fixed time span. A basic example of such functionals is the integrated variance-covariance matrix ([3, 4]). General integrated volatility functionals have received much attention in the recent literature of high-frequency econometrics and statistics; see, for example, [14, 18] and [25]. These functionals are broadly useful for measuring risk-related quantities ([2, 17, 20]) for which the  $g(\cdot)$  function transforms the spot covariance to quantities such as spot betas, correlations, idiosyncratic variances and eigenvalues. More generally, integrated volatility functionals can be used as “moment conditions” for estimating economic models [22]. In this case,  $g(\cdot)$  involves general nonlinear functions implied by the underlying economic theory (e.g., option pricing theory or market microstructure theory), as well as weight functions (i.e., instruments) in the estimation procedure.

---

Received March 2017; revised September 2017.

<sup>1</sup>Supported in part by NSF Grant SES-1326819.

<sup>2</sup>Supported in part by the IBM Corporation Faculty Scholar Fund at the University of Chicago Booth School of Business.

*MSC2010 subject classifications.* 60F05, 60G44, 62F12.

*Key words and phrases.* High-frequency data, jackknife, semimartingale, spot volatility.

A natural “plug-in” estimator of  $\int_0^T g(c_s) ds$  can be formed by replacing the latent spot covariance matrix process with its nonparametric estimate. When the volatility path is sufficiently smooth, [18] shows the “plug-in” estimator is consistent and asymptotically mixed Gaussian. However, the smoothness requirement of [18] does not hold for typical stochastic volatility models (e.g., jump-diffusions). In a recent work, [14] show that under more general volatility dynamics, the “raw” plug-in estimator carries high-order asymptotic biases that are not negligible in the second-order asymptotics. Jacod and Rosenbaum [14] also propose a bias-correction procedure and show the bias-corrected estimator admits a (feasible) central limit theorem. Moreover, this estimator attains the semiparametric efficiency bound established by [9] and [25]. The theory of [14] is further extended by [21] to allow for broader classes of test functions (i.e.,  $g$ ) via a spatial-localization technique.

Although the aforementioned bias-corrected estimator is known in explicit form, implementing in empirical applications can be fairly cumbersome. The main reason is that the correction term involves the second partial derivatives of the test function  $g(\cdot)$  with respect to all elements of the spot covariance matrix. Consequently, to implement the bias correction, the empirical researcher typically needs to prepare a large number of analytical formulas of second partial derivatives. Although such calculations are feasible “in principle,” they are quite costly in empirical research, especially when the researcher may experiment with various choices of  $g(\cdot)$  in search of a good specification. Such a cost may already be substantial even for problems with moderate dimensions. For example, [20] study a volatility-spanning problem with  $d = 10$ , in which the bias-correction term involves 990 distinct second partial derivatives for various highly nonlinear functions with 45 arguments (i.e., the number of distinct elements in the spot covariance matrix). In general, the computational complexity grows at the rate of  $O(d^4)$ .

The task of bias correction can further be complicated when the function  $g(\cdot)$  itself is not known in an analytical form, but is calculated using rather nontrivial numerical procedures. For example, [22] demonstrate integrated volatility functionals can be used as integrated moment conditions for estimating option-pricing models with high-frequency data. In such applications,  $g(\cdot)$  involves an option-pricing formula, which in turn needs to be numerically evaluated by solving ordinary differential equations and Fourier transforms (see, e.g., [10]). The computational complexity of evaluating the second partial derivatives of  $g(\cdot)$  is “two orders of magnitude” higher than that of  $g(\cdot)$  with lower numerical precision, which renders a reliable implementation of the bias correction very challenging.

Set against this background, we propose an easy-to-implement jackknife procedure for bias correction, which avoids the estimation of bias terms altogether. Jackknife methods have been widely used in statistics (see, e.g., [11]). Here, we consider two uncorrected estimators formed using two *different* sequences of local windows in the estimation of spot volatility. Because these estimators have similar bias terms but with different loadings, we can eliminate their biases by forming a proper linear combination between them. We refer to the resulting estimator as the

*two-scale jackknife estimator*. We show this estimator is asymptotically centered mixed Gaussian. Importantly, its conditional asymptotic variance is the same as that of the estimator in [14], and hence the jackknife estimator is also semiparametrically efficient.

To further facilitate inference in practice, we propose an easy-to-implement and consistent estimator for the asymptotic variance via a simulation technique. Our motivation is that existing estimators of the asymptotic variance involve the first partial derivatives of  $g(\cdot)$ , and hence can potentially be cumbersome to implement in practice for reasons mentioned above. In fact, our estimator is simply the sample variance of uncorrected estimators formed using simulated data. This estimator does not require analytical or numerical evaluation of the derivatives of  $g(\cdot)$ , and the simulation can easily be parallelized. In a recent paper, [23] proposed a general approach for estimating asymptotic variances via the “observed asymptotic variance” that is formed as the quadratic variation of “miniature” versions of the original (full-sample) estimator. Our simulation-based estimator for the asymptotic variance is conceptually different from that of [23] because the former is constructed by recomputing the original full-sample estimator for simulated data. In particular, we do not need to divide the sample into blocks for estimating miniature versions of the integrated volatility functional. That said, both estimators share the advantage that the user does not have to know the form of the asymptotic variance, which is clearly convenient in practice. Hence, our method provides a useful complement to that of [23] in empirical research.

We note that both the bias-corrected estimator of [14] and the two-scale jackknife estimator are designed to correct the nonlinearity bias that arises from the (squared) statistical error in the nonparametric estimation of spot volatility, for which the asymptotic justification relies on “undersmoothing.” That is, the local window size for the spot volatility estimation is relatively “small,” so that other biases due to the boundary effect, volatility of volatility, and volatility jumps become asymptotically negligible. By contrast, [15] consider a specific choice of the local window sequence in which all sources of biases are balanced at the same order, and propose explicit correction for each of them.

In the same vein, we show this “complete” bias correction can also be achieved via a *multiscale jackknife estimator* formed as a linear combination of three (or more) uncorrected estimators. The underlying idea is, again, to cancel the biases using uncorrected estimators without estimating the biases explicitly. Theoretically, we show the multiscale jackknife estimator admits the same central limit theorem while allowing for a broad range of growth behavior of the local windows. We show the multiscale bias correction is sufficiently accurate (for obtaining a central limit theorem) even if all types of aforementioned biases are explosive in each of the uncorrected estimators. In particular, the asymptotic behavior of the jackknife procedure is stable regardless of whether the nonparametric volatility estimation features “undersmoothing” or “oversmoothing.” This stability provides a theoretically guaranteed robustness for the multiscale jackknife method.

Our analysis is limited to the setting in which the high-frequency data are observed without noise. It is well known that financial data at the “ultra high” frequency are contaminated by microstructure noise, which would lead to nontrivial bias in the spot volatility estimates. In the multivariate setting, sampling asynchronicity among the underlying processes leads to another type of bias in the estimation of covariances. Therefore, following standard practice, our method is mainly applicable to data that are sparsely sampled, for which the effect of noise and/or asynchronicity is mild. A large and growing literature exists on the estimation of integrated variance and covariance matrix for noisy irregularly sampled data; see, for example, [1, 5–7, 12, 16, 27, 28] and [26]. The efficiency problem in this context has been addressed by [24] using the equivalence-of-experiment approach. Extending such results to the case of efficient estimation of general integrated volatility functionals is a rather nontrivial task that is beyond the scope of the current paper.

This paper is organized as follows. Section 2 presents the setting. Section 3 contains our main results. Section 4 reports simulation results. Section 5 concludes. The Supplementary Material in [19] contains all proofs.

**2. The setting.**

2.1. *The underlying processes.* We consider a  $\mathbb{R}^d$ -valued process  $(X_t)_{t \geq 0}$  defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ . We observe  $X_{i\Delta_n}$  for  $i = 0, \dots, \lfloor T/\Delta_n \rfloor$  over a fixed time interval  $[0, T]$ , where  $\Delta_n \rightarrow 0$ , as  $n \rightarrow \infty$  asymptotically. Below, for any variable  $Y_t$ , we denote its  $j$ th component by  $Y_t^{(j)}$ ; the same convention also applies to matrix- and tensor-valued variables. We denote  $\Delta_i^n X \equiv X_{i\Delta_n} - X_{(i-1)\Delta_n}$ ,  $i \geq 1$ .

Our basic assumption is that  $X$  is an Itô semimartingale (see, e.g., Section 2.1.4 in [13]) with the following form:

$$(2.1) \quad X_t = x_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_{\mathbb{R}} \delta(s, z) \mu(ds, dz),$$

where  $W$  is a  $d'$ -dimensional Brownian motion and  $\mu$  is a Poisson random measure on  $\mathbb{R}_+ \times E$  for some auxiliary Polish space  $E$  with compensator  $\nu(dt, dz) = dt \otimes \lambda(dz)$  for some  $\sigma$ -finite measure  $\lambda(dz)$ . The stochastic volatility process  $\sigma_t$  takes values in  $\mathbb{R}^{d \otimes d'}$ . We denote the spot covariance matrix  $c_t \equiv \sigma_t \sigma_t^\top$ , which takes value in the space  $\mathcal{M}_d$  of  $d$ -dimensional positive semidefinite matrices. We suppose  $c_t$  is also an Itô semimartingale with the form

$$(2.2) \quad \begin{aligned} c_t = c_0 &+ \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s \\ &+ \int_0^t \int_E \tilde{\delta}(s, z) 1_{\{\|\tilde{\delta}(s, z)\| \leq 1\}} (\mu - \nu)(ds, dz) \\ &+ \int_0^t \int_E \tilde{\delta}(s, z) 1_{\{\|\tilde{\delta}(s, z)\| > 1\}} \mu(ds, dz), \end{aligned}$$

where we use a matrix notation by defining matrix-valued processes component by component. In particular, the  $(j, k)$  component of  $\int_0^t \tilde{\sigma}_s dW_s$  is given by  $\sum_{l=1}^{d'} \int_0^t \tilde{\sigma}_s^{(jkl)} dW_s^{(l)}$ . We now collect some regularity conditions for the process  $X$ .

**ASSUMPTION 1.** The following conditions hold for some constant  $r \in [0, 1)$ . There are a sequence  $(J_m)_{m \geq 1}$  of nonnegative bounded  $\lambda$ -integrable functions on  $E$  and a sequence  $(\tau_m)_{m \geq 1}$  of stopping times increasing to  $\infty$ , such that  $\|\delta(t, \cdot)\|^r \wedge 1 \leq J_m(\cdot)$  and  $\|\tilde{\delta}(t, \cdot)\|^2 \wedge 1 \leq J_m(\cdot)$  on  $\{t \leq \tau_m\}$ . The processes  $b_t, \tilde{b}_t$  and  $\tilde{\sigma}_t$  are càdlàg adapted.

Assumption 1 is fairly standard for analyzing the asymptotic properties of estimators formed using high-frequency data; see [13] for many examples. In particular, the constant  $r$  serves as a bound for the activity of jumps in  $X$ . The restriction  $r < 1$  is needed for deriving central limit theorems for jump-robust estimators of volatility functionals.

The primary interest of the current paper is on the efficient estimation of integrated volatility functionals of the form

$$(2.3) \quad S(g) \equiv \int_0^T g(c_s) ds,$$

where  $g(\cdot) : \mathcal{M}_d \mapsto \mathbb{R}^k$  is a three-time continuously differentiable function. For example, in a bivariate case, we take  $g(c) = c^{(12)}/c^{(11)}, c^{(12)}/\sqrt{c^{(11)}c^{(22)}}$ , and  $c^{(22)} - (c^{(12)})^2/c^{(11)}$  for estimating the integrated beta, correlation and idiosyncratic variance, respectively. More complicated transformations arise in nonparametric specification tests for the covariance process [20] and the estimation of economic models [22]) in which the form of  $g(\cdot)$  is determined by the scientific model under investigation.

**2.2. The uncorrected estimator and its high-order biases.** We now proceed to introducing the uncorrected estimator, which is the building block of the jackknife estimator we propose below. The uncorrected estimator of  $S(g)$  is formed by replacing the latent covariance matrix process  $c$  with its nonparametric estimate. To this end, we choose a sequence  $k_n$  of local windows and associate it with the following spot covariance matrix estimator: for  $N_n \equiv \lfloor T/\Delta_n \rfloor - k_n + 1$ ,

$$(2.4) \quad \hat{c}_i^n(k_n) \equiv \frac{1}{k_n \Delta_n} \sum_{j=0}^{k_n-1} \Delta_{i+j}^n X \Delta_{i+j}^n X^\top 1_{\{\|\Delta_{i+j}^n X\| \leq u_n\}}, \quad 1 \leq i \leq N_n,$$

where  $u_n$  is a thresholding sequence for eliminating jumps in  $X$  that satisfies

$$u_n \asymp \Delta_n^\varpi \quad \text{for some } \varpi \in (0, 1/2).$$

Our notation  $\hat{c}_i^n(k_n)$  emphasizes the dependence of this estimator on the local window sequence  $k_n$ . The uncorrected estimator for  $S(g)$  is then constructed as its sample analogue:

$$S_n(g; k_n) \equiv \Delta_n \sum_{i=1}^{N_n} g(\hat{c}_i^n(k_n)).$$

The asymptotic behavior of this uncorrected estimator is fairly complicated and depends crucially on the growth rate of  $k_n$ . The optimal rate for estimating the spot volatility is known to be attained with  $k_n \asymp \Delta_n^{-1/2}$ . Hence, a (seemingly) natural choice of  $k_n$  is such that

$$k_n \sim \theta \Delta_n^{-1/2} \quad \text{for some } \theta \in (0, \infty).$$

Under this rate condition, [14] characterized the asymptotic behavior of the uncorrected estimator. To build intuition for our later discussions, briefly recalling this result is instructive. We consider the special case with  $d = d' = 1$  for simplicity. Below, we use  $\xrightarrow{\mathcal{L}\text{-}s}$  to denote stable convergence in law, which means the convergence in law is joint with any bounded  $\mathcal{F}$ -measurable random variables.

Theorem 3.1 in [14] shows

$$\Delta_n^{-1/2}(S_n(g; k_n) - S(g)) \xrightarrow{\mathcal{L}\text{-}s} Z + B_1(\theta) + B_2(\theta),$$

where  $Z$  is an  $\mathcal{F}$ -conditionally centered Gaussian variable with conditional variance  $\Sigma(g)$  given by

$$(2.5) \quad \Sigma(g) \equiv 2 \int_0^T \partial g(c_s)^2 c_s^2 ds,$$

and  $B_1(\theta)$  and  $B_2(\theta)$  are bias terms given by

$$(2.6) \quad \begin{aligned} B_1(\theta) &\equiv \underbrace{\frac{1}{\theta} \int_0^T \partial^2 g(c_s) c_s^2 ds}_{\text{nonlinearity bias}}, \\ B_2(\theta) &\equiv \underbrace{-\frac{\theta}{2}(g(c_0) + g(c_T))}_{\text{edge effect}} \\ &\quad - \underbrace{\frac{\theta}{12} \int_0^T \partial^2 g(c_s) \tilde{\sigma}_s^2 ds}_{\text{bias due to diffusive movements in } c} \\ &\quad + \underbrace{\theta \sum_{0 < s \leq T} \int_0^1 (g(c_{s-} + w \Delta c_s) - (1-w)g(c_{s-}) - wg(c_s)) dw}_{\text{bias due to jumps in } c}. \end{aligned}$$

We note that, except for the edge effect, the bias terms shown in (2.6) are present only when the test function  $g(\cdot)$  is *nonlinear*. This result shows an important qualitative departure of our analysis on general volatility functionals from the baseline problem of estimating integrated variance-covariance matrices, as the latter corresponds to  $g(\cdot)$  being the (linear) identity function.

Explicitly de-biasing  $B_2(\theta)$  is clearly very difficult because it depends on the volatility of volatility  $\tilde{\sigma}$  and volatility jumps  $\Delta c$ , which involve a layer of latency in addition to the latent volatility process. A simple and elegant solution proposed by [14] is to eliminate  $B_2(\theta)$  asymptotically via *undersmoothing* (i.e.,  $k_n \Delta_n^{1/2} \rightarrow 0$ ). As a result, the nonlinearity bias term  $B_1(\theta)$  becomes explosive and needs correction. Jacod and Rosenbaum [14] proposed correcting  $B_1(\theta)$  via its sample analogue:

$$(2.7) \quad B_{1,n}(g) \equiv \frac{\Delta_n^{1/2}}{k_n} \sum_{i=1}^{N_n} \partial^2 g(\hat{c}_i^n(k_n)) \hat{c}_i^n(k_n)^2,$$

and showed that

$$(2.8) \quad \Delta_n^{-1/2} (S_n(g; k_n) - S(g)) - B_{1,n}(g) \xrightarrow{\mathcal{L}-s} Z.$$

Furthermore, the asymptotic variance  $\Sigma(g)$  can be consistently estimated via a plug-in estimator, which can be used for conducting feasible inference.

This bias-corrected estimator has generated much empirical interest in the recent literature ([2, 17, 20, 22]). However, as mentioned in the [Introduction](#), estimating the bias-correction term and the asymptotic variance can be cumbersome in many empirically interesting scenarios, because of the large number of derivatives to be calculated. The implementation becomes even more complicated if one also wants to estimate and correct for various bias terms in  $B_2(\theta)$  as considered in [15]. We propose a simple-to-implement jackknife method to address these issues, to which we now turn.

### 3. Main results.

**3.1. Two-scale jackknife estimation.** We first introduce the two-scale jackknife estimator. This estimator eliminates only the nonlinearity bias and hence can be considered a counterpart of the estimator of [14]. The key idea underlying our construction is to use uncorrected estimators with different local windows. To this end, we consider two sequences of local windows  $(k_{1,n}, k_{2,n})$  and weights  $(\psi_1, \psi_2)$  such that

$$(3.1) \quad \sum_{q=1}^2 \psi_q = 1, \quad \sum_{q=1}^2 \psi_q k_{q,n}^{-1} = o(\Delta_n^{1/2}).$$

For example, with  $k_{1,n} = k_n$  and  $k_{2,n} = 2k_n$ , the above condition is satisfied with  $\psi_1 = -1$  and  $\psi_2 = 2$ . The two-scale jackknife estimator is then given by

$$TS_n(g) = \psi_1 S_n(g; k_{1,n}) + \psi_2 S_n(g; k_{2,n}).$$

The idea underlying this construction is quite intuitive. Indeed, we observe from (2.6) that the nonlinearity bias  $B_1(\theta)$  is proportional to  $1/\theta$  (hence  $1/k_n$ ). The second condition in (3.1) thus ensures the bias terms in the two uncorrected estimators cancel each other out up to desired precision.

Theorem 1, below, describes the asymptotic property of the two-scale estimator  $TS_n(g)$ . Like [14], we impose the following undersmoothing condition (i.e.,  $\varsigma < 1/2$ ) on the local window sequences.

ASSUMPTION 2. For  $q = 1, 2$ ,  $k_{q,n} \asymp \Delta_n^{-\varsigma}$  for some  $\varsigma \in (\frac{r}{2} \vee \frac{1}{3}, \frac{1}{2})$ .

We also need some smoothness condition for the function  $g(\cdot)$  coupled with some mild pathwise regularity for the spot volatility process. Below, for a compact set  $\mathcal{K} \subset \mathcal{M}_d$  and  $\varepsilon > 0$ , we denote the “ $\varepsilon$ -enlargement” about  $\mathcal{K}$  by

$$\mathcal{K}^\varepsilon \equiv \left\{ M \in \mathcal{M}_d : \inf_{A \in \mathcal{K}} \|M - A\| < \varepsilon \right\}.$$

ASSUMPTION 3. There exist a sequence of stopping times  $(\tau_m)_{m \geq 1}$  increasing to infinity and a sequence of convex compact subsets  $\mathcal{K}_m \subseteq \mathcal{M}_d$ ,  $m \geq 1$ , such that  $c_t \in \mathcal{K}_m$  for  $t \leq \tau_m$  and  $g$  is three-time continuously differentiable on  $\mathcal{K}_m^\varepsilon$  for some  $\varepsilon > 0$ .

Assumption 3 requires the localized process  $c_{t \wedge \tau_m}$  to be compactly valued, and  $g(\cdot)$  is  $\mathcal{C}^3$  on a slight enlargement of this compact support. Importantly, we do not require  $g(\cdot)$  to satisfy the polynomial growth condition as in [14], which is often violated in economic and financial applications. This generality is achieved by using a spatial localization argument as in [20, 21] and [22]. The intuition underlying the spatial localization argument is as follows. We first observe that the spot covariance estimates uniformly approximate the local averages of the true spot covariances over the corresponding time intervals. Hence, with probability approaching 1, these estimates fall in a compact set that is “slightly larger” than the (convex) set in which the spot covariance process takes values. As a result, for all our probabilistic calculations, we can restrict the test function  $g(\cdot)$  locally on a compact set, which avoids restrictions on the global growth rate of  $g(\cdot)$ .

THEOREM 1. Suppose Assumptions 1, 2 and 3 hold and  $\varpi \in [\frac{1-\varsigma}{2-r}, \frac{1}{2})$ . Then

$$\Delta_n^{-1/2}(TS_n(g) - S(g)) \xrightarrow{\mathcal{L}\text{-}s} \mathcal{MN}(0, \Sigma(g)),$$



where  $\mathcal{MN}$  denotes the mixed normal distribution and

$$\Sigma(g) \equiv \sum_{j,k,l,m=1}^d \int_0^T \partial_{jk}g(c_s) \partial_{lm}g(c_s)^\top (c_s^{(jl)} c_s^{(km)} + c_s^{(jm)} c_s^{(kl)}) ds.$$

Theorem 1 shows the two-scale jackknife estimator  $TS_n(g)$  is a  $\Delta_n^{-1/2}$ -consistent estimator of  $S(g)$ . After normalization,  $TS_n$  admits a central limit theorem with  $\mathcal{F}$ -conditional asymptotic covariance matrix  $\Sigma(g)$ , which attains the efficiency bound established in [9] and [25].

To improve the finite-sample performance, [14] suggest removing the boundary effect by adjusting the uncorrected estimator as

$$S'_n(g; k_n) \equiv S_n(g; k_n) + \frac{k_n \Delta_n}{2} (g(\hat{c}_1^n(k_n)) + g(\hat{c}_{N_n}^n(k_n))).$$

Likewise, we can adjust the jackknife estimator as

$$TS'_n(g) \equiv \psi_1 S'_n(g; k_{1,n}) + \psi_2 S'_n(g; k_{2,n}).$$

Because  $\Delta_n^{-1/2}(S'_n(g; k_n) - S_n(g; k_n)) = O_p(k_n \Delta_n^{1/2})$ , which vanishes asymptotically in the current case with undersmoothing, these finite-sample adjustments do not result in any change for Theorem 1.

3.2. *Consistent estimation of  $\Sigma(g)$  via simulation.* We need a consistent estimator for  $\Sigma(g)$  so as to conduct feasible inference based on Theorem 1. A natural choice is the “plug-in” estimator  $\Sigma_n(g)$  given by

$$\sum_{j,k,l,m=1}^d \Delta_n \sum_{i=1}^{N_n} \partial_{jk}g(\hat{c}_i^n) \partial_{lm}g(\hat{c}_i^n)^\top (\hat{c}_i^{n,(jl)} \hat{c}_i^{n,(km)} + \hat{c}_i^{n,(jm)} \hat{c}_i^{n,(kl)}),$$

where  $\hat{c}_i^n = \hat{c}_i^n(k_n)$  and  $k_n$  satisfies  $k_n \rightarrow \infty$  and  $k_n \Delta_n \rightarrow 0$ ; see, for example, Theorem 9.4.1 in [13].

As mentioned in the Introduction, this plug-in estimator can be cumbersome to implement because it requires calculations of the partial derivatives of  $g$ . For this reason, we propose a simulation-based estimator for the asymptotic covariance matrix that avoids the calculation of  $\partial g$  altogether. Algorithm 1, below, describes this estimator, which we denote by  $\Sigma_n^*(g)$ .

ALGORITHM 1. The procedure comprises of four steps:

1. Estimate  $\hat{c}_i^n(k_n)$  for some  $k_n$ .
2. For each block  $b \in \{1, \dots, \lfloor T/k_n \Delta_n \rfloor\}$ , simulate  $(\Delta_{(b-1)k_n+i}^n X^*)_{1 \leq i \leq k_n}$  as i.i.d. draws from  $\mathcal{N}(0, \hat{c}_{(b-1)k_n+1}^n \Delta_n)$ .

3. Compute  $\bar{S}_n^*(g)$  as

$$k_n \Delta_n \sum_{b=1}^{\lfloor T/k_n \Delta_n \rfloor} g \left( \frac{1}{k_n \Delta_n} \sum_{j=0}^{k_n-1} (\Delta_{(b-1)k_n+1+j}^n X^*) (\Delta_{(b-1)k_n+1+j}^n X^*)^\top \right).$$

4. Repeat Steps 2 and 3 for a large number of times. Compute  $\Sigma_n^*(g)$  as the sample covariance matrix of the simulated  $\Delta_n^{-1/2} \bar{S}_n^*(g)$ .

Implementing Algorithm 1 is easy because it only requires recomputing the uncorrected estimator using simulated returns. We note that to implement this algorithm, one only needs to re-estimate the spot covariance matrix in each simulation for  $\lfloor T/k_n \Delta_n \rfloor$  nonoverlapping blocks (instead of for  $N_n$  overlapping windows). Theorem 2, below, shows  $\Sigma_n^*(g)$  is a consistent estimator of  $\Sigma(g)$ .

**THEOREM 2.** *Suppose Assumptions 1 and 3 hold,  $\varpi \in [\frac{1-\varsigma}{2-r}, \frac{1}{2})$ ,  $k_n \rightarrow \infty$  and  $k_n \Delta_n \rightarrow 0$ . Then  $\Sigma_n^*(g) \xrightarrow{\mathbb{P}} \Sigma(g)$ .*

**3.3. Multiscale jackknife estimation and its robustness property.** The two-scale jackknife estimator and the bias-corrected estimator of [14], implicitly or explicitly, correct the nonlinearity bias [i.e., the  $B_1(\theta)$  term in (2.6)]. However, other sources of biases [i.e., the  $B_2(\theta)$  term in (2.6)], including those driven by volatility-of-volatility and volatility jumps, are “eliminated” essentially by making the undersmoothing assumption. Although this type of theoretical argument is quite common in nonparametric statistics, these bias terms may still have a nontrivial effect in finite samples. This effect should be relatively large during sample periods with large fluctuations in volatility, such as crisis and macro news-announcement times. Such periods are of important empirical interest in economic applications because the sizable variation in volatility helps in identifying and testing economic models.

Motivated by this concern, we consider next a “complete” bias correction that eliminates all bias terms identified in (2.6). Along this line, [15] made an interesting contribution by constructing an estimator for each bias term and then correcting the biases explicitly. However, implementing the estimator of [15] is difficult in practice because the (direct) estimation of volatility-of-volatility and volatility jumps are notoriously difficult in finite samples. As a result, the undersmoothing approach has been preferred in economic and financial applications; see, for example, [2, 17, 20, 21] and [22].

We propose an easy-to-implement alternative by using a multiscale jackknife estimator that is formed as a linear combination of three (or more) uncorrected estimators associated with different local windows. To this end, we consider local windows  $k_{q,n}$  and weights  $\psi_q$ ,  $1 \leq q \leq Q$ , such that

$$(3.2) \quad \sum_{q=1}^Q \psi_q = 1, \quad \sum_{q=1}^Q \psi_q k_{q,n}^{-1} = o(\Delta_n^{1/2}), \quad \sum_{q=1}^Q \psi_q k_{q,n} = o(\Delta_n^{-1/2}).$$

For example, when  $Q = 3$ ,  $k_{j,n} = a_j k_n$ , the weights are solved as

$$\begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ a_1 & a_2 & a_3 \\ 1/a_1 & 1/a_2 & 1/a_3 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

The multiscale jackknife estimator is then given by

$$MS_n(g) \equiv \sum_{q=1}^Q \psi_q S_n(g; k_{q,n}).$$

Compared with the weights in the two-scale jackknife [recall (3.1)], the weights in the multiscale construction (3.2) satisfy an additional condition  $\sum_{q=1}^Q \psi_q k_{q,n} = o(\Delta_n^{-1/2})$ . This condition is used to ensure bias terms such as  $B_2(\theta)$  (which is proportional to  $\theta$  and  $k_n$ ) cancel each other out in the linear combination of uncorrected estimators.

Theorem 3, below, describes the asymptotic behavior of the multiscale jackknife estimator  $MS_n(g)$ , for which we need the following condition on the local windows.

ASSUMPTION 4. For all  $1 \leq q \leq Q$ ,  $k_{q,n} \asymp \Delta_n^{-\varsigma}$  for some  $\varsigma \in (\frac{r}{2} \vee \frac{1}{3}, \frac{2}{3})$ .

THEOREM 3. Suppose Assumptions 1, 3, 4 and  $\varpi \in [\frac{1-\varsigma}{2-r}, \frac{1}{2})$ . Then

$$\Delta_n^{-1/2}(MS_n(g) - S(g)) \xrightarrow{\mathcal{L}-s} \mathcal{MN}(0, \Sigma(g)),$$

where  $\Sigma(g)$  is defined as in Theorem 1.

Theorem 3 shows the multiscale jackknife estimator has the same asymptotic distribution as its two-scale counterpart shown in Theorem 1. Hence, the multiscale estimator is also semiparametrically efficient and its asymptotic variance can be estimated using the method described in Section 3.2.

Note that Theorem 3 holds under Assumption 4, which is much weaker than Assumption 2 that is used in Theorem 1. The key difference is that Theorem 3 not only holds with undersmoothing (i.e.,  $k_n \ll \Delta_n^{-1/2}$ ), but also holds with oversmoothing (i.e.,  $k_n \gg \Delta_n^{-1/2}$ ). In the latter case, the two-scale jackknife estimator and the bias-corrected estimator of [14] do not admit a central limit theorem, because the biases due to the boundary effect, volatility-of-volatility and volatility jumps are no longer asymptotically negligible (these bias terms are actually explosive). But such biases are implicitly corrected via the multiscale jackknife, which makes the resulting estimator  $MS_n(g)$  much less sensitive to the growth rate of local windows. This result thus provides a well-defined sense of robustness for the multiscale jackknife estimator.

We can further compare Theorem 3 with the result of [15]. Jacod and Rosenbaum [15] also consider correcting all bias terms identified in (2.6) by explicitly estimating each of them. These authors proved a central limit theorem for their estimator under a specific choice of the local window, namely,  $k_n \asymp \Delta_n^{-1/2}$ . With this choice, the bias terms are “balanced” and have the same order as the term that drives the central limit theorem. Theorem 3, however, does not require the local windows to grow at this particular rate, but allows them to exhibit a broad range of asymptotic behavior. Importantly, it shows the multiscale jackknife bias correction is effective even if the biases are explosive. From a technical point of view, to establish such results, we need to use refined techniques to analyze the biases and their correction, which have not been seen in prior work.

3.4. *Discussion on issues related to microstructure noise and sampling asynchronicity.* The estimators proposed above rely on regularly sampled data in a noise-free setting. However, financial transaction data are often contaminated with microstructure noise and are sampled irregularly in time. These complications will result in biases in the spot covariance estimates and, subsequently, biases in the estimate of integrated volatility functionals. In order to mitigate these microstructural effects, a standard practice is to apply the estimator to sparsely sampled data. By doing so, one reduces the relative effect of microstructure noise and asynchronous sampling by increasing the “signal level” contained in the efficient price increments. As a result, the spot covariance estimates based on sparsely sampled noisy data are “closer” to those formed using the efficient price. In this subsection, we discuss sufficient conditions that justify the use of the sparse sampling method for making inference about integrated volatility functionals (although these conditions can likely be further improved upon).

To fix ideas, let  $\hat{v}_i^n(k_n)$  denote the spot covariance estimate formed using sparsely sampled noisy data. As in previous sections, we use  $\hat{c}_i^n(k_n)$  to denote the spot covariance estimate constructed from the efficient price. For simplicity, we refer to  $\hat{v}_i^n(k_n)$  and  $\hat{c}_i^n(k_n)$  as the noisy and the noise-free estimates, respectively. In view of the above heuristics concerning sparse sampling, we assume that the noisy estimate  $\hat{v}_i^n(k_n)$  consistently approximates its noise-free counterpart uniformly:

$$\sup_{1 \leq i \leq N_n} \|\hat{v}_i^n(k_n) - \hat{c}_i^n(k_n)\| = o_p(1).$$

This condition is relatively mild because it does not require any specific rate of convergence. Like in the proof of our main theorems above, this uniform approximation condition allows us to invoke the spatial localization argument so that we can suppose without loss of generality that  $g(\cdot)$  is compactly supported, and hence, Lipschitz continuous.

In order to show that microstructural complications have negligible effect for inference, we need the difference between  $\hat{v}_i^n(k_n)$  and  $\hat{c}_i^n(k_n)$  to be sufficiently

small on average. More precisely, we consider the following high-level condition: for some  $a_n \rightarrow 0$ ,

$$(3.3) \quad \Delta_n \sum_{i=1}^{N_n} \|\hat{v}_i^n(k_n) - \hat{c}_i^n(k_n)\| = O_p(a_n),$$

where  $a_n$  depicts the rate of convergence. The microstructural effect on the uncorrected estimator for the integrated volatility functional can then be bounded as

$$\begin{aligned} & \left\| \Delta_n \sum_{i=1}^{N_n} g(\hat{v}_i^n(k_n)) - \Delta_n \sum_{i=1}^{N_n} g(\hat{c}_i^n(k_n)) \right\| \\ & \leq K \Delta_n \sum_{i=1}^{N_n} \|\hat{v}_i^n(k_n) - \hat{c}_i^n(k_n)\| = O_p(a_n). \end{aligned}$$

Since the jackknife estimators are formed as linear combinations of uncorrected estimators, the microstructural effect is asymptotically negligible for deriving their central limit theorems if

$$(3.4) \quad a_n = o(\Delta_n^{1/2}).$$

This condition can be further interpreted in terms of the extent to which the observation noise in the price and/or the degree of asynchronicity should be “small” for sparsely sampled data. We start with the case in which the price is observed with noise: instead observing the efficient price  $X_{i\Delta_n}$ , we now observe  $Y_{i\Delta_n} = X_{i\Delta_n} + \varepsilon_{n,i}$  where the noise terms  $\varepsilon_{n,i}$  are mutually independent and independent of  $X$ . Suppose that the noise terms are “small” in the sense that their standard deviation is uniformly bounded by some sequence  $\tilde{a}_n \rightarrow 0$ . We then observe that

$$\begin{aligned} \|\hat{v}_i^n(k_n) - \hat{c}_i^n(k_n)\| & \leq \frac{2}{k_n \Delta_n} \left\| \sum_{j=1}^{k_n} \Delta_{i+j}^n X(\varepsilon_{n,i+j} - \varepsilon_{n,i+j-1})^\top \right\| \\ & \quad + \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} \|\varepsilon_{n,i+j} - \varepsilon_{n,i+j-1}\|^2. \end{aligned}$$

It is easy to see that the two terms on the majorant side of the above estimate are of orders  $O_p(\tilde{a}_n/\sqrt{k_n \Delta_n})$  and  $O_p(\tilde{a}_n^2/\Delta_n)$ , respectively, which yields

$$a_n = \max\{\tilde{a}_n/\sqrt{k_n \Delta_n}, \tilde{a}_n^2/\Delta_n\}.$$

For the multiscale jackknife estimator, we can take  $k_n \asymp \Delta_n^{-1/2}$ . As a result, a sufficient condition for (3.4) is  $\tilde{a}_n = o(\Delta_n^{3/4})$ .

Turning to the case with asynchronous sampling, we measure the degree of asynchronicity as the proportion of mismatch between the irregular sampling

scheme and the regular sampling scheme, which we denote by  $\tilde{b}_n$ . For example, if the minimal distance (e.g., measured by Hausdorff distance) between the set of (irregularly spaced) transaction times and the regular 5-minute sampling grid is less than 1 second on average, then  $\tilde{b}_n$  is bounded by  $1/300$ . Measuring the degree of asynchronicity as such, it is easy to see that  $a_n = O(\tilde{b}_n)$ . Hence,  $\tilde{b}_n = o(\Delta_n^{1/2})$  is a sufficient condition for the asymptotic negligibility of asynchronous sampling in the context of the current paper.

**4. Monte Carlo.** In this section, we examine the finite-sample performance of our estimators in Monte Carlo simulations. We also compare them with the bias-corrected estimators of [14, 15] and [18]. To this end, we focus on the estimation of integrated quarticity in a univariate setting [i.e.,  $g(c) = c^2$ ] for which the bias-correction terms of the “completely” bias-corrected estimator of [15] were given in closed-form. Below, we suppress the dependence on  $g$  in our notation for simplicity. Throughout, we fix  $T = 21$  days and consider two sampling frequencies:  $\Delta = 1$  or 5 minutes. We conduct 1000 Monte Carlo trials in total.

We simulate  $X$  and its volatility process  $\sigma$  according to

$$(4.1) \quad \begin{cases} dX_t = \sigma_t dW_t + dJ_t^X, \\ \sigma_t = \exp(-1.6 + F_t), \quad dF_t = -5F_t dt + 2d\tilde{W}_t + dJ_t^F, \end{cases}$$

with  $\mathbb{E}[dW_t d\tilde{W}_t] = -0.75 dt$ . We simulate  $J_t^X$  as a compensated Poisson jump process with jump-size distribution  $\mathcal{N}(-0.01, 0.02^2)$  and intensity  $\lambda = 36$ , and simulate  $J_t^F$  as a compensated tempered-stable process (or CGMY process in [8]) with the Lévy jump measure given by

$$(4.2) \quad \nu(x) = \frac{\alpha}{|x|^{1+\beta}} e^{-\gamma_-|x|} \cdot 1_{\{x < 0\}} + \frac{\alpha}{x^{1+\beta}} e^{-\gamma_+x} \cdot 1_{\{x > 0\}},$$

where  $\gamma_+ = 3$ ,  $\gamma_- = 5$ ,  $\beta = 0.8$ , and  $\alpha = 4.5$ . The percentage of expected quadratic variation of  $F_t$  due to its jumps is about 30%. In this setting, the volatility of volatility in (2.2) is given by  $\tilde{\sigma}_t = 4c_t$ , and the bias terms  $B_1(\theta)$  and  $B_2(\theta)$  in (2.6) can be calculated explicitly as

$$B_1(\theta) = \frac{2}{\theta} \int_0^T c_s^2 ds,$$

$$B_2(\theta) = -\frac{\theta}{2}(c_0^2 + c_T^2) - \frac{\theta}{6} \int_0^T \tilde{\sigma}_s^2 ds - \frac{\theta}{6} \sum_{s \leq T} (\Delta c_s)^2.$$

We consider several estimators for the integrated quarticity, including the two-scale jackknife estimator  $TS_n$  and its boundary-adjusted version  $TS'_n$ , the multiscale jackknife estimator  $MS_n$ , the undersmoothing estimator  $JR_n^{(1)} \equiv S_n - \Delta_n^{1/2} B_{1,n}$  proposed by [14] and its boundary-adjusted version  $JR^{(2)} \equiv S'_n -$

$\Delta_n^{1/2} B_{1,n}$ , the estimator with complete bias correction proposed by [15]: that is,  $JR_n^{(3)} \equiv S_n - \Delta_n^{1/2} (B_{1,n} + B_{2,n})$ , where

$$B_{2,n} \equiv -\frac{k_n \Delta_n^{1/2}}{2} (\hat{c}_1^n(k_n)^2 + \hat{c}_{N_n}^n(k_n)^2) + \frac{\Delta_n^{1/2}}{k_n} \sum_{i=1}^{N_n} \hat{c}_i^n(k_n)^2 - \frac{\Delta_n^{1/2}}{4} \sum_{i=1}^{N_n-k_n} (\hat{c}_{i+k_n}^n(k_n) - \hat{c}_i^n(k_n))^2.$$

In addition, we consider the kernel-based estimator of [18] given by

$$K_n \equiv \Delta_n \sum_{i=h_n+1}^{\lfloor T/\Delta_n \rfloor} g(\hat{k}_i^n(h_n)),$$

where, writing  $K_h(\cdot) \equiv K(\cdot/h)/h$  for the kernel function  $K(z) = 6(1 + 3z + 2z^2)1_{\{-1 \leq z \leq 0\}}$  and bandwidth  $h_n$ ,

$$\hat{k}_i^n(h_n) \equiv \sum_{j=1}^i K_{h_n \Delta_n}((j-i)\Delta_n) (\Delta_j^n X)^2 1_{\{|\Delta_j^n X| \leq u_n\}}.$$

Unlike the other estimators above,  $K_n$  does not involve bias-correction. As is typical in conventional nonparametric statistics, this estimator was designed for settings where the volatility path is smooth, but this is not the case in our stochastic volatility model. Nevertheless, we include this estimator in our comparison for completeness.

The asymptotic variance for all estimators is  $\Sigma = 8 \int_0^T c_s^4 ds$ , which we estimate using Algorithm 1 with 1000 simulated samples.

In Table 1, we find the multiscale estimator  $MS_n$  achieves the smallest relative bias for each sampling frequency: its bias is moderately smaller than that of  $JR_n^{(3)}$  (which also corrects all sources of biases), and is substantially smaller than those of the other estimators. This finding suggests both  $MS_n$  and  $JR_n^{(3)}$  are effective for bias correction, with the former performing better than the latter. That said, the multiscale estimator has a slightly larger relative root-mean-square error than  $JR_n^{(3)}$ , which reflects a bias-variance trade-off in finite samples. Generally speaking, the jackknife estimators perform similarly to the bias-corrected estimators of [14] and [15], and outperform the kernel-based estimator. Note the jackknife method does not involve any explicit estimation of the biases, which would typically require substantial effort in practical implementation.

Figure 1 compares the histograms of these estimators studentized by the estimated asymptotic standard error. We observe that the estimators that rely on ‘‘undersmoothing,’’ namely,  $TS_n$  and  $JR_n^{(1)}$ , are not well centered. This issue is evidently mitigated after adjusting for the boundary effect, as shown by the histograms of  $TS'_n$  and  $JR_n^{(2)}$ , and further improvement is attained by  $MS_n$  and  $JR_n^{(3)}$ .

TABLE 1

In this table, we report the relative biases and relative root-mean-square-errors (in percentage unit) for seven estimators. For the multiscale estimator, we choose  $(\psi_1, \psi_2, \psi_3) = (-2.5, 8, -4.5)$ , and correspondingly  $(k_{1,n}, k_{2,n}, k_{3,n}) = (15, 30, 45)$  for data sampled every 5 minutes, or  $(40, 80, 120)$  for data sampled every minute. For the two-scale estimator, we choose  $(\psi_1, \psi_2) = (-1, 2)$  along with the same  $(k_{1,n}, k_{2,n})$  above. For  $JR_n$  estimators, we choose  $k_n = 30$  for 5-minute data, and 80 for 1-minute data, respectively. For the kernel-based estimator, we choose  $h_n = 3k_n$  for both 5-minute and 1-minute data, which minimizes approximately the RMSE among various choices *ex post*. We fix  $T = 1$  month

	$\Delta_n = 5$ minutes		$\Delta_n = 1$ minute	
	BIAS (%)	RMSE (%)	BIAS (%)	RMSE (%)
	<i>Jackknife estimators</i>			
$TS_n$	-4.76	11.05	-2.69	5.32
$TS'_n$	-1.09	10.24	-0.69	4.42
$MS_n$	-0.35	11.03	-0.34	4.54
	<i>Estimators with explicit bias correction</i>			
$JR_n^{(1)}$	-3.47	10.54	-1.92	4.83
$JR_n^{(2)}$	-0.98	10.28	-0.58	4.40
$JR_n^{(3)}$	-0.69	10.52	-0.36	4.41
	<i>Kernel-based estimator</i>			
$K_n$	18.80	22.71	6.94	8.44

In particular, the histograms of studentized versions of  $MS_n$  and  $JR_n^{(3)}$  match the standard normal density very well, suggesting the feasible central limit theorem works well in finite samples for these estimators.

To evaluate the impact of microstructure noise on these estimators, we add an i.i.d. Gaussian noise to each observed price in our sample. The standard deviation of the noise is  $10^{-4}$ , which is realistically calibrated. The results are reported in Table 2. Compared with the results in Table 1, the impact of the noise is somewhat noticeable at a 1-minute frequency but is less clear for a 5-minute frequency. This finding is not surprising, given that subsampling is the standard and most widely used procedure in practice to deal with the microstructure noise, and that the signal-to-noise ratio increases as sampling becomes more sparse.

All estimators considered above involve a choice of local windows. In practice, the window size should not be chosen “too small” so as to have enough aggregation within each local window, and should not be taken “too big” so that the time-variation of various underlying processes remain moderate. Data-driven choices of local windows, when available, often depend on user-specified loss functions and unknown latent quantities that are hard to estimate. Given this difficulty, it is advisable to experiment with different local window sizes as a robustness check, to which we now turn.



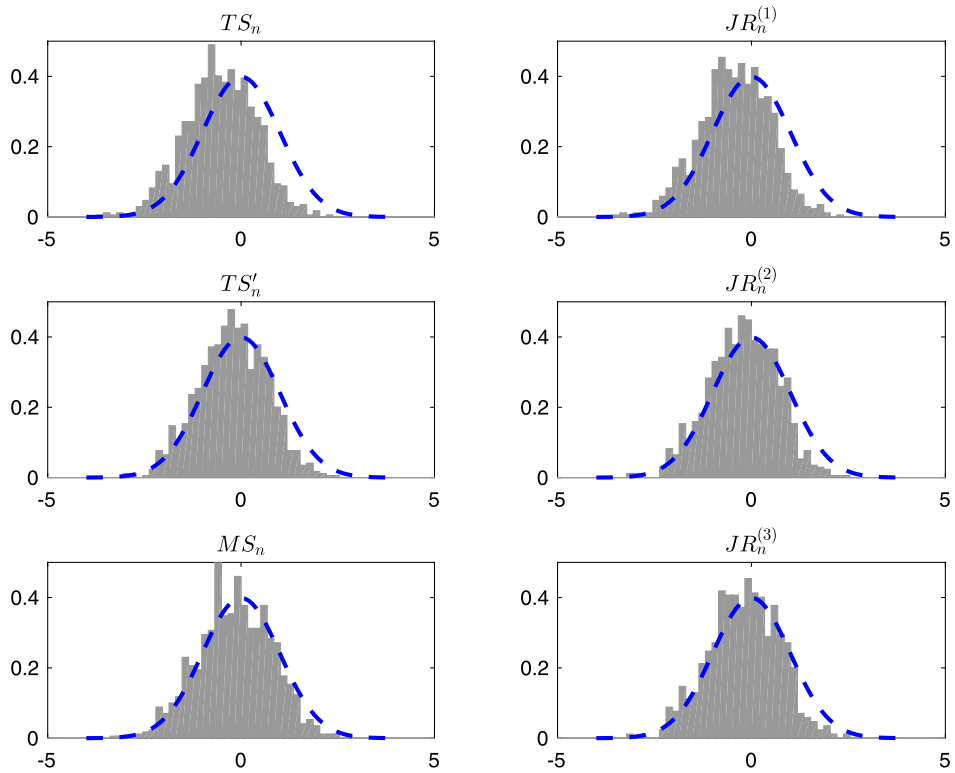


FIG. 1. In this figure, we compare the histograms of the studentized statistics with the standard normal density. The estimators are implemented as described in Table 1. The asymptotic variance is estimated using Algorithm 1 with 1000 simulated samples, for which we set  $k_n = 39$  for 5-minute data, and  $k_n = 78$  for 1-minute data.

In Table 3, we demonstrate the robustness of our multiscale estimator against different choices of local windows (i.e.,  $k_{q,n}$ ). Recall that in Table 1, we choose  $k_{2,n} = 30$  and  $80$  for 5-minute and 1-minute data, respectively, while fixing  $k_{q,n} = a_q k_{1,n}$  for  $(a_1, a_2, a_3) = (1, 2, 3)$ . Here, we use the same values of the  $k_{2,n}$ 's as before, but choose wider ranges of  $(a_1, a_2, a_3)$ :  $(1, 3, 9)$  for 5-minute data and  $(1, 4, 12)$  for 1-minute data. In the case with 1-minute sampling, for example, the jackknife estimator now involves window sizes  $k_{1,n} = 20$  and  $k_{3,n} = 240$ , which clearly span a wide range in practical terms. This robustness check thus poses a nontrivial challenge for the proposed estimators. For comparison, we also report results for the  $JR_n$  estimates corresponding to  $k_n = k_{1,n}$  and  $k_{3,n}$ .

From Table 3, we see that the performance of the multiscale estimator is insensitive to these alternative choices of  $k_{q,n}$ , but the performance of the  $JR_n$  estimators deteriorates. In particular, the benchmark estimator  $JR_n^{(1)}$ , which relies on “under-smoothing,” shows large bias when the local window size is large. In contrast, even

TABLE 2

In this table, we report the relative biases and relative root-mean-square errors (in percentage unit) for seven estimators. The setting is identical to that of Table 1, except that we add i.i.d. Gaussian noises to the observed log prices

	$\Delta_n = 5$ minutes		$\Delta_n = 1$ minute	
	BIAS (%)	RMSE (%)	BIAS (%)	RMSE (%)
<i>Jackknife estimators</i>				
$TS_n$	-3.06	10.54	6.02	9.15
$TS'_n$	0.65	10.42	8.18	11.06
$MS_n$	1.41	11.37	8.53	11.32
<i>Estimators with explicit bias correction</i>				
$JR_n^{(1)}$	-1.75	10.26	6.85	9.83
$JR_n^{(2)}$	0.77	10.49	8.29	11.17
$JR_n^{(3)}$	1.05	10.78	8.51	11.34
<i>Kernel-based estimator</i>				
$K_n$	20.88	24.64	16.42	18.41

TABLE 3

In this table, we report the relative biases and relative root-mean-square errors (in percentage unit) for six estimators. The setting is identical to that of Table 1, except that we choose different choices of bandwidths. For the multiscale estimator, we choose  $(k_{1,n}, k_{2,n}, k_{3,n}) = (10, 30, 90)$  for data sampled every 5 minutes, or  $(20, 80, 240)$  for data sampled every minute. For  $JR_n$  estimators, we choose  $k_n = 10$  and 90 for 5-minute data, and 20 and 240 for 1-minute data, respectively

	$\Delta_n = 5$ minutes		$\Delta_n = 1$ minute	
	BIAS (%)	RMSE (%)	BIAS (%)	RMSE (%)
<i>Jackknife estimators</i>				
$TS_n$	-4.29	10.82	-2.29	5.05
$TS'_n$	-1.00	10.19	-0.62	4.40
$MS_n$	-0.49	10.51	-0.37	4.46
<i>Estimators with explicit bias correction</i>				
	$k_n = 10$		$k_n = 20$	
$JR_n^{(1)}$	-4.91	11.27	-1.68	4.69
$JR_n^{(2)}$	-4.02	11.09	-1.33	4.57
$JR_n^{(3)}$	-5.79	12.07	-1.78	4.77
	$k_n = 90$		$k_n = 240$	
$JR_n^{(1)}$	-8.86	13.49	-4.95	7.16
$JR_n^{(2)}$	-1.80	10.14	-1.17	4.52
$JR_n^{(3)}$	-0.53	10.39	-0.41	4.45

though the multiscale estimator depends on the same choice of local windows, the bias terms from its building blocks are effectively canceled out so that its performance is more robust to the choice of window size.

Overall, the Monte Carlo evidence confirms the effectiveness for bias-correction of the multiscale jackknife and the two-scale jackknife with a boundary adjustment. In addition, we find the standard error generated by Algorithm 1 captures the sampling variabilities of these bias-corrected estimators well. Taken together, the proposed inference procedure shows good performance in finite samples and should provide an easy-to-implement alternative to existing methods.

**5. Conclusion.** We propose jackknife estimators for efficiently estimating general integrated volatility functionals. Such functionals are broadly useful as risk measures or moment conditions in the estimation of scientific models. The proposed jackknife estimator, along with a simulation-based estimator for its asymptotic variance, is easy to implement in practice. In particular, the cumbersome task of calculating (a large number of) partial derivatives is completely avoided. We show the jackknife estimator is semiparametrically efficient and exhibits a type of robustness with respect to the choice of local windows, in that the latter can have both undersmoothing and oversmoothing behaviors.

**Acknowledgments.** We would like to thank Tim Bollerslev, Peter Hansen, Andrew Patton, Jeffrey Russell and George Tauchen for helpful discussions.

## SUPPLEMENTARY MATERIAL

**Supplement to “Efficient estimation of integrated volatility functionals via multiscale jackknife”** (DOI: [10.1214/18-AOS1684SUPP](https://doi.org/10.1214/18-AOS1684SUPP); .pdf). This appendix contains all mathematical proofs.

## REFERENCES

- [1] AÏT-SAHALIA, Y., FAN, J. and XIU, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *J. Amer. Statist. Assoc.* **105** 1504–1517. [MR2796567](#)
- [2] AÏT-SAHALIA, Y. and XIU, D. (2015). Principal component analysis of high frequency data. Technical report.
- [3] ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2003). Modeling and forecasting realized volatility. *Econometrica* **71** 579–625. [MR1958138](#)
- [4] BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72** 885–925. [MR2051439](#)
- [5] BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* **76** 1481–1536. [MR2468558](#)

- [6] BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics* **162** 149–169. [MR2795610](#)
- [7] BIBINGER, M., HAUTSCH, N., MALEC, P. and REISS, M. (2014). Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *Ann. Statist.* **42** 1312–1346. [MR3226158](#)
- [8] CARR, P., GEMAN, H., MADAN, D. B. and YOR, M. (2002). The fine structure of asset returns: An empirical investigation. *J. Bus.* **75** 305–332.
- [9] CLÉMENT, E., DELATTRE, S. and GLOTER, A. (2013). An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility. *Stochastic Process. Appl.* **123** 2500–2521. [MR3054534](#)
- [10] DUFFIE, D., PAN, J. and SINGLETON, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* **68** 1343–1376. [MR1793362](#)
- [11] EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. CBMS–NSF Regional Conference Series in Applied Mathematics **38**. SIAM, Philadelphia, PA. [MR0659849](#)
- [12] HAYASHI, T. and YOSHIDA, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **11** 359–379. [MR2132731](#)
- [13] JACOD, J. and PROTTER, P. (2012). *Discretization of Processes*. *Stochastic Modelling and Applied Probability* **67**. Springer, Heidelberg. [MR2859096](#)
- [14] JACOD, J. and ROSENBAUM, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *Ann. Statist.* **41** 1462–1484. [MR3113818](#)
- [15] JACOD, J. and ROSENBAUM, M. (2015). Estimation of volatility functionals: The case of a  $\sqrt{n}$  window. In *Large Deviations and Asymptotic Methods in Finance* (P. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier and J. Teichmann, eds.). *Springer Proc. Math. Stat.* **110** 559–590. Springer, Cham. [MR3375197](#)
- [16] JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKII, M. and VETTER, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.* **119** 2249–2276. [MR2531091](#)
- [17] KALNINA, I. and XIU, D. (2017). Nonparametric estimation of the leverage effect: A trade-off between robustness and efficiency. *J. Amer. Statist. Assoc.* **112** 384–396. [MR3646579](#)
- [18] KRISTENSEN, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory* **26** 60–93. [MR2587103](#)
- [19] LI, J., LIU, Y. and XIU, D. (2019). Supplement to “Efficient estimation of integrated volatility functionals via multiscale jackknife.” DOI:10.1214/18-AOS1684SUPP.
- [20] LI, J., TODOROV, V. and TAUCHEN, G. (2016). Inference theory for volatility functional dependencies. *J. Econometrics* **193** 17–34. [MR3500174](#)
- [21] LI, J., TODOROV, V. and TAUCHEN, G. (2017). Adaptive estimation of continuous-time regression models using high-frequency data. *J. Econometrics* **200** 36–47. [MR3683660](#)
- [22] LI, J. and XIU, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica* **84** 1613–1633. [MR3537165](#)
- [23] MYKLAND, P. A. and ZHANG, L. (2017). Assessment of uncertainty in high frequency data: The observed asymptotic variance. *Econometrica* **85** 197–231. [MR3611770](#)
- [24] REISS, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *Ann. Statist.* **39** 772–802. [MR2816338](#)
- [25] RENAULT, E., SARISOY, C. and WERKER, B. J. M. (2017). Efficient estimation of integrated volatility and related processes. *Econometric Theory* **33** 439–478. [MR3600049](#)
- [26] SHEPHARD, N. and XIU, D. (2017). Econometric analysis of multivariate realised QML: Estimation of the covariation of equity prices under asynchronous trading. *J. Econometrics* **201** 19–42. [MR3704675](#)

- [27] ZHANG, L. (2011). Estimating covariation: Epps effect, microstructure noise. *J. Econometrics* **160** 33–47. [MR2745865](#)
- [28] ZHANG, L., MYKLAND, P. A. and AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411. [MR2236450](#)

J. LI  
DEPARTMENT OF ECONOMICS  
DUKE UNIVERSITY  
DURHAM, NORTH CAROLINA 27708-0097  
USA  
E-MAIL: [jl410@duke.edu](mailto:jl410@duke.edu)

Y. LIU  
DEPARTMENT OF STATISTICS AND  
OPERATIONS RESEARCH  
UNIVERSITY OF NORTH CAROLINA  
AT CHAPEL HILL  
CHAPEL HILL, NORTH CAROLINA 27599  
USA  
E-MAIL: [yunxiao@live.unc.edu](mailto:yunxiao@live.unc.edu)

D. XIU  
BOOTH SCHOOL OF BUSINESS  
UNIVERSITY OF CHICAGO  
CHICAGO, ILLINOIS 60637  
USA  
E-MAIL: [dacheng.xiu@chicagobooth.edu](mailto:dacheng.xiu@chicagobooth.edu)