Econometric Theory, **33**, 2017, 439–478.
doi:10.1017/S0266466616000013

EFFICIENT ESTIMATION OF INTEGRATED VOLATILITY AND RELATED PROCESSES

ERIC RENAULT
Brown University

CISIL SARISOY
Tilburg University

BAS J.M. WERKER
Tilburg University

We derive nonparametric efficiency bounds for regular estimators of integrated smooth transformations of instantaneous variances, in particular, integrated power variance. We find that realized variance attains the efficiency bound for integrated variance under both regular and irregular sampling schemes. For estimating higher powers such as integrated quarticity, the block-based procedures of Mykland and Zhang (2009) can get arbitrarily close to the nonparametric bounds, when observation times are equidistant. Moreover, the estimator in Jacod and Rosenbaum (2013), whose efficiency was documented for the submodel assuming constant volatility, is efficient also for nonconstant volatility paths. When the observation times are possibly random but predictable, we provide an estimator, similar to that of Kristensen (2010), which can get arbitrarily close to the nonparametric bound. Finally, parametric information about the functional form of volatility leads to a lower efficiency bound, unless the volatility process is piecewise constant.

1. INTRODUCTION

The availability of high-frequency data has led to the development of new estimators of integrated volatility and their asymptotic properties (see Andersen and Bollerslev (1998), Andersen et al. (2001, 2003), Barndorff-Nielsen and Shephard (2001, 2002a,b), among many others). Inference from high-frequency data is not only confined to integrated volatility. There is a strand of literature on estimating

The authors thank Peter C.B. Phillips (the editor), Victor Chernozhukov and Guido Kuersteiner (co-editors), anonymous referee and Willem Haemers, Nikolaus Hautsch, George Tauchen, Viktor Todorov, Arthur van Soest, Ramon van den Akker, Dacheng Xiu, Lan Zhang and seminar participants at Boston College, Harvard-MIT, the 3rd Humboldt–Copenhagen Conference, Monash University, the SoFiE 2013 conference, and University of Tasmania for their helpful comments. Address correspondence to Bas J.M. Werker, Econometrics and Finance Groups, CentER, Tilburg University; e-mail: B.J.M.Werker@TilburgUniversity.edu.

power variances, or other smooth transformations, based on the intraday price data (see Barndorff-Nielsen and Shephard (2003, 2004), Jacod (2008), Mykland and Zhang (2009), Kristensen (2010), and Jacod and Rosenbaum (2013) to name just a few).

We focus in this paper on the issue of *efficient* estimation of integrated smooth transformations of instantaneous variances from two perspectives. First, we analyze the efficiency of estimators proposed in the literature and propose a new estimator that can deal with unequally-spaced (random, but predictable) observation times. Secondly, from the perspective of modeling, we detail the efficiency gains possible when a researcher is willing to make parametric assumptions on the volatility path or, equivalently, we characterize which volatility specifications are adaptive. For this, we analyze, in a concrete probabilistic setting, limiting experiments concerning inference about integrated functions of volatility. To be precise, for the derivation of the efficiency bounds, we consider returns, conditionally on the realizations of both the volatility function and the sampling times, to be normally distributed. This setting is deliberately much simpler than the assumptions that are usually imposed in the literature about realized quantities that often consider general Itô semimartingales, possibly contaminated by some micro structure noise.

Our work complements two recent papers in various ways. First, Clément, Delattre, and Gloter (2013) derive a locally asymptotically mixed normal (LAMN) limiting experiment, assuming that volatility follows a diffusion process. Their Proposition 2 is closely related to our Theorem 3.3, though conceptually different. To be precise, we consider the pathwise properties of volatility such that the model is locally asymptotically normal (LAN), i.e., we consider the inference problem conditionally on the realized intraday path of volatility. This does not mean that we assume volatility to be deterministic, but the statistical experiments we consider are conditional on the realization of the volatility path. This approach is also taken in Reiß (2011). We precisely identify the pathwise properties needed for LAN and show that, for instance, jumps in volatility are not excluded. Moreover, we do not need to assume that the volatility process is a semimartingale. It might also entail, for instance, some fractional Brownian motion component to accommodate long memory in volatility (see, e.g., Comte and Renault (1998)). Although in their more abstract results, Clément et al. (2013) allow for (deterministic) irregularly spaced observation times, their discussion about the efficiency of existing estimators of integrated power variance focuses on regularly spaced data only. We consider even random, albeit predictable, irregularly spaced observation times. Moreover, we also provide a new (nearly) efficient estimator in this case. Second, our paper complements Reiß (2011) as that paper focuses on the limiting experiments arising when prices are contaminated with (market micro structure) noise. This turns out to lead to fundamentally different limiting experiments and, even, to different optimal rates of convergence. In the absence of noise, Reiß (2011)'s limit experiments are no longer valid. Moreover, the absence of noise leads to less restrictive assumptions on the sample paths of volatility, in particular, allowing jumps. We discuss these links in more detail in the remainder of the paper.

The local asymptotic normality result we derive in our simple setting (which allows us to explicitly analyze likelihood ratio processes of a fairly simple form) leads to a well-defined optimality concept for asymptotic inference using the so-called convolution theorem. As stated in Jacod and Rosenbaum (2013), "... even for the simpler problem of estimating integrated volatility, the concept of efficiency in the general nonparametric or semiparametric setting is not well established so far". Accordingly, they decide to generally call "efficient" a procedure which is efficient in the usual sense for the submodel assuming a time-invariant volatility, i.e., what they call a "toy" model where observations are generated by a constant volatility Brownian motion. However, this simple concept of efficiency is not sufficient for at least two reasons. First, if a nonparametric estimator attains a bound induced by a parametric submodel, the nonparametric estimator can indeed be called efficient, but only for data generating processes that belong to this parametric submodel. One of the consequences of our paper is that the Jacod and Rosenbaum (2013) estimator is even nonparametrically efficient for nonconstant volatility paths, at least when the observation times are equally spaced. Second, indeed, if the observation times are irregular, like transactions times or times of quote changes, then these simple parametric submodels may be misleading about nonparametric efficiency. Hayashi, Jacod, and Yoshida (2011) give a counterexample considering the estimation of a submodel with time-invariant volatility with irregular sampling times. For this simple model, the maximum likelihood estimator (MLE) is asymptotically efficient and easy to compute as an average of squared returns divided by corresponding durations. Unfortunately, this MLE formula does not even deliver a consistent estimator of integrated variance if volatility is time varying and the sampling times are irregular. In other words, the quasi-maximum likelihood estimation approach of Xiu (2010) with regularly sampled high-frequency data, based on a quasi-likelihood method as if the volatility were constant, does not work with irregular sampling. Our paper generalizes the counterexample of Hayashi et al. (2011) by showing that even in the case of a parametric model defined by a piecewise constant volatility, the parametric efficiency bound for estimating integrated powers of the volatility does not coincide with the nonparametric bound. It takes an ad hoc assumption (namely, the specific power of volatility properly rescaled according to the density of observation times being piecewise constant) in order to obtain equality of the parametric and nonparametric efficiency bounds.

The present paper offers three important contributions. First of all, we extend the analysis of efficiency of estimators for integrated transformations of instantaneous variance to the situation of irregularly spaced, random but predictable, sampling times. It follows that, for integrated variance, realized variance remains nonparametrically efficient in this case. Our results also show how the denseness of observations throughout the day affects the possible precision of estimators. We provide an estimator, similar to the one proposed in Kristensen (2010), that is nearly efficient also when observation times are irregularly spaced. The "near" efficiency signifies that the limiting variance of our estimator can get arbitrarily close to the nonparametric lower bound, just like the block-based procedure in Mykland and Zhang (2009). We have not yet been able to derive a fully efficient estimator when observation times are irregularly spaced in analogy to the one in Jacod and Rosenbaum (2013) which is based on regularly sampled observation times.

Second, we detail the pathwise properties of volatility needed to obtain our LAN result. This is summarized in the new concept of (sample paths of) locally bounded variance, a concept that does not rule out jumps in volatility and is satisfied, e.g., by the sample paths of Brownian motion. We expect, though we were unable to prove formally, that this condition is much more generally satisfied by sample paths of (Brownian) semimartingales.

Third, we show for which volatility paths and parametric volatility specifications, the nonparametric and parametric lower bound coincide, i.e., when the nonparametric model is adaptive. This is useful if in empirical work, one is willing to take misspecification risk in return for efficiency gains. We show that these gains indeed can be sizable. This has sometimes been overlooked in the literature, in particular due to the fact that realized variance achieves the parametric lower bound for constant volatility specifications (with and without regularly spaced observations).

The rest of the paper is organized as follows. Section 2 introduces our model setup and provides the local properties of volatility paths which are assumed for our analysis. In particular, the functional parameter of interest is made more explicit to address the efficiency issue and the distribution of possibly random observation times is discussed. Section 2 states the set of maintained assumptions about both the return process and the sampling scheme sufficient to derive our asymptotic results. In Section 3, we obtain our lower bounds for (smooth) functionals of volatility and show, for equidistant data, when the nonparametric efficiency bound is attained by existing estimators as those proposed by Mykland and Zhang (2009) and Jacod and Rosenbaum (2013). In Section 4, we discuss an (nearly) efficient estimator for integrated smooth functions of volatility, also when observations are irregularly spaced in time. Subsequently, Section 5 shows in which circumstances parametric volatility models lead to additional information that can be exploited statistically, i.e., when MLE improves upon the nonparametric efficiency bound. Finally, Section 6 concludes and the appendix gathers proofs together.

2. SETTING AND PATHWISE PROPERTIES

We are interested in the pathwise properties of a univariate (instantaneous variance) process $\sigma^2 = \{\sigma^2(t) : 0 \le t \le 1\}$. All processes are assumed to be adapted to a (given) filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0\le t\le 1}, \mathbb{P})$. Inference about (the paths of) σ^2 will be based on observations on a, say, log-price process S at random sampling times $t_{i,n}$. Section 2.1 introduces the properties required on

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

these sampling times, while Section 2.2 discusses further assumptions on S. In order to analyze the impact of the denseness of observations on efficiency bounds, Section 2.3 introduces a process which is closely related to the quadratic variation of time by Mykland and Zhang (2006).

2.1. The sampling scheme

At stage n, we consider a strictly increasing sequence of stopping times $t_{i,n}$, with $i = 0, 1, \ldots, N_n$ and $0 = t_{0,n} < t_{1,n} < \cdots < t_{N_n,n} \leq 1$. The deterministic sequence $n = 1, 2, \ldots$ is used to index the experiments whose limit we will consider. Note that N_n stands for the actual number of observations available at stage n, and this number may be random. Also, even though the volatility process σ^2 is never observed directly, we call, for now with some abuse of terminology, the times $t_{i,n}$ "observation times". In addition, the double array $\left(t_{i,n}\right)_{0 \leq i \leq N_n; n \geq 1}$ of stopping times forms the "sampling scheme".

We assume the mesh of the sampling scheme to converge to zero at some deterministic rate and N_n/n to be bounded from above, almost surely. As the sequence n is an index to define the asymptotic setup, we can always choose N_n to be bounded by n. A standard assumption would be to assume the mesh of the sampling scheme to converge to zero at rate n^{-1}. However, we relax this assumption in order to include, for instance, Poisson sampling with intensity of the order $O(n)$ for which N_n/n converges to 1 but the mesh can only be bounded by $O(\log(n)/n)$.

Assumption 1. We suppose that the sampling scheme is a double array $\left(t_{i,n}\right)_{0 \leq i \leq N_n; n \geq 1}$ of stopping times with respect to $\{\mathcal{F}_t\}_{0 \leq t \leq 1}$ such that

- $0 = t_{0,n} < t_{1,n} < \cdots < t_{N_n,n} \leq 1$;
- for all n, $N_n \leq n$ a.s.;
- $\sqrt{n} \max \left| t_{i,n} - t_{i-1,n} \right| = o_P(1)$, as $n \to \infty$;
- $n \sum_{i=1}^{n} \left(t_{i,n} - t_{i-1,n}\right)^2 = O_P(1)$ as $n \to \infty$.

Moreover, we maintain the (restrictive) assumption that the stopping times are strongly predictable. Obviously, this does cover the case of irregularly, but deterministically, spaced observation times.

Assumption 2. The sampling scheme $\left(t_{i,n}\right)_{0 \leq i \leq N_n; n \geq 1}$ is $\{\mathcal{F}_t\}_{0 \leq t \leq 1}$-predictable; that is $t_{i,n}$ is $\mathcal{F}_{t_{i-1,n}}$ measurable, for all i, n.

It is worth acknowledging that in the spirit of Assumption (C) in Hayashi et al. (2011), it would be possible to relax Assumption 2 so that, conditionally on $\mathcal{F}_{t_{i-1,n}}$, $t_{i,n}$ is independent of the processes of interest. Allowing for genuinely endogenous sampling times, albeit realistic for transaction times, much complicates the specification of the conditional distribution of returns given the observation

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

times. As extensively discussed by Li et al. (2014), it significantly changes the asymptotic distribution of, e.g., realized volatility. A study of nonparametric efficiency in this more general setting appears to be a daunting task, beyond the scope of this paper.

The pathwise behavior of the volatility process σ of interest has to be restricted by a regularity condition. To formalize this, we introduce the concept of locally bounded variance.

DEFINITION 1. *Let f be a real-valued cadlag function on the interval* $[0,1]$. *We say that f is of* locally $\{\mathcal{F}_t\}_{0\le t\le 1}$-bounded variance *if, for any sampling scheme $\left(t_{i,n}\right)_{0\le i\le N_n;n\ge 1}$ satisfying Assumptions 1 and 2,*

$$\max_{n\in\mathbb{N}}\left\{\sum_{i=1}^{N_n} V\left(f|_{t_{i-1,n}}^{t_{i,n}}\right)\right\} \tag{2.1}$$

is almost surely bounded, where

$$V\left(f|_{t_{i-1,n}}^{t_{i,n}}\right)=\frac{1}{t_{i,n}-t_{i-1,n}}\int_{t_{i-1,n}}^{t_{i,n}}\left(f(u)-\frac{1}{t_{i,n}-t_{i-1,n}}\int_{t_{i-1,n}}^{t_{i,n}}f(v)dv\right)^2du. \tag{2.2}$$

Throughout the paper, we assume that the sample paths of σ^2 are of locally $\{\mathcal{F}_t\}_{0\le t\le 1}$-bounded variance. Observe that Lemma A.1 and the observation that $\mathrm{V}(f|_{t_{i-1,n}}^{t_{i,n}})$ equals $\mathrm{V}(f(U))$ when U is uniformly distributed over $(t_{i-1,n},t_{i,n}]$, implies that then also the sample paths of σ^{-2} are of locally bounded variance. As this assumption is key to the asymptotic theory developed in the next sections, it is worth analyzing whether the condition holds for paths of often-used volatility processes.

First of all, note that if f and g are of locally $\{\mathcal{F}_t\}_{0\le t\le 1}$-bounded variance, so are their sum and scalar products. The functions of locally $\{\mathcal{F}_t\}_{0\le t\le 1}$-bounded variance thus form a vector space. Second, we have the bound, for $0\le i\le N_n$,

$$\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right)\le\frac{1}{4}\left[\max_{t_{i-1,n}\le u\le t_{i,n}}f(u)-\min_{t_{i-1,n}\le u\le t_{i,n}}f(u)\right]^2, \tag{2.3}$$

so that

$$\sum_{i=1}^{N_n}\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right)\le\frac{1}{4}\sum_{i=1}^{N_n}\left[\max_{t_{i-1,n}\le u\le t_{i,n}}f(u)-\min_{t_{i-1,n}\le u\le t_{i,n}}f(u)\right]^2. \tag{2.4}$$

As a result, monotonic functions f are of locally $\{\mathcal{F}_t\}_{0\le t\le 1}$-bounded variance and, hence, so are functions of finite variation. In particular, our analysis does not rule out jumps in volatility. For functions not of finite variation, the analysis is more subtle but diffusions are not excluded; see Appendix A for details.

2.2. Functional parameter and model

Inference about (functionals of) the path of σ^2 will be based on observations of a process $S = \{S(t) : 0 \le t \le 1\}$, which can be thought of as a log-price process, at the sampling times $\left(t_{i,n}\right)_{0\le i\le N_n; n\ge 1}$. We throughout assume that S is $\{\mathcal{F}_t\}_{0\le t\le 1}$-adapted. In this section, we formalize the underlying data generating mechanism for the derivation of the efficiency bound. Essentially, we consider returns, conditionally on both the realizations of the volatility and the sampling times, to be normally distributed. We stress that our kernel-based estimator in Section 4 does not require Gaussianity of the returns.

Assumption 3. Suppose that $R_{i,n} := \log\left\{S_{t_{i,n}}/S_{t_{i-1,n}}\right\}$ is the log-return over the time interval $(t_{i-1,n}, t_{i,n}]$. Then, conditionally on $\mathcal{F}_{t_{i-1,n}}$, $R_{i,n}$ is distributed as

$$N\left(\left[\mu(t_{i-1,n}, t_{i,n}) + \gamma(t_{i-1,n}, t_{i,n})\sigma^2(t_{i-1,n}, t_{i,n})\right]\Delta t_{i,n}, \sigma^2(t_{i-1,n}, t_{i,n})\Delta t_{i,n}\right), \quad \textbf{(2.5)}$$

where $\mu(t_{i-1,n}, t_{i,n})$, $\sigma^2(t_{i-1,n}, t_{i,n})$, and $\gamma(t_{i-1,n}, t_{i,n})$ are bounded and $\mathcal{F}_{t_{i-1,n}}$-measurable.

It is worth stressing that for our LAN result it is not needed that *investors* know quantities as $\mu(t_{i-1,n}, t_{i,n})$, $\sigma^2(t_{i-1,n}, t_{i,n})$, or $\Delta t_{i,n}$ at time $t_{i-1,n}$. They may see the volatility (and the subsequent times of trades) as stochastic and their conditional variance at time $t_{i-1,n}$ of the return $R_{i,n}$ can be computed as the projection of the above moments on the σ-field subset of $\mathcal{F}_{t_{i-1,n}}$ that describes the investors' information at time $t_{i-1,n}$. The sequence of σ-fields $\mathcal{F}_{t_{i-1,n}}$, $i = 1, \ldots, N_n$ is a modeling tool for the sake of specifying a statistical model and should not necessarily be interpreted in terms of investors' information. Note that, precisely contrary to the limiting approximations discussed in Reiß (2011), we exclude possible (micro structure) contamination of observed returns. As explained in the introduction, the resulting limiting experiments are materially different.

We assume that the information brought by asset returns is exogenous in the following sense.

Assumption 4. The filtration $\{\mathcal{F}_t\}_{0\le t\le 1}$ is the natural one of $\{(S(t), Z(t)) : 0 \le t \le 1\}$, where $\{Z(t) : 0 \le t \le 1\}$ is a (possibly multivariate) stochastic process of exogenous variables such that, given $\mathcal{F}_0$ and N_n, the joint conditional density of $\{(S(t), Z(t)) : 0 \le t \le 1\}$ can be written

$$p\left(\left\{Z_{t_{i,n}}\right\}_{1\le i\le N_n}\right)\prod_{i=1}^{N_n} p\left(R_{i,n}|\mathcal{F}_{t_{i-1,n}}\right). \quad \textbf{(2.6)}$$

Remark 2.1. Assumption 4 requires non-causality (in the Sims sense) from the return process $(R_{i,n})_{1\le i\le N_n}$ to the $(Z_{t_{i,n}})_{1\le i\le N_n}$. Otherwise, the conditioning information in $p(R_{i,n}|\mathcal{F}_{t_{i-1,n}})$ should also involve future values $Z_{t_{j,n}}$, $j > i$. We also preclude instantaneous causality in order to erase the contemporaneous

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

value $Z_{t_{i,n}}$ in the conditioning information that defines the probability distribution of $R_{i,n}$. Up to discussions regarding initial values, Sims' non-causality is known to be equivalent to Granger non-causality. In other words, we basically assume that returns do not Granger-cause state variables, like stochastic volatility.

It is generally convenient to study the return variance as an integral over the corresponding event interval of the so-called spot volatility process $\sigma^2 = \{\sigma(t) : 0 \leq t \leq 1\}$, so that:

$$\sigma^2(t_{i-1,n}, t_{i,n})\left[t_{i,n} - t_{i-1,n}\right] = \int_{t_{i-1,n}}^{t_{i,n}} \sigma^2(u)\mathrm{d}u. \tag{2.7}$$

In other words, $\sigma^2(t_{i-1,n}, t_{i,n})$ is the arithmetic mean of the function σ^2 over the interval $(t_{i-1,n}, t_{i,n}]$. This situation occurs, for instance, when the returns are generated from log-prices S_t satisfying the differential equation

$$\mathrm{d}\log S_t = a_t \mathrm{d}t + \sigma(t)\mathrm{d}W_t, \tag{2.8}$$

for some appropriate drift a_t and Brownian motion W_t.

We introduce the following notations and assumption. Let $D[0,1]$ denote the set of real-valued cadlag functions on the interval $[0,1]$ and let $D_+[0,1]$ be the subset of functions that take strictly positive values only. Both spaces are equipped with the supremum norm $\|\cdot\|$. Assumption 5 will not be imposed in the limiting results of our kernel-based estimator in Section 4.

Assumption 5. The parameter space Ξ is the set of all elements σ^2 in $D_+[0,1]$ such that σ^2 is bounded away from zero and of locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance.

2.3. On the denseness of events

In order to analyze the consequences of irregularly spaced event times, we introduce a process representing effectively the time change induced by the observation times $t_{i,n}$. We define

$$T_n(u) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{t_{i,n} - t_{i-1,n}}\left[u \wedge t_{i,n} - u \wedge t_{i-1,n}\right], \qquad 0 \leq u \leq 1. \tag{2.9}$$

Observe that the function T_n is piecewise linear and increases over each time interval $(t_{i-1,n}, t_{i,n}]$ exactly an amount $1/n$. As a result $T_n(t_{i,n}) = i/n$, $i = 0, \ldots, n$. Thus, at the observation times $u = t_{i,n}$, $T_n(u)$ coincides with the empirical distribution function of these observation times. For regularly spaced data $t_{i,n} = i/n$ and $T_n(u) = u$. The function T_n is also closely related to what Mykland and Zhang (2006) call the quadratic variation of time; a relation that we will make precise below. Also, note that the sum in (2.9) is taken to $i = n$, and not, as before, to $i = N_n$. This is an abuse of notation that we will maintain throughout the remainder of the paper and is warranted in view of (A.3).

We impose the following additional assumption on the observation times $t_{i,n}$.

Assumption 6. The function T_n as defined in (2.9) converges almost surely to a distribution function T on $[0,1]$ in the topology of weak convergence. That is, for any function $f \in D[0,1]$,

$$\int_{u=0}^{1} f(u)\mathrm{d}T_n(u) \to \int_{u=0}^{1} f(u)\mathrm{d}T(u),$$

almost surely, as $n \to \infty$. Moreover, T admits a strictly positive and bounded density T' and $T_n' \to T$, almost surely.

Remark 2.2. Assumption 6 has a few strong consequences. As the limit T is continuous, the functions T_n monotone, and $[0,1]$ compact, the convergence of T_n to T is uniform (see Buchanan and Hildebrandt (1908)). Consequently, again using the continuity of T, the quantile functions T_n^{-1} also converge weakly, pointwise, and, by the same argument, uniformly to T^{-1}.

It is informative to relate Assumption 6 to the concept of asymptotic quadratic variation of time (AQVT) as in Mykland and Zhang (2006). Observe

$$T_n'(u) = \left[n\left(t_{i,n} - t_{i-1,n}\right)\right]^{-1} \text{ for } t_{i-1,n} < u < t_{i,n}.$$

Consequently, for $u \in [0,1]$,

$$H_n(u) := \int_0^u \frac{1}{T_n'(z)}\mathrm{d}z = n \sum_{t_{i,n} \le u} \left(t_{i,n} - t_{i-1,n}\right)^2 + n\left(t_{i^*+1,n} - t_{i^*,n}\right)\left(u - t_{i^*,n}\right),$$

where $t_{i^*,n} \le u \le t_{i^*+1,n}$. In general, convergence of T_n and H_n cannot be related, but, under appropriate additional smoothness, one would have

$$H(u) := \lim_{n\to\infty} H_n(u) = \lim_{n\to\infty} \int_0^u \frac{1}{T_n'(z)}\mathrm{d}z, \tag{2.10}$$

and thus $H'(u) = 1/T'(u)$. We represent the denseness of events in terms of T_n and T, rather than H_n and H, as this notion arises naturally in the study of the likelihood ratios in the next section. Following Mykland and Zhang (2012), it is easy to check that $H'(u) \equiv 1$ (or, equivalently, $T'(u) \equiv 1$) if and only if

$$\sum_{i=1}^{N_n} \left(t_{i,n} - t_{i-1,n} - \frac{1}{n}\right)^2 = o_P(1).$$

We call such a sampling scheme (asymptotically) regular.

3. LOWER BOUNDS FOR INTEGRATED FUNCTIONS OF VARIANCE

As indicated in the introduction, we base our optimality criteria on the Hájek-Le Cam theory of convergence of experiments. We refer to, e.g., van der Vaart (2000) for details. We show in this section that appropriately parametrized local versions of the model described in Assumptions 1–6 converge, as n tends to infinity, to a Gaussian shift experiment, i.e., our experiment is locally asymptotically normal (see van der Vaart (2000), Section 7). Using by now standard arguments, the least-favorable of these parametric submodels describes the lower bound for estimating functionals of the volatility path, see Section 3.

As is to be expected, realized variance plays a key role in the analysis. The idea of estimating volatility of returns over a fixed interval as the sum of squared realizations given the availability of sufficiently high sampling frequency was noted already in Merton (1980). More recently, realized variance measures constructed from intraday data have been exploited by Taylor and Xu (1997) and Andersen, Bollerslev, and Lange (1999), among others. We define the realized variance process as

$$RV_n(u) = \sum_{i=1}^{n} R_{i,n}^2 I\left\{t_{i,n} \le u\right\}, \qquad 0 \le u \le 1. \tag{3.1}$$

Observe that $RV_n(1)$ coincides with the standard definition of realized variance, but we need this process version below to clarify the role of unequally spaced time points $t_{i,n}$ later. It turns out that in the local asymptotic normality result, another process plays an important role. We define the duration-weighted realized variance process as

$$RV_n^*(u) = \frac{1}{n}\sum_{i=1}^{n} \frac{R_{i,n}^2}{\left(t_{i,n} - t_{i-1,n}\right)^2}\left[u \wedge t_{i,n} - u \wedge t_{i-1,n}\right], \qquad 0 \le u \le 1. \tag{3.2}$$

Note that RV_n^* is a piecewise linear process that increases by an amount of $R_{i,n}^2/[n\left(t_{i,n} - t_{i-1,n}\right)]$ over the interval $(t_{i-1,n}, t_{i,n}]$. We could have used a piecewise constant definition, similar to the definition of the realized variance process RV, but the continuity induced by the linear interpolation turns out to be mathematically convenient.

We first provide a joint functional central limit result for the realized variance RV_n and the duration-weighted realized variance RV_n^* above. A proof is again provided in the appendix. As our nearly efficient estimator in Section 4 takes this result as input, we formulate it here as a condition. Thus, once more, our estimator is valid under much more general conditions than Assumptions 1–6. In particular, it does not rely on returns being normally distributed.

Condition 1. *The processes*

$$\sqrt{n}\left[RV_n(u) - \int_{v=0}^{u} \sigma^2(v)dv\right] \tag{3.3}$$

and

$$\sqrt{n}\left[RV_n^*(u) - \int_{v=0}^{u} \sigma^2(v)dT_n(v)\right] \tag{3.4}$$

jointly converge to $\sqrt{2}\int_{v=0}^{u}\sigma^2(v)T'(v)^{-1/2}dW(v)$ *and* $\sqrt{2}\int_{v=0}^{u}\sigma^2(v)T'(v)^{+1/2}dW(v)$*, respectively, where* W *denotes a standard Brownian motion. The convergence is weak in* $(D[0,1],\|\cdot\|)$.

LEMMA 3.1. *Under Assumptions 1–6, Condition 1 holds.*

As mentioned before, the nonparametric analysis where σ^2 denotes a functional parameter is, in line with classical reasoning, reduced by considering parametric submodels and, then (see Theorem 3.3), considering the least-favorable among them. Thus, fix $\sigma_0^2 \in D_+[0,1]$ and define local alternatives for $h \in D[0,1]$, with $\|h\| \le 1$, by

$$\sigma_{\alpha/\sqrt{n}}^{-2}(u) = \sigma_0^{-2}(u)\left[1+\frac{\alpha}{\sqrt{n}}h(u)\right], \qquad \alpha \in (-1,1). \tag{3.5}$$

We write shorthand $\mathbb{P}_\alpha^{(n)}$ for the probability measure induced by $R_{1n},\ldots,R_{nn}$ under $\sigma_{\alpha/\sqrt{n}}^2$, i.e., $\mathbb{P}_\alpha^{(n)} = \mathbb{P}_{\sigma_{\alpha/\sqrt{n}}^2}^{(n)}$.

THEOREM 3.2. *Under Assumptions 1–6, the experiment* $\left\{\mathbb{P}_\alpha^{(n)} : \alpha \in (-1,1)\right\}$ *is asymptotically normal with*

$$\log\frac{\mathrm{d}\mathbb{P}_\alpha^{(n)}}{\mathrm{d}\mathbb{P}_0} = -\frac{\alpha}{2}\int_0^1 \sigma_0^{-2}(u)h(u)d\sqrt{n}\left[RV_n^*(u)-\int_{v=0}^{u}\sigma_0^2(v)dT_n(v)\right] \tag{3.6}$$
$$-\frac{\alpha^2}{4}\int_{u=0}^{1}h^2(u)dT(u).$$

The Fisher information is given by $\frac{1}{2}\int_{u=0}^{1}h^2(u)dT(u)$*. In particular, the probability measures* $\mathbb{P}_\alpha^{(n)}$ *and* $\mathbb{P}_0^{(n)}$ *are contiguous.*

For more details on local asymptotic normality results as in Theorem 3.2, we refer the reader to van der Vaart (2000). In the following section, we will use it to establish lower bounds on the precision of regular estimators for integrals of smooth transformations of instantaneous variances, in the absence of micro structure noise.

Assume now that we are interested in estimating the following generalized version of the standard integrated volatility

$$\psi_g\left(\sigma^2\right) = \int_0^1 g(u,\sigma^2(u))\omega(u)\mathrm{d}u, \tag{3.7}$$

where ω is a known weighting function on $[0,1]$ and $g:[0,1]\times\mathbb{R}\to\mathbb{R}$ is a known time-dependent transformation. This setup includes most standard measures.

(i) For $\omega(u)\equiv 1$ and $g(u,\sigma^2(u)) = \sigma^{2p}(u)$ we have the so-called "power variation" as studied in particular by Jacod (2008). These power variations are popular in particular to assess the asymptotic variance of estimators of power variation of

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

lower order. For instance, quarticity ($p = 2$) is informative about the asymptotic variance of realized variance ($p = 1$). A nonflat weighting function ω may be necessary to accommodate the effect of irregular sampling.

(ii) For $\omega(u) \equiv 1$ and $g(u, \sigma^2(u)) = \exp\left(-s\sigma^2(u)\right)$ for some given $s \in \mathbb{R}_+$, we have the empirical Laplace transform function of the volatility process. A consistent asymptotically (mixed) normal estimator of the Laplace transform has been provided by Todorov and Tauchen (2012). Li, Tauchen, and Todorov (2013) subsequently use this estimator to estimate the volatility occupation time corresponding to $g(u, \sigma^2(u)) = 1_{]0,x]}(\sigma^2(u))$ for some given $x \in \mathbb{R}_+$. However, this latter example will not be covered here since, in order to compute Cramér-Rao efficiency bounds, we always assume that $g(u, \sigma^2)$ is a continuously differentiable function of the underlying spot variance σ^2.

(iii) Time-dependent transformations of volatility $g(u, \sigma^2(u))$ may be relevant when computing implied volatilities from option prices. These option prices typically depend not only on the underlying spot volatility but also on the time to maturity.

Along the paths induced by $\sigma^2_{\alpha/\sqrt{n}}$ defined in (3.5) we have

$$\begin{aligned}\psi_g\left(\sigma^2_{\alpha/\sqrt{n}}\right) &= \int_0^1 g\left(u, \sigma_0^2(u)\left(1 + \frac{\alpha}{\sqrt{n}} h(u)\right)^{-1}\right)\omega(u)\mathrm{d}u \\ &= \psi_g\left(\sigma_0^2\right) - \frac{\alpha}{\sqrt{n}}\int_0^1 \frac{\partial g}{\partial \sigma^2}\left(u, \sigma_0^2(u)\right)\sigma_0^2(u)h(u)\omega(u)\mathrm{d}u + o\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

as $n \to \infty$. This expansion is valid uniformly in h for $\|h\| \leq 1$ if g fulfills the following assumption.

Assumption 7. $g\left(u, \sigma^2\right)$ is continuous in u and continuously differentiable in σ^2.

As a result of the above expansion, the Fréchet derivative of our parameter of interest $\psi_g\left(\sigma^2\right)$ with respect to $\alpha/\sqrt{n}$) is given by

$$-\int_0^1 \frac{\partial g}{\partial \sigma^2}\left(u, \sigma_0^2(u)\right)\sigma_0^2(u)h(u)\omega(u)\mathrm{d}u.$$

We can now proceed as usual and derive the nonparametric lower bound for estimating $\psi_g\left(\sigma^2\right)$ as the largest bound obtained in the parametric models indexed by h.

More precisely, given h, we apply the Convolution Theorem as, for instance, stated in Bickel et al. (1993) Theorem 2.3.1. Consider a regular[1] estimator $\hat{\psi}_g^{(n)}$ for $\psi_g\left(\sigma^2\right)$ in the sense that, under $\mathbb{P}_\alpha^{(n)}$,

$$\sqrt{n}\left(\hat{\psi}_g^{(n)} - \psi_g\left(\sigma^2_{\alpha/\sqrt{n}}\right)\right) \to_L N(0, V). \tag{3.8}$$

Then, we know that V is at least equal to the squared derivative of $\psi_g\left(\sigma^2\right)$ (with respect to α) times the inverse of the Fisher information, thus

$$V \geq \frac{\left(\int_0^1 \frac{\partial g}{\partial \sigma^2}\left(u,\sigma_0^2(u)\right)\sigma_0^2(u)h(u)\omega(u)\mathrm{d}u\right)^2}{\int_0^1 h^2(u)dT(u)/2}. \tag{3.9}$$

We thus have the following nonparametric bound.

THEOREM 3.3. *Under Assumptions 1–7, any regular estimator for $\psi_g\left(\sigma^2\right)$ as defined in (3.7) based on observations on the grid $t_{i,n}$, $i = 1,\ldots,n$, has, under $\mathbb{P}^{(n)}_{\sigma_0^2}$, a limiting variance of at least*

$$V(g) = 2\int_0^1 \left(\frac{\partial g}{\partial \sigma^2}\left(u,\sigma_0^2(u)\right)\right)^2 \sigma_0^4(u)\frac{\omega^2(u)}{T'(u)}du \tag{3.10}$$

Proof. The least-favorable submodel is obtained by choosing h in the local alternatives (3.5) such that the parametric lower bound (3.9) is maximized. From Cauchy-Schwarz, we know that this happens for

$$h(u) \propto \frac{\frac{\partial g}{\partial \sigma^2}\left(u,\sigma_0^2(u)\right)\sigma_0^2(u)\omega(u)}{T'(u)}.$$

Plugging this least-favorable h into (3.9) gives the result. ■

In order to discuss the relationship between Theorem 3.3 and the extant literature, we examine flat weights ($\omega(u) \equiv 1$) and discuss two separate cases: first (asymptotically) regular sampling ($T'(u) \equiv 1$) and, second, general sampling schemes (arbitrary T').

3.1. Regular sampling without weighting

When the sampling times are regular and $\omega(u) \equiv 1$, the efficiency bound (3.10) reduces to

$$V(g) = 2\int_0^1 \left(\frac{\partial g}{\partial \sigma^2}\left(u,\sigma_0^2(u)\right)\right)^2 \sigma_0^4(u)\mathrm{d}u. \tag{3.11}$$

This formula is a univariate version of formula (30) in Clément et al. (2013) for the case $g(u,\sigma^2) = g(\sigma^2)$. Recall, however, that Clément et al. (2013) derive this bound from a LAMN property, assuming that σ is generated by an Itô process, independent of the Brownian motion defining the return innovations. The validity of Theorem 3.3 is more general. If one wants to see σ as a stochastic process, independent of the leading Brownian motion, it can be any process whose sample paths are almost surely of locally bounded variance.

As far as the time-independency $g(u,\sigma^2) = g(\sigma^2)$ is concerned, it is worth considering both examples above.

Example 1
Power variation: The case of power variation is obtained using $g_p(\sigma^2) = \sigma^{2p}$ and leads to the nonparametric lower bound

$$V(g_p) = 2p^2 \int_0^1 \sigma_0^{4p}(u)\mathrm{d}u. \tag{3.12}$$

Practical implications of this result have been known at least since Mykland and Zhang (2009). For $p = 1$, that is for the estimation of integrated variance, empirical quadratic variation is an (asymptotically) efficient estimator. Note that this instance precisely corresponds to linear g and, hence, smoothing operations and the transformation g commute.

By contrast, for $p > 1$, the realized power variation does not deliver an efficient estimator. Indeed, Mykland and Zhang (2009) also study

$$\hat{\psi}_{g_p} = \frac{1}{\mathrm{E}|N(0,1)|^{2p}} \sum_{i=1}^{n} \left(t_{i,n} - t_{i-1,n}\right)^{1-p} \left|R_{i,n}^2\right|^p \tag{3.13}$$

and give, under somewhat different conditions, the limiting variance[2]

$$\frac{\mathrm{V}|N(0,1)|^{2p}}{\left(\mathrm{E}|N(0,1)|^{2p}\right)^2} \int_{u=0}^{1} \sigma_0^{4p}(u)\mathrm{d}u. \tag{3.14}$$

But it can easily be shown that the coefficient in front of the integral exceeds $2p^2$ if (and only if) $p > 1$. For instance, for $p = 2$ (integrated quarticity), the coefficient equals 10.67, while the lower bound is $2p^2 = 8$— an ARE of only 75%. This inefficiency is also noted in Jacod and Rosenbaum (2013).

Mykland and Zhang (2009) also provide a block-based estimator for integrated powers of volatility whose limiting variance, for large block size (and still equally spaced data) is arbitrarily close to the efficiency bound (see their formula (63)). Even though Mykland and Zhang (2009) do not formally derive an efficiency bound, they give a clear intuition of the reason why their block-based estimator is nearly efficient. Within each block, one computes the maximum likelihood estimator of the variance of returns seen as approximately homoskedastic within the blocks. Then the sum across blocks of power p of these estimators delivers a smaller asymptotic variance than the naive estimator $\hat{\psi}_{g_p}$ when both block size and number of blocks go to infinity. In this sense, their estimator is *nearly* efficient in the sense of Section 4 below. Recently, Jacod and Rosenbaum (2013) introduced an estimator that also has a limiting variance (3.10). Indeed, consider their Theorem 3.2 with the notation $d = 1$, $t = 1$, $s = u$, $g(c) = c^p$, $s_s = \sigma^2(u)$. Then, their limiting variance (3.12) leads to

$$\int_{u=0}^{1} p\sigma_0^{2(p-1)}(u)\, p\sigma_0^{2(p-1)}(u) 2\sigma^4(u)\mathrm{d}u = 2p^2 \int_{u=0}^{1} \sigma_0^{4p}(u)\mathrm{d}u,$$

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

which equals (3.10) when observation times are regular, i.e., $T' = 1$. Jacod and Rosenbaum (2013) note that their estimator is efficient at, what they call, the constant volatility toy model, i.e., $\sigma(u) = \sigma$. We show that their estimator achieves the efficiency bound also at nonconstant volatility within our nonparametric model.

Example 2

Laplace transform: Consider, for some given $s \in \mathbb{R}_+$, the transformation $g_s(\sigma^2) = \exp\left(-s\sigma^2\right)$ which leads to the efficiency bound

$$V(g_s) = 2s^2 \int_0^1 \exp\left(-2s\sigma_0^2(u)\right)\sigma_0^4(u)\mathrm{d}u. \tag{3.15}$$

Todorov and Tauchen (2012) estimate $\psi_{g_s}\left(\sigma^2\right)$ with $\omega(u) \equiv 1$) by the realized Laplace transform

$$\hat{\psi}_{g_s} = \sum_{i=1}^{n} \left(t_{i,n} - t_{i-1,n}\right)\cos\left(\sqrt{2s}\frac{R_{i,n}}{\sqrt{t_{i,n} - t_{i-1,n}}}\right), \tag{3.16}$$

which gives them (see their Theorem 1 with $u = v$), the asymptotic variance

$$V = 2\int_0^1 \exp\left(-2s\sigma^2(u)\right)\left[\frac{\exp\left(s\sigma^2(u)\right) - \exp\left(-s\sigma^2(u)\right)}{2}\right]^2 \mathrm{d}u. \tag{3.17}$$

Using the series expansion

$$\frac{\exp(x) - \exp(-x)}{2} = x\left[1 + \sum_{j=1}^{\infty}\frac{x^{2j}}{(2j+1)!}\right],$$

we find

$$V = 2s^2 \int_0^1 \exp\left(-2s\sigma^2(u)\right)\sigma_0^4(u)\left[1 + \sum_{j=1}^{\infty}\frac{\left(s\sigma^2(u)\right)^{2j}}{(2j+1)!}\right]^2 \mathrm{d}u.$$

As all terms in the series are nonnegative, the realized Laplace transform does not attain the efficiency bound for estimation of the Laplace transform $\psi_{g_s}\left(\sigma^2\right)$ (unless $s = 0$), although in applications the difference may be small. In Section 4, we provide a nearly efficient estimator which is also applicable in this case.

Even though they only address the efficiency issue "in the toy model" of i.i.d. homoskedastic normal returns (constant volatility σ), Jacod and Rosenbaum (2013) give a general statement about inefficiency of naive sample counterparts like realized power variation or the realized Laplace transform. They explain that efficiency requires in general to use "estimators for the spot volatility and approximating the integral $\psi_g\left(\sigma^2\right)$ by Rieman sums, in which the spot volatility is replaced by its estimator". As they rightly mention, this idea can be seen as a generalization of the block-based estimation idea in Mykland and Zhang (2009). It should be added that in earlier work, Kristensen (2010) had a germane idea by plugging in a kernel-based estimator of spot volatility. Both Jacod and Rosenbaum (2013), with a block-based approach, and Kristensen (2010), with a kernel-based approach, get asymptotic variances that attain the efficiency bound, even though they do not present them as the efficiency bound. Also, these papers do not consider the effect of irregular sampling. We turn to this issue now.

3.2. Irregular sampling without weighting

With a flat weighting function but possibly irregular sampling, we get the efficiency bound

$$V(g) = 2\int_0^1 \left(\frac{\partial g}{\partial \sigma^2}\left(u, \sigma_0^2(u)\right)\right)^2 \frac{\sigma_0^4(u)}{T'(u)} \mathrm{d}u \tag{3.18}$$

that is new in the literature. In the particular case of integrated volatility ($g_1(\sigma^2) = \sigma^2$), the efficiency bound above corresponds to the asymptotic variance of realized variance, first derived by Barndorff-Nielsen and Shephard (2002a), at least when $H'(u) = 1/T'(u)$, a.e. This is a confirmation of the aforementioned intuition about the linearity of the transformation g_1, that allows commutation of integration and smoothing operations. This efficiency result must be contrasted with two seemingly opposite claims in the literature.

First, Hayashi et al. (2011) claim (see their p. 1206) that "the realized volatility is asymptotically efficient only when the sampling scheme is asymptotically a regular sampling", that is $T'(u) \equiv 1$. However, our result proves the semiparametric efficiency of realized variance, even with irregular sampling. This seemingly contradictory result comes from the fact that Hayashi et al. (2011) define efficiency through the toy parametric model $\sigma^2(u) \equiv \sigma^2$, constant. Their remark thus means that the model with irregular sampling is not adaptive; see Section 5 for a more comprehensive discussion.

Second, similar to the situation when the sampling times are regular, $p = 1$ is obviously the only case for which ψ_{g_p} is efficiently estimated by its naive sample counterpart. While Mykland and Zhang (2009) study the estimation of power variation only under regular sampling scheme, they refer to Mykland and Zhang (2012) for an extension to irregular sampling. Mykland and Zhang (2012) do discuss the interaction between the block-based estimation strategy and the irregular sampling scheme, but they do not provide explicitly the semiparametric efficiency bound or the way to reach it.

4. A NEARLY EFFICIENT ESTIMATOR

The advantage of a LAN result as in Theorem 3.2 is that it indicates ways to construct efficient estimators. More precisely, the likelihood expansion, which in

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

our analysis is based on the RV^* process, provides an asymptotically sufficient statistic for the parameter of interest, i.e., the path of σ^2. Hence, it is natural to base estimators on RV^*, a route we will follow in this section. This will lead to, what we call, a *nearly* efficient estimator. That is, we will base our estimator on a fixed smoothing kernel K (to be introduced formally below). The estimator then is nearly efficient in the sense that its limiting variance can get arbitrarily close to the nonparametric lower bound (3.10), by taking a kernel K close to the point mass at zero. In this sense, the Mykland and Zhang (2009) estimator is also nearly efficient, while Jacod and Rosenbaum (2013) provide explicit convergence rates for their smoothing parameter to achieve simultaneous convergence, i.e., (full) efficiency. However, note that in both cases (near) efficiency is reached only with asymptotically regular sampling schemes—the case where the two sequences RV_n and RV_n^* are asymptotically equivalent. Our near efficiency result below will be more general since it applies to irregular, even random (albeit predictable), observation times. The trick is to base estimation on a smoothed version of the process RV_n^*.

To get the main intuition, it is worth starting with the kernel-based estimator put forward by Kristensen (2010). That paper introduces a kernel-based estimator of spot volatility

$$\hat{\sigma}^2(u) = \sum_{i=1}^{n} k_h(t_{i,n} - u) R_{i,n}^2 = \int_0^1 k_h(v - u) \mathrm{d}RV_n(v) \tag{4.19}$$

$$k_h(v) = \frac{1}{h} k(v/h), \qquad \int_{-\infty}^{+\infty} k(v)\mathrm{d}v = 1.$$

Kristensen (2010) uses standard realized variation RV, instead of RV^*, but, as mentioned before, the two are asymptotically equivalent under a regular sampling scheme. Kristensen (2010) shows that, under some regularity conditions (including continuous differentiability of the kernel function k)

$$\sup_{a \le u \le 1-a} \left| \hat{\sigma}^2(u) - \sigma^2(u) \right| = O_P(h^m) + O_P(\log(n)/\sqrt{nh}), \tag{4.20}$$

as $h \downarrow 0$, $a \downarrow 0$, and $a/h \to 0$, where the spot volatility function σ is assumed to be m times differentiable. We use Kristensen (2010)'s intuition and define a smoothed version of realized variance, which is pathwise differentiable, as

$$RV_n^S(u) = \int_{v=0}^{1} K(u - v) \mathrm{d}RV_n^*(v), \quad 0 \le u \le 1, \tag{4.21}$$

where K satisfies the following condition.

Assumption 8. The kernel K is a nonnegative real-valued function on $[-1, 1]$ which is twice continuously differentiable with K' and K'' bounded.

One can take, for instance, K as the cumulative distribution function of a probability distribution whose density corresponds to the function k defined in (4.19) and assume it to be continuously differentiable. In particular, we will keep the intuition that, with regular sampling and a probability distribution converging towards point mass at zero (irrespective of its definition through the density function k or through the cumulative distribution function K), we would have $\hat{\sigma}^2(u)$ close to $\sigma^2(u)$ and, for the same reason, $RV_n^S(u)$ close to $RV_n^*(u)$. It is then natural to extend also Kristensen (2010)'s idea of a plug-in estimator for $\psi_g\left(\sigma^2\right)$, defined (up to a local bias correction) as

$$\hat{\psi}_g = \int_0^1 g(u, \hat{\sigma}^2(u))\mathrm{d}u. \tag{4.22}$$

In our notation, this would give under regular sampling

$$\hat{\psi}_g(K|\mathrm{reg}) = \int_0^1 g(u, RV_n^{S\prime}(u))du. \tag{4.23}$$

However, this estimator would not be efficient in general with irregular sampling. To see that, it is worth recalling that, as stressed by Hayashi et al. (2011), see their p. 1205, the maximum likelihood estimator of σ^2, in the toy parametric model $\sigma^2(u) \equiv \sigma^2$ constant, is

$$\frac{1}{N_n}\sum_{i=1}^{n} \frac{R_{i,n}^2}{t_{i,n} - t_{i-1,n}}. \tag{4.24}$$

This suggests that in the plug-in estimator (4.22), the instantaneous variance should be divided by the corresponding time increment, that is multiplied by $T_n'(u)$. This is precisely what RV_n^*, or rather its smoothed version $RV_n^S(u)$, does. Hence, we propose the following estimator.

$$\hat{\psi}_g(K) := \int_{u=0}^1 g\left(u, \frac{RV_n^{S\prime}(u)}{T_n'(u)}\right)\mathrm{d}u. \tag{4.25}$$

As already explained, we are not able, under our weak assumptions, to provide an analysis along the lines of Kristensen (2010). He proves, still under regular sampling, that his estimator (4.22), with a convenient bandwidth sequence h_n converging to zero and a convenient bias correction, is consistent and asymptotically normal with a variance that coincides with our efficiency bound $V(g)$. By contrast, our approach amounts to first considering the limiting behavior of the estimator (4.26) for fixed kernel K (Theorem 4.1 below). Then, in the second stage (Proposition 4.1 below) we derive the asymptotic behavior of the bias and variance of our estimator when the kernel itself converges weakly to the point mass at zero, that is

$K(u) \to K^0(u)$ for all $u \neq 0$ with $K^0(u) = 1_{\{[0,\infty)\}}(u)$.

It is important to stress that the validity of these two results is (much) more general than the setup presented in this paper to derive the efficiency bound: It is only Condition 1 that is needed, not the sufficient conditions as provided in Lemma 3.1.

THEOREM 4.1. *Under Condition 1 and Assumption 7–8, we have, under* $\mathbb{P}^{(n)}_{\sigma_0^2}$,

$$\begin{aligned}\sqrt{n}\left(\hat{\psi}_g(K) - \psi_g\left(\sigma_0^2|K, T_n\right)\right) \\ \to_L \sqrt{2}\int_{u=0}^{1}\frac{\partial g}{\partial\sigma^2}\left(u, T'(u)^{-1}\int_{w=0}^{1}K'(u-w)\sigma_0^2(w)dT(w)\right) \\ \times T'(u)^{-1}\int_{w=0}^{1}K'(u-w)\sigma_0^2(w)T'(w)^{1/2}dW(w)du, \qquad \textbf{(4.26)}\end{aligned}$$

where

$$\psi_g\left(\sigma^2|K, T\right) = \int_{u=0}^{1} g\left(u, -T'(u)^{-1}\int_{w=0}^{1}T'(w)\sigma^2(w)dK(u-w)\right)du. \qquad \textbf{(4.27)}$$

Theorem 4.1 gives the limiting behavior for fixed kernel K. However, to get a consistent estimator of the true unknown value of $\psi_g\left(\sigma_0^2\right)$, we need to consider estimators computed with a kernel K close to K^0. Thus, we consider, subsequently, sequences K_n such that $\lim_{n\to\infty} K_n(u) = K^0(u)$, for all $u \neq 0$. With some abuse of terminology, we will say that the sequence K_n converges weakly to the point mass at zero. Our near efficiency result is then formalized in the following proposition.

PROPOSITION 4.1. *The limit* $\psi_g\left(\sigma^2|K, T\right)$ *defined in (4.27) satisfies*

$$\psi_g\left(\sigma_0^2|K, T\right) \to \psi_g\left(\sigma_0^2\right), \qquad \textbf{(4.28)}$$

as K converges weakly to the point mass at zero. Moreover, the limiting variance of (4.26) equals

$$\begin{aligned}2\int_{w=0}^{1}\left[\int_{u=0}^{1}T'(u)^{-1}\frac{\partial g}{\partial\sigma^2}\left(u, T'(u)^{-1}\int_{v=0}^{1}K'(u-v)\sigma_0^2(v)dT(v)\right)dK(u-w)\right]^2 \\ \times\sigma_0^4(w)T'(w)dw \\ \to 2\int_{w=0}^{1}\left(\frac{\partial g}{\partial\sigma^2}\left(w, \sigma_0^2(w)\right)\right)^2\frac{\sigma_0^4(w)}{T'(w)}dw, \qquad \textbf{(4.29)}\end{aligned}$$

as K converges weakly to the point mass at zero.

In view of the above proposition, we call our estimator *nearly* efficient: the limiting distribution for given kernel K can get arbitrarily close to the lower bound. This is, clearly, a weaker result than the estimator provided in Jacod and Rosenbaum (2013). However, we obtain it within a framework allowing for random, though predictable, irregularly spaced observation times. This complicates the analysis significantly as also can be seen from a closer inspection of Theorem 4.1 where the centering in the central limit theorem is at $\psi_g\left(\sigma_0^2|K,T_n\right)$, i.e., using the observation times represented by T_n. It will, in general, not be the case that $\psi_g\left(\sigma_0^2|K,T\right)$ and $\psi_g\left(\sigma_0^2|K,T_n\right)$ differ in the order of $o_P(n^{-1/2})$ only, unless K tends to the point mass at zero at an appropriate rate. We leave such a construction for further research.

5. ON PARAMETRIC INFORMATION ABOUT THE VOLATILITY PROCESS

The lower bound in Theorem 3.3 constitutes the *nonparametric* lower bound for estimating integrated functions of variance. In this section, we focus on the example of power variation, namely $g_p(\sigma^2)=\sigma^{2p}$ and the simplified notation $\psi_{g_p}=\int_0^1\sigma_0^{2p}(u)\mathrm{d}u$. Instead of considering a nonparametric lower bound, practitioners may prefer to specify a parametric functional for $\sigma^2(u)$. In this section, we study the effect of imposing such parametric information about the time-variation of σ^2, that is, when $\sigma^2(u)=\sigma^2(u|\theta)$ for some (sufficiently smooth) parametrization $\theta\mapsto\sigma^2(\cdot|\theta)$. We will show, at odds with what is sometimes considered common wisdom in this setting, that there are gains from such added information and these efficiency gains can be large.

The case of assuming constant volatility, that is $\sigma^2(u|\theta)=\theta>0$ for $u\in[0,1]$, has been studied extensively in Xiu (2010) and Jacod and Rosenbaum (2013). When observations are equally spaced, and still ignoring market microstructure noise, the Gaussian QMLE estimator for θ equals $RV_n(1)$. As a result, its limiting variance equals $2\int_0^1\sigma_0^4(u)\mathrm{d}u=2\theta_0^2$ and thus the parametric and nonparametric lower bounds coincide.

The results above, in particular that the parametric and nonparametric bounds for estimating integrated variance are equal when volatility is constant, should not be interpreted as that no inference gains are possible if parametric information about the form of the volatility process is available. When such information is available, the MLE estimator, and thus the QMLE estimator, may have a variance strictly smaller than $2\int_0^1\sigma_0^4(u)\mathrm{d}u$ under data generating processes for which the true underlying volatility σ_0^2 is not constant.

Consider a general parametric model $\sigma^2(\cdot|\theta)$ which is assumed to be sufficiently smooth such that the induced maximum-likelihood estimator satisfies the standard asymptotic expansion

$$\sqrt{n}\left(\hat{\theta}^{(n)}-\theta_0\right)=\frac{-1}{2\sqrt{n}}\sum_{i=1}^{n}\left[\frac{R_{i,n}^2}{\left(t_{i,n}-t_{i-1,n}\right)^2/\int_{t_{i-1,n}}^{t_{i,n}}\sigma^{-2}(u|\theta_0)\mathrm{d}u}-1\right]$$
$$\times I(\theta_0)^{-1}\frac{\partial}{\partial\theta}\log\sigma^2\left(t_{i-1,n}|\theta_0\right)+o_{\mathbb{P}}(1), \qquad \textbf{(5.1)}$$

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

under $\mathbb{P}^{(n)}_{\sigma^2(\cdot|\theta_0)}$, with Fisher information

$$I(\theta) = \frac{1}{2}\int_{u=0}^{1}\left(\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta)\right)\left(\frac{\partial}{\partial\theta'}\log\sigma^2(u|\theta)\right)\mathrm{d}T(u). \tag{5.2}$$

Consequently, the implied estimator for $\psi_p(\theta) = \int_0^1 \sigma^{2p}(u|\theta)\mathrm{d}u$ has limiting variance $\dot\psi_p(\theta_0)' I(\theta_0)^{-1}\dot\psi_p(\theta_0)$, under $\mathbb{P}^{(n)}_{\sigma^2(\cdot|\theta_0)}$, with $\dot\psi_p(\theta) = p\int_0^1 \sigma^{2p}(u|\theta)\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta)\mathrm{d}u$. We now have the following result.

THEOREM 5.1. *Let Θ be an open subset of $\mathbb{R}^k$. Consider a parametric model $\{\sigma^2(\cdot|\theta):\ \theta\in\Theta\}$ for which $\psi_p(\theta)$ is differentiable and the maximum-likelihood estimator satisfies (5.1)–(5.2). Then we have the following.*

(i) *The limiting variance of the maximum likelihood estimator is at most (3.10).*

(ii) *Equality holds in case $\sigma^{2p}(\cdot|\theta)/T'(u)$ is piecewise constant with at most $k = \dim(\theta)$ different values, i.e., in case we can partition $[0,1]$ into subsets $A_1,\ldots,A_k$ such that*

$$\frac{\sigma^{2p}(u|\theta)}{T'(u)} = \sum_{j=1}^{k}\exp(d(\theta))_j I_{\{u\in A_j\}}, \tag{5.3}$$

for some C_1-diffeomorphism $d:\Theta\to\mathbb{R}^k$.

Proof. In order to prove (i), fix $\theta\in\Theta$ and project $\sigma^{2p}(u|\theta)/T'(u)$ on the space spanned by the elements of $\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta)$, i.e., write $\sigma^{2p}(u|\theta)/T'(u) = \beta'\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta)+\eta(u)$ where $\int_{u=0}^1\eta(u)\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta)\mathrm{d}T(u) = 0$. Plugging this decomposition in the limiting variance of the maximum likelihood estimator yields

$$\begin{aligned}
\dot\psi_p(\theta_0)'I(\theta_0)^{-1}\dot\psi_p(\theta_0) &= 2p^2\int_0^1\frac{\sigma^{2p}(u|\theta)}{T'(u)}\frac{\partial}{\partial\theta'}\log\sigma^2(u|\theta)\mathrm{d}T(u)\\
&\quad\times\left[\int_{u=0}^1\left(\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta)\right)\left(\frac{\partial}{\partial\theta'}\log\sigma^2(u|\theta)\right)\mathrm{d}T(u)\right]^{-1}\\
&\quad\times\int_0^1\frac{\sigma^{2p}(u|\theta)}{T'(u)}\frac{\partial}{\partial\theta}\log\sigma^2(u|\theta)\mathrm{d}T(u)\\
&= 4p^2\beta' I(\theta)\beta \le 4p^2\beta' I(\theta)\beta + 2p^2\int_{u=0}^1\eta^2(u)\mathrm{d}T(u)\\
&= 2p^2\int_{u=0}^1\left(\frac{\sigma^{2p}(u)}{T'(u)}\right)^2\mathrm{d}T(u) = 2p^2\int_{u=0}^1\frac{\sigma^{4p}(u)}{T'(u)}\mathrm{d}u.
\end{aligned}$$

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

Concerning Part (ii), note that equality holds if and only if $\eta = 0$, that is, if and only if $\sigma^{2p}(u|\theta)/T'(u) = \beta' \frac{\partial}{\partial\theta} \log \sigma^2(u|\theta) = \left(p^{-1}\beta\right)' \frac{\partial}{\partial\theta} \log\left(\sigma^{2p}(u|\theta)/T'(u)\right)$ for some $\beta \in \mathbb{R}^k$ (possibly dependent on θ, but not on u). Clearly, (5.3) indeed provides a sufficient condition for equality of both bounds by taking $p^{-1}\beta = \left[\frac{\partial}{\partial\theta'} d(\theta)\right]^{-1} \exp(d(\theta))$, where the exponential is applied componentwise. ■

The above theorem has an interesting implication, even for equally-spaced data, i.e., $T'(u) = 1$ for all $u \in [0,1]$. Then, the above theorem implies that for piecewise constant parametrizations of the volatility function σ (still with at most $k = \dim(\theta)$ different values) the parametric and nonparametric lower bounds coincide. In that case, for $p = 1$, we thus have that realized variance is even a parametrically efficient estimator of integrated variance.

However, for other (not piecewise constant) parametric specifications, this is not true. For example, consider the situation where a researcher would specify $\sigma^2(u|\theta) = \exp(\theta_1 + \theta_2 u)$, $u \in [0,1]$, still with equally spaced data. Figure 1 shows the ratio of the nonparametric and the parametric lower bounds for estimating integrated power variance for $p = 1$, 2, and 3 in the simple case where $\theta_1 = 0$ and $-5 \le \theta_2 \le 5$. Focusing on estimating integrated variance, the inference gain from the information on the parametric form of the volatility function exceeds 15% when $\theta_2 = 5$. Moreover, the gain from information goes up considerably as p increases. For example, the ratio of the nonparametric and the parametric lower bounds is as much as 1.71 for estimating integrated quarticity ($p = 2$) and 2.30

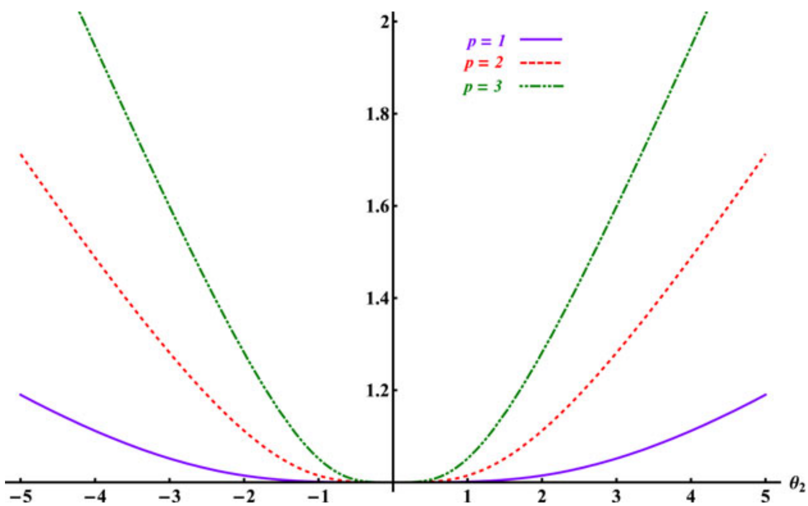

FIGURE 1. The ratio of nonparametric and parametric lower bounds for estimating integrated powers of variance where $\sigma^2(u|\theta) = \exp(\theta_1 + \theta_2 u)$, $u \in [0,1]$, with equally spaced data, i.e., $T'(u) = 1$. This figure considers the powers $p = 1, 2$, and 3 with $\theta_1 = 0$ and $-5 \le \theta_2 \le 5$.

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

for estimating integrated third power variance ($p = 3$) when $\theta_2 = 5$. These values are both theoretically and practically significant.

Clearly, the appropriateness of specifying a parametric model for the time-evolution of intraday volatility is generally an empirical question, with the classical trade-off between possible misspecification and efficiency.

6. CONCLUSIONS

The results in the present paper complement those of Reiß (2011) by focusing on nonparametric lower bounds for integrated powers of volatility, in the absence of market microstructure noise. In line with Clément et al. (2013), we find locally and asymptotically normal limiting experiments at rate $\sqrt{n}$, i.e., the limiting experiments with and without microstructure noise are materially different. Unlike Clément et al. (2013), we focus on the pathwise properties of the volatility process that are needed to obtain this limit. Using these results, we establish the (near) efficiency of the estimator put forward in Mykland and Zhang (2009). Moreover, we demonstrate the efficiency of the Jacod and Rosenbaum (2013) estimator, also at nonconstant volatility within the nonparametric model.

Second, we detail the role of random, though predictable, unequally spaced observation times. We establish precisely how these affect the efficiency bounds. For integrated variance, classical realized variance is efficient, also under these irregular sampling schemes. For higher powers, we provide an estimator that is nearly efficient, i.e., whose limiting variance can get arbitrarily close to the nonparametric lower bound.

Finally, we provide a simple condition under which there are no gains from assuming a parametric specification of the volatility function. This is important as, in applied work, one may prefer the risk of misspecification of a parametric form over the loss of efficiency of fully nonparametric procedures. We show that, with the exception of some very particular volatility functions in relation to the observation scheme, significant efficiency gains are possible in general.

NOTES

1. Regularity is needed to exclude pointwise superefficient estimators. Its definition does not necessarily require a Gaussian limiting distribution, but for our setting this situation suffices.
2. The link between their and our notation is $T = 1$, $r = 2p$, and $H(u) = \int_{v=0}^{u} T'(v)^{-1}\,\mathrm{d}v$.

REFERENCES

Andersen, T.G. & T. Bollerslev (1998) Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.

Andersen, T.G., T. Bollerslev, & S. Lange (1999) Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance* 6, 457–477.

Andersen, T.G., T. Bollerslev, F.X. Diebold, & P. Labys (2001) The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55.

Andersen, T.G., T. Bollerslev, F.X. Diebold, & P. Labys (2003) Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.

Barndorff-Nielsen, O.E. & N. Shephard (2001) Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63, 167–241.

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

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

Barndorff-Nielsen, O.E. & N. Shephard (2003) Realized power variation and stochastic volatility models. *Bernoulli* 9, 243–265.

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

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, & J.A. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press.

Brown, L.D. & M.G. Low (1996) Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics* 6, 2384–2398.

Buchanan, H.E. & T.H. Hildebrandt (1908) Note on the convergence of a sequence of functions of a certain type. *Annals of Mathematics* 9, 123–126.

Clément, E., S. Delattre, & A. Gloter (2013) An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility. *Stochastic Processes and their Applications* 123, 2500–2521.

Comte, F. & E. Renault (1998) Long-memory in continuous time stochastic volatility models. *Mathematical Finance* 8, 291–323. Reprinted in Stochastic Volatility: Selected Readings, N. Shephard (ed.), Oxford University Press, 2005.

Hayashi, T., J. Jacod, & N. Yoshida (2011) Irregular sampling and central limit theorems for power variations: The continuous case. *Annales de l'Institut Henri Poincaré* 47, 1197–1218.

Jacod, J. (2008) Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Processes and their Applications* 118, 517–559.

Jacod, J. & M. Rosenbaum (2013) Quarticity and other functionals of volatility: Efficient estimation. *The Annals of Statistics* 41, 1462–1484.

Jacod, J. & A.N. Shiryaev (2002) *Limit Theorems for Stochastic Processes*, 2nd ed. Springer-Verlag.

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

Le Cam, L.L. (1986) *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag.

Li, Y., P.A. Mykland, E. Renault, L. Zhang, & X. Zheng (2014) Realized volatility when sampling times are possibly endogenous. *Econometric Theory* 30, 580–605.

Li, J., G. Tauchen, & V. Todorov (2013) Volatility occupation times. *The Annals of Statistics* 41, 1865–1891.

Merton, R.C. (1980) On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8, 323–361.

Mykland, P.A. & L. Zhang (2006) ANOVA for diffusions and Itô processes. *The Annals of Statistics* 34, 1931–1963.

Mykland, P.A. & L. Zhang (2009) Inference for continuous semimartingales observed at high frequency. *Econometrica* 77, 1403–1445.

Mykland, P.A. & L. Zhang (2012) The econometrics of high frequency data. In M. Kessler, A. Lindner, & M. Sørensen (eds.), *Statistical Methods for Stochastic Differential Equations*, pp. 109–190. CRC Press, Chapmann and Hall.

Nussbaum, M. (1996) Asymptotic equivalence of density estimation and Gaussian white noise. *The Annals of Statistics* 24, 2399–2430.

Protter, P. (1995) *Stochastic Integration and Differential Equations: A New Approach.* Springer-Verlag.

Reiß, M. (2011) Asymptotic equivalence for inference on the volatility from noisy observations. *The Annals of Statistics* 39, 772–802.

Revuz, D. & M. Yor (1991) *Continuous Martingales and Brownian Motion.* Springer-Verlag.
Taylor, S.J. & X. Xu (1997) The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance* 4, 317–340.
Todorov, V. & G. Tauchen (2012) The realized Laplace transform of volatility. *Econometrica* 80, 1105–1127.
Van der Vaart, A.W. (2000) *Asymptotic Statistics.* Cambridge University Press.
Xiu, D. (2010) Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics* 159, 235–250.

APPENDIX

A. Some lemmas and details on Locally Bounded Variance

We start with a lemma that bounds the variance of the inverse of positive random variables.

LEMMA A.1. *Let* $X \geq c > 0$*, then* $V\{X^{-1}\} \leq c^{-4} Var\{X\}$.

Proof. We have the following inequalities:

$$\begin{aligned} \mathrm{V}\left\{\frac{1}{X}\right\} &\leq \mathrm{E}\left(\frac{1}{X} - \frac{1}{\mathrm{E}\{X\}}\right)^2 \\ &\leq c^{-4}\mathrm{E}\left(X - \mathrm{E}\{X\}\right)^2, \end{aligned}$$

as, for $x, y > c$, $|x^{-1} - y^{-1}| \leq |x-y|/c^2$ since the derivative of $x \mapsto x^{-1}$ is bounded by c^{-2}. ■

The bound (2.4) relates the assumption of locally bounded variance for the sample paths of a stochastic process to its quadratic variation. We know that if $\{X(t) : 0 \leq t \leq 1\}$ is a semimartingale, then its quadratic variation is defined as the probability limit

$$P\lim_{n\to\infty} \sum_{i=1}^{N_n} \left[X(t_{i,n}) - X(t_{i-1,n})\right]^2 = \langle X, X\rangle_1, \quad \textbf{(A.1)}$$

see, e.g., Protter (1995), Theorem II.22. It is a common assumption to consider the volatility process to be a semimartingale or, even smoother, a fractional Brownian motion. Thus, at least for convenient sampling schemes with

$$\left[\max_{t_{i-1,n}\leq u\leq t_{i,n}} X(u) - \min_{t_{i-1,n}\leq u\leq t_{i,n}} X(u)\right]^2 = \left[X(t_{i,n}) - X(t_{i-1,n})\right]^2,$$

the existence of quadratic variation may ensure the convergence in probability of an upper bound of $\sum_{i=1}^{N_n} \mathrm{V}\big(X|_{t_{i-1,n}}^{t_{i,n}}\big)$. However, convergence in probability does not ensure the required almost sure boundedness of $\sum_{i=1}^{N_n} \mathrm{V}\big(X|_{t_{i-1,n}}^{t_{i,n}}\big)$. For instance, it is known that if $X_{t\in[0,1]}$ is a Brownian motion, $\sup \sum_{i=1}^{N_n} \left[X(t_{i,n}) - X(t_{i-1,n})\right]^2$, where the supremum is computed over all possible conformable sampling schemes, is almost surely infinite (see, e.g., the remark below Definition I(2.3) in Revuz and Yor (1991)). One way to circumvent

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

the above issue would be to ensure that the convergence in (A.1) is not only in probability but also almost surely. Almost sure convergence would hold if $\{X(t):0\leq t\leq 1\}$ is a Brownian motion and we consider sequences of sampling schemes that are refining in the sense that

$$\left\{t_{i,n}:\ i=1,\ldots,N_n\right\}\subset\left\{t_{i,n+1}:\ i=1,\ldots,N_{n+1}\right\}, \tag{A.2}$$

see Protter (1995), Theorem I.28. Nevertheless, fortunately, for a process X that is an $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-Brownian motion in the sense of Definition III(2.20) in Revuz and Yor (1991), we can show directly that it is of locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance by applying a convenient law of large numbers to an upper bound of the process $\mathrm{V}\left(X|_{t_{i-1,n}}^{t_{i,n}}\right)$.

PROPOSITION A.1. *If the process $X=\{X(t):0\leq t\leq 1\}$ is a $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-Brownian motion, its sample paths are almost surely of locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance.*

Proof. First, note that we can assume without loss of generality that $N_n=n$. Indeed, if $N_n<n$, we can always complete the sampling scheme as follows

$$i\geq N_n\Rightarrow t_{i+1,n}=t_{i,n}, \tag{A.3}$$

so that, for any f,

$$\sum_{i=1}^{N_n}\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right)=\sum_{i=1}^{n}\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right). \tag{A.4}$$

Moreover, the sampling scheme extended by A.3 still fulfills Assumptions 1 and 2.

We use the following lemma.

LEMMA A.2. *If $X=\{X(t):0\leq t\leq 1\}$ is a $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-Brownian motion and the sampling scheme $\left(t_{i,n}\right)_{0\leq i\leq n}$ satisfies Assumption 2, then we can write*

$$V\left(X|_{t_{i-1,n}}^{t_{i,n}}\right)=\left(t_{i,n}-t_{i-1,n}\right)V\left(Z_n^{(i)}\Big|_0^1\right), \tag{A.5}$$

where $Z^{(i)}$, $i=1,\ldots,n$, are independent standard Brownian motions on $[0,1]$, with $Z_n^{(i)}$ independent of $\mathcal{F}_{t_{i-1,n}}$.

Proof. The only nontrivial case is when $\Delta t_{i,n}=t_{i,n}-t_{i-1,n}>0$. In that case, define $Z_n^{(i)}$ by

$$Z_n^{(i)}(s)=\frac{X(t_{i-1,n}+s\,\Delta t_{i,n})-X(t_{i-1,n})}{\sqrt{\Delta t_{i,n}}},\quad 0\leq s\leq 1.$$

Now, note that since $X=\{X(t):0\leq t\leq 1\}$ is a $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-Brownian motion and the sampling scheme $\left(t_{i,n}\right)_{0\leq i\leq n;n\geq 1}$ is $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-measurable, $Z_n^{(i)},i=1,\ldots,n$, are

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

independent stochastic processes on $[0,1]$. The fact that the $Z_n^{(i)}$'s are standard Brownian motions follows from the self-similarity property of Wiener processes as well as from the fact that $\Delta t_{i,n}$ is known at time $t_{i-1,n}$.

Now, observe that since $\mathrm{V}\big(f|_{t_{i-1,n}}^{t_{i,n}}\big)$ equals $\mathrm{V}(f(U))$ with U uniformly distributed over $[t_{i-1,n}, t_{i,n}]$, we find

$$\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right) = \int_0^1 f(t_{i-1,n} + s\Delta t_{i,n})^2 \mathrm{d}s - \left[\int_0^1 f(t_{i-1,n} + s\Delta t_{i,n})\mathrm{d}s\right]^2.$$

Then,

$$\begin{aligned}
\mathrm{V}\left(X|_{t_{i-1,n}}^{t_{i,n}}\right) &= \int_0^1 \left[X_{t_{i-1,n}} + \sqrt{\Delta t_{i,n}} Z_n^{(i)}(s)\right]^2 \mathrm{d}s - \left[\int_0^1 X_{t_{i-1,n}} + \sqrt{\Delta t_{i,n}} Z_n^{(i)}(s)\mathrm{d}s\right]^2 \\
&= X_{t_{i-1,n}}^2 + \Delta t_{i,n} \int_0^1 Z_n^{(i)}(s)^2 \mathrm{d}s + 2X_{t_{i-1,n}}\sqrt{\Delta t_{i,n}} \int_0^1 Z_n^{(i)}(s)\mathrm{d}s \\
&\quad - X_{t_{i-1,n}}^2 - \Delta t_{i,n}\left[\int_0^1 Z_n^{(i)}(s)\mathrm{d}s\right]^2 - 2X_{t_{i-1,n}}\sqrt{\Delta t_{i,n}} \int_0^1 Z_n^{(i)}(s)\mathrm{d}s \\
&= \Delta t_{i,n}\left\{\int_0^1 Z_n^{(i)}(s)^2 \mathrm{d}s - \left[\int_0^1 Z_n^{(i)}(s)\mathrm{d}s\right]^2\right\} \\
&= \Delta t_{i,n} \mathrm{V}\left(Z_n^{(i)}\Big|_0^1\right).
\end{aligned}$$

■

We continue the proof of Proposition A.1. It is sufficient to prove that

$$\sum_{i=1}^n \Delta t_{i,n} \mathrm{V}\left(Z_n^{(i)}\Big|_0^1\right)$$

is almost surely bounded. From Cauchy–Schwarz, this follows as both

$$n\sum_{i=1}^n \left(\Delta t_{i,n}\right)^2 \text{ and } n^{-1}\sum_{i=1}^n \mathrm{V}\left(Z_n^{(i)}\Big|_0^1\right)^2$$

are bounded almost surely by Assumption 1 and a standard strong law of large numbers. ■

Reiß (2011) imposes smoothness conditions on the sample paths of volatility in terms of Hölder balls. Note that if for some $\alpha > 0$, $\sup_{u \neq v} |f(u) - f(v)| / |u - v|^\alpha \leq R$, we have from (2.3)

$$\mathrm{V}\left(f|_{t_{i-1,n}}^{t_{i,n}}\right) \leq \frac{1}{4} \sup_{t_{i-1,n} \leq u \neq v \leq t_{i,n}} [f(u) - f(v)]^2 \leq \frac{1}{4} R^2 \left(\Delta t_{i,n}\right)^{2\alpha}.$$

As a result, in view of Assumption 1, f is of locally bounded variance for $\alpha \geq 1/2$ (apply, e.g., the Cauchy–Schwarz inequality and use $\sum (\Delta t_{i,n})^2 = O_P(n^{-1})$). Reiß (2011) needs

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

$\alpha \geq (1+\sqrt{5})/4 \approx 0.81$ in his main Theorem 6.2. This, again, shows that the analysis with and without (microstructure) noise, leads to materially different limiting experiments.

Given these considerations, we consider throughout that the assumption of a process being almost surely of locally $\{\mathcal{F}_t\}_{0\leq t\leq 1}$-bounded variance is not overly restrictive. However, besides the cases discussed above, whether paths of general semimartingales are of locally bounded variance is still an open problem.

We also provide an auxiliary lemma to bound certain expressions that can be interpreted as expectations and arise in the various proofs.

LEMMA A.3. *Let X and Y be two strictly positive random variables bounded from above by M. Then, we have the following bounds*

$$0 \leq E\{X\} - \left(E\left\{X^{-1}\right\}\right)^{-1} \leq M\sqrt{V\{X\}V\left\{X^{-1}\right\}}, \tag{A.6}$$

$$0 \leq E\left\{X^2\right\} - \left(E\left\{X^{-1}\right\}\right)^{-2} \leq V\{X\} + 2M^2\sqrt{V\{X\}V\left\{X^{-1}\right\}}, \tag{A.7}$$

$$\left|(E\{XY\})^2 - (E\{X\})^2 E\left\{Y^2\right\}\right| \leq 4M^2\sqrt{V\{X\}V\{Y\}} + M^2 V\{Y\}. \tag{A.8}$$

Proof. The left-hand side inequality of (A.6) is well known. For the right-hand side observe

$$0 \geq 1 - \mathrm{E}\{X\}\mathrm{E}\left\{X^{-1}\right\} = \mathrm{Cov}\left\{X, X^{-1}\right\}.$$

As the harmonic mean of X is bounded by M, (A.6) follows from Cauchy–Schwarz.

To prove (A.7), write

$$\mathrm{E}\left\{X^2\right\} - \left(\mathrm{E}\left\{X^{-1}\right\}\right)^{-2} = \mathrm{V}\{X\} + \left(\mathrm{E}\{X\} + \left(\mathrm{E}\left\{X^{-1}\right\}\right)^{-1}\right)\left(\mathrm{E}\{X\} - \left(\mathrm{E}\left\{X^{-1}\right\}\right)^{-1}\right).$$

Finally, concerning (A.8), observe

$$\begin{aligned}
&\left|(\mathrm{E}\{XY\})^2 - (\mathrm{E}\{X\})^2\mathrm{E}\left\{Y^2\right\}\right| \\
&= \left|(\mathrm{Cov}\{X,Y\})^2 + 2\mathrm{E}\{X\}\mathrm{E}\{Y\}\mathrm{Cov}\{X,Y\} - (\mathrm{E}\{X\})^2\mathrm{V}\{Y\}\right| \\
&\leq |\mathrm{Cov}\{X,Y\}[\mathrm{Cov}\{X,Y\} + 2\mathrm{E}\{X\}\mathrm{E}\{Y\}]| + M^2\mathrm{V}\{Y\} \\
&\leq 2M^2(\mathrm{V}\{X\}\mathrm{V}\{Y\})^{1/2} + M^2\mathrm{V}\{Y\}.
\end{aligned}$$

■

B. Proofs

B.1. An asymptotically equivalent model. The specification (2.7) is classical in applications, but inconvenient for the likelihood calculations underlying the proof of the LAN property. We, therefore, introduce here an alternative volatility specification, using harmonic means instead of arithmetic means, and show that both define asymptotically the same statistical experiments.

Thus, consider the alternative specification

$$\frac{1}{\sigma^2(t_{i-1,n}, t_{i,n})} = \frac{1}{t_{i,n} - t_{i-1,n}} \int_{t_{i-1,n}}^{t_{i,n}} \sigma^{-2}(u)\mathrm{d}u. \tag{B.1}$$

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

In this case, we write

$$\sigma^2(t_{i-1,n},t_{i,n}) = H_\sigma(t_{i-1,n},t_{i,n}) := \left[\frac{1}{t_{i,n}-t_{i-1,n}}\int_{t_{i-1,n}}^{t_{i,n}}\sigma^{-2}(u)\mathrm{d}u\right]^{-1}.$$

This specification differs from the arithmetic mean specification considered in (2.7)

$$\sigma^2(t_{i-1,n},t_{i,n}) = \frac{1}{t_{i,n}-t_{i-1,n}}\int_{t_{i-1,n}}^{t_{i,n}}\sigma^{2}(u)\mathrm{d}u = A_\sigma(t_{i-1,n},t_{i,n}).$$

The same functional parameter σ^2 will, in general, not produce the same value for the arithmetic and harmonic means. It is well known that the arithmetic mean $A_\sigma(t_{i-1,n},t_{i,n})$ and the harmonic mean $H_\sigma(t_{i-1,n},t_{i,n})$ coincide only if the function σ^2 is constant over the interval $(t_{i-1,n},t_{i,n}]$. However, as the length of each observation interval $(t_{i-1,n},t_{i,n}]$ becomes asymptotically negligible, one may hope that a model defined in terms of the harmonic means $H_\sigma(t_{i-1,n},t_{i,n})$ is equivalent to the financially more meaningful model defined in terms of the arithmetic means $A_\sigma(t_{i-1,n},t_{i,n})$. Of course, this takes some smoothness on the volatility function σ. We formalize this in Proposition B.1 below using precisely the smoothness condition of locally bounded variance for σ^2.

Recall that cadlag functions on a compact set like $[0,1]$ are bounded, so that, under Assumption 5, both σ^2 and σ^{-2} will be bounded away from zero and infinity. Hence, for all $\sigma^2 \in \Xi$,

$$0 < m_\sigma := \inf_{0\le u\le 1}\left\{\sigma^2(u),\sigma^{-2}(u)\right\} \le M_\sigma := \sup_{0\le u\le 1}\left\{\sigma^2(u),\sigma^{-2}(u)\right\} < +\infty.$$

This allows us to uniformly control the difference between arithmetic and harmonic means of σ^2. The proof of the following result is immediate from Lemma A.3.

LEMMA B.1. *For all $\sigma^2 \in \Xi$, we have*

$$0 \le A_\sigma(t_{i-1,n},t_{i,n}) - H_\sigma(t_{i-1,n},t_{i,n}) \le M_\sigma\sqrt{V\left(\sigma^2\big|_{t_{i-1,n}}^{t_{i,n}}\right)}\sqrt{V\left(\sigma^{-2}\big|_{t_{i-1,n}}^{t_{i,n}}\right)}.$$

Lemma B.1 allows us to study the asymptotic equivalence of the two following statistical experiments in the sense of Le Cam (1986).

Experiment 1. *Suppose that observed returns satisfy Assumption 3 with, for $i = 1,\dots,N_n$,*

$$\sigma^2(t_{i-1,n},t_{i,n}) = A_\sigma(t_{i-1,n},t_{i,n}), \tag{B.2}$$

for some $\sigma^2 \in \Xi$.

Experiment 2. *Suppose that observed returns satisfy Assumption 3 with, for $i = 1,\dots,N_n$,*

$$\sigma^2(t_{i-1,n},t_{i,n}) = H_\sigma(t_{i-1,n},t_{i,n}), \tag{B.3}$$

for some $\sigma^2 \in \Xi$.

Observe that the Experiments 1 and 2 have the same parameter space Ξ. We establish asymptotic equivalence by showing below that Le Cam's deficiency pseudodistance between the Gaussian experiments converges to zero (almost surely in the sampling times), when $n \to \infty$. Since we deal with Gaussian distributions, it is convenient to use the known fact (see, e.g., Nussbaum (1996), Formula (12)) that Le Cam's squared pseudodistance is bounded by four times the value of the squared Hellinger distance between the corresponding densities. In order to show that this Hellinger distance converges to zero, we use Assumption 4 to conclude (see, e.g., Lemma 2.4. in Nussbaum (1996)) that the squared Hellinger distance between Experiments 1 and 2 is bounded by

$$2\sum_{i=1}^{N_n} D^2\left(P^A_{\sigma,i,n}; P^H_{\sigma,i,n}\right),$$

where $P^A_{\sigma,i,n}$ and $P^H_{\sigma,i,n}$ are, respectively, the normal probability distribution with variance $A_\sigma(t_{i-1,n}, t_{i,n})\Delta t_{i,n}$ and that with variance $H_\sigma(t_{i-1,n}, t_{i,n})\Delta t_{i,n}$ and the means as in Assumption 3. Here, $D^2(P^A_{\sigma,i,n}; P^H_{\sigma,i,n})$ stands for the squared Hellinger distance between $P^A_{\sigma,i,n}$ and $P^H_{\sigma,i,n}$, conditionally on the observation times.

It turns out that the squared Hellinger distance between two normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ is easily computed (see, e.g., Brown and Low (1996)) as

$$D^2\left(N\left(\mu_1,\sigma_1^2\right), N\left(\mu_2,\sigma_2^2\right)\right) = 2\left\{1 - \left[\frac{2\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2}\right]^{1/2} \exp\left[-\frac{(\mu_1-\mu_2)^2}{4\left(\sigma_1^2+\sigma_2^2\right)}\right]\right\}. \quad \textbf{(B.4)}$$

We now have the following result.

PROPOSITION B.1. *Under Assumptions 1–5, the statistical Experiments 1 and 2 are asymptotically equivalent in the sense that, almost surely,*

$$\lim_{n\to\infty}\sum_{i=1}^{N_n} D^2\left(P^A_{\sigma,i,n}; P^H_{\sigma,i,n}\right) = 0.$$

Proof. From (B.4) we have

$$D^2\left(P^A_{\sigma,i,n}; P^H_{\sigma,i,n}\right) = 2\left\{1 - \left[\frac{2\sigma_{i1}\sigma_{i2}}{\sigma_{i1}^2+\sigma_{i2}^2}\right]^{1/2} \exp\left[-\frac{(\mu_{i1}-\mu_{i2})^2}{4\left(\sigma_{i1}^2+\sigma_{i2}^2\right)}\right]\right\},$$

where

$$\sigma_{i1}^2 = A_\sigma(t_{i-1,n}, t_{i,n}),$$
$$\sigma_{i2}^2 = H_\sigma(t_{i-1,n}, t_{i,n}),$$
$$(\mu_{i1}-\mu_{i2})^2 = \gamma^2(t_{i-1,n}, t_{i,n})\left(\sigma_{i1}^2 - \sigma_{i2}^2\right)^2 \Delta t_{i,n}.$$

We rewrite

$$\frac{2\sigma_{i1}\sigma_{i2}}{\sigma_{i1}^2+\sigma_{i2}^2} = 1 - \frac{(\sigma_{i1}-\sigma_{i2})^2}{\sigma_{i1}^2+\sigma_{i2}^2} = 1 - \frac{\left(\sigma_{i1}^2-\sigma_{i2}^2\right)^2}{\left(\sigma_{i1}^2+\sigma_{i2}^2\right)(\sigma_{i1}+\sigma_{i2})^2},$$

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

and define

$$\varepsilon_{i,n} := \frac{\left(\sigma_{i1}^2 - \sigma_{i2}^2\right)^2}{\sigma_{i1}^2 + \sigma_{i2}^2},$$

so that

$$D^2\left(P_{\sigma,i,n}^A; P_{\sigma,i,n}^H\right) = 2\left\{1 - \left[1 - \frac{\varepsilon_{i,n}}{(\sigma_{i1}+\sigma_{i2})^2}\right]^{1/2} \exp\left[-\frac{\gamma^2(t_{i-1,n},t_{i,n})\Delta t_{i,n}}{4}\varepsilon_{i,n}\right]\right\}.$$

Observe that the derivative of the map $\varepsilon_{i,n} \mapsto D^2\left(P_{\sigma,i,n}^A; P_{\sigma,i,n}^H\right)$, for given σ_{i1}, σ_{i2}, and γ, is

$$\begin{aligned}
&\left|\frac{\partial}{\partial \varepsilon_{i,n}} D^2\left(P_{\sigma,i,n}^A; P_{\sigma,i,n}^H\right)\right| \\
&\quad = \left|\left[1 - \frac{\varepsilon_{i,n}}{(\sigma_{i1}+\sigma_{i2})^2}\right]^{-1/2} (\sigma_{i1}+\sigma_{i2})^{-2} + 2\left[1 - \frac{\varepsilon_{i,n}}{(\sigma_{i1}+\sigma_{i2})^2}\right]^{1/2} \frac{\gamma^2(t_{i-1,n},t_{i,n})\Delta t_{i,n}}{4}\right| \\
&\qquad \times \exp\left[-\frac{\gamma^2(t_{i-1,n},t_{i,n})\Delta t_{i,n}}{4}\varepsilon_{i,n}\right] \\
&\quad \le \left[1 - \frac{\varepsilon_{i,n}}{(\sigma_{i1}+\sigma_{i2})^2}\right]^{-1} (\sigma_{i1}+\sigma_{i2})^{-2} + \frac{\gamma^2(t_{i-1,n},t_{i,n})\Delta t_{i,n}}{2} \\
&\quad = \frac{1}{(\sigma_{i1}+\sigma_{i2})^2 - \varepsilon_{i,n}} + \frac{\gamma^2(t_{i-1,n},t_{i,n})\Delta t_{i,n}}{2},
\end{aligned}$$

and thus is bounded by, say, K as $(\sigma_{i1}+\sigma_{i2})^2 - \varepsilon_{i,n} = 2\sigma_{i1}\sigma_{i2} + 4\sigma_{i1}^2\sigma_{i2}^2/\left(\sigma_{i1}^2+\sigma_{i2}^2\right) \ge 2m_\sigma$ and γ is bounded by Assumption 3.

Thus, by the mean-value theorem,

$$\begin{aligned}
\sum_{i=1}^{N_n} D^2\left(P_{\sigma,i,n}^A; P_{\sigma,i,n}^H\right) &\le K \sum_{i=1}^{N_n} \varepsilon_{i,n} \\
&\le \frac{1}{2m_\sigma} \max_{1\le i\le N_n} \left|\sigma_{i1}^2 - \sigma_{i2}^2\right| \sum_{i=1}^{N_n} \left|\sigma_{i1}^2 - \sigma_{i2}^2\right|.
\end{aligned}$$

But we also know

$$\left|\sigma_{i1}^2 - \sigma_{i2}^2\right| = \left|A_\sigma(t_{i-1,n},t_{i,n}) - H_\sigma(t_{i-1,n},t_{i,n})\right| \le M_\sigma \sqrt{V\left(\sigma^2\big|_{t_{i-1,n}}^{t_{i,n}}\right)}\sqrt{V\left(\sigma^{-2}\big|_{t_{i-1,n}}^{t_{i,n}}\right)},$$

so that by the assumption of locally bounded variance we obtain

$$\sum_{i=1}^{N_n} D^2\left(P_{\sigma,i,n}^A; P_{\sigma,i,n}^H\right) \le K \sum_{i=1}^{N_n} \varepsilon_{i,n} \qquad \to 0, \text{ a.s.} \qquad \blacksquare$$

Proposition B.1 establishes the asymptotic equivalence of using the arithmetic or harmonic average of spot variance in the description of our experiment.

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

B.2. Proof of Lemma 3.1. First, we verify that the centering proposed in (3.3) and (3.4) is indeed appropriate. Observe, using (A.6),

$$\sup_{0\le u\le 1}\sqrt{n}\left|\sum_{t_{i,n}\le u}\mathrm{E}_{\sigma_0^2}\left\{R_{i,n}^2\middle|\mathcal{F}_{t_{i-1,n}}\right\}-\int_{v=0}^{u}\sigma_0^2(v)\mathrm{d}v\right|$$

$$\le\sqrt{n}\sum_{i=1}^{n}\left[\mu\left(t_{i-1,n},t_{i,n}\right)+\gamma\left(t_{i-1,n},t_{i,n}\right)\frac{1}{\left(t_{i,n}-t_{i-1,n}\right)^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(v)\mathrm{d}v}\right]^2\left(t_{i,n}-t_{i-1,n}\right)^2$$

$$+\left(t_{i,n}-t_{i-1,n}\right)\left[\left(t_{i,n}-t_{i-1,n}\right)^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^2(v)\mathrm{d}v-\frac{1}{\left(t_{i,n}-t_{i-1,n}\right)^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(v)\mathrm{d}v}\right].$$

We show that both terms in the summation above converges to zero. The second term can be bounded, using (A.6) once more, as

$$\sqrt{n}\max_{i=1,\ldots,n}\left|t_{i,n}-t_{i-1,n}\right|\left\|\sigma_0^2\right\|\sum_{i=1}^{n}\left(\mathrm{V}\left(\sigma_0^2\Big|_{t_{i-1,n}}^{t_{i,n}}\right)\mathrm{V}\left(\sigma_0^{-2}\Big|_{t_{i-1,n}}^{t_{i,n}}\right)\right)^{1/2},$$

which converges to zero in view of Assumption 1, the Cauchy–Schwarz inequality, and Assumption 5. Concerning the first term, recall that μ, γ, and σ_0^2 are all bounded. Hence it suffices to study

$$\sqrt{n}\sum_{i=1}^{n}\left(t_{i,n}-t_{i-1,n}\right)^2, \tag{B.5}$$

which also converges to zero in the view of Assumption 1.

With respect to RV^*, we note that without loss of generality we may study the slightly redefined version

$$RV_n^*(u)=\sum_{i=1}^{n}\frac{R_{i,n}^2}{n\left(t_{i,n}-t_{i-1,n}\right)}I\left\{t_{i,n}\le u\right\}, \tag{B.6}$$

which differs at most $R_{i,n}^2/\left[n\left(t_{i,n}-t_{i-1,n}\right)\right]=O_{\mathbb{P}}(1/n)$ from (3.2). For this version we find, again using Lemma A.3,

$$\sup_{0\le u\le 1}\sqrt{n}\left|\sum_{t_{i,n}\le u}\frac{1}{n\left(t_{i,n}-t_{i-1,n}\right)}\mathrm{E}_{\sigma_0^2}\left\{R_{i,n}^2\middle|\mathcal{F}_{t_{i-1,n}}\right\}-\int_{v=0}^{u}\sigma_0^2(v)\mathrm{d}T_n(v)\right|$$

$$\le\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\mu\left(t_{i-1,n},t_{i,n}\right)+\gamma\left(t_{i-1,n},t_{i,n}\right)\frac{1}{\left(t_{i,n}-t_{i-1,n}\right)^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(v)\mathrm{d}v}\right]^2\left(t_{i,n}-t_{i-1,n}\right)$$

$$+\left[\left(t_{i,n}-t_{i-1,n}\right)^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^2(v)\mathrm{d}v-\frac{1}{\left(t_{i,n}-t_{i-1,n}\right)^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(v)\mathrm{d}v}\right].$$

Again, we show that both terms converge to zero. The second term can be bounded, in view of (A.6), by

$$\frac{1}{\sqrt{n}} \left\| \sigma_0^2 \right\| \sum_{i=1}^{n} \left(\mathrm{V}\left(\sigma_0^2 \Big|_{t_{i-1,n}}^{t_{i,n}} \right) \mathrm{V}\left(\sigma_0^{-2} \Big|_{t_{i-1,n}}^{t_{i,n}} \right) \right)^{1/2}, \tag{B.7}$$

which converges to zero in view of Assumption 5. For the first term, it is sufficient to consider

$$\sqrt{n} \max_{i=1,\ldots,n} \left| t_{i,n} - t_{i-1,n} \right|, \tag{B.8}$$

which converges to zero in view of Assumption 1.

Using the above results, we can now prove the claim by an application of Theorem VIII.3.33 in Jacod and Shiryaev (2002) to the exactly centered versions of the bivariate process (RV_n, RV_n^*). In the notation of Jacod and Shiryaev (2002), we have $k = i$, $t = u$, and $\sigma_t^n = nT_n(u)$ [recall $t_{i,n} = T_n^{-1}(i/n)$] and we consider the martingale difference sequence

$$\begin{aligned} U_i^n &= \sqrt{n} \left[\frac{1}{\left[n\left(t_{i,n} - t_{i-1,n} \right) \right]^{-1}} \right] \left(R_{i,n}^2 - \mathrm{E}_{\sigma_0^2} \left\{ R_{i,n}^2 \Big| \mathcal{F}_{t_{i-1,n}} \right\} \right) \\ &= \sqrt{n} \left[\frac{1}{\left[n\left(t_{i,n} - t_{i-1,n} \right) \right]^{-1}} \right] \frac{\left(t_{i,n} - t_{i-1,n} \right)^2}{\int_{v=t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(v) \mathrm{d}v} \left(Z_{i,n}^2 - 1 \right), \end{aligned} \tag{B.9}$$

with $Z_{i,n}$ i.i.d. standard normal. For the required Lindeberg condition, note that for $0 < M \le \inf_{0 \le u \le 1} \sigma_0^{-2}(u)$ we have

$$\left| U_i^n \right| \le \left| \left[\begin{matrix} \sqrt{n}\left(t_{i,n} - t_{i-1,n} \right) \\ 1/\sqrt{n} \end{matrix} \right] \right| \frac{\left| Z_{i,n}^2 - 1 \right|}{M} = \sqrt{n\left(t_{i,n} - t_{i-1,n} \right)^2 + 1/n} \frac{\left| Z_{i,n}^2 - 1 \right|}{M}.$$

Thus, for all $\varepsilon > 0$,

$$\begin{aligned} &\sum_{t_{i,n} \le u} \mathrm{E}_{\sigma_0^2} \left\{ \left| U_i^n \right|^2 I_{\{|U_i^n| > \varepsilon\}} \Big| \mathcal{F}_{t_{i-1,n}} \right\} \\ &\quad \le \sum_{i=1}^{n} \frac{n\left(t_{i,n} - t_{i-1,n} \right)^2 + 1/n}{M^2} \mathrm{E} \left\{ \left(Z_{i,n}^2 - 1 \right)^2 I_{\left\{ \sqrt{n\left(t_{i,n} - t_{i-1,n} \right)^2 + 1/n} \left| Z_{i,n}^2 - 1 \right| > M\varepsilon \right\}} \right\} \\ &\quad \le \max_{i=1,\ldots,n} \mathrm{E} \left\{ \left(Z_{i,n}^2 - 1 \right)^2 I_{\left\{ \sqrt{n\left(t_{i,n} - t_{i-1,n} \right)^2 + 1/n} \left| Z_{i,n}^2 - 1 \right| > M\varepsilon \right\}} \right\} \sum_{i=1}^{n} \frac{n\left(t_{i,n} - t_{i-1,n} \right)^2 + 1/n}{M^2}, \end{aligned} \tag{B.10}$$

which converges to zero in view of Assumption 6 and the fact that the mapping $x \mapsto \mathrm{E}\left\{ (Z_{i,n}^2 - 1)^2 I_{\{\sqrt{x} |Z_{i,n}^2 - 1| > M\varepsilon\}} \right\}$ is continuous at $x = 0$.

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

Finally, the quadratic variation of the limiting Gaussian process follows from the law of large numbers by

$$\sum_{t_{i,n}\le u} \mathrm{E}_{\sigma_0^2}\left\{U_i^n U_i^{n\prime} \middle| \mathcal{F}_{t_{i-1,n}}\right\} \tag{B.11}$$

$$= 2\sum_{t_{i,n}\le u} n \begin{bmatrix} 1 & [n(t_{i,n}-t_{i-1,n})]^{-1} \\ [n(t_{i,n}-t_{i-1,n})]^{-1} & [n(t_{i,n}-t_{i-1,n})]^{-2} \end{bmatrix} \left(\frac{(t_{i,n}-t_{i-1,n})^2}{\int_{v=t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(v)\mathrm{d}v}\right)^2$$

$$= 2\sum_{t_{i,n}\le u} \begin{bmatrix} n(t_{i,n}-t_{i-1,n}) & 1 \\ 1 & [n(t_{i,n}-t_{i-1,n})]^{-1} \end{bmatrix} \int_{v=t_{i-1,n}}^{t_{i,n}} \sigma_0^4(v)\mathrm{d}v$$

$$+2\sum_{t_{i,n}\le u} \begin{bmatrix} n(t_{i,n}-t_{i-1,n}) & 1 \\ 1 & [n(t_{i,n}-t_{i-1,n})]^{-1} \end{bmatrix} (t_{i,n}-t_{i-1,n})$$

$$\times\left(\left[\frac{1}{t_{i,n}-t_{i-1,n}}\int_{v=t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(v)\mathrm{d}v\right]^{-2} - \frac{1}{t_{i,n}-t_{i-1,n}}\int_{v=t_{i-1,n}}^{t_{i,n}} \sigma_0^4(v)\mathrm{d}v\right)$$

$$= 2\sum_{t_{i,n}\le u} \int_{v=t_{i-1,n}}^{t_{i,n}} \sigma_0^4(v)\mathrm{d}\begin{bmatrix} T_n^{-1}(v) & v \\ v & T_n(v) \end{bmatrix} + o_P(1)$$

$$\to 2\int_{v=0}^{u} \sigma_0^4(v)\mathrm{d}\begin{bmatrix} T^{-1}(v) & v \\ v & T(v) \end{bmatrix},$$

where the $o_P(1)$-term follows from (A.7) combined with Assumption 5 and where the final convergence follows from the weak convergence condition in Assumption 6.

B.3. Proof of Theorem 3.2. The proof consists of showing that the likelihood ratio satisfies the appropriate quadratic expansion. First, observe

$$\frac{t_{i,n}-t_{i-1,n}}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})} = \frac{t_{i,n}-t_{i-1,n}}{\sigma_0^2(t_{i-1,n},t_{i,n})} + \frac{\alpha}{\sqrt{n}}\int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(u)h(u)\mathrm{d}u. \tag{B.12}$$

Now, the log-likelihood ratio of $\mathbb{P}_\alpha^{(n)}$ with respect to $\mathbb{P}_0^{(n)}$ is, in view of Assumption 3 and 4, given by

$$\log\frac{\mathrm{d}\mathbb{P}_\alpha^{(n)}}{\mathrm{d}\mathbb{P}_0^{(n)}} = \frac{1}{2}\sum_{i=1}^{n}\log\frac{\sigma_0^2(t_{i-1,n},t_{i,n})}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})}$$

$$+\frac{1}{2}\sum_{i=1}^{n}\frac{\left[R_{i,n}-\left[\mu(t_{i-1,n},t_{i,n})+\gamma(t_{i-1,n},t_{i,n})\sigma_0^2(t_{i-1,n},t_{i,n})\right](t_{i,n}-t_{i-1,n})\right]^2}{\sigma_0^2(t_{i-1,n},t_{i,n})(t_{i,n}-t_{i-1,n})}$$

$$-\frac{1}{2}\sum_{i=1}^{n}\frac{\left[R_{i,n}-\left[\mu(t_{i-1,n},t_{i,n})+\gamma(t_{i-1,n},t_{i,n})\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})\right](t_{i,n}-t_{i-1,n})\right]^2}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})(t_{i,n}-t_{i-1,n})}$$

$$= \frac{1}{2}\sum_{i=1}^{n}\log\frac{\sigma_0^2(t_{i-1,n},t_{i,n})}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})}$$

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

$$
\begin{aligned}
&-\frac{1}{2}\sum_{i=1}^{n}\frac{R_{i,n}^2}{t_{i,n}-t_{i-1,n}}\left[\frac{1}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})}-\frac{1}{\sigma_0^2(t_{i-1,n},t_{i,n})}\right]\\
&+\sum_{i=1}^{n}R_{i,n}\mu(t_{i-1,n},t_{i,n})\left[\frac{1}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})}-\frac{1}{\sigma_0^2(t_{i-1,n},t_{i,n})}\right]\\
&+\frac{1}{2}\sum_{i=1}^{n}\mu(t_{i-1,n},t_{i,n})^2\left(t_{i,n}-t_{i-1,n}\right)\left[\frac{1}{\sigma_0^2(t_{i-1,n},t_{i,n})}-\frac{1}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})}\right]\\
&+\frac{1}{2}\sum_{i=1}^{n}\gamma\left(t_{i-1,n},t_{i,n}\right)^2\left(t_{i,n}-t_{i-1,n}\right)\left[\sigma_0^2(t_{i-1,n},t_{i,n})-\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})\right]\\
=&\,\frac{1}{2}\sum_{i=1}^{n}\log\left[1+\frac{\alpha}{\sqrt{n}}\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u}\right]\\
&-\frac{\alpha}{2\sqrt{n}}\sum_{i=1}^{n}\frac{R_{i,n}^2}{(t_{i,n}-t_{i-1,n})^2}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u+u_n\\
=&-\frac{\alpha}{2}\int_0^1\sigma_0^{-2}(u)h(u)\mathrm{d}\sqrt{n}\left[RV_n^*(u)-\int_{v=0}^{u}\sigma_0^2(v)\mathrm{d}T_n(v)\right]\\
&-\frac{\alpha^2}{4}\int_{u=0}^{1}h^2(u)\mathrm{d}T(u)+u_n+r_n,
\end{aligned}
\qquad \textbf{(B.13)}
$$

with

$$
\begin{aligned}
u_n=&\sum_{i=1}^{n}R_{i,n}\mu(t_{i-1,n},t_{i,n})\left[\frac{1}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})}-\frac{1}{\sigma_0^2(t_{i-1,n},t_{i,n})}\right]\\
&+\frac{1}{2}\sum_{i=1}^{n}\mu(t_{i-1,n},t_{i,n})^2\left(t_{i,n}-t_{i-1,n}\right)\left[\frac{1}{\sigma_0^2(t_{i-1,n},t_{i,n})}-\frac{1}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})}\right]\\
&+\frac{1}{2}\sum_{i=1}^{n}\gamma\left(t_{i-1,n},t_{i,n}\right)^2\left(t_{i,n}-t_{i-1,n}\right)\left[\sigma_0^2(t_{i-1,n},t_{i,n})-\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n},t_{i,n})\right],
\end{aligned}
$$

and

$$
\begin{aligned}
r_n=&\,\frac{1}{2}\sum_{i=1}^{n}\log\left[1+\frac{\alpha}{\sqrt{n}}\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u}\right]\\
&-\frac{\alpha}{2}\sqrt{n}\int_0^1 h(u)\mathrm{d}T_n(u)+\frac{\alpha^2}{4}\int_{u=0}^{1}h^2(u)\mathrm{d}T(u).
\end{aligned}
\qquad \textbf{(B.14)}
$$

We need to show that each term in u_n and r_n converges to zero.

The first term of u_n can be bounded, using the Cauchy–Schwarz inequality, by

$$
\left(\frac{1}{\sqrt{n}}\left[\sqrt{n}\sum_{i=1}^{n}\frac{R_{i,n}^2}{n\left(t_{i,n}-t_{i-1,n}\right)}\right]\right)^{1/2}
$$

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

$$\times \left(\sum_{i=1}^{n} \mu(t_{i-1,n}, t_{i,n})^2 \alpha^2 \left[\frac{\int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(u) h(u) \mathrm{d}u}{(t_{i,n} - t_{i-1,n})} \right]^2 (t_{i,n} - t_{i-1,n}) \right)^{1/2},$$

which converges to zero as $\sqrt{n}\sum_{i=1}^{n} R_{i,n}^2/\left(n\left(t_{i,n}-t_{i-1,n}\right)\right)$ is $O_P(1)$ by Lemma 3.1 and μ and σ_0^{-2} are bounded.

The (absolute value of the) second term of u_n can be bounded by

$$\frac{1}{2\sqrt{n}} \sum_{i=1}^{n} \left| \mu(t_{i-1,n}, t_{i,n})^2 \alpha \int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(u) h(u) \mathrm{d}u \right|$$

which converges to zero as μ is bounded and $\sum_{i=1}^{n} \int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(u)h(u)\mathrm{d}u = \int_0^1 \sigma_0^{-2}(u)h(u)\mathrm{d}u$.

The third term of u_n can be bounded using $|x^{-1} - y^{-1}| \le c^{-2}|x-y|$ for $x \ge c$, $y \ge c$, and $c > 0$. Indeed, for n sufficiently large, both $\sigma_0^2(t_{i-1,n}, t_{i,n})$ and $\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n}, t_{i,n})$ are larger than $c = \frac{1}{2}\min_{0\le u\le 1}\sigma_0^2(u)$. Thus, we can bound the third term of u_n by

$$\frac{1}{2} c^{-2} \sum_{i=1}^{n} \gamma\left(t_{i-1,n}, t_{i,n}\right)^2 \left(t_{i,n} - t_{i-1,n}\right) \left| \frac{1}{\sigma_0^2(t_{i-1,n}, t_{i,n})} - \frac{1}{\sigma_{\alpha/\sqrt{n}}^2(t_{i-1,n}, t_{i,n})} \right|,$$

and we may proceed as for the second term of u_n.

Now, let's concentrate on r_n. Using $\|h\| \le 1$ and $\left|\log(1+x) - x + x^2/2\right| < |x|^3$ for $|x| < 1/2$, we have, for $n \ge 4$,

$$\begin{aligned} |r_n| \le{}& \frac{\alpha}{2\sqrt{n}} \sum_{i=1}^{n} \left| \frac{\int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(u) h(u) \mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(u) \mathrm{d}u} - \frac{1}{t_{i,n} - t_{i-1,n}} \int_{t_{i-1,n}}^{t_{i,n}} h(u) \mathrm{d}u \right| \\ &+ \frac{\alpha^2}{4n} \sum_{i=1}^{n} \left| \left(\frac{\int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(u) h(u) \mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(u) \mathrm{d}u} \right)^2 - n \int_{t_{i-1,n}}^{t_{i,n}} h^2(u) \mathrm{d}T_n(u) \right| \\ &+ \frac{\alpha^2}{4} \left| \int_0^1 h^2(u) \mathrm{d}[T_n - T](u) \right| \\ &+ \frac{\alpha^3}{2n\sqrt{n}} \sum_{i=1}^{n} \left| \frac{\int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(u) h(u) \mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}} \sigma_0^{-2}(u) \mathrm{d}u} \right|^3. \end{aligned} \tag{B.15}$$

We need to show that each of these four remainder terms converge to zero. For the last term, this is obvious as each element in the sum is bounded by 1 since $\|h\| \le 1$. Convergence of the third term follows from the weak convergence in Assumption 6. The second term can

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

be bounded, using (A.8), as

$$\frac{\alpha^2}{4n}\left\|\sigma_0^2\right\|\sum_{i=1}^{n}\left|\left(\left(t_{i,n}-t_{i-1,n}\right)^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u\right)^2\right.$$
$$\left.-\left(\left(t_{i,n}-t_{i-1,n}\right)^{-1}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u\right)^2\left(t_{i,n}-t_{i-1,n}\right)^{-1}\int_{t_{i-1,n}}^{t_{i,n}}h^2(u)\mathrm{d}u\right|$$
$$\le\frac{\alpha^2}{n}\left\|\sigma_0^2\right\|^3\sum_{i=1}^{n}\mathrm{V}\left(\sigma^{-2}\Big|_{t_{i-1,n}}^{t_{i,n}}\right)^{1/2}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)^{1/2}$$
$$+\frac{\alpha^2}{4n}\left\|\sigma_0^2\right\|^3\sum_{i=1}^{n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)$$
$$\le\alpha^2\left\|\sigma_0^2\right\|^3\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathrm{V}\left(\sigma^{-2}|_{t_{i-1,n}}^{t_{i,n}}\right)\frac{1}{n}\sum_{i=1}^{n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)}$$
$$+\frac{\alpha^2}{4n}\left\|\sigma_0^2\right\|^3\sum_{i=1}^{n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right),$$

which converges to zero by Assumption 3 and since

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)\le\max_{i=1,\ldots,n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)\to 0, \tag{B.16}$$

due to the fact that h is cadlag on [0, 1] and $\max_{i=1,\ldots,n}\left(t_{i,n}-t_{i-1,n}\right)\to 0$.

Finally, for the first term in (B.15), observe

$$\frac{\alpha}{2\sqrt{n}}\sum_{i=1}^{n}\left|\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)h(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u}-\frac{1}{t_{i,n}-t_{i-1,n}}\int_{t_{i-1,n}}^{t_{i,n}}h(u)\mathrm{d}u\right|$$
$$=\frac{\alpha}{2\sqrt{n}}\sum_{i=1}^{n}\left|\frac{\int_{t_{i-1,n}}^{t_{i,n}}\left[\sigma_0^{-2}(u)-\frac{1}{t_{i,n}-t_{i-1,n}}\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(v)\mathrm{d}v\right]h(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^{-2}(u)\mathrm{d}u}\right|$$
$$\le\frac{\alpha}{2\sqrt{n}}\left\|\sigma_0^2\right\|\sum_{i=1}^{n}\mathrm{V}\left(\sigma^{-2}\Big|_{t_{i-1,n}}^{t_{i,n}}\right)^{1/2}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)^{1/2}$$
$$\le\left[\max_{i=1,\ldots,n}\mathrm{V}\left(h|_{t_{i-1,n}}^{t_{i,n}}\right)^{1/2}\right]\frac{\alpha}{2}\left\|\sigma_0^2\right\|\left[\sum_{i=1}^{n}\mathrm{V}\left(\sigma^{-2}\Big|_{t_{i-1,n}}^{t_{i,n}}\right)\right]^{1/2}\to 0,$$

in view of (B.16).

To complete the proof of this LAN result, observe, using Lemma 3.1,

$$-\frac{1}{2}\int_{u=0}^{1}\sigma_0^{-2}(u)h(u)\mathrm{d}\sqrt{n}\left[RV_n^*(u)-\int_{v=0}^{u}\sigma_0^2(v)\mathrm{d}T_n(v)\right]$$
$$\to_L-\frac{1}{\sqrt{2}}\int_{u=0}^{1}h(u)\sqrt{T'(u)}\mathrm{d}W(u)\sim N\left(0,\frac{1}{2}\int_{u=0}^{1}h^2(u)\mathrm{d}T(u)\right).$$

The stated contiguity is a well-known consequence of Le Cam's first lemma, see, e.g., van der Vaart (2000), Lemma 6.4.

B.4. Proof of Theorem 4.1. We start this proof with a lemma that describes the limiting behavior of the derivative of the smoothed realized variance RV_n^S.

LEMMA B.2. *Under Assumption 8, we have*

$$RV_n^{S\prime}(u) = K'(u-1)RV_n^*(1) + \int_{v=0}^{1} K''(u-v)RV_n^*(v)dv. \tag{B.17}$$

Moreover, under Condition 1 and Assumption 7–8, we have the following convergence of the process $RV_n^{S\prime}$, under $\mathbb{P}^{(n)}_{\sigma_0^2}$,

$$\sqrt{n}\left[RV_n^{S\prime}(u) - \int_{w=0}^{1} K'(u-w)\sigma_0^2(w)dT_n(w)\right]$$
$$\to_L \sqrt{2}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)T'(w)^{1/2}dW(w). \tag{B.18}$$

Proof. First, note that in the remainder of this appendix, we, with some abuse of notation, specify a process using its value at time u. All statements should be read as weak convergence in $D[0,1]$ with respect to the Skorohod topology.

Relation (B.17) follows directly from partial integration and $RV_n^*(0) = 0$. Relation (B.18) follows by applying the continuous mapping theorem to $f \mapsto K'(\cdot - 1)f(1) + \int_{v=0}^{1} K''(\cdot - v)f(v)\mathrm{d}v$ (which is linear and bounded in view of the boundedness of both K' and K''). More precisely, from Condition 1,

$$\begin{aligned}
&\sqrt{n}\left[RV_n^{S\prime}(u) - \int_{w=0}^{1} K'(u-w)\sigma_0^2(w)\mathrm{d}T_n(w)\right]\\
&= \sqrt{n}\left[RV_n^{S\prime}(u) - K'(u-1)\int_{v=0}^{1}\sigma_0^2(v)\mathrm{d}T_n(v) - \int_{w=0}^{1}\int_{v=w}^{1} K''(u-v)\mathrm{d}v\,\sigma_0^2(w)\mathrm{d}T_n(w)\right]\\
&= \sqrt{n}\left[RV_n^{S\prime}(u) - K'(u-1)\int_{v=0}^{1}\sigma_0^2(v)\mathrm{d}T_n(v) - \int_{v=0}^{1} K''(u-v)\int_{w=0}^{v}\sigma_0^2(w)\mathrm{d}T_n(w)\mathrm{d}v\right]\\
&\to_L \sqrt{2}K'(u-1)\int_{v=0}^{1}\sigma_0^2(v)T'(v)^{1/2}\mathrm{d}W(v)\\
&\quad + \sqrt{2}\int_{v=0}^{1} K''(u-v)\int_{w=0}^{v}\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w)\mathrm{d}v\\
&= \sqrt{2}K'(u-1)\int_{v=0}^{1}\sigma_0^2(v)T'(v)^{1/2}\mathrm{d}W(v)\\
&\quad + \sqrt{2}\int_{w=0}^{1}\int_{v=w}^{1} K''(u-v)\mathrm{d}v\,\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w)\\
&= \sqrt{2}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w).
\end{aligned}$$

■

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

In view of Lemma B.2 and the almost sure convergence of T_n' in Assumption 6

$$\sqrt{n}\left[T_n'(u)^{-1}RV_n^{S\prime}(u) - T_n'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)\mathrm{d}T_n(w)\right]$$
$$\rightarrow_L \sqrt{2}T'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w).$$

Hence, applying the delta method for the transformation $(u, x(u)) \mapsto g(u, x(u))$, we find

$$\sqrt{n}\left[g\left(u, T_n'(u)^{-1}RV_n^{S\prime}(u)\right) - g\left(u, T_n'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)\mathrm{d}T_n(w)\right)\right]$$
$$\rightarrow_L \sqrt{2}\frac{\partial g}{\partial \sigma^2}\left(u, T'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)\mathrm{d}T(w)\right)$$
$$\times T'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w).$$

Applying the continuous mapping theorem to the bounded linear functional $x(\cdot) \mapsto \int_{u=0}^{1} x(u)\mathrm{d}u$ leads to (4.26).

B.5. Proof of Proposition 4.1. If K converges weakly to the distribution function $I\{\cdot \geq 0\}$, we have that

$$\int_{w=0}^{1} T'(w)\sigma_0^2(w)\mathrm{d}K(u-w),$$

converges pointwise to

$$\int_{w=0}^{1} T'(w)\sigma_0^2(w)\mathrm{d}I(u-w \geq 0) = -T'(u)\sigma_0^2(u).$$

Consequently, from the bounded convergence theorem (recall that T' is bounded away from zero and that T', σ_0^2, g, and K' are bounded), we find that

$$\int_{u=0}^{1} g\left(u, -T'(u)^{-1}\int_{w=0}^{1} T'(w)\sigma_0^2(w)\mathrm{d}K(u-w)\right)\mathrm{d}u$$

converges to

$$\int_{u=0}^{1} g\left(u, \sigma_0^2(u)\right)\mathrm{d}u = \psi_g\left(\sigma_0^2\right).$$

Let $g^{(2)}$ denote the derivative with respect to the second argument of g, i.e., $g^{(2)}(u, \sigma^2) = \frac{\partial g}{\partial \sigma^2}(u, \sigma^2)$. Concerning the limiting distribution, observe that we can

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press

rewrite it as

$$\sqrt{2}\int_{u=0}^{1} g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right)$$
$$\times T'(u)^{-1}\int_{w=0}^{1} K'(u-w)\sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w)\mathrm{d}u$$
$$=\sqrt{2}\int_{w=0}^{1}\int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right)\mathrm{d}K(u-w)$$
$$\times \sigma_0^2(w)T'(w)^{1/2}\mathrm{d}W(w),$$

from which the variance (4.29) follows. The convergence of the limiting variance when K converges weakly to point mass at zero, follows as above. In particular, we have

$$\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v) \to -\sigma_0^2(u)T'(u),$$

pointwise in u. Hence, we also have the pointwise convergence

$$g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right) \to g^{(2)}\left(u, \sigma_0^2(u)\right).$$

Now, observe that

$$\int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right)\mathrm{d}K(u-w)$$
$$-\int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, \sigma_0^2(u)\right)\mathrm{d}I\{u-w\geq 0\}$$
$$=\int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right)\mathrm{d}[K(u-w)-I\{u-w\geq 0\}]$$
$$+\int_{u=0}^{1} T'(u)^{-1}\left[g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right)-g^{(2)}\left(u, \sigma_0^2(u)\right)\right]\mathrm{d}I\{u-w\geq 0\}$$

converges to zero, pointwise in w. Indeed, this follows as the integrand in the first term on the right-hand side is bounded (and the weak convergence of K), while for the second term on the right-hand side we can, again, apply the dominated convergence theorem. Thus, we have established, pointwise in w,

$$\int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, T'(u)^{-1}\int_{v=0}^{1} K'(u-v)\sigma_0^2(v)\mathrm{d}T(v)\right)\mathrm{d}K(u-w)$$
$$\to \int_{u=0}^{1} T'(u)^{-1}g^{(2)}\left(u, \sigma_0^2(u)\right)\mathrm{d}I\{u-w\geq 0\}$$
$$= T'(w)^{-1}g^{(2)}\left(w, \sigma_0^2(w)\right),$$

as K converges weakly to pointmass at zero. Now, the convergence in (4.29) follows from another application of the bounded convergence theorem.

https://doi.org/10.1017/S0266466616000013 Published online by Cambridge University Press