Realized Range-Based Estimation of Integrated Variance[*]

Kim Christensen[†]   Mark Podolskij[‡]

This Print/Draft: September 15, 2005

Abstract

We provide a set of probabilistic laws for estimating quadratic variation of continuous semimartingales with the realized range-based variance; a statistic that replaces every squared return of realized variance with a normalized squared range. If the entire sample path of the process is available - and given weak conditions - our statistic is consistent and has a mixed Gaussian limit with five times the precision of realized variance. In practice, of course, inference is drawn from discrete data and true ranges are unobserved, leading to downward bias. We solve this problem to give a consistent, mixed normal estimator, irrespective of non-trading. It has varying degrees of efficiency over realized variance, depending on how many observations that are used to construct the high-low. The methodology is applied to TAQ data and compared with realized variance. Our findings suggest the empirical path of quadratic variation is also estimated better with the intraday high-low statistic.

JEL Classification: C10; C22; C80.

Keywords: Central Limit Theorem; Continuous Semimartingale; High-Frequency Data; Integrated Variance; Quadratic Variation; Realized Range-Based Variance; Realized Variance; Stochastic Volatility; Volatility Measurement.

[†]Aarhus School of Business, Dept. of Marketing and Statistics, Fuglesangs Allé 4, 8210 Aarhus V, Denmark. Phone: (+45) 89 48 63 74, fax: (+45) 86 15 37 92, e-mail: kic@asb.dk.

[‡]Ruhr University of Bochum, Dept. of Probability and Statistics, Universitätstrasse 150, 44801 Bochum, Germany. Phone: (+49) 234 / 23283, fax: (+49) 234 / 32 14559, e-mail: podolski@cityweb.de.

1 Introduction

The latent security price volatility is an essential measure of unexpected return variation and a key ingredient in several pillars of financial economics. Some years ago, academia customarily adopted constant volatility models (e.g., Black & Scholes (1973) and related literature), despite that the data argued against this assumption (e.g., Mandelbrot (1963)). Today, the gathering of empirical evidence makes us recognize that the conditional variance is both time-varying and highly persistent. Such stylized facts were uncovered by the development and application of strict parametric models, such as ARCH (see, e.g., Bollerslev, Engle & Nelson (1994)), through stochastic volatility models (e.g., Ghysels, Harvey & Renault (1996)), and more recently non-parametric methods based on high-frequency data, the most conspicuous idea being *realized variance* (RV), see, e.g., Andersen, Bollerslev, Diebold & Labys (2001) or Barndorff-Nielsen & Shephard (2002a); henceforth abbreviated as ABDL and BN-S.

RV is the sum of squared returns over non-overlapping intervals within a sampling period. The theory states that, given weak regularity conditions, RV converges in probability to the *quadratic variation* (QV) of all semimartingales as the sampling frequency tends to infinity (e.g., Protter (2004)).

In practice, the consistency result of RV breaks down as data limitations prevent the sampling frequency from rising without bound. Most notably, market microstructure effects contaminate high-frequency asset prices. This invalidates the asymptotic properties of RV, and in the presence of noise it is both biased and inconsistent (e.g., Bandi & Russell (2004, 2005), Aït-Sahalia, Mykland & Zhang (2005), and Hansen & Lunde (2006)). Though current research seeks to develop methods of making RV robust against microstructure noise, getting the most accurate estimates of price variations remain unresolved. Set against this backdrop, we suggest another proxy: *realized range-based variance* (RRV).

Range-based estimation of volatility (developed in, e.g., Feller (1951), Garman & Klass (1980), Parkinson (1980), Rogers & Satchell (1991) and Kunitomo (1992)) is very informative, since the extremes are formed from the entire curve of the process and therefore reveal more information than points sampled at fixed intervals. Indeed, the daily squared range is about five times more efficient at estimating the scale of Brownian motion than the daily squared return. However, Andersen & Bollerslev (1998, footnote 20) remark that "...compared to the measurement errors reported in Table 3, this puts the accuracy of the high-low estimator around that afforded by the intraday sample variance based on two- or three-hour returns." As a consequence of the daily range's measurement error against RV, the class of range-based proxies remains neglected.

Nonetheless, one subject - with the potential of markedly reducing this error - remains uncharted territory: intraday range-based estimation of stochastic volatility. That is, while the range is recognized as being highly efficient vis-à-vis the return on a daily basis, no one

has explored the properties of price ranges sampled within the trading day in the context of estimating QV. With access to high-frequency data, however, low-frequency measures are also obsolete in the range-based context. For example, with exchange rate data available around the clock, what can we expect by using, properly transformed, high-frequency ranges? Direct extrapolation suggests that if daily ranges are as accurate as RV based on two- or three-hour returns, then hourly ranges, say, achieve the accuracy of RV sampled at five- or ten-minute intervals.

We propose to sample and sum intraday price ranges to construct more efficient estimates of QV. Our contribution is four-fold. First, we develop a non-parametric method for measuring QV with RRV. Second, unlike the existing time-invariant theory for the high-low, we deal with estimation of time-varying volatility, when the driving terms of the price process are (possibly) continuously evolving random functions. Such a model is capable of - but arguably also necessary for - fitting the facts of financial markets data; in particular the second moment structure of the conditional return distribution. Third, we derive a set of probabilistic laws for sampling intraday high-lows. Fourth, we remove the problems with downward bias reported in the previous range-based literature.

The paper is structured as follows. In the next section, we unfold the necessary diffusion theory, present various ways of measuring volatility and advance our methodological contribution by suggesting RRV and a version thereof without problems of non-trading effects. Under mild conditions, we prove consistency for the estimation methods and find mixed Gaussian central limit theorems. Section 3 illustrates the approach with Monte Carlo analysis to uncover the finite sample properties, and we progress in section 4 with some empirical results. Rounding up, section 5 offers conclusions and sketches several directions for future research.

2   A Semimartingale Framework

In this section, we propose a new method based on the price range for consistently estimating QV. The theory is developed for the log-price of a univariate asset evolving in continuous time over some interval, say $p = \{p_t\}_{t \in [0,\infty)}$. p is defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0,\infty)}, \mathbb{P})$ and adapted to the filtration $\{\mathcal{F}_t\}_{t \in [0,\infty)}$; i.e. a family of σ-fields with $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for all $s \leq t < \infty$.

The basic building block is that p constitutes a continuous sample path semimartingale.[1] Hence, we write the time t log-price in the generic form:

$$p_t = p_0 + \int_0^t \mu_s \mathrm{d}s + \int_0^t \sigma_{s-} \mathrm{d}W_s, \quad \text{for } 0 \leq t < \infty, \tag{2.1}$$

where $\mu = \{\mu_t\}_{t \in [0,\infty)}$ (the instantaneous mean) is a locally bounded predictable process, $\sigma =$

---

[1]We adopt the continuity assumption as a starting point only. In concurrent work, we are analyzing the properties of our estimators, when p exhibits jumps.

$\{\sigma_t\}_{t\in[0,\infty)}$ (the spot volatility) is a càdlàg process, $W = \{W_t\}_{t\in[0,\infty)}$ is a standard Brownian motion, and $\sigma_{s-} = \lim_{t\to s, t<s} \sigma_t$.

Much work in both theoretical and empirical finance or time series econometrics is cast within this setting (see, e.g., Andersen, Bollerslev & Diebold (2002) or BN-S (2005c) for excellent reviews and several references). Except for the continuity of the local martingale, we impose little structure on the model. In fact, for semimartingales with continuous martingale component as above, the form $\{\int_0^t \mu_s \mathrm{d}s\}_{t\in[0,\infty)}$ is implicit, when the drift term is predictable (in the absence of arbitrage).[2] Note in passing that, without loss of generality, we can restrict the functions μ and σ to be bounded (e.g., Barndorff-Nielsen, Graversen, Jacod, Podolskij & Shephard (2005)).

The objective is to estimate a suitable measure of the return variation over a subinterval $[a,b] \subseteq [0,\infty)$, termed the sampling period or measurement horizon. We assume $[a,b] = [0,1]$; this will be thought of as representing a trading day, but the choice is arbitrary and can serve as a normalization. At any two distinct sampling times t_{i-1} and t_i, with $0 \le t_{i-1} < t_i \le 1$, the intraday return over $[t_{i-1}, t_i]$ is denoted by:

$$r_{t_i, \Delta_i} = p_{t_i} - p_{t_{i-1}}, \qquad (2.2)$$

where $\Delta_i = t_i - t_{i-1}$. As a convention, we call the return stretching $[0,1]$ for the interday or daily return (i.e., taking $t_i = \Delta_i = 1$).

From the theory of stochastic integration, it is well-known that QV is a natural measure of sample path variability for the class of semimartingales. Specifically, for every semimartingale, $X = \{X_t\}_{t\in[0,\infty)}$, there exists a unique increasing QV process, $[X] = \{[X]_t\}_{t\in[0,\infty)}$, given by:

$$[X]_t = X_t^2 - 2 \int_0^t X_{s-} \mathrm{d}X_s, \qquad (2.3)$$

with $X_{s-} = \lim_{t\to s, t<s} X_t$.

In the absence of jumps in p, QV is entirely induced by innovations to the local martingale. Moreover, in our framework QV coincides with the *integrated variance* (IV) that is central to financial economics, whether in asset and derivatives pricing, portfolio selection or risk management (e.g., Andersen, Bollerslev & Diebold (2002)). IV is the object of interest here, and we recollect its definition:

$$IV = \int_0^1 \sigma_s^2 \mathrm{d}s. \qquad (2.4)$$

The econometrical problem is that IV is latent, which renders empirical estimation of this quantity a crucial issue in practice. We shall briefly review the literature on existing methods for measuring IV, before carrying on to suggest a new approach.

---

[2]Moreover, all local martingales, whose QV (to be defined in a moment) is absolutely continuous, has the martingale representation of the second term in Equation (2.1), e.g., Doob (1953). We refer to BN-S (2004, footnote 6) for the details of this aspect.

2.1 Return-Based Estimation of Integrated Variance

Not long ago, the daily squared return was employed as a non-parametric ex-post measure of IV. Although the estimator is (conditionally) unbiased under some auxiliary conditions, the interday return is a noisy indicator of volatility. With the advent of high-frequency data, more recent work computes RV, being the sum of squared intraday returns sampled over non-overlapping intervals (see, e.g., ABDL (2001) or BN-S (2002a)).[3]

More formally, consider a deterministic partition $0 = t_0 < t_1 < \cdots < t_n = 1$. Then, adopting the notation of Hansen & Lunde (2005), we define RV at sampling times $\Xi = \{t_i \mid i = 0, 1, \ldots, n\}$, or sampling frequency n, by setting:

$$RV^\Xi = \sum_{i=1}^n r_{t_i, \Delta_i}^2. \tag{2.5}$$

The daily squared return is the least efficient member of this class of estimators, and the justification for the RV procedure builds directly on the theory of QV. An equivalent definition of QV is the probability limit of RV^Ξ, as the diameter of Ξ tends to zero (e.g., Protter (2004)). That is, in our model:

$$RV^\Xi \xrightarrow{p} IV, \tag{2.6}$$

provided $\max_{1 \le i \le n}\{\Delta_i\} \to 0$ as $n \to \infty$.[4] The convergence is locally uniform in time. Thus, by definition RV^Ξ is consistent and given a complete record of p, IV is estimated with arbitrary accuracy, effectively making it observed.

Of course, we are forced to work with discretely sampled data in applications, but the theory encourages using high-frequency proxies to reduce the measurement error. And though an irregular partition of the sampling period suffices for consistency, an equidistant time series of intraday returns is often computed in practice by various approaches, such as linear interpolation (see, e.g., Andersen & Bollerslev (1997a, 1997b, 1998), and Andersen, Bollerslev, Diebold & Ebens (2001)) or the previous-tick method suggested in Wasserfallen & Zimmermann (1985).[5] Oomen (2005) gives a characterization of RV under alternative sampling schemes.

The equidistant RV based on n high-frequency returns sampled over non-overlapping intervals of length $\Delta = 1/n$ is defined as:

$$RV^\Delta = \sum_{i=1}^n r_{i\Delta, \Delta}^2. \tag{2.7}$$

---

[3]The stimulating work of Rosenberg (1972) was an early pioneer in the empirical construction of a volatility proxy that exploited data sampled at a higher frequency than the measurement horizon.

[4]In general, the non-normed qth order *realized power variation* is defined as $\sum_{i=1}^n |r_{t_i, \Delta_i}|^q$, with $q > 0$, see BN-S (2004). For Brownian semimartingales, only $q = 2$ leads to a nontrivial limit (IV).

[5]A side-effect of the linear interpolation method is that - with a fixed number of data - $RV^\Xi \xrightarrow{p} 0$ as $n \to \infty$, because the limit of the interpolated process is of continuous bounded variation, see Hansen & Lunde (2006). Intuitively, a straight line is the minimum variance path between two points.

BN-S (2002a) found a distribution theory for RV^Δ in relation to IV. The law of the scaled difference between RV^Δ and IV has a mixed Gaussian limit,

$$\sqrt{n}\left(RV^\Delta - IV\right) \xrightarrow{d} MN\left(0, 2IQ\right),$$ (2.8)

where

$$IQ = \int_0^1 \sigma_s^4 \mathrm{d}s,$$ (2.9)

is the *integrated quarticity*. Thus, the size of the error bounds for RV^Δ is positively related to the level of σ, so it is more difficult for RV to estimate IV, when σ is high. BN-S (2002a) also derived a feasible central limit theorem (CLT), where all quantities except IV can be computed directly from the data. This was done by simply replacing the latent IQ by a consistent estimator, like *realized quarticity* (RQ):

$$RQ^\Delta = \frac{n}{3}\sum_{i=1}^{n} r_{i\Delta,\Delta}^4,$$ (2.10)

making it possible to construct approximate confidence bands for RV^Δ to measure the size of the estimation error involved with finite sampling.

2.2 Range-Based Estimation of Integrated Variance

In practice, the choice of volatility proxy is less obvious, as financial markets are not frictionless and microstructure bias sneaks into RV, when n is too large. With noisy prices, for instance, RV is biased and inconsistent, see, e.g., Zhou (1996), Bandi & Russell (2004, 2005), Aït-Sahalia et al. (2005), and Hansen & Lunde (2006).[6] Academia has recognized this by developing bias reducing techniques (e.g., prewhitening of the high-frequency return series with moving average or autoregressive filters as in Andersen, Bollerslev, Diebold & Ebens (2001) and Bollen & Inder (2002), or kernel-based estimation as in Zhou (1996) and Hansen & Lunde (2006)). In empirical work, the benefits of more frequent sampling is traded off against the damage caused by cumulating noise, and - using various criteria for picking the optimal sampling frequency - the result is often moderate sampling (e.g., at the 5-, 10-, or 30-minute frequency), whereby data are discarded.

The pitfalls of RV motivate our choice of another proxy with a long history in finance: the price range or high-low. Using a terminology similar to the above, we define the intraday range at sampling times t_{i-1} and t_i as:

$$s_{p_{t_i,\Delta_i}} = \sup_{t_{i-1}\leq s,t\leq t_i}\left\{p_t - p_s\right\}.$$ (2.11)

Compared to the return over $[t_{i-1}, t_i]$, r_{t_i,Δ_i}, the extra subscript p indicates that we are taking supremum of the price process. Below, we also need the range of a standard Brownian motion

---

[6]Technically, with IID noise, RV diverges to infinity almost surely, i.e. $RV^\Xi \xrightarrow{a.s.} \infty$ as $n \to \infty$.

over $[t_{i-1}, t_i]$, which is denoted by:

$$s_{W_{t_i, \Delta_i}} = \sup_{t_{i-1} \leq s, t \leq t_i} \{W_t - W_s\}. \tag{2.12}$$

We use the shorthand notation s_p and s_W for the interday, or daily, ranges (again by taking $t_i = \Delta_i = 1$).

2.2.1 The Interday Range

The logic of the range is appealing: suppose the asset price fluctuates wildly within the sampling period, but happens to end near the starting point; then an interday high-low, unlike the return, correctly reports the level of volatility as high.

Its attractiveness is not based on just intuitive grounds, however. The theoretical under-pinnings go a long way back.[7] Feller (1951) found the distribution of the range by using the theory of Brownian motion. According to his work, the density of $s_{W_{t_i, \Delta_i}}$ is given by:

$$\Pr\left[s_{W_{t_i, \Delta_i}} = r\right] = 8 \sum_{x=1}^{\infty} (-1)^{x-1} \frac{x^2}{\sqrt{\Delta_i}} \phi\left(\frac{xr}{\sqrt{\Delta_i}}\right), \tag{2.13}$$

with $\phi(y) = \exp\left(-y^2/2\right)/\sqrt{2\pi}$. The infinite series is evaluated by a suitable truncation. In Figure 1, we plot the probability function of s_W.

[INSERT FIGURE 1 ABOUT HERE]

In a historical context, a major reason for selecting the daily high-low to estimate IV related to its sampling stability. Viewed separatively, the density function does not reveal this feature. Therefore, the figure also displays the distribution of the daily absolute return. By comparing these proxies, the efficiency of the daily range, or in other words its lower variance vis-à-vis the daily return, is more evident.

Parkinson (1980) advanced Feller's insights by deriving the moment generating function of the range of a scaled Brownian motion, $p_t = \sigma W_t$.[8] For the rth moment:

$$\mathbb{E}\left[s_{p_{t_i, \Delta_i}}^r\right] = \lambda_r \Delta_i^{r/2} \sigma^r, \quad \text{for } r \geq 1, \tag{2.14}$$

with $\lambda_r = \mathbb{E}[s_W^r]$. In particular, $\lambda_2 = 4\ln(2)$ and $\lambda_4 = 9\zeta(3)$ are needed below.[9] Noting that $IV = \sigma^2$ in this model, we thus get an unbiased estimate for daily sampling by scaling s_p^2 down with λ_2, which is the classic high-low estimator.

---

[7]There are basically two branches in the context of range-based volatility: i) relies purely on the high-low, while ii) adds the open-close, e.g., Garman & Klass (1980) or Rogers & Satchell (1991). Brown (1990) and Alizadeh, Brandt & Diebold (2002) argue against inclusion of the latter on the grounds that they are highly contaminated by microstructure effects. Thus, throughout we only report on the high-low estimator.

[8]Note, σ does double-duty; representing either the process $\sigma = \{\sigma_t\}_{t \in [0,\infty)}$ or a constant diffusion parameter $\sigma_t = \sigma$. The meaning is clear from the context.

[9]The explicit formula for λ_r is: $\lambda_r = \frac{4}{\sqrt{\pi}}(1 - \frac{4}{2^r})2^{r/2}\Gamma(\frac{r+1}{2})\zeta(r-1)$, for $r \geq 1$; where $\Gamma(x)$ and $\zeta(x)$ denote the Gamma and Riemann's zeta function, respectively.

If $\mu \neq 0$, the daily range cannot distinguish drift from volatility and is biased. A few methods were suggested to allow nonzero - but constant - μ. Rogers & Satchell (1991) used the exponential distribution and Wiener-Hopf factorization of a Lévy process, while Kunitomo (1992) suggested the range of a Brownian bridge; both to produce estimators that are independent of μ.

Arguably, a process with constant μ and σ is irrelevant from an empirical point of view. The most critical aspect of range-based theory is perhaps the homoscedasticity constraint forced upon σ. An overwhelming amount of research indicates that the conditional variance is time-varying, see, e.g., Ghysels et al. (1996). Nonetheless, to our knowledge there exists little theory about range-based estimation of IV in the presence of a continually evolving diffusion parameter.[10] Previous work achieve (randomly) changing volatility by holding σ_t fixed within the trading day, while allowing for (stochastic) shifts between them (e.g., Alizadeh et al. (2002), Brunetti & Lildholdt (2002)).[11] Still, there are strong intraday movements in σ_t (e.g., Andersen & Bollerslev (1997b)).

A major objective of this paper is therefore to extend the theoretical domain of the extreme value method to a more general class of stochastic processes. Contrary to the extant research, we develop a statistical framework for the Brownian semimartingale in Equation (2.1), featuring less restrictive dynamics for μ and σ.

2.2.2 A Realized Range-Based Estimator

As stated earlier, the (transformed) daily price range is less efficient than RV for moderate values of n; two- or three-hour returns suffices. But with tick-by-tick data at hand, we can exploit the insights of RV to construct more precise range-based estimates of IV by sampling high-lows within the trading day. Curiously, a rigorous analysis of intraday ranges is missing in the volatility literature.

Accordingly, consider again the partition $0 = t_0 < t_1 < \cdots < t_n = 1$. We then propose an RRV estimator of IV at sampling times Ξ, or sampling frequency n:

$$RRV^{\Xi} = \frac{1}{\lambda_2} \sum_{i=1}^{n} s^2_{p_{t_i, \Delta_i}}. \tag{2.15}$$

This procedure has two advantages over the previous return- and range-based methods suggested in the extant research on volatility measurement. First, the approach inspects all data points (regardless of the sampling frequency), whereby we avoid neglecting information about

---

[10]A notable exception is Gallant, Hsu & Tauchen (1999), who estimate two-factor stochastic volatility models in a general continuous time framework. They derive the density function of the range in this setting, but do not otherwise explore its theoretical properties.

[11]Brunetti & Lildholdt (2002) consider a model with GARCH dynamics and show that the scaled squared range is unbiased for the unconditional variance.

IV. Second, the efficiency of RRV^Ξ is several times that of RV^Ξ, leading to narrower confidence intervals for IV (see below).

We denote the equidistant version of the intraday high-low statistic by RRV^Δ.

2.2.3 Properties of Realized Range-Based Variance

At a minimum, the estimator should be consistent for IV. With the time-invariant scaled Brownian motion, proving this property of RRV^Δ is trivial.[12] As the infill asymptotics start operating by letting $n \to \infty$, we achieve an increasing sequence of IID random variables, $\{s_{p_{i\Delta,\Delta}}\}_{i=1,\dots,n}$. Suitably transformed to unbiased measures of σ^2 using (2.14), the consistency follows from a standard law of large numbers by averaging. To see this, note that $\mathbb{E}\left(RRV^\Delta\right) = \sigma^2$ and $\text{var}\left(RRV^\Delta\right) = \Lambda n^{-1}\sigma^4$, with:

$$\Lambda = \frac{\lambda_4 - \lambda_2^2}{\lambda_2^2}. \tag{2.16}$$

Hence, MSE $\to 0$ as $n \to \infty$, which is sufficient. Also, for this process a CLT is easily found as:

$$\sqrt{n}\left(RRV^\Delta - IV\right) \xrightarrow{d} N\left(0, \Lambda\sigma^4\right). \tag{2.17}$$

If μ and σ are stochastic, establishing the large sample properties of RRV^Δ is more involved, but nonetheless possible. Overall, the basic idea extends to general Brownian semimartingales, given some regularity on μ and σ. To justify our approach, we therefore progress by deriving limit theorems for RRV^Δ. Its probability limit is stated first.[13]

Theorem 1 *Let μ and σ fulfil the conditions following Equation (2.1). As $n \to \infty$*

$$RRV^\Delta \xrightarrow{p} IV. \tag{2.18}$$

This mirrors the consistency of RV that by definition converges, in probability, to the limit process IV. Theorem 1 allows for general specifications of μ as a consequence of the fact that for continuous time arbitrage-free price processes, the expected move in p is an order of magnitude lower than variation induced by the local martingale; here comprised by the stochastic volatility component $\left\{\int_0^t \sigma_{s-}\mathrm{d}W_s\right\}_{t\in[0,\infty)}$. Thus, while the daily range is sensitive to drift, the mean component vanishes (sufficiently fast) as $n \to \infty$. But our analysis extends the theory of the range much further from the perspective of volatility measurement. Except for weak technical conditions on σ, no knowledge about its dynamics is needed. Hence, we allow for very general continuous time processes, including, but not limited to, models for σ with leverage, long-memory, jumps or diurnal effects. This is certainly not true in the previous range-based literature.

---

[12]Henceforth, we use equidistant estimation, as this simplifies the notational burden in the proofs. All results generalize to irregular subdivisions of the sampling period, so long as $\max\{\Delta_i\}_{1\leq i\leq n} \to 0$. We just need a slight modification of the conditional variance in the CLT, as spelled out below.

[13]Throughout the paper, proofs of the theorems are reserved for the Appendix.

2.2.4 Asymptotic Distribution Theory

In empirical work, the consistency result of RRV^Δ becomes unreliable due to microstructure frictions, when n is too large. Theorem 1 hides the precision IV is estimated with by fixing n at a moderate level, and econometricians often compute confidence bands as a guide to the error made from estimation based on finite sampling. To strengthen the convergence in probability, we now develop a distribution theory for RRV^Δ.

The above weak assumptions on μ and σ are too general to prove a CLT, and we need slightly stronger conditions, collectively referred to as Assumption **(M)** and **(V)**:

(M) μ is continuous.

(V) σ is everywhere invertible (V$_1$) and satisfies:

$$\sigma_t = \sigma_0 + \int_0^t \mu_s' \mathrm{d}s + \int_0^t \sigma_s' \mathrm{d}W_s + \int_0^t v_s \mathrm{d}B_s', \quad \text{for } 0 \le t < \infty, \tag{V$_2$}$$

where $\mu' = \{\mu_t'\}_{t\in[0,\infty)}$ is locally bounded, $\sigma' = \{\sigma_t'\}_{t\in[0,\infty)}$ and $v = \{v_t\}_{t\in[0,\infty)}$ are càdlàg, and $B' = \{B_t'\}_{t\in[0,\infty)}$ is a Brownian motion independent of W.

We also need a special mode of convergence, namely stable convergence in law. It is standard in the RV literature. But to avoid any confusion about our terminology, we present the definition here, as it is probably not widely familiar to people in econometrics and finance.

Definition 1 *A sequence of random variables, $\{X_n\}_{n\in\mathbb{N}}$, converges stably in law with limit X, defined on an appropriate extension of $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\in[0,\infty)}, \mathbb{P})$, if and only if for every $\mathcal{F}$-measurable, bounded random variable Y and any bounded, continuous function g, the convergence $\lim_{n\to\infty} \mathbb{E}[Yg(X_n)] = \mathbb{E}[Yg(X)]$ holds.*

Throughout the paper, the symbol $X_n \xrightarrow{d_s} X$ is used to denote stable convergence in law (see, e.g., Rényi (1963) or Aldous & Eagleson (1978) for more details). Note that it implies weak convergence, which may equivalently be defined by taking $Y = 1$. We need this stronger version to prove a standard CLT, as the conditional variance of the limit processes below is not deterministic.

We now state the main result, which is a (nonstandard) CLT.

Theorem 2 *If (M) and (V) are satisfied, then we have:*

$$\sqrt{n}\left(RRV^\Delta - IV\right) \xrightarrow{d_s} \sqrt{\Lambda} \int_0^1 \sigma_s^2 \mathrm{d}B_s, \tag{2.19}$$

where $B = \{B_t\}_{t\in[0,1]}$ is a standard Brownian motion, independent of $\mathcal{F}$ (written $B \perp\!\!\!\perp \mathcal{F}$).

A critical feature of this theorem is that the left-hand side converges stably to a stochastic integral with respect to B, which in turn is unrelated to the driving terms μ, σ and W.

This means $\sqrt{n}\left(RRV^{\Delta} - IV\right)$ has a mixed normal limit, with σ governing the mixture.[14] In general, this gives the unconditional distribution a higher peak and heavier tails than for Gaussian random variables. To summarize, conditional on σ:

$$\sqrt{n}\left(RRV^{\Delta} - IV\right) \xrightarrow{d} MN\left(0, \Lambda IQ\right). \tag{2.20}$$

Remark 1 *The Λ scalar in front of IQ in Equation (2.20) is roughly 0.4. In contrast, the number appearing in the CLT for RV^{Δ} is 2.*

Hence, measurement errors of RRV^{Δ} are about one-fifth of those extracted with RV^{Δ}. This is not surprising: RRV^{Δ} uses all the data, whereas RV^{Δ} is based on high-frequency returns sampled at fixed points in time. As, for the moment, p is assumed fully observed, RV^{Δ} is neglecting a lot of information.

IQ on the right-hand side in (2.20) is statistically infeasible, i.e. it cannot be computed from the data. We estimate it with *realized range-based quarticity (RRQ)*:

$$RRQ^{\Delta} = \frac{n}{\lambda_4} \sum_{i=1}^{n} s_{p_{i\Delta,\Delta}}^4. \tag{2.21}$$

With techniques similar to the proof of Theorem 1, we have that $RRQ^{\Delta} \xrightarrow{p} IQ$. Thus, by the properties of stable convergence (e.g., Jacod (1997)), we get the next corollary.

Corollary 1 *Given (M) and (V), it holds that:*

$$\frac{\sqrt{n}\left(RRV^{\Delta} - IV\right)}{\sqrt{\Lambda RRQ^{\Delta}}} \xrightarrow{d} N(0,1). \tag{2.22}$$

Remark 2 *With irregular sampling schemes, the distributional result in (2.20) - and those in the next sections - changes slightly (the stochastic limit is unchanged). Set,*

$$H_{n,s}^{\Xi} = n \sum_{i=1}^{j:t_j \leq s} \left(t_i - t_{i-1}\right)^2, \tag{2.23}$$

and assume that a pointwise limit H_s^{Ξ} of $H_{n,s}^{\Xi}$ exists and is continuously differentiable. Then, as $n \to \infty$ such that $\max_{1 \leq i \leq n}\{\Delta_i\} \to 0$:

$$\sqrt{n}\left(RRV^{\Xi} - IV\right) \xrightarrow{d} MN\left(0, \Lambda \int_0^1 \frac{\partial H_s^{\Xi}}{\partial s} \sigma_s^4 \mathrm{d}s\right). \tag{2.24}$$

The derivative $\partial H_s^{\Xi}/\partial s$ is small, when sampling runs quickly. Hence, there are potential gains in having more frequent observations, when σ is high. Obviously, for equidistant subdivisions

---

[14]Earlier drafts of this paper had a non-mixed Gaussian CLT and the stronger conditions, A: $\mu = 0$ and B: σ is Hölder continuous of order $\gamma > 1/2$, i.e. $\sigma_t - \sigma_s = \mathrm{O}_p\left(\mid t - s \mid^{\gamma}\right)$ for $t \to s$. We have substantially weakened these restrictions and also proved the mixed Gaussian CLT. Svend E. Graversen was helpful in pointing our attention to a result that enabled us to remove A and B (Lemma 1 in the Appendix).

$H_s^\Xi = s$, *so the extra term drops out. The theory is made feasible with*

$$RRQ^\Xi = \frac{n}{\lambda_4} \sum_{i=1}^{n} s_{p_{t_i,\Delta_i}}^4 \xrightarrow{p} \int_0^1 \frac{\partial H_s^\Xi}{\partial s} \sigma_s^4 \mathrm{d}s. \tag{2.25}$$

2.2.5 Discrete High-Frequency Data

In practice, we draw inference about IV from discretely sampled data. As p is not continuously monitored, we cannot extract the true supremum of the increments to the semimartingale. If unaccounted for, the intraday high-low statistic becomes downward biased, and with a fixed number of data the error will be progressively more severe as n gets larger.

Building on the earlier work and simulation evidence of Garman & Klass (1980), Rogers & Satchell (1991) proposed a technique for bias correcting the range; a method that largely removed the error from a numerical perspective.

Nonetheless, it turns out to be misleading thinking about ranges as downward biased. The source of the problem is λ_2, which is constructed on the grounds that p is fully observed. Therefore, we now develop an estimator that accounts for the number of high-frequency data used in forming the high-low, in order to scale properly. To formalize this idea, a bit more notation is required. Assume, without loss of generality, that $mn + 1$ equidistant price data are available, giving mn increments. They are split into n intervals with m innovations each and we denote the observed maximum over the ith interval by:

$$om_{p_{i\Delta,\Delta}} = \max_{0 \le s,t \le m} \left\{ p_{(i-1)/n + t/mn} - p_{(i-1)/n + s/mn} \right\}. \tag{2.26}$$

Also, let:

$$om_W = \max_{0 \le s,t \le m} \left\{ W_{t/m} - W_{s/m} \right\}. \tag{2.27}$$

Then, we define the new intraday high-low statistic by setting:

$$RRV_m^\Delta = \frac{1}{\lambda_{2,m}} \sum_{i=1}^{n} om_{p_{i\Delta,\Delta}}^2, \tag{2.28}$$

where $\lambda_{r,m} = \mathbb{E}\left[om_W^r\right]$. The constant appearing in this expression is nothing more than (the reciprocal of) the rth moment of the range of a standard Brownian motion over a unit interval, when we only observe m increments to the underlying continuous time process.

To the best of our knowledge, there is no explicit formula for $\lambda_{r,m}$, but it is easily computed to any degree of accuracy from simple simulations. Figure 2 details this for the example $r = 2$ and all values m that integer divides 23,400.

[INSERT FIGURE 2 ABOUT HERE]

Of course, $\lambda_{2,m} \to \lambda_2$ as $m \to \infty$, but note also that $\lambda_{2,1} = 1$, which defines RV^Δ. The downward bias reported from previous simulation studies on the range is a consequence of

the fact that $1/\lambda_2$ was incorrectly applied in place of $1/\lambda_{2,m}$, as the bias is in one-to-one correspondence with the difference.

Having completed these preliminaries, we prove consistency and asymptotic normality for the estimator in equation (2.28) by letting $n \to \infty$. Note, m is not required to approach infinity; convergence to any natural number is sufficient.

Theorem 3 *Assume $n \to \infty$ and $m \to c \in \mathbb{N} \cup \{\infty\}$. Then,*

$$RRV_m^\Delta \overset{p}{\to} IV. \tag{2.29}$$

Moreover, if (M) and (V) are satisfied:

$$\sqrt{n}\left(RRV_m^\Delta - IV\right) \overset{d_s}{\to} \sqrt{\Lambda_c} \int_0^1 \sigma_s^2 \mathrm{d}B_s, \tag{2.30}$$

where $\Lambda_c = \left(\lambda_{4,c} - \lambda_{2,c}^2\right)/\lambda_{2,c}^2$ and $B \perp\!\!\!\perp \mathcal{F}$. Finally,

$$\frac{\sqrt{n}\left(RRV_m^\Delta - IV\right)}{\sqrt{\Lambda_m RRQ_m^\Delta}} \overset{d}{\to} N(0,1), \tag{2.31}$$

with $\Lambda_m = \left(\lambda_{4,m} - \lambda_{2,m}^2\right)/\lambda_{2,m}^2$ and

$$RRQ_m^\Delta = \frac{n}{\lambda_{4,m}} \sum_{i=1}^n om_{p_{i\Delta,\Delta}}^4. \tag{2.32}$$

Remark 3 *The distribution theory in Theorem 3 nests RV, in the sense that for the special case $m = 1$, it provides a CLT for RV^Δ, as discussed in, e.g., BN-S (2002a) or Barndorff-Nielsen et al. (2005).*

[INSERT FIGURE 3 ABOUT HERE]

To provide an impression of the efficiency of RRV_m^Δ, Figure 3 depicts Λ_m on the y-axis, as a function of the number of intraday returns m along the x-axis. Several hundred recordings are needed for a good fit to the asymptotic value of 0.4, but the steep initial decline renders the advantage of RRV_m^Δ large compared to RV^Δ even for moderate m. For $m = 10$, say, the scalar appearing in front of IQ in the CLT for the realized range-based estimator equals roughly 0.7, making the confidence intervals for IV much smaller. In our experience, $m = 10$ or higher values are usually obtained for moderately liquid assets at empirically relevant frequencies, like 5-minute sampling.

3 Monte Carlo Exploration

In this section, we illustrate the workings of our theory by using repeated samples from a stochastic volatility model to further study the finite sample performance of RRV_m^Δ and document

its asymptotic properties. The following bivariate system of stochastic differential equations is simulated:

$$\mathrm{d}p_t = \sigma_t \mathrm{d}W_t$$
$$\mathrm{d}\ln\sigma_t^2 = \theta(\omega - \ln\sigma_t^2)\mathrm{d}t + \eta\mathrm{d}B_t, \tag{3.1}$$

where W and B are independent Brownian motions, while (θ, ω, η) are parameters.[15] Thus, spot log-variance evolves as a mean reverting Ornstein-Uhlenbeck process with mean ω, mean reversion parameter θ and volatility η (see, e.g., Gallant et al. (1999), Alizadeh et al. (2002), and Andersen, Benzoni & Lund (2002)). The vector $(\theta, \omega, \eta) = (0.032, -0.631, 0.115)$ is from Andersen, Benzoni & Lund (2002), who apply Efficient Method of Moments (EMM) to calibrate numerous continuous time models.

The initial conditions are set to $p_0 = 0$ and $\ln\sigma_0^2 = \omega$, and our simulation design is completed by generating $T = 1,000,000$ daily replications from this model with mn price increments each, where mn depends on the setting (see below). Throughout, we ignore the irregular spacing of empirical high-frequency data and work with equidistant data.

3.1 Simulation Results

The distributional result for RRV_m^Δ is detailed by selecting $m = 10$. Reported results are not that sensitive to specific choices of m, but in general higher values make the size properties of the asymptotic confidence bands better. We simulate $n = 10$, 50, 100 for a total of $mn = 100$, 500, 1000 increments each day, allowing us to show the gradual convergence in distribution to the standard normal for daily high-frequency sample sizes that resemble those of moderately liquid assets. With a total of $T = 1,000,000$ replications generated, we get very accurate estimates of the actual finite sample density.

[INSERT FIGURE 4 ABOUT HERE]

Figure 4, upper panel, graphs kernel densities for the standardized errors of RRV_m^Δ; cf. the ratio in Equation (2.31). It details that for $n = 10$, the distribution is left-skewed with a poor behavior in both the center and tail areas compared to the superimposed N(0,1) reference density. The size distortions diminish by progressively increasing the sample and with $n = 100$ the tails are tracked quite closely.

BN-S (2005a) showed that log-based inference via standard linearization methods improved the raw distribution theory for RV^Δ. They found a better finite sample behavior for the errors of the log-transform than those extracted with the feasible version of the CLT outlined in Equation (2.8). The shape of the actual densities for RRV_m^Δ suggests this also applies in our

---

[15]A discrete time version of the continuous time model in (3.1) is obtained with a standard Euler approximation scheme, i.e., $p_{t+\Delta} = p_t + \sigma_t\sqrt{\Delta}\phi_t$ and $\ln\sigma_{t+\Delta}^2 = \theta\omega\Delta + \ln\sigma_t^2(1 - \theta\Delta) + \eta\sqrt{\Delta}\epsilon_t$, where ϕ_t and ϵ_t are orthogonal N(0,1) variates.

setting. By the delta method, the log-version of the CLT for RRV_m^Δ takes the form:

$$\sqrt{n}\left(\ln RRV_m^\Delta - \ln IV\right) \xrightarrow{d} MN\left(0, \frac{\Lambda_c IQ}{IV^2}\right). \tag{3.2}$$

In the lower panel of Figure 4, we plot the density functions of the feasible log-based t-statistics. Apparently, the coverage probabilities of Equation (3.2) are a much better guide for small values of n; with $n = 100$ providing a near perfect fit to the N(0,1). Hence, the results for RRV_m^Δ are consistent with the findings for RV^Δ.

This technique is also applicable to study other (differentiable) functions of RRV_m^Δ. For convenience, we will state the CLT of a particularly useful transformation, namely standard deviations/volatilities obtained by taking square roots:

$$\sqrt{n}\left(\sqrt{RRV_m^\Delta} - \sqrt{IV}\right) \xrightarrow{d} MN\left(0, \frac{\Lambda_c IQ}{4IV}\right). \tag{3.3}$$

4 Empirical Application: Alcoa Aluminium

We investigate the empirical properties of intraday ranges by analyzing an equity from the Dow Jones Industrial Average as of the reconfiguration April 8, 2004. Out of the thirty stocks composing the index, Alcoa, a relatively illiquid security trading under the ticker symbol AA, was selected at random.

High-frequency data were extracted from the TAQ database, which is a recording of trades and quotes from the securities listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and the National Association of Securities Dealers Automated Quotation (NASDAQ). The sample period covers January 2, 2001 through December 31, 2004; a total of 1,004 trading days. We restrict attention to NYSE updates and only report the results of the quotation data, for which the midquote is used. The analysis of transaction data is available upon request. All raw data were filtered for irregularities (e.g., a price of zero, entries posted outside the official NYSE opening hours, or quotes with negative spreads), and a second algorithm handled a few clusters of outliers found afterwards by visual inspection of the time series. Dividend payments were adjusted and no stock splits took place.

[INSERT TABLE 1 ABOUT HERE]

Summary statistics for the average number of data points remaining after filtering is given in Table 1. The column $\#\Delta p_{\tau_i} \neq 0$, where $\Delta p_{\tau_i} = p_{\tau_i} - p_{\tau_{i-1}}$ and τ_i is the arrival time of the ith tick, counts the number of price changes relative to the previous posting. $\#\Delta^2 p_{\tau_i} \neq 0$ does the same on second differences, but after having removed updates with $\Delta p_{\tau_i} = 0$. These numbers are of special importance to the realized range-based estimator for the purpose of calculating the $\lambda_{2,m}$ scalars, but also to construct confidence intervals. Initially, we found m on the basis of all nonzero increments; i.e. the $\#\Delta p_{\tau_i} \neq 0$ numbers within each intraday sampling

interval. This clearly meant m was too high, because of instantaneous reversals (e.g., bid-ask bounce behavior). We judged that a fair method of determining m was to only count repeated reversals once. Thus, to compute the normalizing constants and confidence bands, we took the $\#\Delta^2 p_{\tau_i} \neq 0$ numbers.

The estimation of RV^Δ and RRV_m^Δ proceeds with 5-minute sampling through the trading session starting 9:30a.m. EST until 4:00p.m. EST.; i.e. by setting $n = 78$ or $\Delta = 300$ seconds.[16] We use previous-tick method to compute returns for RV^Δ. This is not required for the intraday high-low statistic, where maximum is taken over all increments in the sampling interval. Notice that as the empirical high-frequency data are irregularly distributed, this means there are, in general, different values of m in the 5-minute intervals. This does not cause any problems, however, for the theory extends directly to this setting, provided we use the individual values of m in the estimation, which - after all - is the most natural thing to do.

[INSERT TABLE 2 ABOUT HERE]

Summary statistics of the resulting time series are printed in Table 2. RV^Δ achieves a lower minimum and higher maximum than RRV_m^Δ, while its overall mean is higher. Kurtosis figures are consistent with the mixed Gaussian limit theory, and the variance is lower for RRV_m^Δ, though somewhat less than predicted by the distribution theory. We expected this, as the data from the empirical price process are, in all likelihood, not drawn from a Brownian semimartingale (e.g., there are jumps and microstructure frictions). Our intraday high-low statistic, in turn, behaves differently for other specifications, which we will address elsewhere. Still, the variance of RRV_m^Δ is only 65% that of RV^Δ. The correlation between RV^Δ and RRV_m^Δ is 0.973, pointing towards little gain - at relevant frequencies - in taking linear combinations of the estimators to reduce sampling variation. In fact, from the joint distribution of RV^Δ and RRV_m^Δ, the conditional covariance matrix at time s is given by:

$$\Sigma_s = \sigma_s^4 \begin{pmatrix} 2 & \\ \frac{1}{\lambda_{2,m}}\mathrm{cov}\left(W_1^2, om_W^2\right) & \Lambda_m \end{pmatrix}. \tag{4.1}$$

The covariance term appearing in Σ_s is hard to tackle analytically. We used some simulations to inspect the structure of the correlation matrix around a grid of values for m that roughly matches our sample (unreported results). Based on this, we found that the empirical correlation corresponds closely with the theoretical level.

[INSERT FIGURE 5 ABOUT HERE]

---

[16]This choice was guided by signature plots, i.e. averages of the estimators across different sampling frequencies n. We found increasing signs of microstructure noise by moving beyond the 5-minute frequency.

In Figure 5, IV estimates are drawn for the two methods: RV^Δ and RRV_m^Δ. The time series agree on the size and direction-of-change for IV. The key point is that the sample path of the high-low statistic is less volatile compared to RV^Δ (but still quite erratic). Again, this suggests that the measurement errors of RV^Δ are larger compared to RRV_m^Δ, and that the theoretical gains of the realized range-based estimator also hold for the empirical identification of σ; at least for the 5-minute frequency.

[INSERT FIGURE 6 ABOUT HERE]

To underscore these insights, we extracted quote data for the subsample period January 2, 2002 - December 31, 2002 to plot in Figure 6 the IV estimates together with 95% confidence intervals, constructed from the log-based theory.[17] We see the widening of the confidence bands, when σ goes up. Nevertheless, the stability of the realized range-based estimator feeds into much smaller intervals, consistent with the theoretical relationship between the m and Λ_m scalars from Figure 3 that implies very few increments are required for RRV_m^Δ to gain a significant advantage in efficiency over RV^Δ.[18]

[INSERT FIGURE 7 ABOUT HERE]

These empirical findings transfer into more persistent time series behavior for the realized range-based estimator, as shown by the autocorrelation functions in Figure 7. We included the first 75 lags and report Bartlett's two standard error bands for testing a white noise null hypothesis (that is immensely rejected by the data). All autocorrelations are positive, starting at about 0.70 - 0.80 and ending around 0.10 - 0.15. The decay pattern of the series is nearly identical but evolves more smoothly and at higher levels for RRV_m^Δ. Combined, these observations might be put to work in a forecasting exercise, although we do not pursue this idea here.

All told, realized range-based estimation of IV offers several advantages compared to RV; both from a theoretical and practical viewpoint. But, as a final remark, we acknowledge that the probabilistic aspects delivered in this paper needs further refinement at higher frequencies, because microstructure frictions contaminate the data. It is central with statistical tools for controlling the impact of the noise and getting consistent estimates of IV. These techniques are already being developed for RV, see, e.g., Barndorff-Nielsen, Hansen, Lunde & Shephard (2004) or Aït-Sahalia et al. (2005). It presents a topic for future research to verify if our method extends along these lines, and we are currently devoting a separate paper for a formal analysis of the realized range-based estimator and market microstructure noise.

---

[17]This procedure is recommended by BN-S (2002*b*), who argue for using the intervals of the log-based theory - converted back to levels - on empirical data. In the process of exponentiation, these confidence bands become asymmetric and assign more probability of ending in the upper regions.

[18]With $\#\Delta^2 p_{\tau_i} \neq 0$ equal to 1,002 on average for the quote data, we have roughly $m = 13$ price increments within each of the 78 5-minute intervals during the trading day.

5   Conclusions and Directions for Future Research

The RRV estimator is an approach based on intraday price ranges for non-parametric measurement of QV of continuous semimartingales. Under weak regularity conditions, we have shown it extracts the latent volatility more accurately than previous methods. Another contribution of this paper, particularly useful for empirical analysis, is the solution to the problems with downward bias that has haunted the range-based literature for decades.

The finite sample distributions of the estimator were inspected with Monte Carlo analysis. For moderate samples, the coverage probabilities of the confidence bands for the t-statistics correspond with the limit theory, in particular for log-based inference.

We highlighted the empirical potential of realized range-based vis-à-vis return-based variances by applying our method to a set of 5-minute high-frequency data for Alcoa. Consistent with theory, the intraday range-based statistic has smaller confidence bands than RV. Although empirical price processes are very different from diffusion models and that real data are noisy objects, we feel the results support our theory quite well and opens up alternative practical routes for estimating QV.

In future projects, we envision several extensions of the current framework. First, there is plenty of evidence against the continuous sample path diffusion. We are convinced that an intraday high-low statistic can estimate QV, when the price exhibits jumps. This theory is being developed in another paper, along with realized range-based bipower variation. Second, with microstructure noise in observed asset prices, further comparisons of the high-low and RV are needed. Finally, we can handle the bivariate case with the polarization identities, so multivariate range-based analysis constitutes a promising future application.

A Appendix of Proofs

A.1 Proof of Theorem 1

First, we define:

$$
\begin{aligned}
\xi_i^n &= \frac{1}{\lambda_2}\sigma_{\frac{i-1}{n}}^2 s_{W_{i\Delta,\Delta}}^2, \\
U_n &= \sum_{i=1}^n \xi_i^n.
\end{aligned}
$$

Note that:

$$
\mathbb{E}\left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}}\right] = \frac{1}{n}\sigma_{\frac{i-1}{n}}^2,
$$

so,

$$
\sum_{i=1}^n \mathbb{E}\left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}}\right] \xrightarrow{p} IV. \tag{A.1}
$$

Now, by setting

$$
\eta_i^n = \xi_i^n - \mathbb{E}\left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}}\right],
$$

we get:

$$
\mathbb{E}\left[(\eta_i^n)^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] = \Lambda\frac{1}{n^2}\sigma_{\frac{i-1}{n}}^4.
$$

Therefore,

$$
\sum_{i=1}^n \mathbb{E}\left[(\eta_i^n)^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] \xrightarrow{p} 0.
$$

Hence, the assertion $U_n \xrightarrow{p} IV$ follows directly from (A.1). As a sufficient condition in the next step, we deduce $RRV^\Delta - U_n \xrightarrow{p} 0$. Note the equality:

$$
\begin{aligned}
RRV^\Delta - U_n &= \frac{1}{\lambda_2}\sum_{i=1}^n \left(s_{p_{i\Delta,\Delta}} - \sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\right)\left(s_{p_{i\Delta,\Delta}} + \sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\right) \\
&\equiv R_n^1 + R_n^2,
\end{aligned}
$$

with R_n^1 and R_n^2 defined by:

$$
\begin{aligned}
R_n^1 &= \frac{2}{\lambda_2}\sum_{i=1}^n \sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\left(s_{p_{i\Delta,\Delta}} - \sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\right) \\
R_n^2 &= \frac{1}{\lambda_2}\sum_{i=1}^n \left(s_{p_{i\Delta,\Delta}} - \sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\right)^2.
\end{aligned}
$$

We decompose the second term further:

$$
\begin{aligned}
R_n^2 &\leq \frac{1}{\lambda_2} \sum_{i=1}^{n} \left(\sup_{(i-1)/n \leq s, t \leq i/n} \left| \int_s^t \mu_u \mathrm{d}u + \int_s^t \left(\sigma_u - \sigma_{\frac{i-1}{n}} \right) \mathrm{d}W_u \right| \right)^2 \\
&\leq \frac{2}{\lambda_2} \sum_{i=1}^{n} \left(\sup_{(i-1)/n \leq s, t \leq i/n} \left| \int_s^t \mu_u \mathrm{d}u \right| \right)^2 + \frac{2}{\lambda_2} \sum_{i=1}^{n} \left(\sup_{(i-1)/n \leq s, t \leq i/n} \left| \int_s^t \left(\sigma_u - \sigma_{\frac{i-1}{n}} \right) \mathrm{d}W_u \right| \right)^2 \\
&\equiv R_n^{2.1} + R_n^{2.2}.
\end{aligned}
$$

It is straightforward to verify the estimation $\mathbb{E}\left[R_n^{2.1}\right] = \mathrm{O}\left(n^{-1}\right)$. For the latter term, we exploit the Burkholder inequality (e.g., Revuz & Yor (1998)):

$$
\begin{aligned}
\mathbb{E}\left[R_n^{2.2}\right] &\leq \frac{2C}{\lambda_2} \sum_{i=1}^{n} \mathbb{E}\left[\int_{\frac{i-1}{n}}^{\frac{i}{n}} \left(\sigma_u - \sigma_{\frac{i-1}{n}} \right)^2 \mathrm{d}u \right] \\
&= \frac{2C}{\lambda_2} \mathbb{E}\left[\int_0^1 \left(\sigma_u - \sigma_{\frac{[nu]}{n}} \right)^2 \mathrm{d}u \right] \\
&= \mathrm{o}(1),
\end{aligned}
$$

for some constant $C > 0$. Thus, $R_n^2 = \mathrm{o}_p(1)$. With a decomposition as above and the Cauchy-Schwarz inequality, we have $R_n^1 = \mathrm{o}_p(1)$. By assembling the parts, $RRV^\Delta - U_n \xrightarrow{p} 0$. ∎

A.2 Proof of Theorem 2

We need the following lemma from analysis.

Lemma 1 *Given two continuous functions $f, g : I \to \mathbb{R}$ on compact $I \subseteq \mathbb{R}^n$, assume t^* is the only point where the maximum of the function f on I is achieved. Then, it holds:*

$$
M_\epsilon(g) \equiv \frac{1}{\epsilon} \left[\sup_{t \in I} \{f(t) + \epsilon g(t)\} - \sup_{t \in I} \{f(t)\} \right] \to g(t^*) \quad as \quad \epsilon \downarrow 0.
$$

Proof

Construct the set

$$
\bar{G} = \left\{ h \in C(I) \mid h \text{ is constant on } B_\delta(t^*) \cap I \text{ for some } \delta > 0 \right\}.
$$

As usual, $C(I)$ is the set of continuous functions on I and $B_\delta(t^*)$ is an open ball of radius δ centered at t^*. Take $\bar{g} \in \bar{G}$ and recall $\bar{g}$ is bounded on I. Thus, for ϵ small enough:

$$
\begin{aligned}
\sup_{t \in I} \{f(t) + \epsilon \bar{g}(t)\} &= \max \left\{ \sup_{t \in I \cap B_\delta(t^*)} \{f(t) + \epsilon \bar{g}(t)\}, \sup_{t \in I \cap B_\delta^c(t^*)} \{f(t) + \epsilon \bar{g}(t)\} \right\} \\
&= \sup_{t \in I \cap B_\delta(t^*)} \{f(t) + \epsilon \bar{g}(t)\} \\
&= f(t^*) + \epsilon \bar{g}(t^*).
\end{aligned}
$$

So,

$$M_\epsilon(\bar{g}) \to \bar{g}(t^*),$$

$\forall \bar{g} \in \bar{G}$. Now, let $g \in C(I)$. As $\bar{G}$ is dense in $C(I)$, $\exists \bar{g} \in \bar{G} : \bar{g}(t^*) = g(t^*)$ and $||\bar{g} - g||_\infty < \epsilon'$ ($||\cdot||_\infty$ is the sup-norm). We see that $||M_\epsilon(\bar{g}) - M_\epsilon(g)||_\infty < \epsilon'$, and

$$||M_\epsilon(g) - g(t^*)||_\infty \leq ||M_\epsilon(\bar{g}) - \bar{g}(t^*)||_\infty + ||M_\epsilon(g) - M_\epsilon(\bar{g})||_\infty \to 0.$$

Thus, the assertion is established. ∎

With this lemma at hand, we proceed with a three-stage proof of Theorem 2. In the first part, a CLT is proved for the quantity

$$\bar{U}_n = \sqrt{n} \sum_{i=1}^n \eta_i^n.$$

The second step is to define a new sequence:

$$U_n' = \sqrt{n} \frac{1}{\lambda_2} \sum_{i=1}^n \left(s_{p_{i\Delta,\Delta}}^2 - \mathbb{E}\left[s_{p_{i\Delta,\Delta}}^2 \mid \mathcal{F}_{\frac{i-1}{n}} \right] \right),$$

and show the result

$$U_n' - \bar{U}_n \xrightarrow{p} 0.$$

The interested reader may note that Assumptions **(M)** and **(V)** are not needed for Part I and II. Finally, in Part III, the theorem follows from:

$$\sqrt{n} \sum_{i=1}^n \left(\frac{1}{\lambda_2} \mathbb{E}\left[s_{p_{i\Delta,\Delta}}^2 \mid \mathcal{F}_{\frac{i-1}{n}} \right] - \mathbb{E}\left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}} \right] \right) \xrightarrow{p} 0, \text{ and}$$

$$\sqrt{n} \left(\sum_{i=1}^n \mathbb{E}\left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}} \right] - IV \right) \xrightarrow{p} 0.$$

Proof of Part I

Notice that:

$$n \sum_{i=1}^n \mathbb{E}\left[(\eta_i^n)^2 \mid \mathcal{F}_{\frac{i-1}{n}} \right] \xrightarrow{p} \Lambda IQ,$$

and by the scaling property of Brownian motion,

$$\sqrt{n} \sum_{i=1}^n \mathbb{E}\left[\eta_i^n \left(W_{\frac{i}{n}} - W_{\frac{i-1}{n}} \right) \mid \mathcal{F}_{\frac{i-1}{n}} \right] \xrightarrow{p} \frac{\nu}{\lambda_2} IV,$$

where $\nu = \mathbb{E}\left[W_1 s_W^2 \right]$. Quite trivially, $\{W_t\}_{t \in [0,1]} \overset{d}{=} \{-W_t\}_{t \in [0,1]}$, with the consequence $\nu = -\nu$ and, hence, $\nu = 0$.

Next, let $N = \{N_t\}_{t \in [0,1]}$ be a bounded martingale on $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0,1]}, P \right)$, which is orthogonal to W (i.e., with quadratic covariation $[W, N]_t = 0$). Then,

$$\sqrt{n} \sum_{i=1}^n \mathbb{E}\left[\eta_i^n \left(N_{\frac{i}{n}} - N_{\frac{i-1}{n}} \right) \mid \mathcal{F}_{\frac{i-1}{n}} \right] = 0. \tag{A.2}$$

For this result, we use Clark's Representation Theorem (see, e.g., Karatzas & Shreve (1998, Appendix E)):

$$s_{W_{i\Delta,\Delta}}^2 - \frac{1}{n}\lambda_2 = \int_{\frac{i-1}{n}}^{\frac{i}{n}} H_s^n \mathrm{d}W_s, \tag{A.3}$$

for some predictable function H_s^n. Notice $\mathbb{E}\left[\int_a^b f_s \mathrm{d}W_s \left(N_b - N_a\right) \mid \mathcal{F}_a\right] = 0$, for any $[a,b]$ and predictable f. To prove this assertion, take a partition $a = t_0^* < t_1^* < \ldots < t_n^* = b$ and compute:

$$\mathbb{E}\left[\sum_{i=1}^n f_{t_{i-1}^*}\left(W_{t_i^*} - W_{t_{i-1}^*}\right)\left(N_b - N_a\right) \mid \mathcal{F}_a\right] = \mathbb{E}\left[\sum_{i=1}^n f_{t_{i-1}^*}\left(W_{t_i^*} - W_{t_{i-1}^*}\right) N_b \mid \mathcal{F}_a\right]$$

$$= \mathbb{E}\left[\sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[f_{t_{i-1}^*}\left(W_{t_i^*} - W_{t_{i-1}^*}\right) N_b \mid \mathcal{F}_{t_i^*}\right] \mid \mathcal{F}_{t_{i-1}^*}\right] \mid \mathcal{F}_a\right]$$

$$= 0.$$

From Equation (A.3), (A.2) is attained. Finally, stable convergence in law follows by Theorem IX 7.28 in Jacod & Shiryaev (2002):

$$\sqrt{n}\left(RRV^\Delta - IV\right) \xrightarrow{d_s} \sqrt{\Lambda} \int_0^1 \sigma_s^2 \mathrm{d}B_s.$$

$\blacksquare$

Proof of Part II

We begin by setting

$$\zeta_i^n = \sqrt{n}\left(\frac{1}{\lambda_2} s_{p_{i\Delta,\Delta}}^2 - \xi_i^n\right),$$

and obtain the identity:

$$U_n' - \bar{U}_n = \sum_{i=1}^n \left(\zeta_i^n - \mathbb{E}\left[(\zeta_i^n)^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right]\right).$$

To complete the second step, it suffices that

$$\sum_{i=1}^n \mathbb{E}\left[(\zeta_i^n)^2\right] \to 0.$$

We can show this result with the same methods applied to the estimates of R_n^1 and R_n^2 in the proof of Theorem 1. $\blacksquare$

Proof of Part III

It holds that:

$$\sqrt{n}\left(\sum_{i=1}^n \mathbb{E}\left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}}\right] - IV\right) = \sqrt{n}\sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}}\left(\sigma_{\frac{i-1}{n}}^2 - \sigma_s^2\right)\mathrm{d}s.$$

Exploiting the results of Barndorff-Nielsen et al. (2005), we find, under Assumption V_2,

$$\sqrt{n}\left(\sum_{i=1}^{n}\mathbb{E}\left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}}\right] - IV\right) \xrightarrow{p} 0.$$

Now, we prove the first convergence of Part III stated above. After some tedious computations - identical to the methods in Theorem 1 - we get that, using V_2,

$$\sqrt{n}\sum_{i=1}^{n}\left(\frac{1}{\lambda_2}\mathbb{E}\left[s_{p_{i\Delta,\Delta}}^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] - \mathbb{E}\left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}}\right]\right)$$

$$= 2\sqrt{n}\frac{1}{\lambda_2}\sum_{i=1}^{n}\mathbb{E}\left[\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\left(s_{p_{i\Delta,\Delta}-\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}}\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right] + \mathrm{o}_p(1)$$

$$= 2\sqrt{n}\frac{1}{\lambda_2}\sum_{i=1}^{n}\mathbb{E}\left[\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\left(\sup_{s,t\in\left[\frac{i-1}{n},\frac{i}{n}\right]}\left(\sigma_{\frac{i-1}{n}}(W_t - W_s) + \int_s^t \mu_u\mathrm{d}u + \int_s^t\left(\sigma_u - \sigma_{\frac{i-1}{n}}\right)\mathrm{d}W_u\right)\right.\right.$$

$$\left.\left. -\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right] + \mathrm{o}_p(1).$$

By appealing to Assumption V_2 again, we get the decomposition:

$$\sqrt{n}\sum_{i=1}^{n}\left(\frac{1}{\lambda_2}\mathbb{E}\left[s_{p_{i\Delta,\Delta}}^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] - \mathbb{E}\left[\xi_i^n \mid \mathcal{F}_{\frac{i-1}{n}}\right]\right) = V_n^1 + V_n^2 + \mathrm{o}_p(1),$$

with the random variables V_n^1 and V_n^2 defined by

$$\begin{aligned}
V_n^1 &= 2\frac{1}{\lambda_2}\sum_{i=1}^{n}\mathbb{E}\left[\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\left\{\sup_{s,t\in\left[\frac{i-1}{n},\frac{i}{n}\right]}\left(\sqrt{n}\,\sigma_{\frac{i-1}{n}}(W_t - W_s) + \sqrt{n}\int_s^t \mu_{\frac{i-1}{n}}\mathrm{d}u\right.\right.\right. \\
&+ \left.\left.\left.\sqrt{n}\int_s^t\left\{\sigma'_{\frac{i-1}{n}}\left(W_u - W_{\frac{i-1}{n}}\right) + v_{\frac{i-1}{n}}\left(B'_u - B'_{\frac{i-1}{n}}\right)\right\}\mathrm{d}W_u - \sqrt{n}\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\right\} \mid \mathcal{F}_{\frac{i-1}{n}}\right],
\end{aligned}$$

and

$$\begin{aligned}
V_n^2 &= 2\sqrt{n}\frac{1}{\lambda_2}\sum_{i=1}^{n}\mathbb{E}\left[\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\left\{\sup_{s,t\in\left[\frac{i-1}{n},\frac{i}{n}\right]}\left(\sigma_{\frac{i-1}{n}}(W_t - W_s) + \int_s^t \mu_u\mathrm{d}u + \int_s^t\left(\sigma_u - \sigma_{\frac{i-1}{n}}\right)\mathrm{d}W_u\right)\right.\right. \\
&- \left.\left. \sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\right\} \mid \mathcal{F}_{\frac{i-1}{n}}\right] - V_n^1 \\
&\leq 2\sqrt{n}\frac{1}{\lambda_2}\sum_{i=1}^{n}\mathbb{E}\left[\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\left\{\sup_{s,t\in\left[\frac{i-1}{n},\frac{i}{n}\right]}\left(\int_s^t\left(\mu_u - \mu_{\frac{i-1}{n}}\right)\mathrm{d}u + \int_s^t\left\{\int_{\frac{i-1}{n}}^u \mu'_r\mathrm{d}r\right.\right.\right.\right. \\
&+ \left.\left.\left.\left. \int_{\frac{i-1}{n}}^u\left(\sigma'_r - \sigma'_{\frac{i-1}{n}}\right)\mathrm{d}W_r + \int_{\frac{i-1}{n}}^u\left(v_r - v_{\frac{i-1}{n}}\right)\mathrm{d}B'_r\right\}\mathrm{d}W_u\right)\right\} \mid \mathcal{F}_{\frac{i-1}{n}}\right].
\end{aligned}$$

From the Cauchy-Schwarz and Burkholder inequalities, we find that

$$V_n^2 = \mathrm{o}_p(1).$$

At this point, we invoke the Lemma 1 by setting:

$$
\begin{aligned}
f_{in}(s,t) &= \sqrt{n}\,\sigma_{\frac{i-1}{n}}(W_t - W_s), \\
g_{in}(s,t) &= n\int_s^t \mu_{\frac{i-1}{n}}\mathrm{d}u + n\int_s^t \left\{\sigma'_{\frac{i-1}{n}}\left(W_u - W_{\frac{i-1}{n}}\right) + v_{\frac{i-1}{n}}\left(B'_u - B'_{\frac{i-1}{n}}\right)\right\}\mathrm{d}W_u \\
&= \mu_{\frac{i-1}{n}}g_{in}^1(s,t) + \sigma'_{\frac{i-1}{n}}g_{in}^2(s,t) + v_{\frac{i-1}{n}}g_{in}^3(s,t).
\end{aligned}
$$

Note that $\epsilon = 1/\sqrt{n}$. Through Assumption V_1, we get the following identity:

$$
\begin{aligned}
(t_{in}^*(W), s_{in}^*(W)) &= \underset{s,t\in\left[\frac{i-1}{n},\frac{i}{n}\right]}{\arg\sup}\; f_{in}(s,t) \\
&= \underset{s,t\in\left[\frac{i-1}{n},\frac{i}{n}\right]}{\arg\sup}\; \sqrt{n}(W_t - W_s) \\
&\overset{d}{=} \underset{s,t\in[0,1]}{\arg\sup}\,(W_t - W_s).
\end{aligned}
$$

A standard result then states that the points $t_{in}^*(W)$ and $s_{in}^*(W)$ are unique, almost surely, so the lemma applies. Hence, by imitating the proof of the lemma, we get the decomposition:

$$
V_n^1 = 2\frac{1}{\lambda_2}\sum_{i=1}^n \mathbb{E}\left[\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\left(\frac{1}{\sqrt{n}}g_{in}(t_{in}^*(W), s_{in}^*(W)) + R_{in}\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right],
$$

where the term R_{in} satisfies:

$$
\mathbb{E}\left[(R_{in})^2\right] = \mathrm{o}\left(n^{-1}\right),
$$

(uniformly in i). By the Cauchy-Schwarz inequality, we have the estimation:

$$
2\frac{1}{\lambda_2}\sum_{i=1}^n \mathbb{E}\left[\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}R_{in} \mid \mathcal{F}_{\frac{i-1}{n}}\right] = \mathrm{o}_p(1).
$$

As $g_{in}^1(s,t)$, $g_{in}^2(s,t)$ and $g_{in}^3(s,t)$ are independent of $\mathcal{F}_{\frac{i-1}{n}}$, we obtain:

$$
\begin{aligned}
\mathbb{E}\left[\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\frac{1}{\sqrt{n}}g_{in}^1(t_{in}^*(W), s_{in}^*(W)) \mid \mathcal{F}_{\frac{i-1}{n}}\right] &\equiv \frac{1}{\sqrt{n}}\sigma_{\frac{i-1}{n}}\mu_{\frac{i-1}{n}}\nu_1 \\
\mathbb{E}\left[\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\frac{1}{\sqrt{n}}g_{in}^2(t_{in}^*(W), s_{in}^*(W)) \mid \mathcal{F}_{\frac{i-1}{n}}\right] &\equiv \frac{1}{\sqrt{n}}\sigma_{\frac{i-1}{n}}\sigma'_{\frac{i-1}{n}}\nu_2 \\
\mathbb{E}\left[\sigma_{\frac{i-1}{n}}s_{W_{i\Delta,\Delta}}\frac{1}{\sqrt{n}}g_{in}^3(t_{in}^*(W), s_{in}^*(W)) \mid \mathcal{F}_{\frac{i-1}{n}}\right] &\equiv \frac{1}{\sqrt{n}}\sigma_{\frac{i-1}{n}}v_{\frac{i-1}{n}}\nu_3,
\end{aligned}
$$

with

$$
\nu_k = \mathbb{E}\left[s_{W_{i\Delta,\Delta}}g_{in}^k(t_{in}^*(W), s_{in}^*(W))\right], \quad \text{for } k = 1, 2, \text{ and } 3.
$$

Note that,

$$
(t_{in}^*(W), s_{in}^*(W)) = (s_{in}^*(-W), t_{in}^*(-W)). \tag{A.4}
$$

Using (A.4) and the relationship $(W, B) \overset{d}{=} (-W, -B)$, it follows that $\nu_k = -\nu_k$ and, hence, $\nu_k = 0$ for $k = 1, 2$, and 3. This yields the estimation:

$$
V_n^1 = \mathrm{o}_p(1),
$$

and the proof is complete. ∎

A.3   Proof of Theorem 3

The result is shown in the same manner as the proofs of Theorem 1 and 2.   ■

References

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2005), 'How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise', *Review of Financial Studies* **18**(2), 351–416.

Aldous, D. J. & Eagleson, G. K. (1978), 'On Mixing and Stability of Limit Theorems', *Annals of Probability* **6**(2), 325–331.

Alizadeh, S., Brandt, M. W. & Diebold, F. X. (2002), 'Range-Based Estimation of Stochastic Volatility Models', *Journal of Finance* **57**(3), 1047–1092.

Andersen, T. G., Benzoni, L. & Lund, J. (2002), 'An Empirical Investigation of Continuous-Time Equity Return Models', *Journal of Finance* **57**(4), 1239–1284.

Andersen, T. G. & Bollerslev, T. (1997*a*), 'Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High-Frequency Returns', *Journal of Finance* **57**(3), 975–1005.

Andersen, T. G. & Bollerslev, T. (1997*b*), 'Intraday Periodicity and Volatility Persistence in Financial Markets', *Journal of Empirical Finance* **4**(2), 115–158.

Andersen, T. G. & Bollerslev, T. (1998), 'Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts', *International Economic Review* **39**(4), 885–905.

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2002), Parametric and Nonparametric Volatility Measurement, *in* L. P. Hansen & Y. Ait-Sahalia, eds, 'Handbook of Financial Econometrics', North-Holland. Forthcoming.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The Distribution of Realized Stock Return Volatility', *Journal of Financial Economics* **61**(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The Distribution of Realized Exchange Rate Volatility', *Journal of the American Statistical Association* **96**(453), 42–55.

Bandi, F. M. & Russell, J. R. (2004), Microstructure Noise, Realized Variance, and Optimal Sampling, Working Paper, Graduate School of Business, University of Chicago.

Bandi, F. M. & Russell, J. R. (2005), 'Separating Microstructure Noise from Volatility', *Journal of Financial Economics* (Forthcoming).

Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J., Podolskij, M. & Shephard, N. (2005), A Central Limit Theorem for Realized Power and Bipower Variations of Continuous Semimartingales, *in* Y. Kabanov & R. Lipster, eds, 'From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev', Springer-Verlag. Forthcoming.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2004), Regular and Modified Kernel-Based Estimators of Integrated Variance: The Case with Independent Noise, Working Paper, Nuffield College, University of Oxford.

Barndorff-Nielsen, O. E. & Shephard, N. (2002*a*), 'Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models', *Journal of the Royal Statistical Society: Series B* **64**(2), 253–280.

Barndorff-Nielsen, O. E. & Shephard, N. (2002*b*), 'Estimating Quadratic Variation using Realized Variance', *Journal of Applied Econometrics* **17**(5), 457–477.

Barndorff-Nielsen, O. E. & Shephard, N. (2003), 'Realized Power Variation and Stochastic Volatility', *Bernoulli* **9**, 243–265.

Barndorff-Nielsen, O. E. & Shephard, N. (2004), 'Power and Bipower Variation with Stochastic Volatility and Jumps', *Journal of Financial Econometrics* **2**(1), 1–48.

Barndorff-Nielsen, O. E. & Shephard, N. (2005*a*), How Accurate is the Asymptotic Approximation to the Distribution of Realized Volatility, *in* D. W. F. Andrews, J. L. Powell, P. A. Ruud & J. H. Stock, eds, 'Identification and Inference for Econometrics Models', Cambrigde University Press, pp. 306–331.

Barndorff-Nielsen, O. E. & Shephard, N. (2005*b*), 'Power Variation and Time Change', *Theory of Probability and Its Applications* (Forthcoming).

Barndorff-Nielsen, O. E. & Shephard, N. (2005*c*), Variation, Jumps, Market Frictions and High Frequency Data in Financial Econometrics, Working Paper, Nuffield College, University of Oxford.

Black, F. & Scholes, M. (1973), 'The Pricing of Options and Corporate Liabilities', *Journal of Political Economy* **81**(3), 637–654.

Bollen, B. & Inder, B. (2002), 'Estimating Daily Volatility in Financial Markets Utilizing Intraday Data', *Journal of Empirical Finance* **9**(5), 551–562.

Bollerslev, T., Engle, R. F. & Nelson, D. B. (1994), ARCH Models, *in* R. F. Engle & D. McFadden, eds, 'Handbook of Econometrics: Volume IV', North-Holland, pp. 2959–3038.

Brown, S. (1990), Estimating Volatility, *in* S. Figlewski, W. Silber & M. Subrahmanyam, eds, 'Financial Options', Business One Irwin.

Brunetti, C. & Lildholdt, P. M. (2002), Return-Based and Range-Based (Co)Variance Estimation - With an Application to Foreign Exchange Markets, Working Paper, University of Pennsylvania.

Doob, J. L. (1953), *Stochastic Processes*, 1 edn, John Wiley and Sons.

Doornik, J. A. (2002), *Object-Oriented Matrix Programming Using Ox*, 3 edn, Timberlake Consultants Press.

Feller, W. (1951), 'The Asymptotic Distribution of the Range of Sums of Independent Random Variables', *Annals of Mathematical Statistics* **22**(3), 427–432.

Gallant, A. R., Hsu, C.-T. & Tauchen, G. E. (1999), 'Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance', *Review of Economics and Statistics* **81**(4), 617–631.

Garman, M. B. & Klass, M. J. (1980), 'On the Estimation of Security Price Volatilities from Historical Data', *Journal of Business* **53**(1), 67–78.

Ghysels, E., Harvey, A. C. & Renault, E. (1996), Stochastic Volatility, *in* G. S. Maddala & C. R. Rao, eds, 'Handbook of Statistics: Volume 14', North-Holland, pp. 119–191.

Hansen, P. R. & Lunde, A. (2005), 'A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data', *Journal of Financial Econometrics* (Forthcoming).

Hansen, P. R. & Lunde, A. (2006), 'Realized Variance and Market Microstructure Noise', *Journal of Business and Economic Statistics* (Forthcoming).

Jacod, J. (1997), On Continuous Conditional Gaussian Martingales and Stable Convergence in Law, Seminaire de Probabilities XXXI, 232-246.

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

Karatzas, I. & Shreve, S. E. (1998), *Methods of Mathematical Finance*, 1 edn, Springer-Verlag.

Kunitomo, N. (1992), 'Improving the Parkinson Method of Estimating Security Price Volatilities', *Journal of Business* **64**(2), 295–302.

Mandelbrot, B. B. (1963), 'The Variation of Certain Speculative Prices', *Journal of Business* **36**(4), 394–419.

Oomen, R. C. A. (2005), Properties of Realized Variance under Alternative Sampling Schemes, Working Paper, Warwick Business School.

Parkinson, M. (1980), 'The Extreme Value Method for Estimating the Variance of the Rate of Return', *Journal of Business* **53**(1), 61–65.

Protter, P. (2004), *Stochastic Integration and Differential Equations*, 1 edn, Springer-Verlag.

Rényi, A. (1963), 'On Stable Sequences of Events', *Sankhya: The Indian Journal of Statistics; Series A* **25**(3), 293–302.

Revuz, D. & Yor, M. (1998), *Continuous Martingales and Brownian Motion*, 3 edn, Springer-Verlag.

Rogers, L. C. G. & Satchell, S. E. (1991), 'Estimating Variances from High, Low, and Closing Prices', *Annals of Applied Probability* **1**(4), 504–512.

Rosenberg, B. (1972), The Behavior of Random Variables with Nonstationary Variance and the Distribution of Security Prices, Working Paper, University of California, Berkeley.

Wasserfallen, W. & Zimmermann, H. (1985), 'The Behavior of Intraday Exchange Rates', *Journal of Banking and Finance* **9**(1), 55–72.

Zhou, B. (1996), 'High-Frequency Data and Volatility in Foreign-Exchange Rates', *Journal of Business and Economic Statistics* **14**(1), 45–52.

Table 1: Data Cleaning, Tick Data with Price Effects and without Instantaneous Reversals

| Company	No. Tick Data Pr. Trading Day					
	Trades			Quotes		
	All	$\#\Delta p_{\tau_i} \neq 0$	$\#\Delta^2 p_{\tau_i} \neq 0$	All	$\#\Delta p_{\tau_i} \neq 0$	$\#\Delta^2 p_{\tau_i} \neq 0$
Alcoa	2285	966	549	5393	1370	1002

The table gives information about the filtering of Alcoa high-frequency data. All numbers are averages. $\#\Delta p_{\tau_i} \neq 0$ is the daily amount of tick data left, after counting out price repetitions in consecutive ticks. $\#\Delta^2 p_{\tau_i} \neq 0$ also removes instantaneous price reversals.

Table 2: Summary Statistics for Realized Estimators of Integrated Variance

	Mean	Var.	Skew.	Kurt.	Min.	Max.	Correlation	
							RV^{Δ}	RRV_m^{Δ}
RV^{Δ}	0.093	0.533	2.686	12.656	0.011	0.600	1.000	
RRV_m^{Δ}	0.082	0.349	2.306	9.592	0.013	0.421	0.973	1.000

We report summary statistics for the annualized realized return- and range-based estimators for Alcoa during the sample period covering January 2, 2001 up to December 31, 2004. The table prints the mean, variance, skewness, kurtosis, mininum and maximum of the various time series, plus their correlation. Variance figures are multiplied by 100.

Figure 1: The Distribution of the Absolute Return and Range
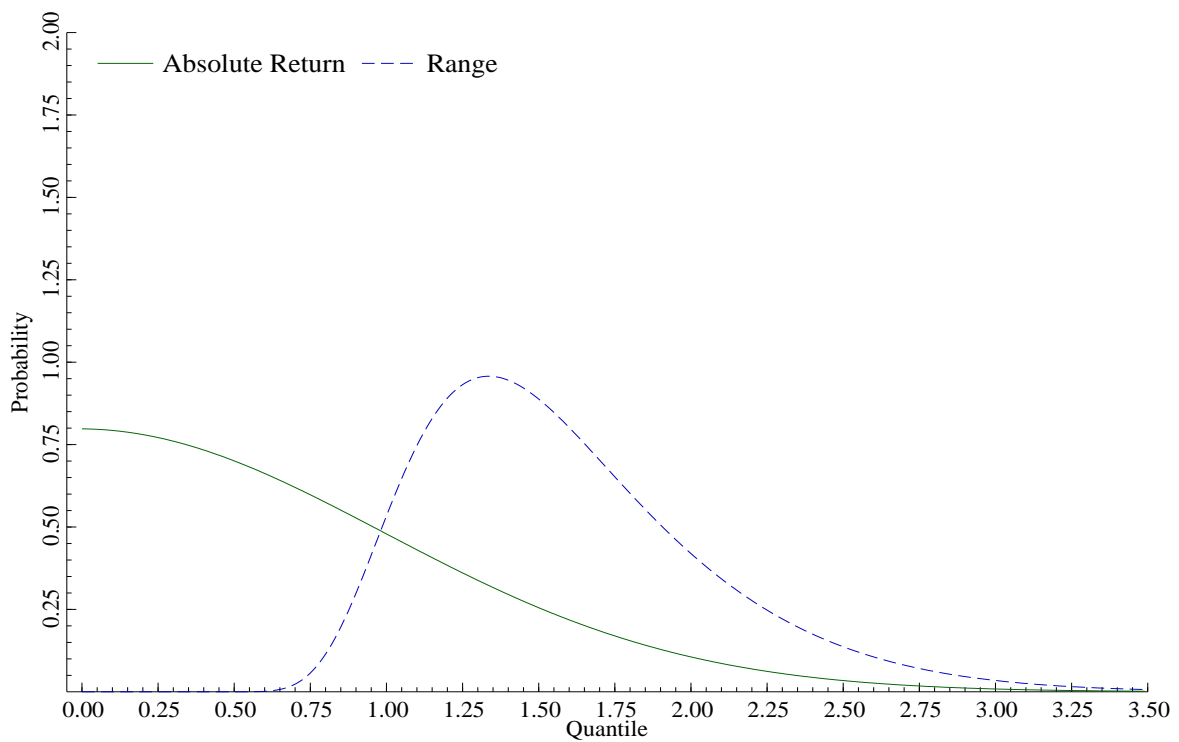
We plot the distribution of the absolute return and range of a standard Brownian motion over an interval of unit length.

Figure 2: The Finite Sample Expectation of the Squared Range
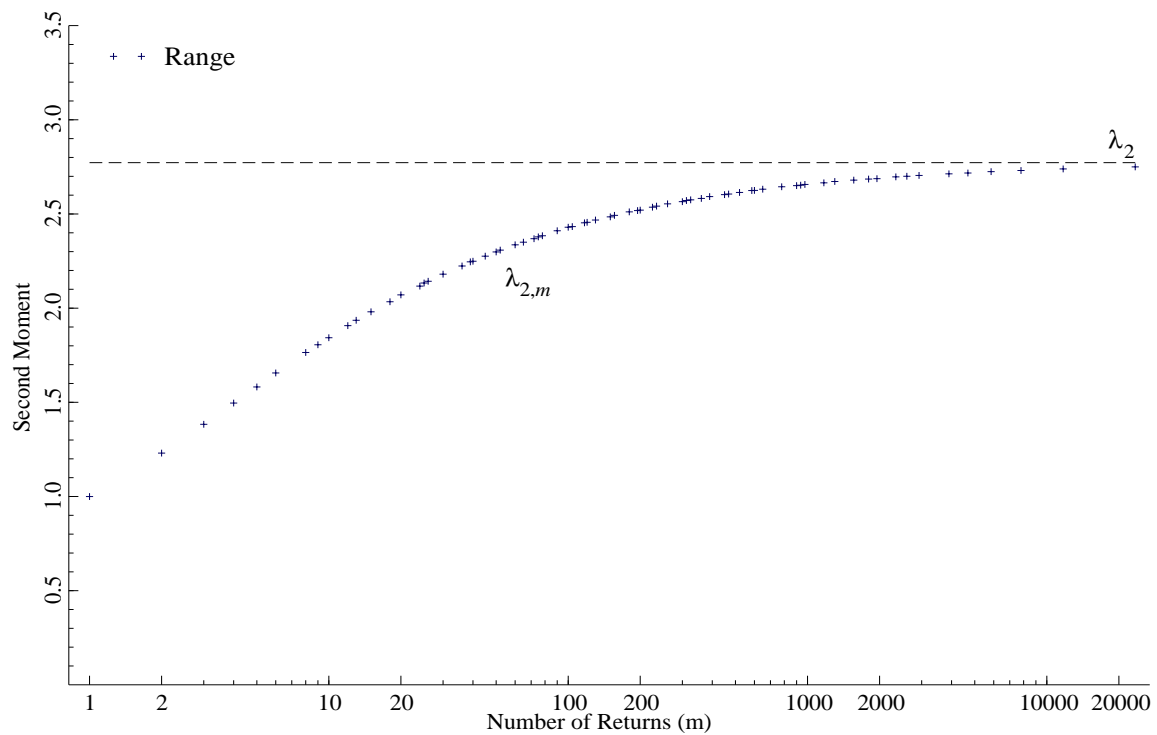
This figure details the second moment of the range of a standard Brownian motion over an interval of unit length, when increments to the underlying continuous time process is only observed at m equidistant points in time. The dashed line is the asymptotic value, and the figure is based on a simulation with 1,000,000 repetitions.

Figure 3: The Variance Factor of the Realized Range-Based Estimator
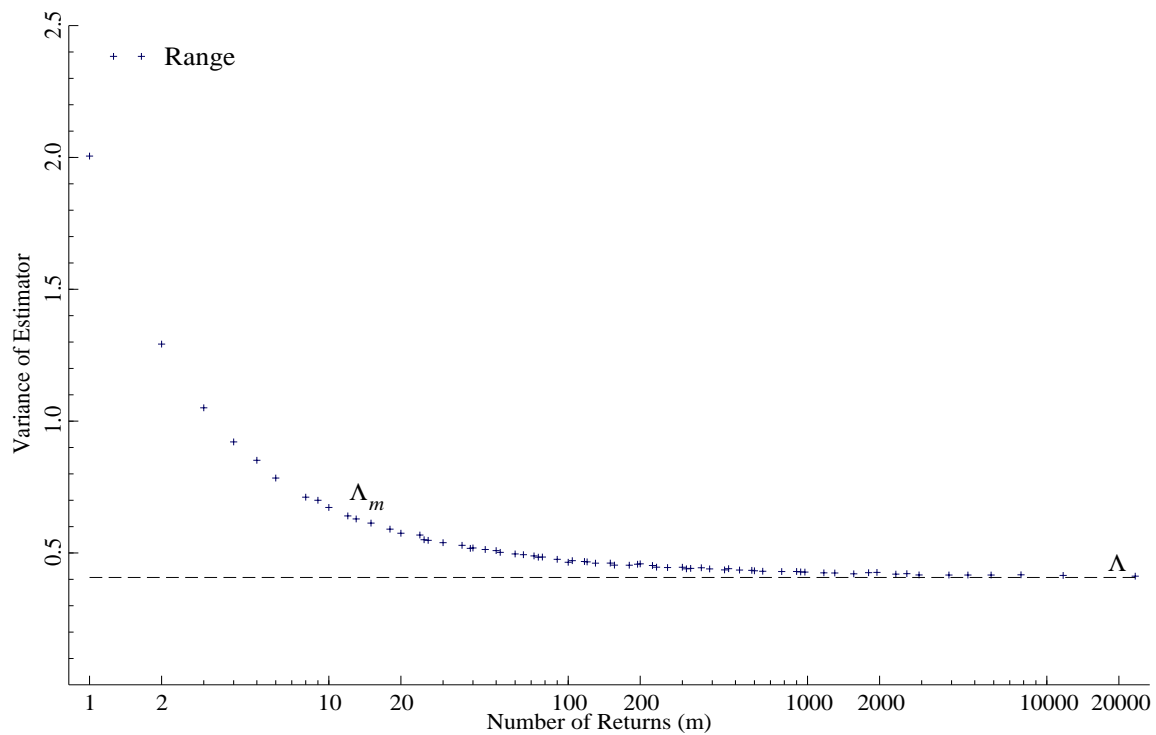
We plot the variance of the realized range-based estimator of integrated variance of a standard Brownian motion over an interval of unit length, when increments to the underlying continuous time process is only observed at m equidistant points in time. The dashed line is the asymptotic value, and the figure is based on a simulation with 1,000,000 repetitions.

Figure 4: Asymptotic Normality for the Standardized Realized Range-Based Statistic in Estimating Integrated Variance
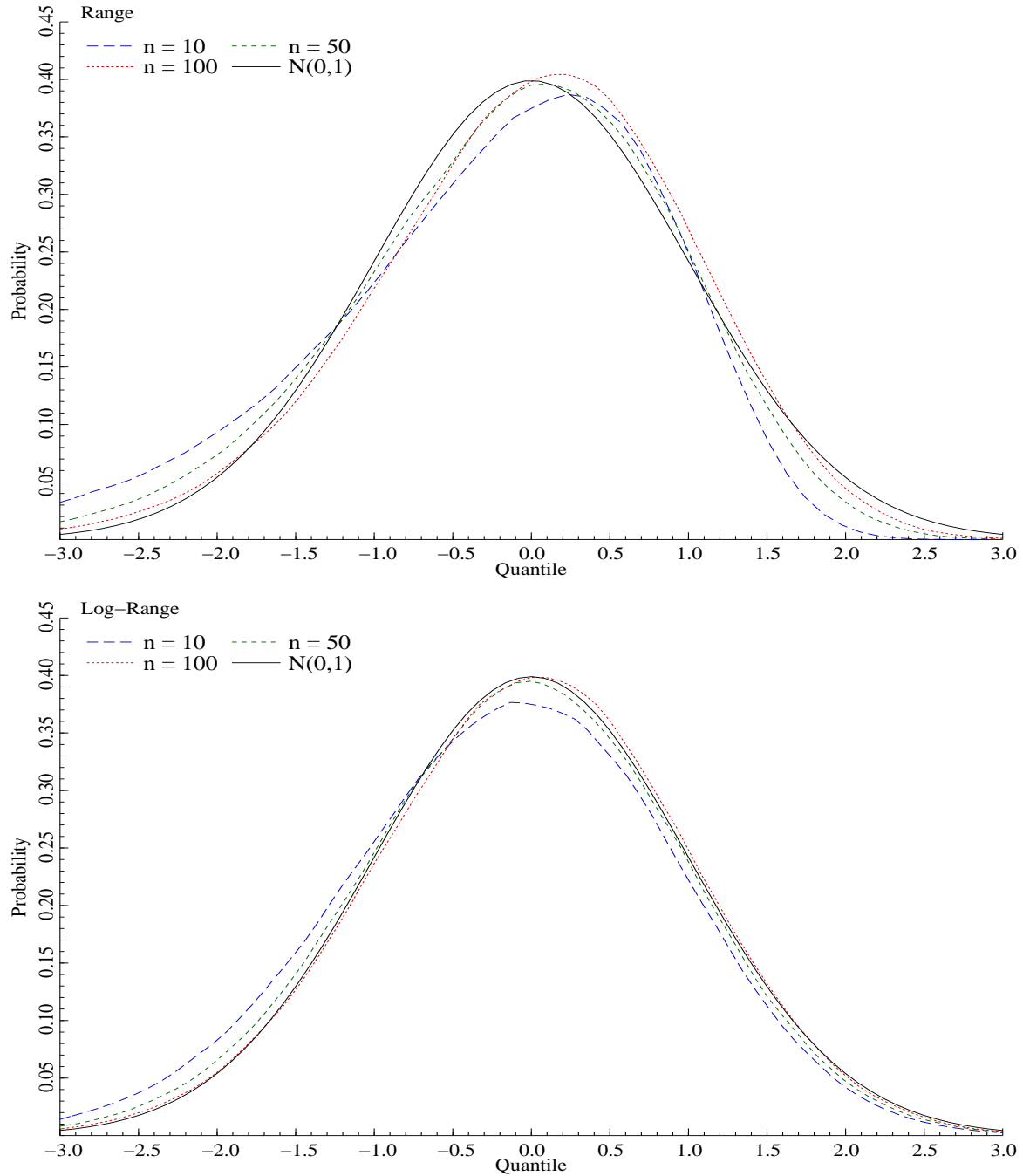

The figure shows kernel densities of the measurement errors for the studentized RRV_m^Δ in estimating IV. We show the finite sample settings $n = 10, 50, 100$ and $m = 10$. All plots are based on a simulation with 1,000,000 repetitions from a log-normal diffusion for σ, as detailed in the main text. The upper panel depicts t-statistics of the feasible CLT for RRV_m^Δ, while the lower panel is the corresponding log-based version. The solid line is the N(0,1) density.

Figure 5: Estimates of Integrated Variance for Alcoa
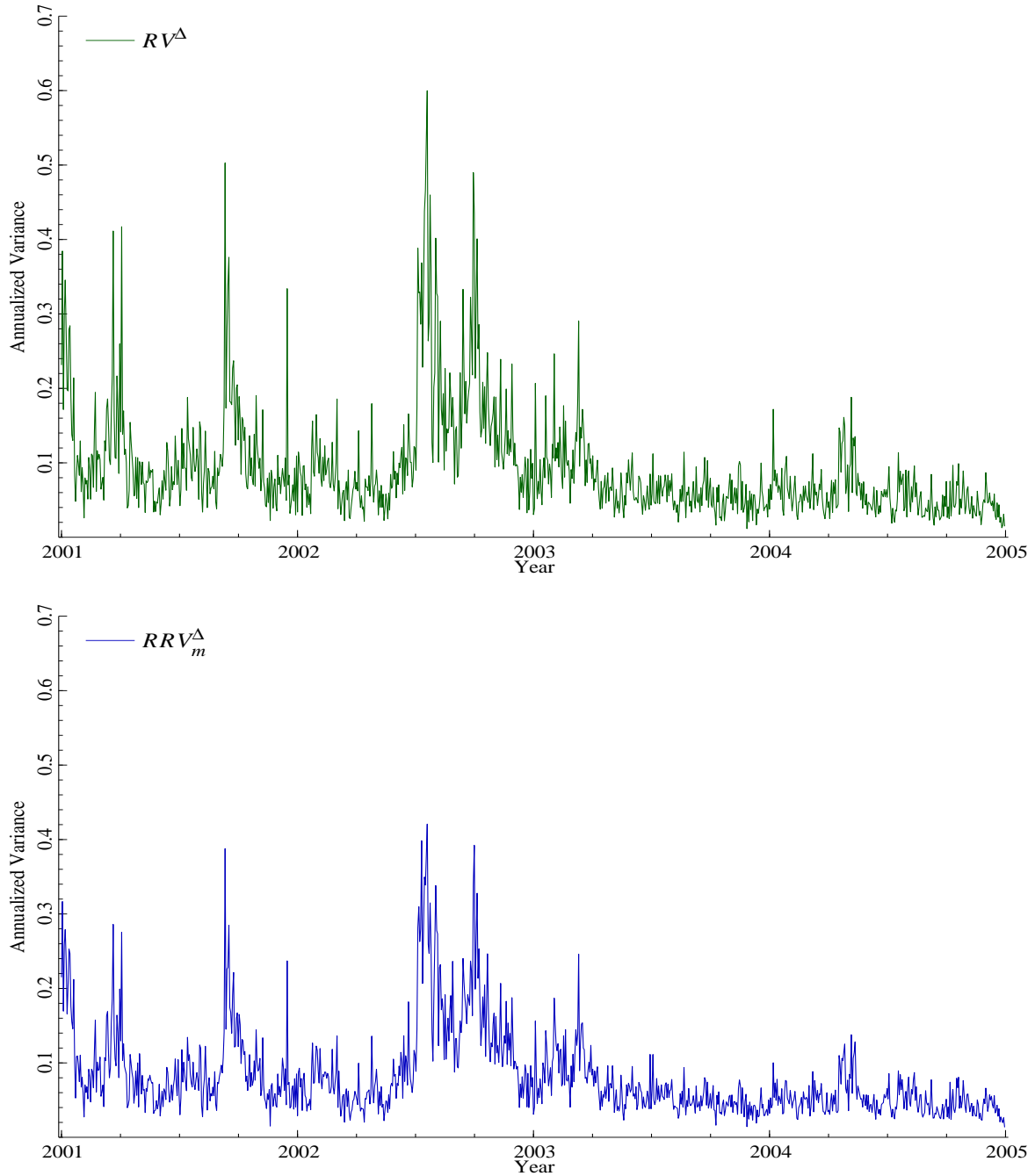
Time series for the daily *IV* estimates for Alcoa are plotted for the sample period January 2, 2001 to December 31, 2004. All series are constructed from 5-minute midquote returns or ranges for quotation data extracted from the TAQ database.

Figure 6: Confidence Intervals for Integrated Variance
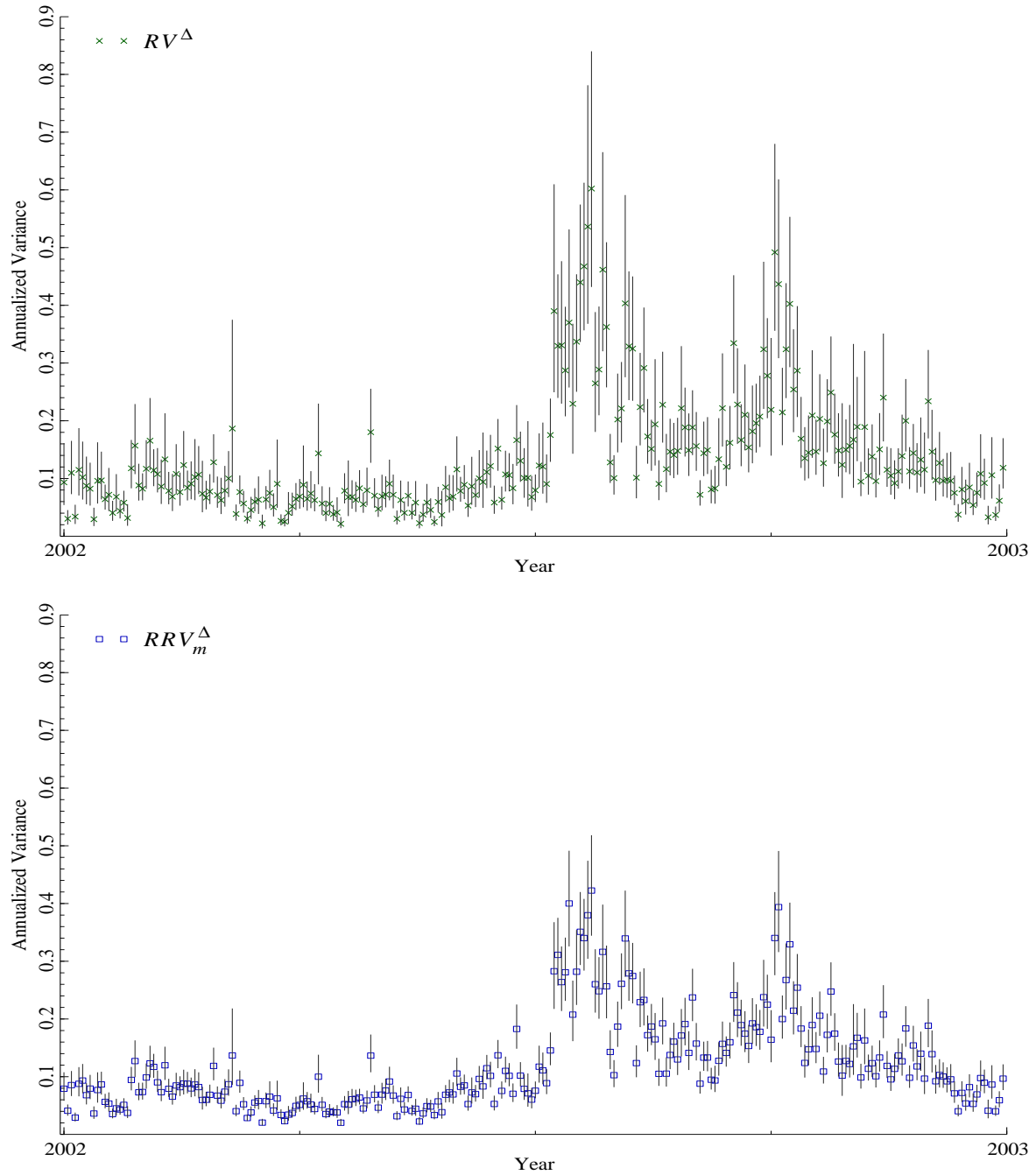
Daily *IV* estimates for Alcoa are drawn together with 95% confidence intervals for the subsample period January 2, 2002 through December 31, 2002. The confidence bands are constructed from the log-based limit theory and converted back to variances. The upper (lower) dashed line is the 97,5% (2,5%) level.

Figure 7: Autocorrelation Functions for Estimators of Integrated Variance
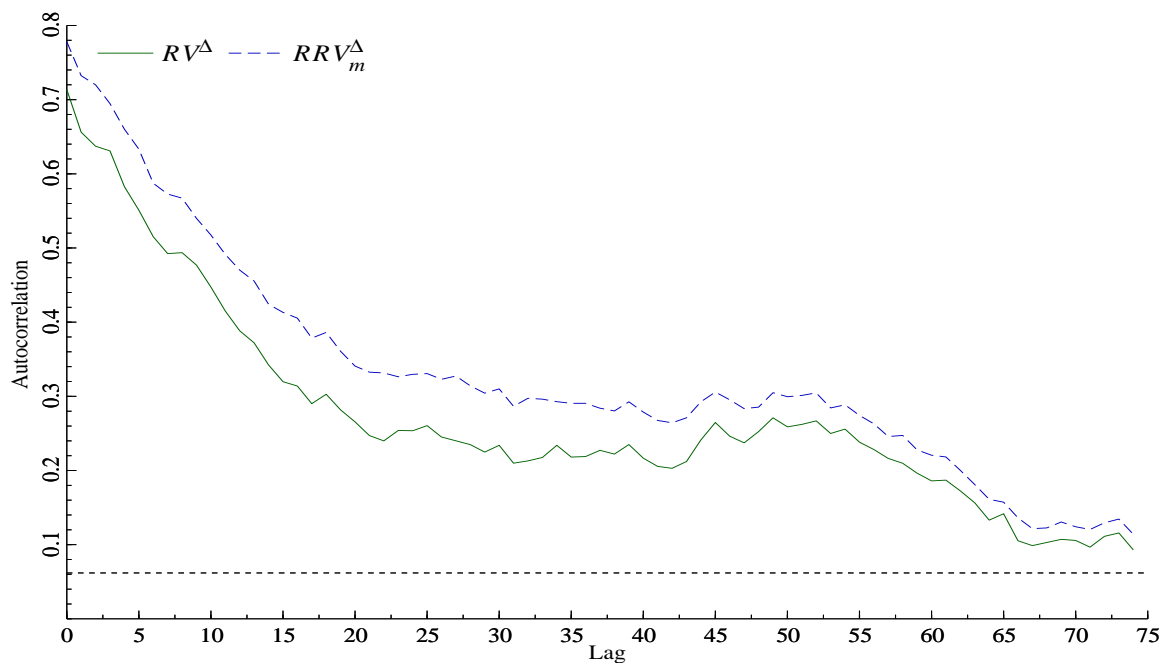
We plot the autocorrelation functions of RV^Δ and RRV_m^Δ constructed from 5-minute midquote returns or ranges for Alcoa through the sample period January 2, 2001 to December 31, 2004; or 1,004 trading days in total. The first 75 lags are shown and the dashed vertical line is Bartlett's standard errors for testing a white noise hypothesis.