CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Finite Sample Results of Range-Based Integrated Volatility Estimation

Eduardo Rossi and Filippo Spazzini

CEA@Cass Working Paper Series

WP–CEA–04-2009

Finite sample results of Range-based integrated volatility estimation *

Eduardo Rossi †
University of Pavia, Italy

Filippo Spazzini ‡
University of Milano, Italy

First draft: June 5, 2009
This version: October 28, 2009

Abstract

In this paper we consider the finite-sample properties of Realized Range estimators of integrated volatility and we compare them to those of the Realized Volatility estimators when a sample of high-frequency data is observed. Simulated data are obtained from different generating mechanisms for the instantaneous volatility process, e.g. Ornstein-Uhlenbeck, long memory and jump processes. We analyze the impact that missing observations have on the Realized Range measures and we propose a simple correction in order to reduce the bias. We also evaluate the robustness of the different approaches considered when high-frequency prices are affected by bid-ask bounce and price discreteness. Simulation results confirm that realized range corrected for irregular sampling has lower bias while not increasing the estimator variance. The simulations also show how the degree of persistence in the estimated Integrated Variance series crucially depends on the sampling frequency adopted in the estimation and thus on the precision of the estimators. A brief empirical application with high-frequency IBM data is also included.

*****Preliminary version**. Part of this work was completed while the first author was visiting the CEA@CASS, whose hospitality he gratefully acknowledges.

†Dipartimento Economia politica e metodi quantitativi. Via San Felice 5, 27100 Pavia, Italy. Tel.: +39 0382/986207 Fax: +39 0382/304226. E-mail: erossi@eco.unipv.

‡Dipartimento Scienze economiche, aziendali e statistiche. Via Conservatorio 7, 20122 Milano, Italy. E-mail: filippo.spazzini@gmail.com

1 Introduction

One feature of financial returns volatility is that actual realizations are not directly observable. A common approach to deal with the latency of return volatility is to adopt parametric models for the volatility, e.g., (G)ARCH or stochastic volatility models. Alternately one can exploit option pricing models to extract from option prices the "implied volatility" over a fixed future horizon. Such procedures remain model-dependent and incorporate a potentially time-varying volatility risk-premium in the measure so they do not provide unbiased forecasts of the volatility of the underlying asset. High-frequency financial prices have recently become available for a host of different instruments and markets, allowing the use of realized variation, or realized volatility, (RV hereafter) in volatility measurement which does not rely on any specific parametric characterization of the volatility process. RV, which is a sum of squared intraday returns, yields a perfect estimate of the Integrated Variance (or integrated volatility, IV hereafter) in the ideal situation where prices are observed continuously and without measurement error. This result suggests that the RV should be based on intraday returns sampled at the highest possible frequency (tick-by-tick data), e.g., Andersen, Bollerslev, and Diebold (2002), Andersen, Bollerslev, Diebold, and Labys (2003), Barndorff-Nielsen and Shephard (2002). See for an updated survey Andersen and Benzoni (2009). However, the presence of market microstructure noise in high-frequency financial data complicates the estimation of financial volatility and makes RV unreliable. The theoretical validity of RV estimation procedures hinges on the observability of the true price process. When the sample frequency increases, the market microstructure dynamics generate a divergence between the observed price and the true or *noiseless* price process, whose quadratic variation is the object of interest. This divergence can be due to transaction price changes occurring as multiples of ticks (price discreteness, see e.g. Harris, (1990, 1991)) or to the properties of the trading mechanism (as in Black (1976), Amihud and Mendelson (1987)). Bandi and Russell (2008) provide a general treatment of the effect of market microstructure noise on realized variance estimates. In particular, they show that the realized variance estimates are asymptotically swamped by noise as the number of squared return data increases over a fixed time period. They also propose a methodology to optimally choose the sampling frequency as the number of squared return data increases over a fixed time period. A simple solution to this problem often adopted in empirical work is to sample at moderate frequencies, e.g., every 5−, 10−, or 30− minutes. In this way the benefit of more frequent sampling is traded off against the damage caused by cumulating noise.

Several bias correction procedures have been put forward, see e.g., Zhou (1996), Aït-Sahalia, Mikland, and Zhang (2005), Bandi and Russell (2008), Bandi and Russell (2006), Zhang, Mikland, and Aït-Sahalia (2005) and Hansen and Lunde (2006); for an update survey on this topic see McAleer and Medeiros (2008). In general, as highlighted in Hansen and Lunde (2006), the best remedy for market microstructure noise depends on the properties of the noise. Alternatively some filtering techniques were used, e.g. Andersen, Bollerslev, Diebold, and Ebens (2001) and Maheu and McCurdy (2002) adopted a moving average filter, while Bollen and Inden (2002) used an autoregressive filter. See Nielsen and Frederiksen (2008) for a finite sample analysis of high-frequency IV estimators in the presence of microstructure noise.

Meanwhile, a rivaling approach to the estimation of Integrated Volatility, based on the aggregation of intra-daily ranges, the Realized Range (RRG hereafter), has been proposed by Martens and van Dijk (2007), Christensen and Podolskij (2007) and Christensen, Podolskij, and Vetter (2009). The Realized Range estimator looks particularly appealing when microstructure noise prevents from the use of the whole record of high-frequency prices to compute Realized Volatility, since it exploits a larger amount of information and therefore it is in principle able to attain a higher precision.

The aim of this paper is to examine the finite sample properties of several versions of the Realized

Range estimator by means of a fairly extended Monte Carlo simulation and compare them to those of some standard IV estimators stemming from the Realized Volatility literature. We analyze the robustness of these estimators to long memory and jumps in the instantaneous volatility, and to bid-ask bounce, discreteness and missing observations in the log-price process. In particular, we adopt Ornstein-Uhlenbeck processes, long memory processes (Comte and Renault (1996), Comte and Renault (1998)) and jump processes (Andersen, Benzoni, and Lund (2002), Eraker, Johannes, and Polson (2003), Eraker (2004)) for the instantaneous volatility. We also model the bid-ask bounce, the price discreteness and the missing observations in the spirit of Nielsen and Frederiksen (2008).

The Realized Range estimators turn out to be strongly downward biased under all Data Generating Processes when we consider missing observations leading to irregular sampling. In order to deal with this problem we introduce an heuristically justified correction which is able to strongly reduce this bias without affecting the estimators' precision.

With bid-ask bounce the bias-corrected Realized Range estimators are steadily downward biased across all processes and all parameter configurations considered, but the bias is unaffected by the simulation setup. However, the bias-corrected RRGs are three times more efficient than the Hansen and Lunde estimator (Hansen and Lunde (2006)).

We devote some attention to the estimation of the long memory parameter of time series of daily IV estimates obtained from the competing estimators. Our analysis shows how the degree of persistence in the estimated IV series crucially, but not surprisingly, depends on the sampling frequency adopted in the IV estimation and thus on the precision of the estimators. These results shed some shadows on the common practice of drawing inferences on the long memory features of the instantaneous volatility process based on long memory estimates obtained from estimated IV series.

IBM tick-by-tick transaction prices are used to empirically illustrate the behavior of the competing estimators. Volatility signature plots suggest how the microstructure noise significantly affects the non-bias corrected estimators for sampling frequencies higher than 15 minutes, while the bias-corrected versions seem reliable starting from the 5 minutes frequency. We propose a long-memory signature plot to highlight the effect of the sampling frequency chosen in the IV estimation on the estimated long memory parameter. This shows how the long memory estimates based on Realized Volatility and Hansen and Lunde estimators are importantly affected by the sampling frequency, while the estimates based on Realized Ranges are more stable across the sampling frequencies.

The remainder of the paper is organized as follows. Section 2 and 3 introduce the Realized Volatility and Realized Range estimators used in the simulations. Section 4 describes the Monte Carlo setup while Section 5 illustrates the simulation results. Section 6 presents our empirical illustration with IBM data while Section 7 concludes the paper.

2 Realized volatility

The objective is to estimate a suitable measure of the return variation over the trading day which is normalized to be $[0, 1]$. Suppose the log-price of an asset, $p(t)$, follows a stochastic volatility model (SV):

$$\mathrm{d}p(t) = \mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W(t), \quad t \geq 0 \tag{1}$$

or

$$p(t) = p(0) + \int_0^t \mu(u)\mathrm{d}u + \int_0^t \sigma(u)\mathrm{d}W(u), \quad t \geq 0 \tag{2}$$

where the mean process $\mu(t)$ is locally bounded and predictable, $\sigma(t)$ is assumed to be independent of Standard Brownian motion $W(\cdot)$ and càdlàg. This set up is quite common in the literature on

RV , e.g., Barndorff-Nielsen and Shephard (2007), Andersen, Bollerslev, and Diebold (2002).

Consider an equidistant partition $0 = t_0 < t_1 < \ldots < t_n = 1$, where $t_i = i/n$, and $\Delta = 1/n$. Adopting the notation of Hansen and Lunde (2005), the RV at sampling frequency n is

$$RV^{\Delta} = \sum_{i=1}^{n} r^2_{i\Delta,\Delta}. \tag{3}$$

From the theory of stochastic integration, when $n \to \infty$ the quadratic variation:

$$\langle p \rangle = p \lim_{n\to\infty} \sum_{i=1}^{n} r^2_{i\Delta,\Delta}$$

coincides with the IV:

$$IV = \int_0^1 \sigma(u)^2 \mathrm{d}u \tag{4}$$

which is the object of interest. Then $RV^{\Delta} \xrightarrow{p} IV$. Barndorff-Nielsen and Shephard (2002) derived a distribution theory for RV^{Δ}.

When microstructure noise is present the RV is both biased and inconsistent, see, e.g. Zhou (1996), Hansen and Lunde (2006), Bandi and Russell (2006, 2008).

Suppose the price is contaminated by microstructure noise, and the noise is modeled as i.i.d. sequence of random variables with mean zero and finite variance ω^2,

$$\widetilde{p}_t = p_t + \eta_t \tag{5}$$

with η independent of p_t. This noise introduces spurious volatility and a negative serial correlation in $\mathrm{d}\widetilde{p}_t$. Hansen and Lunde (2006) (Lemma 2) show that, when $E(|\eta_t|^4 < \infty \, \forall t$, then $E(RV^{\Delta}) = IV + 2n\omega^2$. In this case independent market microstructure noise leads to a bias that diverges to infinity. Hansen and Lunde (2006) consider the following estimator:

$$RV^{\Delta}_{HL} = \sum_{i=1}^{n} r^2_{i\Delta,\Delta} + 2\frac{M}{M-1} \sum_{i=2}^{M} r_{i\Delta,\Delta} r_{(i-1)\Delta,\Delta}. \tag{6}$$

which is robust to the first order serial correlation in returns. This estimator (HL hereafter) is adopted to estimate IV in simulations (see Section 5) when microstructure noise is present.

3 Range estimation of integrated volatility

3.1 Parkinson estimator

Realized range can be considered the development of the daily range estimator proposed by Parkinson (1980). Thus before illustrating the RRG we briefly outline the Parkinson's method. Parkinson assumes that the log-price follows a scaled Brownian motion

$$\mathrm{d}p(t) = \sigma \mathrm{d}W(t) \tag{7}$$

the daily range can be defined as

$$s_p = \sup_{0\le t,s \le 1} \{p_t - p_s\}^2 .$$

Parkinson (1980) derived, using the Feller's results, the moment generating function of the range of a scaled Brownian motion ($p(t) = \sigma W(t)$). The r-th moment of s_p is:

$$E[s_p^r] = \lambda_r \sigma^r \quad r \geq 1,$$

where

$$\lambda_r = E[s_W^r] = E[\sup\{p_1 - p_0\}^r] = \frac{4}{\sqrt{\pi}} \left(1 - \frac{4}{2^r}\right) 2^{\frac{r}{2}} \Gamma\left(\frac{r+1}{2}\right) \zeta(r-1) \quad r \geq 1$$

$\Gamma(\cdot)$ and $\zeta(\cdot)$ represent the Gamma and zeta functions, respectively.

So, under the assumption of a fully observed continuous time log-price path, as an estimator of the daily integrated volatility Parkinson proposed:

$$RG_P = \frac{1}{\lambda^2 \sigma^2} s_p^2 = \frac{s_p^2}{4 \log(2)}. \tag{8}$$

The RG_P estimator is particularly interesting because simply requires the observation of the maximum and the minimum price over the trading day. We include it in our experiment since it can be considered as a simple benchmark for RV and RRG.

3.2 Realized range estimator

We assume that we have tick-by-tick data at hand, namely p is assumed fully observed. Consider the equidistant partition $0 = t_0 < t_1 < \ldots < t_n = 1$ where $t_i = i/n$ and $\Delta = 1/n$ for $i = 1, \ldots, n$. The intraday range at sampling times t_{i-1} and t_i ($i = 1, 2, \ldots, n$) is

$$s_{p,\Delta_i} = \sup_{t_{i-1} \leq t, s \leq t_i} \{p_t - p_s\}. \tag{9}$$

The RRG estimator for the interval $[0, 1]$ is defined as:

$$RRG^\Delta = \frac{1}{\lambda_2} \sum_{i=1}^{n} s_{p_{i\Delta},\Delta}^2 \tag{10}$$

Under the assumption that the log-price process is (7) as in Parkinson (1980), and provided that $E\left[s_{p\Delta,\Delta}^2\right] = \frac{\lambda_2}{n}\sigma^2$, where $\lambda_2 = 4\log(2)$:

$$E[RRG^\Delta] = \sigma^2 \quad Var[RRG^\Delta] = \Lambda n^{-1} \sigma^4$$

with

$$\Lambda = \frac{\lambda_4 - \lambda_2^2}{\lambda_2} \simeq 0.4073.$$

As $n \to \infty$, $RRG^\Delta \xrightarrow{p} \sigma^2$ and by a CLT

$$\sqrt{n}(RRG^\Delta - \sigma^2) \xrightarrow{d} N(0, \Lambda\sigma^4).$$

If $\mu(t)$ and $\sigma(t)$ are stochastic, Christensen and Podolskij (2007) show that

$$RRG^\Delta \xrightarrow{p} IV.$$

The result is obtained for very general continuous time processes, including models with leverage, long-memory, diurnal effects or jumps (in $\sigma(t)$).

In order to prove a CLT, Christensen and Podolskij (2007) assume that the stochastic volatility process satisfies:

$$\sigma(t) = \sigma(0) + \int_0^t \varpi(u)\mathrm{d}u + \int_0^t \nu(u)\mathrm{d}W_u + \int_0^t v(u)\mathrm{d}B_u \quad t \geq 0 \tag{11}$$

where $\varpi(t)_{t\geq 0}$, $\nu(t)_{t\geq 0}$, and $v(t)_{t\geq 0}$ are cadlag. $B(t)$ is a Brownian Motion independent of $W(t)$. They establish that RRG^{Δ} converges in law to a mixed normal with σ governing the mixture, i.e.:

$$\sqrt{n}(RRG^{\Delta} - IV) \xrightarrow{d} MN(0, \Lambda IQ)$$

with $IQ = \int_0^1 \sigma(u)^4 \mathrm{d}u$. For realized range Λ is approximately 0.4 while for realized volatility is 2. Hence the sampling error of RRG^{Δ} are about one-fifth of those based on realized volatility. RRG^{Δ} uses all the data, whereas realized volatility is based on high-frequency returns sampled at fixed points in time. It follows that

$$\frac{\sqrt{n}(RRG^{\Delta} - IV)}{\sqrt{\Lambda RRQ^{\Delta}}} \xrightarrow{d} N(0,1)$$

where $RRQ^{\Delta} = \frac{n}{\lambda_4} \sum_{i=1}^{n} s^4_{p_{i\Delta},\Delta}$ consistently estimates the *Integrated Quarticity*:

$$IQ = \int_0^1 \sigma(u)^4 du.$$

3.3 Realized-range estimator with discretely sampled data

When the inference about IV is based on a finite sample the intraday high-low statistic will be progressively more downward biased as n gets larger, since the number of prices in each Δ decreases. The source of bias is λ_2, which is constructed on the presumption that p is fully observed. Christensen and Podolskij (2007) develop an estimator that accounts for the number of high-frequency data used in forming the high-low, in order to scale properly. Christensen and Podolskij (2007) assume that $mn+1$ equally spaced observations of the price are available, giving mn returns. Thus there are n intervals each with m returns. The log-price for each time in the interval $(0,1)$ is denoted as $p_{\frac{i-1}{n}+\frac{t}{mn}}$, where $i = 1\ldots, n$ and $t = 0, \ldots, m$. The observed range over the i-th interval is defined as

$$s_{p_{i\Delta},\Delta,m} = \max_{0\leq s,t\leq m} \left\{ p_{\frac{i-1}{n}+\frac{t}{mn}} - p_{\frac{i-1}{n}+\frac{s}{mn}} \right\}. \tag{12}$$

Moreover

$$s_{W,m} = \max_{0\leq s,t\leq m} \left\{ W_{t/m} - W_{s/m} \right\}$$

and

$$\lambda_{r,m} = E[s^r_{W,m}].$$

$\lambda_{r,m}$ is the r-th moment of the range of a standard Brownian motion over a unit interval when we observe only m increments of the underlying continuous time process. The value of $\lambda_{r,m}$ is obtained through numerical simulation of a standard Brownian motion observed m times over the unit interval and $\lambda_{2,m} \to \lambda_2$ as $m \to \infty$. The realized-range estimator based on discrete observations is

$$RRG^{\Delta}_m = \frac{1}{\lambda_{2,m}} \sum_{i=1}^{n} s^2_{p_{i\Delta},\Delta,m}. \tag{13}$$

RRG_m^Δ is a consistent estimator of IV as $n \to \infty$. Moreover if we assume that the stochastic volatility process is of the kind in (11) and $m \to c \in \mathbb{N} \cup \infty$:

$$\frac{\sqrt{n}(RRG_m^\Delta - IV)}{\sqrt{\Lambda_m RRQ_m^\Delta}} \xrightarrow{d} N(0,1)$$

with $\Lambda_m = \frac{\lambda_{4,m} - \lambda_{2,m}^2}{\lambda_{2,m}^2}$, and

$$RRQ_m^\Delta = \frac{n}{\lambda_{4,m}} \sum_{i=1}^{n} s^4_{p_{i\Delta,\Delta,m}}.$$

3.4 Realized-Range estimation in presence of microstructure noise

When the price is contaminated by microstructure noise as in (5), and the noise is modeled as i.i.d. sequence of random variables with mean zero and finite variance ω^2, Christensen, Podolskij, and Vetter (2009) show that the estimator of the integrated variance is

$$RRG_{m,BC}^\Delta = \frac{1}{\widetilde{\lambda}_{2,m}} \sum_{i=1}^{n} (s_{\widetilde{p}_{i\Delta,\Delta,m}} - 2\widehat{\omega}_N)^2 \tag{14}$$

where

$$\widetilde{\lambda}_{r,m} = E\left[\left| \max_{t:\eta_{\frac{t}{m}}=\omega, s:\eta_{\frac{s}{m}}=-\omega} \left(W_{\frac{t}{m}} - W_{\frac{s}{m}} \right) \right|^r \right]. \tag{15}$$

The variance of the noise process ω^2 can be consistently estimated with

$$\widehat{\omega}_N^2 = \frac{RV^N}{2N}, \tag{16}$$

where $N = nm$, i.e. the total number of log-returns, and

$$N^{1/2}\left(\widehat{\omega}_N^2 - \omega^2\right) \xrightarrow{d} \mathcal{N}(0, \omega^4).$$

However, when $\mu = 0$ in eq. (1), that is the log-price process has no drift, $\widehat{\omega}_N^2$ is biased. Oomen (2005) suggested an alternative unbiased estimator based on the negative of first order autocovariance of returns:

$$\widetilde{\omega}_N^2 = -\frac{1}{N-1} \sum_{i=1}^{N-1} r_{i\Delta,\Delta} r_{(i+1)\Delta,\Delta} \tag{17}$$

and

$$N^{1/2}\left(\widetilde{\omega}_N^2 - \omega^2\right) \xrightarrow{d} \mathcal{N}(0, 5\omega^4).$$

In the implementation of $RRG_{m,BC}$ we adopt $\widetilde{\omega}_N^2$ to estimate, on a day by day basis, the variance of the noise except when it is negative, in this case we switch to $\widehat{\omega}_N^2$.

When the distribution of η_t is supposed to be

$$P(\eta_t) = \frac{1}{2}(\delta_\omega - \delta_{-\omega})$$

where δ is the Dirac measure and ω is a positive constant, Christensen, Podolskij, and Vetter (2009) show that, as $m, n \to \infty$

$$RRG_{m,BC}^\Delta \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u.$$

and they provide the limiting distribution.

3.5 Realized-range estimator with irregularly sampled data

Before any sampling scheme is used all financial prices appear to be irregularly spaced, i.e. transactions do not take place at regular intervals. We have two different time scales, namely, transaction time where prices are sampled with every transaction and tick time where prices are sampled with every price change. However they both can be transformed in evenly spaced series simply sampling at fixed intervals, using the previous tick method (Wasserfallen and Zimmermann (1985)). Oomen (2005, 2006) points out that it is not only the choice of sampling frequency but also the choice of sampling scheme which has an important impact on the properties of RV, both in the absence and in the presence of noise. Griffin and Oomen (2008) suggest that there can be a considerable benefit of sampling in tick time. However, because the noise turns out to be highly dependent in tick time, any bias correction to RV that is based on *i.i.d.* noise assumption can fail. The definition of RV imposes no particular requirement on the way in which prices are sampled as long as the corresponding returns are nonoverlapping and span the interval of interest. Transaction time sampling where prices are recorded with each transaction are considered in Barndorff-Nielsen and Shephard (2004), Frijns and Lehnert (2004), and Hansen and Lunde (2006). While Aït-Sahalia, Mikland, and Zhang (2005), Corsi, Zumbach, Muller, and Dacorogna (2001), and Zhou (1996) adopt tick time sampling.

Christensen and Podolskij (2007) develop the realized-range estimator assuming discretely and evenly sampled observations. In this paper we analyze the effect that microstructure noise and irregularly spaced observations have on the bias and the variance of realized-range estimators. To this end we need to modify the realized range estimator in order to take into account the irregularly spaced case. In particular, an important element in the estimator implementation is the normalizing constant $\lambda_{r,m}$ which is computed by simulation, assuming to observe m increments of the underlying standard Brownian motion over the interval $(0, \Delta)$. So, to compute the $\lambda_{r,m}$ we simulate the Brownian motion as:

$$W_i = W_{i-1} + \frac{1}{\sqrt{m}}\epsilon_i, \quad i = 1, \ldots, m \tag{18}$$

where $\epsilon_i \sim i.i.d.N(0,1)$.
When irregular sampling occurs the realized-range estimator is implemented by dividing each subinterval range by the appropriate λ_{2,m_i}. This is computed by simulating the Brownian motion, e.g. the (18) is simulated with m_i steps, assuming that the discrete realizations of the price are fully observed while they are not.

Irregular sampling implies that some of the prices are missed. This in turn produces a necessarily downward bias because the observed range is always smaller or equal than the actual underlying range.

There is a possibility of reducing this bias by modifying the normalizing factor λ_{2,m_i}.

1. To compute a new $\lambda_{r,m_{RS,i}}$ we need to make an assumption on the frequency of discrete sampling from which the observed prices have been drawn. For instance, if on average we observe prices every five seconds (see for example IBM transactions data in Section 6) it can be realistic to assume a frequency of 1 second for the underlying discrete process.

2. Let us define M being the maximum number of price increments observable in an interval of length Δ. This implies that if we let Δ to increase, M gets larger.

3. In this case, a convenient strategy (as we show in Section 5), is to replace λ_{r,m_i} with $\lambda_{r,m_{RS,i}}$ which is the rth moment of the range of a standard Brownian motion with diffusion coefficient equal to $\sqrt{M}$ where only $m_{RS,i}$ randomly uneven increments of the underlying continuous

time process out of M steps are observed, with $m_{RS,i} \leq M$:

$$\lambda_{r,m_{RS,i}} = E[s^r_{W,m_{RS,i}}].$$

This quantities can be simulated, using (18), over S trajectories, where only $m_{RS,i}$ out of M increments are observed in each trajectory.

When the sampling is irregular, we modify the realized range estimator by using a different normalizing constant $\lambda_{2,m_{RS,i}}$ for each intraday range:

$$RRG^{\Delta}_{m_{RS}} = \sum_{i=1}^{n} \frac{s^2_{p_{i\Delta,\Delta},m}}{\lambda_{2,m_{RS,i}}}. \tag{19}$$

The plots in Figure 1 report the ratios between $\lambda_{2,m_{RS,i}}$ and λ_{2,m_i} for 4 different choice of Δ.

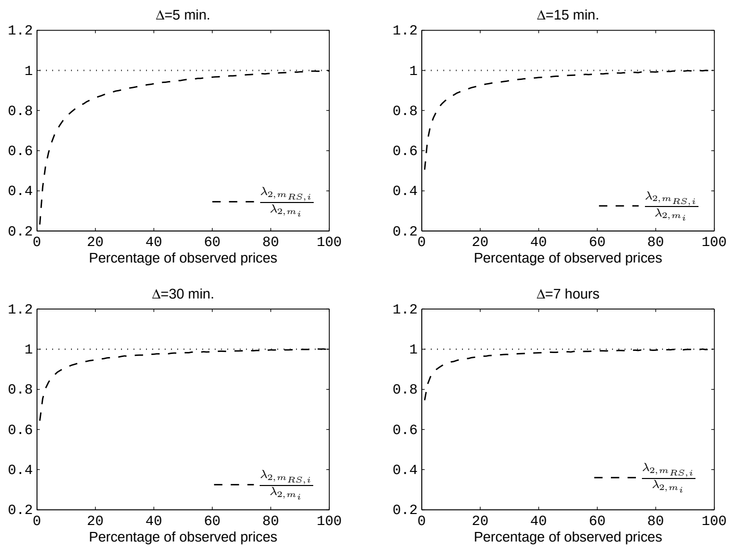

Figure 1: Figure plots the ratio between $\lambda_{2,m_{RS,i}}$ and λ_{2,m_i} for 4 different choice of Δ. $\lambda_{2,m_{RS,i}}$ is computed simulating a discrete standard Brownian motion over the interval $[0,\Delta]$ where only $m_{RS,i}$ out of M randomly placed increments are observed. λ_{2,m_i} is computed from the simulation of a standard Brownian motion with exactly m_i increments over the same interval. In abscissa the fraction of observed increments over the maximum number of increments assumed to be equal to M

They reveal that, keeping fixed the percentage of observed prices used in computation of $\lambda_{2,m_{RS,i}}$ and λ_{2,m_i}, the impact of the correction due to the irregular sampling is a decreasing function of Δ. This can be seen considering that, keeping fixed the percentage of observed prices and letting Δ to increase we are approaching in both cases the continuous Brownian motion. The second moment of the ranges of these continuous time Brownian motions equal in the limit $4\log(2)$.

Moreover, similarly to what reported in Christensen, Podolskij, and Vetter (2009), the numerical stability of the simulation of $\lambda_{r,m_{RS}}$ is quite poor when we observe a percentage of data points too low. This is particularly true when we compute $\widetilde{\lambda}_{r,m_{RS}}$. When irregularly sampled data are contaminated by microstructure noise, the estimator in (14) can be modified along the procedure so far described as:

$$RRG^{\Delta}_{m_{RS},BC} = \sum_{i=1}^{n} \frac{(s_{\widetilde{p}_{i\Delta,\Delta},m} - 2\widehat{\omega}_N)^2}{\widetilde{\lambda}_{2,m_{RS,i}}}. \tag{20}$$

In order to compute the relevant normalizing constant $\widetilde{\lambda}_{2,m_{RS,i}}$ when we have data contaminated with bid-ask effects, the simulation proceeds as follows:

1. We fix the fraction of prices actually observed, namely m_i/M. We also assume that in the interval $[0, \Delta]$ one could observe at most m_i prices if all the underlying price path were observed.

2. We simulate 1 million paths of discrete standard Brownian motions of length $M+1$ and steps $1/M$.

3. We coarse with the bid-ask effect the simulated prices (the variance of the bid-ask bounce does not affect the lambda so we choose a reasonable value).

4. We randomly select m_i prices out of $M+1$ and then compute the range imposing that the max occurs when the price is a bid and the in is an ask price.

5. We average these ranges over the Monte Carlo replications with the relevant exponent.

Clearly there are stability problems when m_i is really small since the probability of not observing a bid or an ask increases as we let m_i to decrease: this explains the strong downward bias when we compute RRG measure with random sampling at extremely high frequencies (higher that 5 minutes). This problem has also been reported by Christensen, Podolskij, and Vetter (2009). On the other hand, there is no need to increase the sampling frequency too much given the relatively high efficiency compared to RV.

3.6 Realized-Range estimation in presence of jumps in volatility

Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2001) show that under very general processes for the instantaneous volatility, which include the possibility of jumps, long memory or non-stationarity the quadratic variation of the price process coincides with the integrated variance.

In a large simulation study, Nielsen and Frederiksen (2008) show that a larger magnitude of volatility jumps weakens the performance of the realized volatility estimators, that is bias and RMSE increase with the jump size. In particular, increasing the sensitivity of the arrival intensity of jumps towards the level of the volatility implies an increase in the relative biases of the realized volatility estimators. An even more pronounced impact is found on the variability of the estimators.

Christensen and Podolskij (2007) show that under (11), the Realized Range estimator is a consistent estimator of IV when $n \to \infty$. However, no finite sample analysis of this class of integrated volatility estimators has been carried out and we address this issue in the simulation study below; see Sections 4 and 5.

4 Simulation set up

4.1 Models

For the Monte Carlo simulations we generate the log-price $p(t)$ as:

$$\mathrm{d}p(t) = \sigma(t)\mathrm{d}W_1(t) \tag{21}$$

and we assume that the instantaneous volatility process $\sigma(t)$ follows, alternatively, two distinct processes:

- *Long memory in volatility*

$$\mathrm{d}\log\sigma^2(t) = \alpha(\beta - \log\sigma^2(t))\mathrm{d}t + \nu\mathrm{d}W_d(t) \tag{22}$$

- *Jumps in volatility*

$$\mathrm{d}\sigma^2(t) = \alpha(\beta - \sigma^2(t))\mathrm{d}t + \nu \mathrm{d}W_2(t) + \kappa(t)\mathrm{d}q(t) \tag{23}$$

where $W_1(t)$ and $W_2(t)$ are independent Wiener processes. Both volatility processes ensure, under regularity, that volatility cannot be negative. In (22) W_d is a fractional Brownian Motion[1] of order d independent of $W_1(t)$. It may be represented by the Holmgren-Riemann-Liouville fractional integral

$$W_d(t) = \int_0^t \frac{(t-s)^2}{\Gamma(d+1)} \mathrm{d}W(s), \quad t > 0,$$

where $W(s)$ is a standard Brownian motion, see Marinucci and Robinson (1999).

We in fact generate instantaneous stochastic volatility process with a population long memory parameter of d, resulting in a integrated volatility with long memory parameter of $1 + d$.

In (23) we allow for jumps in volatility, represented by the jump process $\mathrm{d}q(t)$. This a CIR process (reference), i.e. a square-root model for the volatility process, with the addition of a positive jump process. The arrival of jumps is assumed to follow a Poisson process with intensity given by $\lambda_1 + \lambda_2\sigma^2(t)$. The arrival of jumps is assumed to depend linearly on volatility through the parameter λ_2. We assume that the arrival of jumps is exponentially distributed with mean μ. This setup is adopted in (Eraker, Johannes, and Polson 2003) and (Eraker 2004).

We simulate from Euler discrete approximation of Model A and B a set of discrete trajectories with a time step of 1 second for seven hours per day, which roughly corresponds to the trading period of NYSE. Thus we have a total $60 \times 60 \times 7 = 25.200$ log-prices per day for 10.000 days. The generated series are used to compute the integrated volatility for each day and to calculate the relative bias and variance for each of the estimators considered.

As in Nielsen and Frederiksen (2008), the parameters values used for the simulation from model A are: $\alpha = (0.000, 0.0.0062, 0.0124)$, $\beta = -1$, $\nu = (0.0374, 0.1122)$ and $d = (0.00, 0.15, 0.30, 0.45)$. All these parameters refers to an annualized stochastic volatility process. We initialize each simulated day with $p(0) = \log(100)$ and $\sigma(0) = \exp(\beta)$.

In model B we simulate using the following parameter values: $\alpha = 0.023$, $\beta = 0.943$, $\mu = (0.7515, 1.530)$, $\lambda_0 = (0.002, 0.0.01)$, and $\lambda_1 = (0, 1.298, 2.596)$. As starting values we use $p(0) = \log(100)$ and $\sigma(0) = \beta$.

4.2 Irregularly spaced sampling

In order to obtain randomly sampled series, we draw from the generated log-price series at random times. The interval between two successive available log-prices is modeled as an random variable exponentially distributed of mean τ, which is set to 5 seconds, in order to mimic the observed trade intensity of main stocks (for instance General Electric as in (Christensen and Podolskij 2007) and IBM in our empirical application (see Section 6), while Δ is let to vary, i.e. $\Delta = 5, 15, 30$ minutes.

We assume that the day is split into n intervals of constant length $\Delta = 1/n$, and hence we observe a random number of prices in each interval which is denoted by m_i, which is distributed according to a Poisson law with intensity parameter Δ/τ. Thus, each observed path is constituted by a random number of log-prices: $\sum_{i=1}^n m_i$, on average $25200/\tau$. As a consequence each interval i contains a random number of observed log-prices, which is a decreasing function of τ. The realized volatility estimator is computed using the imputation method, namely we use the closest price realization to obtain equally spaced log-price series. For what concerns the Realized Range

[1]To simulate increments from the fractional Brownian motion we implement the Matlab routine by Yingchun Zhou and Stilian Stoev which is based on the circulant embedding algorithm for the values of interest of the Hurst's exponent.

estimator, differently from Christensen and Podolskij (2007) we use irregularly spaced prices data. This means that in every sub-interval m_i with $i = 1 : n$ we observe a different number of prices. As a consequence we compute the Realized Range estimator either using the appropriate λ_{2,m_i} as suggested by Christensen and Podolskij (2007) or using $\lambda_{2,m_i,RS}$ as in eq. (19).

4.3 Microstructure noise

Two different noise processes are considered:

- *Bid-ask bounce*:
$$\widetilde{p}(t) = p(t) + \frac{\xi}{2}\mathbb{I}(t) \tag{24}$$
where ξ is the percentage spread, and the order-driven indicator variables $\mathbb{I}(t)$ are independently across p and t and identically distributed with $Pr\{\mathbb{I}(t) = 1\} = Pr\{\mathbb{I}(t) = -1\} = \frac{1}{2}$. This variable takes value 1 when the transaction is buyer-initiated, and -1 when it is seller-initiated. We adopt the simplest bid-ask bounce specification in order to compare with the existing literature. It is interesting to note that $d\widetilde{p}(t)$ exhibits spurious volatility and negative serial correlation (see Nielsen and Frederiksen (2008)).

- *Price discreteness*
$$\widetilde{p}(t) = \begin{cases} \text{ceil}(p(t) + \xi/2, \text{tick}), & \text{if } \mathbb{I}(t) = 1 \\ \text{floor}(p(t) - \xi/2, \text{tick}), & \text{if } \mathbb{I}(t) = -1 \end{cases} \tag{25}$$
where *tick* is the smallest reported price variation. In the simulations we use tick sizes (referred to prices, and not log-prices) equal to 1/8, 1/16, and 1/100.

4.4 Evaluation criteria

In order to compare the estimators of the daily IV, we calculate for each technique and for each day the relative error statistic as in eq.(26-28) in Nielsen and Frederiksen (2008).
Relative error:
$$\nu_s = \frac{\widehat{IV}_t - \sum_{t=1}^{T}(p_s(t) - p_s(t-1))^2}{\sum_{t=1}^{T}(p_s(t) - p_s(t-1))^2} \qquad s = 1, \ldots, S,$$
where $\widehat{IV}_t$ denotes any of the estimators considered and $\sum_{t=1}^{T}(p_s(t) - p_s(t-1))^2$ is a discrete-time approximation of the quadratic variation of $\log p(t)$. A natural choice for the $E(IV_t)$ would have been the sum over each day of the simulated instantaneous volatility, however to allow a fair comparison with the results reported in Nielsen and Frederiksen (2008) we adopted their measures. Moreover, when we evaluate the long memory estimation of the IV, using different estimator, we will rely on the simulated instantaneous volatilities to compute the IV, e.g. $\sum_{t=1}^{T}\sigma(t)^2$, in order to preserve the memory structure. The daily-average percentage relative bias is:
$$\text{bias} = \overline{\nu}_S = \frac{100}{S}\sum_{s=1}^{S}\nu_s.$$

The Root Mean Square Error
$$RMSE = \left(\frac{1}{S}\sum_{s=1}^{S}\nu_s^2\right)^{1/2} = \sqrt{\overline{\nu}_S^2 + g_\nu^2}$$

where $g_\nu^2 = S^{-1}\sum_{s=1}^{S}(\nu_s - \overline{\nu}_S)^2$ is the sample variance of ν_s.

5 Results

In this Section we present and comment the simulation results. We consider the following cases:

0M Without microstructure noise

1. Model A: Bias in Table 1, RMSE in Table 2
2. Model A II: Bias in Table 3, RMSE in Table 4
3. Model B: Bias in Table 5, RMSE in Table 6

1M With microstructure noise

BA Bid-Ask bounce

1. Model A: Bias in Table 7, RMSE in Table 8
2. Model B: Bias in Table 9, RMSE in Table 10

PD Bid-Ask bounce + Price Discreteness

1. Model A: Bias in Table 11, RMSE in Table 12

Case 0M-Model A: Table (1) reports the relative percentage bias of the competing estimators considering different values of the parameters α, the speed of mean reversion, and d the degree of long memory of the fBM driving the instantaneous volatility process. The results reveal that the estimators are very robust towards strong persistence ($d > 0.30$) and even non stationarity, $\alpha = 0$ and $d = 0$, that is the relative bias does not show any discernible pattern across different parameters values. From Table (1) is evident how the Realized Range is never able to achieve the lowest bias and this can be attributed to the fact that with irregularly spaced data the normalizing constant λ_{2,m_i} used in eq. (13) is too large since it is computed assuming a fully observe discrete time price path. On the other hand, the correction in eq. (19), which is based on the assumption that discrete time price paths are not fully observed, minimizes the relative bias in 5 cases out of 12. Furthermore, the Realized Range estimator seems to be adversely affected by the size of Δ; in fact, it is evident how the relative bias is worsening as we let Δ to decrease. Turning now to Daily Range measures, it is interesting to note how the Parkinson's estimator is largely downward biased for every parameter setup. This is due to the scaling constant computed under the assumption of full observation of the continuous time price process. On the other hand, the Daily Realized Range measures performs much better, with slightly better results for the Random Sampling correction. Finally, we can observe how the RRG_{RS} outperforms RV in terms of relative bias at least in the 30 min. and 15 min. cases.

The RMSEs (Table 2) are unaffected by the degree of persistence in the instantaneous volatility process, differently from what reported in Nielsen and Frederiksen (2008) where the RMSEs of the RV increase with the long memory parameter. It is important to notice that at the same sampling frequency, RRG and RRG_{RS} estimators are twice more efficient than RV.

Case 0M-Model A II: Comparing Table 1 with Table 3, we notice that the increase in the volatility of volatility ν partially offsets the effects of different sampling frequencies on RV and RRG_{RS}. In particular, the benefit of increasing the sampling frequency for RV seems less evident than what observed previously. There are no changes in the RMSEs, reported in Tables 2 and 4.

We also analyze the impact that τ, i.e. the average distance between available observations, has on the bias and the RMSE of the estimators. In fact, in Table 1 we notice how the bias of the RRG estimators progressively increases with the sampling frequency. In Figure 2 we show how this effect is mainly due to the missing observations, and we compare the performances of RV, RRG and our correction for the irregular sampling (RRG_{RS}) using 4 different sampling frequencies and letting τ varying from 1 second (all the prices are observed) to 60 seconds (on average one price

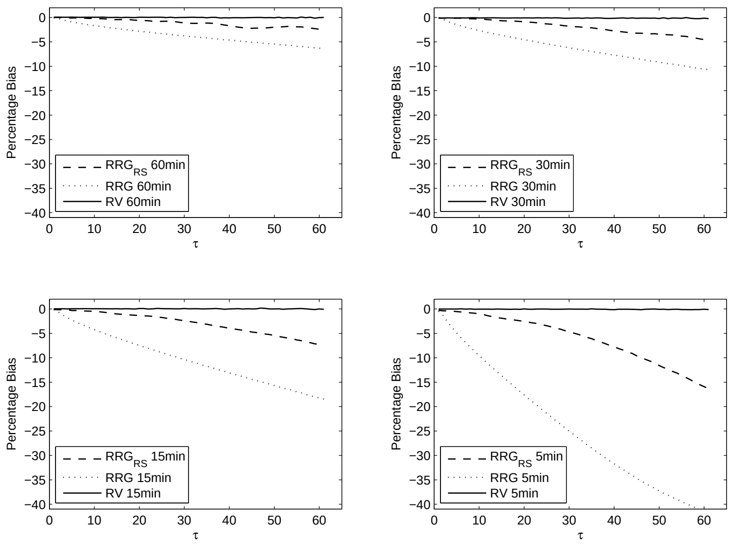

Figure 2: Percentage bias of alternative IV estimators as a function of τ (in seconds) at different sampling frequencies. Each panel reports RV, the Realized Volatility estimator, RRG_{RS} the Realized Range estimator with the Random Sampling correction and RRG, the Realized Range estimator, each computed at 4 different frequencies (60 min., 30 min., 15 min. and 5 min.). The simulated log-prices are obtained from Model A in eq. (22) with $\alpha = 0.0062$ and $d = 0.45$; 10,000 days with 86,400 log-prices each are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean τ.

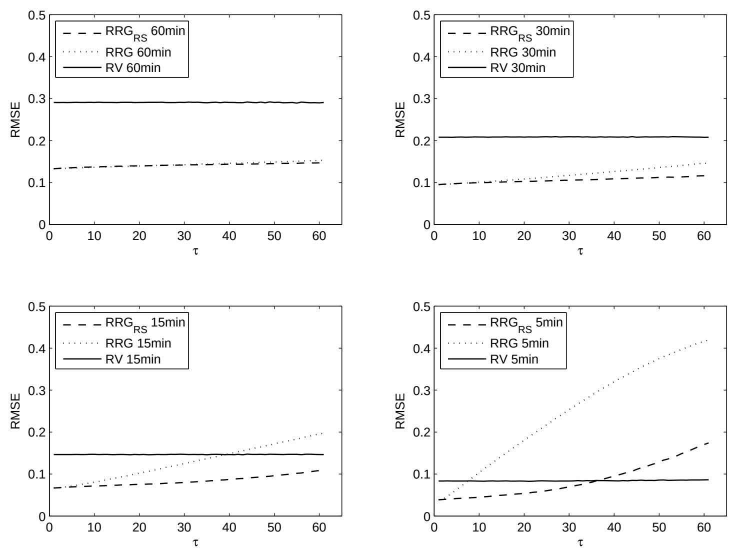

Figure 3: RMSE of alternative IV estimators as a function of τ (in seconds) at different sampling frequencies. Each panel reports RV, the Realized Volatility estimator, RRG_{RS} the Realized Range estimator with the Random Sampling correction and RRG, the Realized Range estimator, each computed at 4 different frequencies (60 min., 30 min., 15 min. and 5 min.). The simulated log-prices are obtained from Model A in eq. (22) with $\alpha = 0.0062$ and $d = 0.45$; 10,000 days with 86,400 log-prices each are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean τ.

Table 1: Simulation results for model A: relative percentage bias

α	d	Realized Volatility			Daily Range			Realized Range			Realized Range RS		
		30 min	15 min	5 min	Park.	Real.	RS	30 min	15 min	5 min	30 min	15 min	5 min
0.000	0.00	-0.116	0.195	**0.021**	-2.360	-0.929	0.073	-1.626	-2.613	-5.873	**-0.021**	-0.114	-0.960
	0.15	0.254	0.504	0.299	-1.390	**0.056**	-0.396	-1.407	-2.487	-5.770	0.032	-0.096	-1.028
	0.30	0.469	0.492	0.155	-1.611	-0.169	0.440	-1.383	-2.448	-5.842	0.147	**-0.039**	-0.921
	0.45	0.252	**0.081**	0.229	-2.792	-1.367	-1.192	-1.524	-2.557	-5.770	-0.149	-0.245	-1.031
0.006	0.00	**-0.067**	0.080	0.140	-2.584	-1.156	0.802	-1.682	-2.614	-5.856	-0.048	-0.224	-1.041
	0.15	0.517	0.394	0.191	-2.616	-1.188	1.250	-1.381	-2.477	-5.799	0.204	**0.068**	-0.929
	0.30	-0.552	**-0.001**	-0.062	-2.247	-0.814	0.428	-1.831	-2.598	-5.916	-0.086	-0.242	-1.009
	0.45	-0.157	0.083	**0.005**	-2.606	-1.178	0.417	-1.719	-2.620	-5.943	0.082	-0.022	-0.940
0.012	0.00	0.350	0.297	**0.029**	-1.977	-0.540	0.193	-1.567	-2.636	-5.879	-0.244	-0.232	-1.027
	0.15	0.123	0.178	0.329	-2.146	-0.712	-0.666	-1.627	-2.592	-5.777	-0.182	**-0.091**	-1.028
	0.30	0.582	0.245	0.175	-1.816	-0.377	0.793	-1.361	-2.501	-5.804	**-0.008**	-0.247	-0.997
	0.45	-0.377	-0.395	-0.114	-3.713	-2.302	**0.006**	-2.088	-2.936	-5.969	-0.183	-0.169	-1.044

Note: The simulated log-prices are obtained from Model A in eq. (22) with $\beta = -1$ and $\nu = 0.0324$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The RV in eq. (3), the RRG in eq. (13) and the RRG with Random Sampling correction in eq. (19) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in three different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$ and the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$. In boldface the minimum bias in absolute value for each parameter configuration.

per minute is observed). We notice how, when the sampling frequency is low, i.e. considering 60 minutes intervals, the effect of τ on the bias of RRG is negligible compared to the cases with higher sampling frequency. The bias of RRG_{RS} is roughly one third of the bias of RRG for every τ and every sampling frequency considered and it is a concave function of τ, while the bias of RRG is convex in τ. If we consider highly liquid assets, i.e. with $\tau < 15$ seconds, the bias of RRG_{RS} is very close to that of RV.

Figure 3 reports the RMSEs as a function of τ. The RMSE for the RRG_{RS} estimator is unrelated to τ for the lowest sampling frequencies. On the other hand, when we consider the highest sampling frequencies, the RMSEs tend to increase with τ as a consequence of the increasing bias. Nonetheless, the RRG_{RS} is more efficient than RRG in every case and more efficient than RV in all but the most extreme cases, i.e. sampling frequency of 5 min. and $\tau > 35$ seconds.

Case 0M-Model B: Table 5 shows how increasing λ_1, i.e. the dependence of the arrival of jumps to the level of volatility, the bias of RV estimators increases, while the impact of λ_0 is unclear. Again, on average, increasing the sampling frequency lowers the bias across almost all cases considered. On the other hand, the biases of RRG_{RS} do not show any discernible pattern, thus we can conclude that this estimator is fairly robust to the jumps in volatility specification. Moreover, the highest sampling frequency induces a larger bias. The two preferred estimators are the RV with 5 minutes returns and the RRG_{RS} with 30 minutes intervals. On average, Daily Ranges over-perform RRG. Table 6 confirms how the RRGs are characterized by a bias-variance trade-off whilst RV are not. However, it is worth noting how the RV 5 minutes and the RRG_{RS} 30 minutes have approximately the same bias and RMSE.

The main message of this analysis without microstructure noise is that the optimal sampling frequency when using RV estimator is the highest possible irrespective of the asset trading activity, i.e. the bias and the RMSE decrease increasing the sampling frequency. On the other hand, the optimal sampling frequency for Realized Range estimators crucially depends on the trading activity of the asset. In fact, increasing the sampling frequency, improves the RMSE but deteriorates the relative bias.

Case 1M-BA Model A: Table 7 illustrates the percentage bias for two Realized Volatility estimators and five Realized Range estimators. It is strongly evident how the bias correction for both classes of estimators is crucial to contain the relative biases to a reasonable level. The best

Table 2: Simulation results for model A: RMSE

α	d	Realized Volatility			Daily Range			Realized Range			Realized Range RS		
		30 min	15 min	5 min	Park.	Real.	RS	30 min	15 min	5 min	30 min	15 min	5 min
0.000	0.00	0.378	0.268	0.153	0.633	0.642	0.653	0.177	0.128	0.094	0.178	0.129	0.079
	0.15	0.376	0.270	0.155	0.641	0.651	0.632	0.176	0.129	0.094	0.178	0.129	0.078
	0.30	0.380	0.269	0.155	0.636	0.645	0.645	0.176	0.128	0.094	0.178	0.128	0.078
	0.45	0.382	0.269	0.154	0.638	0.646	0.623	0.178	0.129	0.093	0.177	0.128	0.078
0.006	0.00	0.375	0.267	0.154	0.635	0.643	0.660	0.177	0.129	0.094	0.179	0.128	0.078
	0.15	0.382	0.268	0.155	0.624	0.633	0.661	0.178	0.128	0.094	0.178	0.129	0.077
	0.30	0.378	0.268	0.154	0.644	0.653	0.648	0.176	0.128	0.094	0.179	0.128	0.078
	0.45	0.378	0.267	0.156	0.621	0.630	0.653	0.177	0.128	0.095	0.179	0.129	0.078
0.012	0.00	0.380	0.269	0.154	0.643	0.652	0.644	0.179	0.128	0.094	0.176	0.127	0.077
	0.15	0.376	0.269	0.154	0.647	0.657	0.634	0.175	0.129	0.093	0.179	0.130	0.079
	0.30	0.386	0.267	0.156	0.643	0.652	0.654	0.178	0.129	0.094	0.178	0.129	0.078
	0.45	0.378	0.264	0.153	0.614	0.623	0.657	0.178	0.128	0.095	0.179	0.130	0.078

Note: The simulated log-prices are obtained from Model A in eq. (22) with $\beta = -1$ and $\nu = 0.0324$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The RV in eq. (3), the RRG in eq. (13) and the RRG with Random Sampling correction in eq. (19) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in three different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$ and the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$.

performances in term of bias are obtained with HL and $RRG_{BC,RS}$. It is worth mentioning the fact the HL results seems to be adversely affected by the increase in the long memory parameter, while the $RRG_{BC,RS}$ is extremely robust towards the parameter specification and the spread magnitude. Moreover, the HL estimator bias worsens with the spread magnitude in high volatility of volatility scenarios and the biases have mixed signs while Realized Range estimators corrected for the bias steadily underestimate the true integrated volatility. Finally, Daily Range with Bias correction show biases of the same order of those observed for the RRG_{BC} sampled at 15 and 5 minutes.

Even in this case we can observe how the RRG class with Bias correction is fairly superior in terms of RMSE to HL and RV estimators, as shown in Table 8. For the HL estimator, it is evident that the spread magnitude increases the RMSE at every frequency considered and that the precision of the estimator does not increase with the sampling frequency. This effect is present for the RRG_{BC} too even though it is less striking. This implies that for both classes of estimators, it is not optimal to increase too much the sampling frequency.

Case 1M-BA Model B: From the simulation results for model B with bid-ask bounce, percentage bias reported in Table 9, it is evident that HL is by far the best estimator. The estimator is robust to the noise specification and to the sensitivity of the jump arrival intensity to the level of volatility. The best choice among the Realized Range class is the $RRG_{BC,RS}$ which however is always downward bias.

So, it seems that the Realized Range Bias Correction by Christensen, Podolskij, and Vetter (2009) introduces a steady downward bias in the estimation; on the other hand, this correction counterbalances the upward bias induced by the bid-ask bounce which is evident in Tables 7 and 9.

RMSEs in Table 10 reveal that for HL, differently from the results of Model A, we can gain in efficiency increasing the sampling frequency. However, it is important to stress the fact that while the $RRG_{BC,RS}$ presents RMSE smaller than those of HL, the latter are mainly due to the variability of the estimator, while the former are principally induced by the bias. Thus we can conclude that the variability of $RRG_{BC,RS}$ is much smaller than the variability of HL.

Results obtained with bid-ask bounce and price discreteness modeled as in eq. (25) are similar to those obtained without price discreteness and discussed so far. This is particularly true when we consider the smallest tick size (1 $ cent) that reflects the decimalization of 2000. Thus, we do

Table 3: Simulation results for model A II: relative percentage bias

α	d	Realized Volatility			Daily Range			Realized Range			Realized Range RS		
		30 min	15 min	5 min	Park.	Real.	RS	30 min	15 min	5 min	30 min	15 min	5 min
0.000	0.00	0.106	-0.069	-0.104	-3.273	-1.855	1.384	-1.744	-2.810	-6.002	**0.059**	-0.113	-0.553
	0.15	-0.287	-0.389	**-0.064**	-2.365	-0.933	-0.502	-1.771	-2.791	-5.925	-0.271	-0.412	-0.644
	0.30	-0.150	-0.072	**0.010**	-2.641	-1.214	0.044	-1.540	-2.587	-5.857	-0.147	-0.414	-0.732
	0.45	0.068	0.067	0.036	-2.058	-0.622	0.308	-1.720	-2.656	-5.917	**-0.003**	-0.237	-0.578
0.006	0.00	0.583	0.104	**0.054**	-1.732	-0.292	-0.456	-1.548	-2.709	-5.906	-0.078	-0.291	-0.639
	0.15	**-0.018**	0.080	0.239	-2.220	-0.786	-0.378	-1.516	-2.578	-5.740	-0.411	-0.447	-0.739
	0.30	-0.456	-0.372	-0.169	-1.604	-0.161	**-0.070**	-1.995	-2.839	-5.995	-0.112	-0.357	-0.548
	0.45	-0.430	-0.059	0.142	-2.819	-1.394	-0.663	-1.907	-2.774	-5.917	**0.038**	-0.259	-0.597
0.012	0.00	**-0.134**	-0.516	-0.315	-2.412	-0.981	-0.816	-1.802	-2.972	-6.003	-0.652	-0.623	-0.636
	0.15	0.301	**-0.047**	-0.134	-1.851	-0.412	-0.425	-1.588	-2.669	-5.869	-0.133	-0.208	-0.547
	0.30	-0.368	**-0.130**	-0.174	-2.906	-1.482	-0.515	-1.838	-2.744	-5.990	-0.238	-0.477	-0.664
	0.45	0.094	**-0.091**	-0.101	-2.430	-1.000	-0.919	-1.648	-2.664	-5.886	-0.636	-0.624	-0.677

Note: The simulated log-prices are obtained from Model A in eq. (22) with $\beta = -1$ and $\nu = 0.1122$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The RV in eq. (3), the RRG in eq. (13) and the RRG with Random Sampling correction in eq. (19) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in three different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$ and the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$. In boldface the minimum bias in absolute value for each parameter configuration.

not report here the results for Model B[2] with bid-ask bounce and price discreteness and we will discuss only the results for Model A.

Case 1M-BA-PD Model A: As in the previous case HL outperforms Realized Range based estimators, as shown in Table 11; however we notice of the performances of this estimators worsen when we introduce long memory in instantaneous volatility. The bias corrections in RV and RRG are able to offset the effect of the tick size which is instead evident in the non corrected versions.

On average, looking at the bias-RMSE trade-off, it seems to be convenient to use the 5 minutes frequency for both the HL and the $RRG_{BC,RS}$ estimators.

5.1 Long memory estimation

In this Section we explore the persistence properties of estimated Integrated Variance series. It is a common finding in the literature that RV estimates exhibit long memory features, see Andersen, Bollerslev, Diebold, and Labys (2001), Martens, van Dijk, and Pooter (2004), Morana and Beltratti (2004) and Thomakos and Wang (2003) among others. Recently, Lieberman and Phillips (2008) show how long memory may arise from the accumulative process underlying realized volatility.

We generate artificial time series of 2000 days with 86400 log-prices each from Model A considering several parameters values for the speed of mean reversion α and for the long memory parameter of the increments of fractional Brownian motion d. As a consequence, the fractional Brownian motion is characterized by a long memory parameter equal to $1 + d$.

We estimate the $(1 + d)$ parameter using the Exact Local Whittle estimator by Shimotsu and Phillips (2005) which is constructed considering the fractional process Y_t generated by

$$(1 - L)^d Y_t = u_t \mathbb{I}(t \geq 1), \quad t = 0, \pm 1, \ldots,$$

where u_t is a stationary process with zero mean and spectral density $f_u(\omega) \sim G_0$ as $\omega \to 0$.

The Exact Local Whittle estimator of d is:

$$\hat{d} = \arg \min_{d \in [\delta_1, \delta_2]} R(d) \tag{26}$$

[2] They are available from the Authors upon request.

Table 4: Simulation results for model A II: RMSE

α	d	Realized Volatility			Daily Range			Realized Range			Realized Range RS		
		30 min	15 min	5 min	Park.	Real.	RS	30 min	15 min	5 min	30 min	15 min	5 min
0.000	0.00	0.383	0.267	0.153	0.616	0.624	0.639	0.178	0.129	0.095	0.179	0.128	0.078
	0.15	0.378	0.264	0.155	0.635	0.644	0.670	0.176	0.128	0.094	0.180	0.129	0.078
	0.30	0.378	0.271	0.157	0.629	0.638	0.662	0.178	0.130	0.095	0.178	0.129	0.079
	0.45	0.383	0.270	0.155	0.638	0.647	0.642	0.179	0.130	0.095	0.179	0.130	0.079
0.006	0.00	0.377	0.268	0.155	0.638	0.647	0.649	0.177	0.128	0.095	0.181	0.130	0.078
	0.15	0.377	0.268	0.157	0.641	0.650	0.656	0.178	0.129	0.094	0.179	0.130	0.079
	0.30	0.377	0.267	0.154	0.654	0.663	0.659	0.177	0.129	0.095	0.178	0.128	0.078
	0.45	0.376	0.267	0.154	0.630	0.639	0.656	0.176	0.128	0.094	0.179	0.128	0.079
0.012	0.00	0.381	0.269	0.153	0.631	0.639	0.658	0.177	0.129	0.095	0.177	0.129	0.079
	0.15	0.376	0.266	0.154	0.639	0.648	0.662	0.178	0.130	0.095	0.180	0.129	0.078
	0.30	0.378	0.268	0.154	0.628	0.636	0.642	0.178	0.129	0.095	0.178	0.128	0.079
	0.45	0.376	0.265	0.154	0.634	0.643	0.646	0.176	0.127	0.094	0.177	0.128	0.077

Note: The simulated log-prices are obtained from Model A in eq. (22) with $\beta = -1$ and $\nu = 0.1122$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The RV in eq. (3), the RRG in eq. (13) and the RRG with Random Sampling correction in eq. (19) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in three different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$ and the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$.

where

$$R(d) = \log\left(\hat{G}(d)\right) - 2d\frac{1}{m}\sum_{j=1}^{m}\log\omega_j, \quad \hat{G}(d) = \frac{1}{m}\sum_{j=1}^{m} I_X(\omega_j).$$

$I_X(\omega_j)$ is the periodogram of $X = (1-L)^d Y$ evaluated at the frequency $\omega_j = (2\pi j)/T$ and $m < T$ where T is the number of observations. Shimotsu and Phillips (2005) show that:

$$m^{1/2}\left(\hat{d} - d\right) \xrightarrow{d} N\left(0, \frac{1}{4}\right), \quad \text{as } n \to \infty.$$

Table 13 reports the estimates of $1+d$ obtained from artificial time series of daily Integrated Volatility computed using RV and RRG estimators at different sampling frequencies. As a benchmark, we also report the estimates of $1+d$ of the time series of the Integrated Volatility constructed, at 1 second frequency, as:

$$IV_t = \sum_{i=1}^{86{,}400} \sigma^2_{i,t}, \quad t = 1, \ldots, 2000.$$

We compute also IV series through the summation of instantaneous volatility every 30 and 300 seconds. As we can see the series of IV constructed using three different aggregation frequencies present exactly the same long memory parameter estimate. This confirms that the IV process can be considered a self-similar fractal. A self-similar fractal has the same H, or Hurst coefficient, where $H = d + 1/2$ for all choices of time intervals (see **?**). If we now turn to the estimates of $1+d$ obtained from RV and RRG series, we can observe that the long memory parameter estimates are severely downward biased, and this worsening as the sample frequency used in the computation decreases. For instance, in the case of continuous-time random walk ($\alpha = 0,\ 1+d = 1$) the estimate of $1+d$ with RV ranges from 0.765, using 30 seconds returns, to 0.239 using 1-hour returns. The same effect can be observed in the estimates from the RRG series, which are still downward biased even though the biases are sensibly smaller for all frequencies considered.

The size of the mean reversion parameter α has no clear cut effects on the estimates of $1+d$. It seems that the driving factor behind this downward bias in the long memory parameter estimates is

Table 5: Simulation results for model B: relative percentage bias

μ	λ_0	λ_1	Realized Volatility			Daily Range			Realized Range			Realized Range RS		
			30 min	15 min	5 min	Park.	Real.	RS	30 min	15 min	5 min	30 min	15 min	5 min
0.751	0.002	0.000	**0.002**	-0.184	0.080	-1.239	0.209	-0.531	-1.803	-2.769	-5.879	-0.127	-0.397	-1.084
		1.298	0.395	0.103	0.229	-2.377	-0.946	0.338	-1.560	-2.643	-5.824	**-0.099**	-0.276	-1.053
		2.596	-0.303	-0.255	0.206	-3.196	-1.777	-1.057	-1.629	-2.646	-5.915	**-0.131**	-0.155	-0.940
	0.010	0.000	-0.260	0.037	-0.009	-2.035	-0.599	-0.788	-1.758	-2.728	-5.939	**0.008**	-0.093	-0.879
		1.298	-0.117	-0.010	**0.006**	-1.153	0.296	-0.203	-1.868	-2.696	-5.882	0.151	-0.087	-0.838
		2.596	0.239	0.113	**0.000**	-2.380	-0.949	-0.092	-1.789	-2.720	-5.936	0.031	-0.102	-1.062
1.530	0.002	0.000	0.372	-0.126	**-0.043**	-1.732	-0.291	0.397	-1.594	-2.763	-5.920	0.090	-0.145	-1.081
		1.298	-0.251	0.070	0.108	-1.849	-0.410	-0.720	-1.657	-2.636	-5.837	**-0.062**	-0.257	-0.991
		2.596	-0.159	**-0.073**	0.322	-1.122	0.327	0.614	-1.696	-2.647	-5.807	0.433	0.128	-0.814
	0.010	0.000	0.077	0.316	0.152	-2.193	-0.759	0.834	-1.541	-2.592	-5.864	0.358	**-0.054**	-0.920
		1.298	-0.235	**-0.125**	0.195	-1.797	-0.358	0.382	-1.652	-2.646	-5.826	-0.127	-0.307	-1.147
		2.596	0.013	-0.171	**0.003**	-1.780	-0.340	-0.569	-1.872	-2.836	-5.967	0.097	-0.021	-0.909

Note: The simulated log-prices are obtained from Model B in eq. (23) with $\alpha = 0.023$ and $\beta = 0.943$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The *RV* in eq. (3), the *RRG* in eq. (13) and the *RRG* with Random Sampling correction in eq. (19) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in three different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$ and the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$. In boldface the minimum bias in absolute value for each parameter configuration.

the variance of the IV estimators. In fact, when we let the sampling frequency to decrease, we are adding progressively more noise to the IV estimation (see e.g. Table 2). This is also confirmed by the fact that RRG estimates, which are more efficient than RV at any given sampling frequency (see e.g. Christensen and Podolskij (2007) for the asymptotic variances), yields long memory parameter estimates less biased.

Notwithstanding, the estimates of $1 + d$ computed using both RV and RRG at any sampling frequency considered are completely misleading about the degree of long memory in the underlying DGP. This observation in in accord with the intuition provided by Lieberman and Phillips (2008) that fitted long memory can be an artifact of accumulation.

Table 6: Simulation results for model B: RMSE

μ	λ_0	λ_1	Realized Volatility			Daily Range			Realized Range			Realized Range RS		
			30 min	15 min	5 min	Park.	Real.	RS	30 min	15 min	5 min	30 min	15 min	5 min
0.751	0.000	0.000	0.384	0.268	0.153	0.649	0.658	0.650	0.177	0.129	0.094	0.179	0.129	0.079
		1.298	0.376	0.263	0.155	0.620	0.628	0.644	0.176	0.127	0.094	0.176	0.126	0.078
		2.596	0.375	0.267	0.156	0.620	0.628	0.633	0.176	0.129	0.095	0.179	0.129	0.077
	0.010	0.000	0.371	0.268	0.154	0.638	0.647	0.646	0.176	0.129	0.094	0.178	0.128	0.078
		1.298	0.381	0.268	0.155	0.653	0.662	0.652	0.179	0.129	0.095	0.179	0.128	0.078
		2.596	0.381	0.270	0.157	0.631	0.640	0.641	0.178	0.130	0.096	0.177	0.127	0.078
1.530	0.000	0.000	0.376	0.263	0.153	0.641	0.650	0.653	0.175	0.127	0.094	0.181	0.130	0.079
		1.298	0.373	0.264	0.153	0.634	0.643	0.650	0.174	0.127	0.094	0.179	0.128	0.078
		2.596	0.379	0.270	0.156	0.643	0.652	0.655	0.178	0.129	0.094	0.180	0.129	0.078
	0.010	0.000	0.379	0.271	0.156	0.647	0.656	0.657	0.178	0.130	0.095	0.179	0.127	0.078
		1.298	0.377	0.268	0.156	0.640	0.650	0.658	0.177	0.128	0.095	0.179	0.128	0.078
		2.596	0.380	0.268	0.155	0.631	0.640	0.639	0.178	0.129	0.095	0.179	0.129	0.078

Note: The simulated log-prices are obtained from Model B in eq. (23) with $\alpha = 0.023$ and $\beta = 0.943$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The RV in eq. (3), the RRG in eq. (13) and the RRG with Random Sampling correction in eq. (19) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in three different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$ and the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$.

Table 7: Simulation results for model A with bid-ask bounce: relative percentage bias

d	ξ	ν	Realized Volatility			Hansen-Lunde			Daily Range				
			30 min	15 min	5 min	30 min	15 min	5 min	Park.	Real.	RS	Real. BC	RS BC
0.0	0.003	0.05	17.53	26.54	71.02	-8.82	4.38	**2.58**	14.22	15.89	15.08	-5.66	-3.84
		0.50	13.28	24.11	70.85	-4.49	-2.48	**-1.19**	13.18	14.84	13.38	-5.55	-5.38
	0.005	0.05	36.31	66.65	199.12	1.70	5.10	0.25	34.12	36.09	25.48	-4.63	-4.38
		0.50	27.58	62.08	190.98	9.18	-7.69	6.62	32.37	34.31	23.87	-5.36	-5.74
	0.010	0.05	141.22	259.62	793.72	**1.68**	7.51	-9.40	59.40	61.74	55.59	-5.32	-4.20
		0.50	128.90	256.15	776.48	1.78	**-1.05**	14.98	53.23	55.48	55.00	-5.99	-4.63
0.3	0.003	0.05	6.93	19.84	69.79	-4.26	-9.36	-6.02	6.04	7.60	15.78	-4.40	-3.20
		0.50	12.04	22.21	66.15	7.30	3.38	**-0.13**	11.22	12.85	15.13	-6.37	-3.77
	0.005	0.05	33.30	71.30	194.14	-10.69	-4.58	4.60	22.73	24.53	25.50	-4.40	-4.36
		0.50	42.07	66.42	198.18	7.17	17.82	1.98	27.11	28.98	25.78	-5.73	-4.10
	0.010	0.05	117.25	248.14	772.26	**0.23**	11.90	-3.19	61.78	64.15	54.90	-4.00	-4.69
		0.50	132.83	267.65	784.15	3.19	-15.57	22.93	50.84	53.06	54.07	-5.67	-5.41

d	ξ	ν	Realized Range			Realized Range RS			Realized Range BC			Realized Range RS BC		
			30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min
0.0	0.003	0.05	60.51	89.13	176.12	61.37	92.78	191.68	-3.82	-5.08	-6.63	-3.78	-3.62	-3.99
		0.50	59.54	88.21	176.07	61.33	92.72	191.70	-3.97	-4.78	-6.63	-3.83	-3.68	-4.00
	0.005	0.05	117.83	178.00	371.80	115.30	179.99	398.40	**0.02**	-1.40	-6.21	-3.68	-3.54	-3.60
		0.50	112.88	175.40	372.27	114.48	179.57	398.17	-4.35	-4.38	-5.94	-4.25	-3.81	**-3.75**
	0.010	0.05	277.16	457.89	1096.65	286.60	473.26	1163.91	-5.57	-4.79	-4.51	-2.95	-2.66	-1.90
		0.50	278.64	458.85	1096.13	285.93	473.19	1163.64	-4.13	-3.51	-4.55	-3.28	-2.65	-1.96
0.3	0.003	0.05	55.57	85.62	176.02	61.55	92.94	191.78	-7.11	-6.75	-6.63	-3.63	**-3.48**	-3.87
		0.50	56.81	85.41	175.76	61.24	92.67	191.64	-6.26	-7.16	-6.78	-3.83	-3.62	-3.87
	0.005	0.05	112.35	174.88	371.86	115.75	180.57	398.68	-4.47	-4.21	-6.14	-3.41	**-3.21**	-3.51
		0.50	115.21	174.46	372.03	115.28	180.39	398.62	**-1.95**	-3.95	-6.06	-3.68	-3.28	-3.49
	0.010	0.05	277.62	457.52	1096.33	285.82	472.87	1163.65	-4.61	-4.06	-4.53	-3.35	-2.81	-1.98
		0.50	285.51	465.50	1096.05	285.85	472.92	1163.84	-0.82	**-0.70**	-4.57	-3.39	-2.84	-1.99

Note: The simulated log-prices are obtained from Model A in eq. (22) with $\beta = -1$ and $\alpha = 0.0062$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The log-prices are contaminated with a bid-ask bounce effect described in eq. (24). The RV in eq. (3), the Hansen and Lunde estimator in eq. (6), the RRG in eq. (13), the RRG with Random Sampling correction in eq. (19), the RRG_{BC} as in eq. (14) and the $RRG_{RS,BC}$ as in eq. (20) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in five different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$, the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$, the Realized Range computed with $\widetilde{\lambda}_{2,m}$, the Realized Range with Random Sampling correction computed with $\widetilde{\lambda}_{2,m_{RS}}$. In boldface the minimum bias in absolute value for each parameter configuration.

Table 8: Simulation results for model A with bid-ask bounce: RMSE

d	ξ	ν	Realized Volatility			Hansen-Lunde			Daily Range				
			30 min	15 min	5 min	30 min	15 min	5 min	Park.	Real.	RS	Real. BC	RS BC
0.0	0.003	0.05	0.48	0.44	0.76	0.59	0.51	0.39	0.60	0.61	0.70	0.61	0.62
		0.50	0.44	0.43	0.75	0.71	0.52	0.35	0.80	0.81	0.69	0.61	0.61
	0.005	0.05	0.57	0.79	2.03	0.90	0.61	0.50	0.96	0.97	0.76	0.62	0.62
		0.50	0.50	0.73	1.95	0.73	0.50	0.47	0.86	0.88	0.76	0.60	0.63
	0.010	0.05	1.63	2.72	8.00	1.08	1.01	1.31	1.08	1.11	0.96	0.63	0.62
		0.50	1.51	2.69	7.84	1.06	0.90	1.32	0.92	0.94	0.97	0.61	0.63
0.3	0.003	0.05	0.42	0.39	0.74	0.65	0.54	0.38	0.56	0.57	0.72	0.63	0.63
		0.50	0.44	0.39	0.70	0.73	0.54	0.35	0.73	0.75	0.72	0.62	0.64
	0.005	0.05	0.60	0.84	1.98	0.68	0.60	0.48	0.66	0.68	0.77	0.62	0.63
		0.50	0.64	0.80	2.02	0.83	0.65	0.50	0.70	0.71	0.77	0.62	0.63
	0.010	0.05	1.45	2.60	7.77	0.89	1.01	1.41	1.02	1.04	0.96	0.63	0.62
		0.50	1.57	2.81	7.90	0.96	0.95	1.30	0.80	0.82	0.94	0.61	0.60

d	ξ	ν	Realized Range			Realized Range RS			Realized Range BC			Realized Range RS BC		
			30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min
0.0	0.003	0.05	0.65	0.91	1.77	0.65	0.94	1.92	0.20	0.15	0.12	0.18	0.14	0.11
		0.50	0.63	0.90	1.76	0.65	0.94	1.92	0.18	0.15	0.12	0.18	0.14	0.11
	0.005	0.05	1.21	1.79	3.72	1.18	1.81	3.99	0.19	0.15	0.14	0.19	0.15	0.14
		0.50	1.15	1.76	3.73	1.17	1.81	3.99	0.16	0.13	0.14	0.18	0.15	0.14
	0.010	0.05	2.79	4.59	10.97	2.89	4.74	11.64	0.20	0.16	0.22	0.19	0.17	0.22
		0.50	2.80	4.60	10.97	2.88	4.74	11.64	0.17	0.14	0.22	0.19	0.17	0.22
0.3	0.003	0.05	0.60	0.87	1.76	0.65	0.95	1.92	0.19	0.15	0.12	0.18	0.14	0.11
		0.50	0.60	0.87	1.76	0.65	0.94	1.92	0.17	0.13	0.12	0.18	0.14	0.11
	0.005	0.05	1.15	1.76	3.72	1.19	1.82	3.99	0.18	0.15	0.14	0.18	0.14	0.14
		0.50	1.18	1.75	3.72	1.18	1.82	3.99	0.17	0.12	0.14	0.19	0.15	0.14
	0.010	0.05	2.80	4.58	10.97	2.88	4.74	11.64	0.22	0.18	0.22	0.19	0.17	0.22
		0.50	2.87	4.66	10.96	2.88	4.74	11.64	0.18	0.15	0.22	0.19	0.17	0.22

Note: The simulated log-prices are obtained from Model A in eq. (22) with $\beta = -1$ and $\alpha = 0.0062$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The log-prices are contaminated with a bid-ask bounce effect described in eq. (24). The RV in eq. (3), the Hansen and Lunde estimator in eq. (6), the RRG in eq. (13), the RRG with Random Sampling correction in eq. (19), the RRG_{BC} as in eq. (14) and the $RRG_{RS,BC}$ as in eq. (20) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in five different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$, the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$, the Realized Range computed with $\widetilde{\lambda}_{2,m}$, the Realized Range with Random Sampling correction computed with $\widetilde{\lambda}_{2,m_{RS}}$.

Table 9: Simulation results for model B with bid-ask bounce: relative percentage bias

ξ	λ_1	Realized Volatility			Hansen-Lunde			Daily Range				
		30 min	15 min	5 min	30 min	15 min	5 min	Park.	Real.	RS	Real. BC	RS BC
0.003	0.000	1.54	3.42	10.55	0.88	**-0.04**	0.17	1.67	3.16	4.11	-5.57	-5.10
	1.298	1.34	3.47	10.45	-0.77	-0.38	**-0.13**	1.60	3.09	4.16	-5.18	-5.06
	2.596	1.40	3.35	10.57	0.60	-0.39	**0.01**	2.64	4.14	3.89	-6.00	-5.32
0.005	0.000	4.50	9.46	29.36	1.36	**0.47**	-0.49	6.86	8.43	9.39	-5.05	-3.95
	1.298	5.40	10.10	29.68	0.60	**0.45**	0.57	7.21	8.78	9.98	-6.33	-3.40
	2.596	4.46	9.52	29.54	0.49	**-0.17**	-0.35	6.29	7.85	8.38	-4.62	-4.88
0.010	0.000	18.83	39.01	118.09	**-0.29**	-0.99	-0.43	16.29	17.99	20.35	-5.44	-3.50
	1.298	19.56	38.74	117.63	0.32	**0.20**	-0.32	16.92	18.63	19.92	-6.79	-3.89
	2.596	19.64	39.54	117.85	0.74	**-0.47**	0.56	16.83	18.55	20.75	-5.35	-3.15

ξ	λ_1	Realized Range			Realized Range RS			Realized Range BC			Realized Range RS BC		
		30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min
0.003	0.000	17.12	24.48	44.04	18.90	27.52	51.94	-4.85	-5.26	-6.70	-4.07	-3.83	-4.22
	1.298	17.02	24.61	43.91	19.01	27.73	52.10	-4.91	-5.09	-6.79	-3.97	-3.66	-4.11
	2.596	16.94	24.60	44.03	19.09	27.79	52.14	-4.95	-5.06	-6.74	-3.90	-3.61	-4.07
0.005	0.000	33.65	49.81	94.68	36.47	53.83	105.78	-5.08	-5.19	-6.77	-3.69	-3.58	-4.12
	1.298	34.38	50.16	94.73	36.56	53.78	105.67	-4.50	-4.97	-6.73	-3.59	-3.58	-4.13
	2.596	33.94	49.92	94.70	36.27	53.70	105.77	-4.87	-5.14	-6.91	-3.87	-3.68	-4.13
0.010	0.000	81.42	124.52	257.53	84.49	130.13	278.03	-4.81	-4.91	-6.41	-3.79	-3.38	-3.50
	1.298	81.36	124.39	257.64	84.35	129.76	277.59	-4.89	-5.07	-6.46	-3.93	-3.67	-3.84
	2.596	81.76	124.58	257.59	84.71	130.18	278.23	-4.52	-4.82	-6.39	-3.67	-3.40	-3.51

Note: The simulated log-prices are obtained from Model B in eq. (23) with $\alpha = 0.023$, $\beta = 0.943$, $\nu = 0.137$, $\mu = 1.530$ and $\lambda_0 = 0.010$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The log-prices are contaminated with a bid-ask bounce effect described in eq. (24). The RV in eq. (3), the Hansen and Lunde estimator in eq. (6), the RRG in eq. (13), the RRG with Random Sampling correction in eq. (19), the RRG_{BC} as in eq. (14) and the $RRG_{RS,BC}$ as in eq. (20) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in five different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$, the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$, the Realized Range computed with $\widetilde{\lambda}_{2,m}$, the Realized Range with Random Sampling correction computed with $\widetilde{\lambda}_{2,m_{RS}}$. In boldface the minimum bias in absolute value for each parameter configuration.

Table 10: Simulation results for model B with bid-ask bounce: RMSE

ξ	λ_1	Realized Volatility			Hansen-Lunde			Daily Range				
		30 min	15 min	5 min	30 min	15 min	5 min	Park.	Real.	RS	Real. BC	RS BC
0.003	0.000	0.38	0.27	0.20	0.66	0.47	0.28	0.63	0.64	0.66	0.61	0.62
	1.298	0.38	0.28	0.20	0.68	0.47	0.28	0.64	0.65	0.65	0.63	0.61
	2.596	0.38	0.28	0.20	0.68	0.48	0.28	0.65	0.66	0.64	0.62	0.60
0.005	0.000	0.40	0.31	0.35	0.69	0.49	0.30	0.67	0.68	0.68	0.62	0.62
	1.298	0.40	0.31	0.36	0.69	0.49	0.30	0.67	0.68	0.69	0.61	0.63
	2.596	0.39	0.31	0.36	0.68	0.48	0.29	0.64	0.65	0.67	0.63	0.62
0.010	0.000	0.48	0.53	1.22	0.71	0.53	0.41	0.70	0.71	0.74	0.62	0.63
	1.298	0.49	0.53	1.22	0.72	0.53	0.41	0.70	0.71	0.72	0.61	0.62
	2.596	0.49	0.54	1.22	0.72	0.53	0.40	0.70	0.71	0.74	0.61	0.63

ξ	λ_1	Realized Range			Realized Range RS			Realized Range BC			Realized Range RS BC		
		30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min
0.003	0.000	0.25	0.28	0.45	0.27	0.31	0.53	0.18	0.14	0.12	0.18	0.14	0.11
	1.298	0.25	0.28	0.45	0.27	0.31	0.53	0.18	0.14	0.12	0.18	0.14	0.11
	2.596	0.25	0.28	0.45	0.27	0.31	0.53	0.18	0.14	0.12	0.18	0.14	0.11
0.005	0.000	0.39	0.52	0.95	0.42	0.56	1.06	0.18	0.14	0.12	0.18	0.14	0.11
	1.298	0.40	0.52	0.95	0.42	0.56	1.06	0.18	0.14	0.12	0.18	0.14	0.11
	2.596	0.39	0.52	0.95	0.42	0.56	1.06	0.18	0.14	0.12	0.18	0.14	0.11
0.010	0.000	0.85	1.26	2.58	0.88	1.31	2.78	0.18	0.14	0.13	0.18	0.14	0.12
	1.298	0.85	1.26	2.58	0.88	1.31	2.78	0.18	0.14	0.13	0.18	0.14	0.12
	2.596	0.85	1.26	2.58	0.88	1.32	2.79	0.18	0.14	0.13	0.18	0.14	0.12

Note: The simulated log-prices are obtained from Model B in eq. (23) with $\alpha = 0.023$, $\beta = 0.943$, $\nu = 0.137$, $\mu = 1.530$ and $\lambda_0 = 0.010$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The log-prices are contaminated with a bid-ask bounce effect described in eq. (24). The RV in eq. (3), the Hansen and Lunde estimator in eq. (6), the RRG in eq. (13), the RRG with Random Sampling correction in eq. (19), the RRG_{BC} as in eq. (14) and the $RRG_{RS,BC}$ as in eq. (20) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in five different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$, the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$, the Realized Range computed with $\widetilde{\lambda}_{2,m}$, the Realized Range with Random Sampling correction computed with $\widetilde{\lambda}_{2,m_{RS}}$. In boldface the minimum bias in absolute value for each parameter configuration.

Table 11: Simulation results for model A with bid-ask bounce and price discreteness: relative percentage bias

d	ξ	$Tick$	Realized Volatility			Hansen-Lunde			Daily Range				
			30 min	15 min	5 min	30 min	15 min	5 min	Park.	Real.	RS	Real. BC	RS BC
0.0	0.003	0.125	23.34	47.80	144.11	**0.01**	-0.39	0.10	18.52	20.26	22.58	-5.47	-3.53
		0.062	16.59	34.16	102.44	**-0.37**	-0.44	-0.46	14.81	16.49	17.18	-6.22	-5.07
		0.010	13.03	24.87	74.52	1.14	0.99	**0.01**	13.82	15.49	14.74	-3.70	-4.63
	0.005	0.125	51.54	102.14	306.88	**0.54**	1.24	0.93	28.84	30.73	32.62	-6.09	-4.56
		0.062	41.53	82.60	246.50	0.46	**-0.07**	0.33	26.43	28.29	28.53	-3.96	-4.84
		0.010	33.82	67.82	201.86	-1.24	-0.85	**-0.43**	22.56	24.36	26.44	-5.92	-4.02
	0.010	0.125	164.91	328.43	985.51	-1.79	**0.05**	-1.68	58.82	61.15	62.60	-6.25	-5.09
		0.062	145.59	293.09	875.18	2.60	**0.27**	-0.31	56.37	58.66	58.71	-5.75	-4.76
		0.010	131.36	263.27	791.99	**0.21**	-1.28	-2.01	51.70	53.92	56.47	-5.82	-4.00
0.3	0.003	0.125	24.58	48.78	144.73	1.21	0.52	**-0.18**	18.66	20.40	21.78	-6.43	-4.25
		0.062	16.79	33.97	102.91	-1.40	-0.70	**-0.47**	14.49	16.17	18.84	-6.68	-3.60
		0.010	12.17	24.82	74.30	-0.51	-0.67	**0.09**	12.42	14.07	15.03	-6.26	-4.35
	0.005	0.125	51.26	102.94	307.64	1.08	**0.52**	-1.00	29.91	31.81	32.14	-5.02	-4.98
		0.062	41.01	82.14	246.15	-1.35	0.12	**0.07**	25.96	27.80	28.91	-6.07	-4.53
		0.010	34.77	68.45	202.09	**-0.08**	1.04	0.42	24.05	25.87	25.07	-6.55	-5.23
	0.010	0.125	164.38	327.83	988.01	-1.10	1.12	**-0.25**	60.26	62.61	63.71	-5.55	-4.18
		0.062	146.06	292.95	875.32	**-0.49**	-0.89	-0.82	53.90	56.16	59.00	-4.80	-4.57
		0.010	131.94	263.21	789.58	**-0.05**	1.56	-0.73	51.73	53.95	55.29	-5.33	-4.82

d	ξ	$Tick$	Realized Range			Realized Range RS			Realized Range BC			Realized Range RS BC		
			30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min
0.0	0.003	0.125	91.15	140.17	294.09	94.55	146.22	316.33	-5.48	-5.79	-7.13	-4.48	-4.40	-4.66
		0.062	74.26	113.33	231.99	77.19	118.17	250.60	-5.14	-5.17	-6.98	-4.21	-4.11	-4.41
		0.010	61.51	92.25	184.58	63.51	96.65	200.53	-4.46	-4.87	-6.69	-4.05	-3.66	-4.03
	0.005	0.125	149.48	235.74	523.74	153.68	244.15	558.30	-4.88	-5.20	-5.46	-3.94	-3.71	-3.33
		0.062	130.45	204.03	444.98	133.97	211.13	475.52	-4.42	-4.64	-6.01	-3.74	-3.48	-3.30
		0.010	114.62	178.34	383.22	118.16	185.12	410.51	-4.60	-4.64	-6.17	-3.72	-3.34	-3.48
	0.010	0.125	329.94	547.79	1331.61	336.88	563.62	1412.43	-4.05	-3.70	-3.45	-3.45	-2.70	-0.90
		0.062	304.62	502.91	1210.72	310.82	517.38	1285.01	-4.14	-4.05	-4.08	-3.24	-2.48	-1.16
		0.010	283.79	466.25	1114.48	290.07	479.90	1182.41	-3.98	-3.90	-4.54	-3.20	-2.84	-2.09
0.3	0.003	0.125	91.40	140.13	293.96	94.26	145.78	315.90	-5.36	-5.91	-7.38	-4.70	-4.72	-4.91
		0.062	74.24	112.90	231.94	77.79	118.64	250.67	-5.15	-5.48	-7.02	-3.74	-3.76	-4.27
		0.010	60.89	91.74	184.53	63.63	96.58	200.55	-4.89	-5.15	-6.64	-3.93	-3.69	-3.96
	0.005	0.125	149.31	235.62	523.48	153.33	244.05	558.38	-5.07	-5.38	-5.75	-4.18	-3.81	-3.30
		0.062	129.77	203.52	444.75	133.95	211.07	475.24	-4.89	-4.95	-6.05	-3.73	-3.48	-3.39
		0.010	115.22	178.63	383.32	117.83	184.91	410.07	-4.19	-4.49	-6.05	-3.95	-3.49	-3.72
	0.010	0.125	329.94	547.69	1331.97	337.18	564.06	1413.07	-4.07	-3.77	-3.05	-3.18	-2.30	-0.29
		0.062	304.01	502.55	1211.19	310.95	517.49	1285.03	-4.22	-3.90	-4.16	-3.26	-2.57	-1.37
		0.010	283.86	466.65	1114.55	289.87	479.96	1182.52	-4.02	-3.86	-4.57	-3.13	-2.50	-1.53

Note: The simulated log-prices are obtained from Model A in eq. (22) with $\beta = -1$, $\alpha = 0.0062$ and $\nu = 0.1122$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The log-prices are contaminated with a bid-ask bounce effect described in eq. (24) and with price discreteness as in eq. (25). The RV in eq. (3), the Hansen and Lunde estimator in eq. (6), the RRG in eq. (13), the RRG with Random Sampling correction in eq. (19), the RRG_{BC} as in eq. (14) and the $RRG_{RS,BC}$ as in eq. (20) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in five different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$, the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$, the Realized Range computed with $\widetilde{\lambda}_{2,m}$, the Realized Range with Random Sampling correction computed with $\widetilde{\lambda}_{2,m_{RS}}$. In boldface the minimum bias in absolute value for each parameter configuration.

Table 12: Simulation results for model A with bid-ask bounce and price discreteness: RMSE

d	ξ	$Tick$	Realized Volatility			Hansen-Lunde			Daily Range				
			30 min	15 min	5 min	30 min	15 min	5 min	Park.	Real.	RS	Real. BC	RS BC
0.0	0.003	0.125	0.51	0.61	1.48	0.72	0.55	0.44	0.71	0.72	0.75	0.61	0.63
		0.062	0.47	0.49	1.07	0.72	0.53	0.38	0.70	0.72	0.70	0.61	0.61
		0.010	0.45	0.41	0.79	0.72	0.52	0.35	0.69	0.71	0.70	0.63	0.62
	0.005	0.125	0.76	1.14	3.12	0.81	0.67	0.67	0.77	0.78	0.81	0.61	0.63
		0.062	0.67	0.95	2.51	0.78	0.63	0.59	0.76	0.78	0.78	0.63	0.62
		0.010	0.60	0.80	2.06	0.76	0.59	0.52	0.74	0.75	0.78	0.60	0.63
	0.010	0.125	1.88	3.43	9.94	1.17	1.24	1.69	0.98	1.00	1.01	0.61	0.61
		0.062	1.68	3.06	8.83	1.13	1.15	1.54	0.97	0.99	1.00	0.62	0.63
		0.010	1.54	2.76	7.99	1.08	1.07	1.42	0.94	0.96	0.97	0.61	0.63
0.3	0.003	0.125	0.53	0.62	1.49	0.74	0.57	0.44	0.72	0.74	0.74	0.61	0.62
		0.062	0.47	0.49	1.07	0.71	0.52	0.38	0.68	0.70	0.73	0.61	0.63
		0.010	0.44	0.41	0.79	0.71	0.51	0.35	0.69	0.71	0.72	0.61	0.63
	0.005	0.125	0.75	1.15	3.12	0.81	0.67	0.67	0.79	0.81	0.80	0.63	0.62
		0.062	0.66	0.94	2.50	0.78	0.62	0.59	0.76	0.78	0.77	0.61	0.62
		0.010	0.61	0.81	2.06	0.76	0.60	0.52	0.74	0.76	0.75	0.61	0.62
	0.010	0.125	1.87	3.42	9.97	1.18	1.23	1.70	0.99	1.02	1.03	0.61	0.62
		0.062	1.68	3.06	8.83	1.13	1.14	1.54	0.95	0.97	1.01	0.63	0.64
		0.010	1.54	2.76	7.97	1.07	1.06	1.40	0.92	0.94	0.98	0.62	0.63

d	ξ	ν	Realized Range			Realized Range RS			Realized Range BC			Realized Range RS BC		
			30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min	30 min	15 min	5 min
0.0	0.003	0.125	0.94	1.42	2.94	0.98	1.48	3.17	0.19	0.15	0.14	0.19	0.15	0.125
		0.062	0.78	1.15	2.32	0.81	1.20	2.51	0.18	0.14	0.13	0.18	0.14	0.12
		0.010	0.65	0.94	1.85	0.67	0.98	2.01	0.18	0.14	0.12	0.19	0.14	0.11
	0.005	0.125	1.52	2.37	5.24	1.56	2.45	5.59	0.19	0.16	0.16	0.19	0.15	0.16
		0.062	1.33	2.05	4.45	1.37	2.12	4.76	0.19	0.15	0.15	0.19	0.15	0.15
		0.010	1.17	1.79	3.84	1.21	1.86	4.11	0.18	0.15	0.14	0.18	0.15	0.14
	0.010	0.125	3.32	5.49	13.32	3.39	5.65	14.13	0.20	0.18	0.24	0.20	0.18	0.24
		0.062	3.07	5.04	12.11	3.13	5.18	12.85	0.19	0.17	0.23	0.20	0.18	0.23
		0.010	2.86	4.67	11.15	2.92	4.81	11.83	0.19	0.17	0.22	0.20	0.17	0.22
0.3	0.003	0.125	0.95	1.41	2.94	0.97	1.47	3.16	0.19	0.15	0.14	0.19	0.15	0.13
		0.062	0.78	1.14	2.32	0.81	1.20	2.51	0.18	0.14	0.13	0.18	0.14	0.12
		0.010	0.65	0.93	1.85	0.67	0.98	2.01	0.18	0.14	0.12	0.18	0.14	0.11
	0.005	0.125	1.52	2.37	5.24	1.56	2.45	5.59	0.19	0.16	0.16	0.19	0.15	0.16
		0.062	1.32	2.05	4.45	1.36	2.12	4.76	0.18	0.15	0.15	0.18	0.15	0.15
		0.010	1.18	1.80	3.84	1.21	1.86	4.10	0.19	0.15	0.14	0.18	0.14	0.14
	0.010	0.125	3.32	5.49	13.32	3.39	5.65	14.13	0.20	0.18	0.24	0.20	0.18	0.24
		0.062	3.06	5.03	12.11	3.13	5.18	12.85	0.20	0.18	0.23	0.20	0.18	0.23
		0.010	2.86	4.68	11.15	2.92	4.81	11.83	0.19	0.17	0.22	0.19	0.17	0.22

Note: The simulated log-prices are obtained from Model A in eq. (22) with $\beta = -1$, $\alpha = 0.0062$ and $\nu = 0.1122$; 10,000 independent days with 25,200 log-prices each (one price per second for 7 hours) are simulated and the interval between two successive available log-prices is modeled as a random variable exponentially distributed with mean $\tau = 5$ seconds. The log-prices are contaminated with a bid-ask bounce effect described in eq. (24) and with price discreteness as in eq. (25). The RV in eq. (3), the Hansen and Lunde estimator in eq. (6), the RRG in eq. (13), the RRG with Random Sampling correction in eq. (19), the RRG_{BC} as in eq. (14) and the $RRG_{RS,BC}$ as in eq. (20) estimators are computed with 30 min., 15 min. and 5 min. sampling frequencies; the Daily Range is implemented in five different versions: the Parkinson's in eq. (8), the Realized Range computed with $\lambda_{2,m}$, the Realized Range with Random Sampling correction computed with $\lambda_{2,m_{RS}}$, the Realized Range computed with $\widetilde{\lambda}_{2,m}$, the Realized Range with Random Sampling correction computed with $\widetilde{\lambda}_{2,m_{RS}}$. In boldface the minimum bias in absolute value for each parameter configuration.

Table 13: Long memory estimation

α	$1+d$	$\widehat{1+d}$ IV 1sec.	30sec.	5min.	RV 30sec.	1min.	5min.	10min.	15min.	20min.	30min.	45min.	60min.
0.0000	1.00	1.012	1.012	1.012	0.765	0.724	0.483	0.437	0.328	0.351	0.286	0.357	0.239
0.0000	1.15	1.082	1.082	1.082	0.782	0.690	0.446	0.372	0.307	0.319	0.235	0.235	0.209
0.0000	1.30	1.278	1.278	1.278	0.717	0.637	0.394	0.329	0.292	0.261	0.235	0.280	0.107
0.0000	1.45	1.565	1.565	1.565	0.809	0.711	0.548	0.460	0.432	0.426	0.330	0.377	0.298
0.0124	1.00	0.949	0.949	0.949	0.724	0.620	0.489	0.417	0.408	0.381	0.334	0.364	0.241
0.0124	1.15	1.140	1.140	1.140	0.863	0.797	0.667	0.602	0.613	0.444	0.436	0.459	0.376
0.0124	1.30	1.346	1.346	1.346	0.819	0.680	0.463	0.399	0.344	0.327	0.302	0.319	0.139
0.0124	1.45	1.438	1.438	1.438	0.683	0.661	0.593	0.569	0.548	0.443	0.408	0.452	0.319
α	$1+d$	$\widehat{1+d}$ IV 1sec.	30sec.	5min.	RRG 30sec.	1min.	5min.	10min.	15min.	20min.	30min.	45min.	60min.
0.0000	1.00	1.012	1.012	1.012	0.884	0.845	0.678	0.612	0.522	0.516	0.465	0.458	0.394
0.0000	1.15	1.082	1.082	1.082	0.948	0.870	0.677	0.593	0.508	0.490	0.419	0.379	0.346
0.0000	1.30	1.278	1.278	1.278	0.899	0.791	0.645	0.594	0.435	0.445	0.386	0.346	0.296
0.0000	1.45	1.565	1.565	1.565	0.947	0.888	0.730	0.631	0.582	0.604	0.503	0.472	0.480
0.0124	1.00	0.949	0.949	0.949	0.858	0.778	0.618	0.572	0.553	0.538	0.465	0.434	0.398
0.0124	1.15	1.140	1.140	1.140	1.013	0.935	0.743	0.698	0.680	0.681	0.657	0.626	0.614
0.0124	1.30	1.346	1.346	1.346	0.975	0.876	0.652	0.628	0.622	0.604	0.456	0.420	0.340
0.0124	1.45	1.438	1.438	1.438	0.715	0.696	0.646	0.632	0.633	0.613	0.622	0.587	0.575

Note: The simulated log-prices are obtained from Model A in eq. (22) with $\beta = -1$ and $\nu = 0.1122$; the artificial time series is made of 2,000 days with 86,400 log-prices each (one price per second for 24 hours). The RV in eq. (3) and the RRG in eq. (13) are computed with 30 sec., 1 min., 5 min., 10 min., 15 min., 20 min., 30 min., 45 min. and 60 min. sampling frequencies. To compute the IV we sum the instantaneous volatilities every second, every 30 seconds and every 5 minutes. With each IV estimator we obtain series of 2000 estimated daily integrated volatilities and then the long memory parameter is estimated with the Exact Local Whittle estimator (Shimotsu and Phillips (2005)). In the estimation $m = 2000^{0.6} \approx 96$ thus the asymptotic standard error is equal to 0.051.

6 Empirical illustration

In order to explore empirically the IV estimators so far considered we compute daily IV estimates using tick-by-tick data for the IBM stock over the period from January 1, 2004 to December 31, 2006 for a total of 755 trading days. The data were extracted from the TAQ database and we report the result for the transaction data. The raw data were filtered for major irregularities (e.g. entries posted outside the NYSE opening hours or coupled with a NYSE error code, for which only the corrected entries were considered). The average number of prices per trading day after filtering is about 4605, about one every 5 seconds.

Table 14: Results for IBM tick-by-tick data

	Mean	Var	Skew	Kurt	Min	Max
RV^{30}	1.932	2.309	2.806	16.555	0.121	15.832
RV^{15}	2.008	2.338	3.510	23.762	0.230	16.730
RV^{5}	2.143	1.625	2.614	16.092	0.383	12.461
HL^{30}	1.941	3.656	2.345	10.699	0.263	13.855
HL^{15}	1.869	2.550	2.476	12.344	0.224	14.069
HL^{5}	1.907	2.036	3.336	21.582	0.303	13.399
$Park$	1.841	2.916	5.451	62.982	0.122	26.195
RRG^{Daily}	1.869	3.003	5.450	62.965	0.123	26.579
RRG^{Daily}_{RS}	1.755	2.676	5.406	62.480	0.111	25.073
RRG^{Daily}_{BC}	1.735	2.613	5.404	62.444	0.110	24.776
$RRG^{Daily}_{BC,RS}$	1.755	2.676	5.406	62.480	0.111	25.073
RRG^{30}	2.037	1.418	2.576	14.704	0.317	11.176
RRG^{15}	2.085	1.350	2.894	19.211	0.340	11.941
RRG^{5}	2.121	1.079	2.081	12.052	0.343	10.042
RRG^{30}_{RS}	2.070	1.461	2.582	14.756	0.321	11.443
RRG^{15}_{RS}	2.144	1.415	2.914	19.428	0.349	12.167
RRG^{5}_{RS}	2.261	1.204	2.052	11.835	0.370	10.624
RRG^{30}_{BC}	1.821	1.122	2.443	13.542	0.278	9.609
RRG^{15}_{BC}	1.822	1.011	2.735	18.201	0.284	10.681
RRG^{5}_{BC}	1.755	0.732	1.583	7.124	0.237	6.667
$RRG^{30}_{BC,RS}$	1.834	1.135	2.443	13.553	0.281	9.689
$RRG^{15}_{BC,RS}$	1.843	1.031	2.736	18.252	0.288	10.825
$RRG^{5}_{BC,RS}$	1.798	0.762	1.571	7.069	0.243	6.799

Note: The Table reports the IV estimates for the sample period from January 1, 2004 to December 31, 2006. The volatilities are expressed in annualized percentages. We provide the mean, variance, skewness, kurtosis, minimum and maximum for each estimator.

We compute RV and HL estimates using the imputation methods, while for the RRG and RRG_{BC} we eliminate, to compute the λ, the price repetitions and the price reversals, as suggested in Christensen and Podolskij (2007). Moreover, to compute the λ for the RRG_{RS} and $RRG_{BC,RS}$ we assume that the underlying DGP has a granularity of 1 second.

Table 14 reports the descriptive statistics for the IV estimators. The sample mean of the bias corrected estimators is lower than the sample mean of the respective non bias corrected counterparts, suggesting that the effect of microstructure noise is not negligible even considering an extremely liquid stock at sampling frequencies not too extreme. Intra-daily Realized Range estimators, exploiting more sample information than Daily Range and Realized Volatility measures, show smaller sample variances, skewness and kurtosis.

The volatility signature plots in Figure 4 graph the average sample values of RV, HL, RRG, RRG_{BC}, RRG_{RS}, $RRG_{BC,RS}$ for different sampling frequencies. These different IV estimates should, in the absence of microstructure frictions, provide the same measure of the IV independently of the sampling frequency. In Figure 4 we notice how the dispersion across the estimators lowers

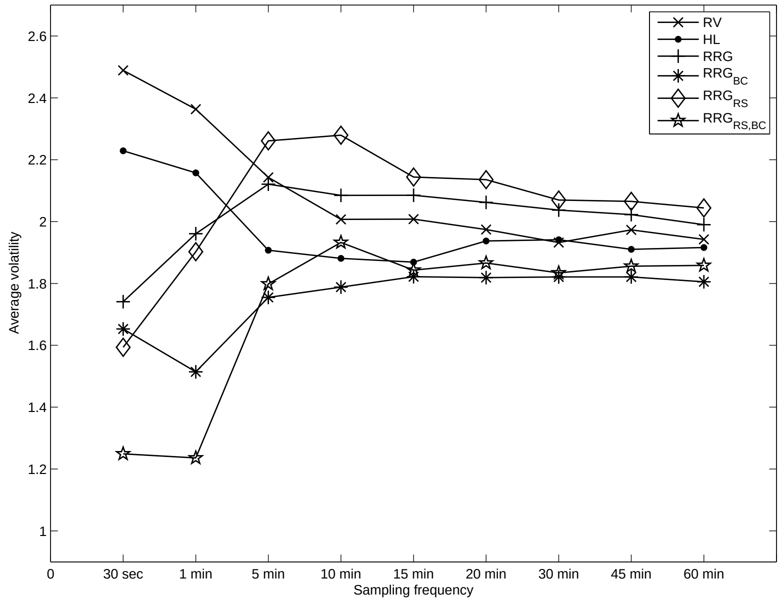

Figure 4: Volatility signature plots for IBM. The Figure displays volatility signature plots for IBM transaction prices using daily averages over the sample period January 1, 2004 - December 31, 2006 for RV, HL, RRG, RRG_{BC}, RRG_{RS}, $RRG_{BC,RS}$. Volatility signature plots graph the average sample values (measured over a long time span) of volatility estimators for different sampling frequencies.

as we decrease the sampling frequency. The bias corrected estimators provide substantially similar results when sampling frequencies lower than 5 minutes are used, while the non bias corrected estimators stabilize at sampling frequencies lower than 30 minutes.

HL estimates are substantially larger than the $RRG_{BC,RS}$ and RRG_{BC} estimates for sampling frequencies of 30 seconds and 1 minute. This seems to suggest that the HL estimator computed at very high frequencies is not totally robust to the microstructure noise. On the other hand, the strong underestimation of $RRG_{BC,RS}$ and RRG_{BC} can be explained with the numerical instability of the λ parameters when there are only a few data point in the sub-intervals considered, as reported in Christensen, Podolskij, and Vetter (2009).

Finally, Figure 5 plots the Exact Local Whittle estimates of the long memory parameter computed for IBM daily IV estimated using RV, HL, RRG, and RRG_{BC} at several sampling frequencies. It appears that the estimates using RV and HL are extremely dependent on the sampling frequency used in the IV estimation, while the estimates using RRG and RRG_{BC} look much more stable in accordance with the simulation results in Table 13. This confirms how the inferences on the long memory features of IV estimates depend crucially on the selected sampling frequency, particularly when using Realized Volatility estimators.

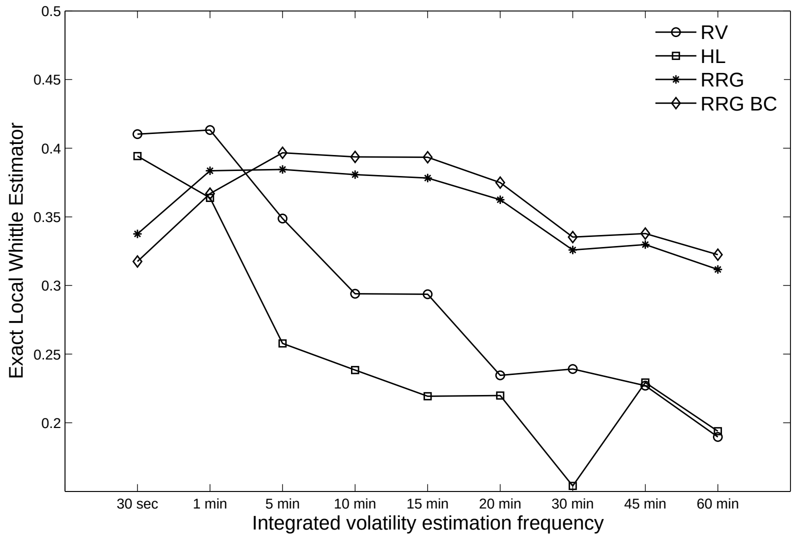

Figure 5: Long memory parameter estimation of IBM Integrated Variance. The Figure displays Exact Local Whittle estimates for IBM IV estimated using RV, HL, RRG, RRG_{BC} at several sampling frequencies.

7 Conclusions

In this paper we evaluate the finite sample properties of the Realized Range estimators class recently introduced by Christensen and Podolskij (2007) and Christensen, Podolskij, and Vetter (2009) through extensive Monte Carlo simulations under different Data Generating Processes for the instantaneous volatility.

The Realized Range estimator is particularly appealing when there is microstructure noise, which prevents from using the whole record of prices to compute Realized Volatility, since it exploits a larger amount of information.

Christensen, Podolskij, and Vetter (2009) prove the asymptotic distribution of a bias-corrected version of the Realized Range estimator under i.i.d. microstructure noise assumptions but the finite sample properties of the estimator are not fully analyzed.

We analyze the robustness of these estimators to long memory and jumps in the volatility, and to bid-ask bounce, discreteness and missing observations in the log-price process. The Realized Range estimators turn out to be strongly downward biased under all DGPs, when we consider missing observations leading to irregular sampling. In order to deal with this problem we introduce an heuristically justified correction which is able to strongly reduce this bias without affecting the estimators precisions.

We compare the Realized Range estimators with Realized Volatility, Hansen and Lunde bias correction, and different implementations of the Daily Range. With no microstructure noise, the simulations results reveal the existence of a trade-off between bias and RMSE, which is due to the fact that we systematically underestimate the true range because the prices are not fully observed. Our correction, following Christensen and Podolskij (2007), eliminates this trade-off, hence allowing to compute the Realized Range using the frequency coupled with the smallest RMSE.

With bid-ask bounce, on the other hand, the Hansen and Lunde correction seems to be preferred in almost all cases, with the notable exception of strong long memory in the instantaneous volatility process. The bias-corrected RRGs are steadily downward biased across all processes and all parameter configurations considered but the bias is unaffected by the simulation setup. However, the bias-corrected RRGs are three times more efficient than the HL estimator.

We focus also on the estimation of the long memory parameter daily IV estimates computed from simulated time series of high frequency data. The interesting result here is that all the estimators considered lead to a strong underestimation of the underlying true long memory parameter. Moreover, this bias seems to be connected to the sampling variance of the estimators, thus to the sampling frequency used in the IV estimation.

Finally, in a brief empirical example using transaction data on IBM stock, we show how the daily IV estimates obtained from Realized Volatility estimators have larger variance, skewness and kurtosis than those obtained from Realized Range estimators. Volatility signature plots confirm the presence of strong microstructure noise at sampling frequencies higher than 5 minutes.

In future work, we intend to look at the option pricing implications of alternative Realized Range measures to fully understand the economic impact of the bias-variance trade-off.

References

Amihud, Y., and H. Mendelson (1987): "Trading mechanism and stock returns: an empirical investigation," *Journal of Finance*, 42, 533–553.

Andersen, T., L. Benzoni, and J. Lund (2002): "An empirical investigation of continuous-time equity returns models," *Journal of Finance*, 57, 1239–1284.

Andersen, T., and T. Bollerslev (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

Andersen, T. G., and L. Benzoni (2009): "Realized Volatility," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch. Springer-Verlag.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2002): "Parametric and Nonparametric Volatility Measurement," NBER Technical Working Papers 0279, National Bureau of Economic Research, Inc.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and Ebens (2001): "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61, 43–76.

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

——— (2003): "Modeling and forecasting realized volatility," *Econometrica*, 71, 579–625.

Aït-Sahalia, Y., P. Mikland, and L. Zhang (2005): "How often to sample a continuous time process in the presence of market microstructure noise," *Review of Financial Studies*, 18, 351–416.

Bandi, F., and J. Russell (2006): "Separating market microstructure noise from volatility," *Journal of Financial Economics*, 79, 655–692.

——— (2008): "Microstructure noise, realized volatility, and optimal sampling," *Review of Economic Studies*, 75(2).

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

——— (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of Royal Statistical Society. Series B*, 64, 253–280.

——— (2004): "Econometric analysis of realised covariation: High frequency based covariance, regression and correlation in financial economics," *Econometrica*, 72.

——— (2007): "Variation, jumps, market frictions and high frequency data in financial econometrics," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson, vol. 3 of *Econometric Society Monographs*. Cambridge University Press.

Black, F. (1976): "Noise," *Journal of Finance*, 41, 529–543.

Bollen, B., and B. Inden (2002): "A Discrete-Time Model for Daily S&P 500 Returns and Realized Variations: Jumps and Leverage Effects," *Journal of Empirical Finance*, 9, 551–562.

Christensen, K., and M. Podolskij (2007): "Realized range-based estimation of integrated variance," *Journal of Econometrics*, (141), 323–349.

Christensen, K., M. Podolskij, and M. Vetter (2009): "Bias-correcting the realized range-based variance in the presence of market microstructure noise," *Finance and Stochastics*, 13(2), 239–268.

Comte, F., and E. Renault (1996): "Long-memory continuous-time models," *Journal of Econometrics*, 73, 101–149.

——— (1998): "Long-memory in continuous-time stochastic volatility models," *Mathematical Finance*, 8, 291–323.

Corsi, F., G. Zumbach, U. A. Muller, and M. M. Dacorogna (2001): "Consistent high-precision volatility from high-frequency data," *Economic Notes*, 30, 183–204.

Eraker, B. (2004): "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance*, 59(2), 1367–1403.

Eraker, B., M. S. Johannes, and N. G. Polson (2003): "The impact of jumps in returns and volatility," *Journal of Finance*, 53(3), 1269–1300.

Frijns, and Lehnert (2004): "Realized variance in the presence of non-IID microstructure noise: A structural approach.," Discussion paper, Limburg Institute of Financial Economics.

Griffin, J., and R. Oomen (2008): "Sampling returns for realized variance calculations: tick time or transcation time?," *Econometric Reviews*, 27(1-3), 230–253.

Hansen, P. R., and A. Lunde (2005): "A Realized Variance for the whole day based on intermittent high-frequency data," *Journal of Financial Econometrics*, 3(4), 525–554.

——— (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24(2), 127–161.

Harris, L. (1990): "Estimation of stock variance and serial covariance from discrete observations," *Journal of Financial & Quantitative Analysis*, (25), 291–306.

——— (1991): "Stock price clustering and discreteness," *Review of Financial Studies*, (4), 389–415.

Lieberman, O., and P. Phillips (2008): "Refined inference on long memory in realized volatility," *Econometric reviews*, (27), 254–267.

Maheu, J., and T. McCurdy (2002): "Nonlinear features of Realized FX Volatility," *Review of Economics & Statistics*, (84), 668–681.

Marinucci, D., and P. Robinson (1999): "Alternative forms of fractional Brownian motion," *Journal of Statistical Planning and Inference*, 80, 111–122.

Martens, M., and D. van Dijk (2007): "Measuring volatility with the realized range," *Journal of Econometrics*, (138), 181–207.

Martens, M., D. van Dijk, and M. D. Pooter (2004): "Modeling and forecasting S&P 500 volatility: Long memory, structural breaks and nonlinearity," Discussion Paper 067, Tinbergen Institute Discussion Paper.

McAleer, M., and M. Medeiros (2008): "Realized Volatility: a review," *Econometric Reviews*, 27(1-3), 10–45.

Morana, C., and A. Beltratti (2004): "Structural change and long-range dependence in volatility of the exchange rates: either, neither or boths?," *Journal of Empirical Finance*, (11), 629–658.

Nielsen, M., and P. Frederiksen (2008): "Finite sample accuracy and choice sampling frequency in integrated volatility estimation," *Journal of Empirical Finance*, (15), 265–286.

Oomen, R. (2005): "Properties of bias-corrected realized variance under alternative sampling schemes," *Journal of Financial Econometrics*, (3), 555–577.

——— (2006): "Properties of realized variance under alternative sampling schemes," *Journal of Business & Economic Statistics*, 24(2), 219–237.

Parkinson, M. (1980): "The extreme value method for estimating the variance of the rate of return," *Journal of Business*, 53(1), 61–65.

Shimotsu, K., and P. Phillips (2005): "Exact Local Whittle estimation of fractional integration," *Annals of Statistics*, 33, 1890–1933.

Thomakos, D., and T. Wang (2003): "Realized volatility in the futures markets," *Journal of Empirical Finance*, (10), 321–353.

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

Zhang, L., P. Mikland, and Y. Aït-Sahalia (2005): "A tale of two time scales: determining integrated volatility with noisy high frequency data," *Journal of the American Statistical Association*, 100, 1394–1411.

Zhou, L. (1996): "High-frequency data and volatility in foreign-exchange rates," *Journal of Business and Economic Statistics*, 14(1), 45–52.