Finance Stoch (2010) 14: 249–283
DOI 10.1007/s00780-009-0120-1

Zero-intelligence realized variance estimation

Jim Gatheral · Roel C.A. Oomen

Received: 19 March 2007 / Accepted: 4 April 2009 / Published online: 26 January 2010

Abstract Given a time series of intra-day tick-by-tick price data, how can realized variance be estimated? The obvious estimator—the sum of squared returns between trades—is biased by microstructure effects such as bid–ask bounce and so in the past, practitioners were advised to drop most of the data and sample at most every five minutes or so. Recently, however, numerous alternative estimators have been developed that make more efficient use of the available data and improve substantially over those based on sparsely sampled returns. Yet, from a practical viewpoint, the choice of which particular estimator to use is not a trivial one because the study of their relative merits has primarily focused on the speed of convergence to their asymptotic distributions, which in itself is not necessarily a reliable guide to finite sample performance (especially when the assumptions on the price or noise process are violated). In this paper we compare a comprehensive set of nineteen realized variance estimators using simulated data from an artificial "zero-intelligence" market that has been shown to mimic some key properties of actual markets. In evaluating the competing estimators, we concentrate on efficiency but also pay attention to implementation, practicality, and robustness. One of our key findings is that for scenarios frequently

The authors wish to thank an anonymous referee and the associate editor, Yacine Aït-Sahalia, Mavira Mancino, Andrew Patton, Simona Sanfelici, Kevin Sheppard, and the participants of the workshop on high-frequency data at Imperial College London for helpful comments and suggestions.

J. Gatheral
Bank of America Merrill Lynch, New York, USA

J. Gatheral
Courant Institute, New York University, New York, USA

R.C.A. Oomen (✉)
Deutsche Bank, London, UK
e-mail: roel.ca.oomen@gmail.com

R.C.A. Oomen
Department of Quantitative Economics, University of Amsterdam, Amsterdam, The Netherlands

encountered in practice, the best variance estimator is not always the one suggested by theory. In fact, an ad hoc implementation of a subsampling estimator, realized kernel, or maximum likelihood realized variance, delivers the best overall result. We make firm practical recommendations on choosing and implementing a realized variance estimator, as well as data sampling.

Keywords Limit order book · Market microstructure noise · Micro-price · Realized variance · Sampling schemes

Mathematics Subject Classification (2000) 62E20 · 62P20

JEL Classification C10 · C80

1 Introduction

Accurate real-time volatility forecasts are needed for many applications, such as the real-time pricing of options and real-time risk management of trading positions. In order to generate a forecast, however, we first need a good estimate of realized variance. On the one hand, without an efficient estimator of realized variance, it is hard to see how the performance of different forecasts can be reliably compared. On the other hand, as shown in a recent elegant paper [9], it seems that a volatility forecasting model generates its best ex ante estimate of realized variance when supplied with its best ex post measure.

Microstructure effects such as bid–ask bounce cause the series of price returns between trades to be autocorrelated; so the obvious estimator of realized variance—the sum of squared returns between trades—is very biased. In the early literature, for example in [5], the general prescription is that high-frequency price series should be sampled at intervals sufficiently long for autocorrelation effects to become insignificant. For a liquid US stock, the recommended interval is typically around 5 minutes. However, such a stock may easily trade 1,000 times or more in a 5-minute interval and in that case, the prescription to sample no more frequently than every 5 minutes would amount to throwing away more than 99.9% of the data. To quote [27], "it is difficult to accept that throwing away data, especially in such quantities, can be an optimal solution." More practically, if our interest is in forecasting volatility over the next five minutes, it seems reasonable to suppose that there would be much more relevant volatility information in the last five minutes of trading than in an equivalently long series of 13 trading days-worth of 5-minute returns.

Many authors have suggested estimators and sampling procedures that are designed to use all of the prices in a tick-by-tick data set. Both the design and the analysis of these estimators are often based on reasonable but perhaps overly simplistic assumptions on the nature of the "microstructure noise" process. A typical assumption is that the efficient price follows a random walk: Trade prices are regarded as observations of the efficient market prices polluted by i.i.d. microstructure noise. However, it is not at all clear that the price process may be neatly decomposed into an "efficient" price and associated noise process, and even if this were possible, it is not clear what the specification of the "microstructure noise" process should be. In addition, much of the emphasis in comparing the relative merits of competing realized variance estimators has been on the rate of convergence to their respective asymptotic distributions.[1] But it is well known that this is not always a reliable guide to finite sample performance. So in practice, when the assumptions on the noise and price process may be violated and the sample size is perhaps not big enough for the asymptotic convergence rate to be the sole criterion for performance, which estimator should we use?

The contribution of the present paper is to shed some light on these issues with the aim to provide practitioners with firm guidelines on how to obtain efficient and robust realized variance estimates. To this end, we study the performance of nineteen different realized variance estimators (chosen as a representative sample of those currently studied in detail elsewhere in the realized volatility literature) in the context of a simulated artificial "zero-intelligence" market that has been shown to mimic some key properties of actual markets. In this market, events—market orders, limit orders and cancelations—arrive randomly in time. As in a real double auction market, a new limit order may be placed at any price level that is worse than the opposite best quote. However, unlike a real market, it is as if market agents choose actions randomly without regard even to the current state of the order book. In this sense, the market is "non-intelligent." Because we can measure the true volatility (i.e., the volatility of returns measured over sufficiently long horizons for the impact of microstructure noise to be absent) in this simulation to any desired accuracy, the relative performance of the competing realized variance measures can be established.

An important aspect of our analysis is the focus on data sampling. We consider three distinct sampling schemes that can all be implemented in practice and give rise to price series with radically different properties, namely (i) the series of trade prices, (ii) the series of mid-quotes, and (iii) the series of micro-prices formed by linear weighting of the best bid and ask price by market depth. The first two of these sampling schemes are self-explanatory. The third, the volume weighted mid-quote or micro-price series, though very familiar to market practitioners, is new in the realized variance literature. Conveniently, the limit-order book model allows us to implement these sampling schemes in an internally consistent manner and gain insights into their properties.

Our main findings can be summarized as follows. In terms of sampling, mid-quote and micro-price data are between 40 to 60 times less noisy than trade data (as measured by the microstructure noise variance), leading to an efficiency gain for realized variance estimation of around 50%. Between the mid-quote and micro-price, the former is weakly preferred: In our zero-intelligence model setup, limit order placement is random so that the information content of market depths is limited and linear weighting may actually add some noise to the sampled price series. Next, with regard to the choice of variance estimators, we find that the realized kernels of [10], the subsampling methods of [26, 27], and the parametric maximum likelihood estimator of [2] perform best overall, both in terms of efficiency and in terms of robustness to

[1]Notable exceptions include [9] that concentrates on the payoff of an option trading strategy based on alternative realized variance measures, and [1, 6], and [16] that consider the forecast accuracy in the presence of noise.

time-varying parameters. However, implementation of the non-parametric methods requires one to choose tuning parameters. If this selection is to be done "optimally" as prescribed by theory, then this requires the evaluation of a highly non-linear function in terms of latent variables such as integrated variance and quarticity which are notoriously difficult to measure. Perhaps more importantly, the method is susceptible to misspecification of the microstructure noise process which constitutes a major drawback from an empirical viewpoint. Similarly, the MLE approach requires a model selection procedure based on the autocorrelation pattern of observed returns. Yet, we find that an ad hoc choice of tuning parameters and a pragmatic model specification upon which to base the MLE gives reasonably efficient, and sometimes even superior, results. Indeed, averaged across sampling schemes and sample sizes, these implementations yield the best all-round performance.

The remainder of this paper is organized as follows. In Sect. 2, we describe the zero-intelligence market. In Sect. 3, we discuss the realized variance measures studied in this paper and provide details on their implementation. In Sect. 4, we outline the simulation design and we present the results, taking care to highlight what we consider to be significant differences in performance as opposed to minor differences that may be disregarded for practical purposes. Section 5 concludes and summarizes our recommendations.

2 The zero-intelligence limit-order book market

Motivated by our desire to model microstructure noise more realistically, we simulate from the limit-order book model of [25], SFGK hereafter. According to the specification of this model, limit orders can be placed at any integer price level p where $-\infty < p < \infty$. It may be natural to think of these price levels as being logarithms of the actual price (which must of course be non-negative). Limit sell orders may be placed at any level greater than the best bid $p^b(t)$ at time t and limit buy orders at any level less than the best offer $p^a(t)$. In particular, just as in real markets, limit orders may be placed inside the spread (if the current spread is greater than one tick). Market orders arrive randomly at rate μ, limit orders (per price level) arrive at rate α and a δ proportion of existing limit orders is canceled. All market orders and limit orders are for one share.

The SFGK model depends on so few parameters that we can use dimensional analysis to predict relationships between statistical properties of the model and its inputs. For example, it is easy to see that the asymptotic book depth far away from the best quote must be given by α/δ. This is because in the steady state, orders arriving into the book must balance orders leaving the book. Far away from the best quote, the probability of a limit order leaving the book as the result of an execution against a market order is very small. Thus we need only to consider new orders arriving balancing existing orders leaving due to cancelation. At a given price level, and for a fixed time interval Δ, the expected number of orders arriving is $\alpha\Delta$ and the expected number of orders canceled is $\delta d\Delta$ where d is the number of shares at the book at that price level. For these to balance, we must have $\alpha\Delta = \delta d\Delta$, or equivalently,

$$d = \frac{\alpha}{\delta}.$$

Although the model is simple to describe, its behavior can be rather complex. For example, suppose there is one share at the best bid when the spread is two ticks. If a new buy order is placed inside the spread one tick below the best offer $p^a(t)$, the best bid $p^b(t)$ increases and the spread decreases to one tick. If a market sell order arrives, the book will revert to its initial state. Now suppose that the same events occur but in the reverse order. The market sell order increases the spread to three ticks and a new limit buy order is placed one tick below the best offer. The spread decreases to one tick. The resulting book shape and even the mid-quote are quite different. Another example that highlights the non-trivial dynamics of the zero-intelligence model is that the distribution of order depth is a function of distance to the best quote. Conditional on no change in the best quotes, order depth would be i.i.d. at any given level of the book deeper than the best quote. In fact, as shown by SFGK, this distribution is Poisson with mean α/δ. However, as the order book evolves, the best quote prices move, and each level in the book retains some memory of having been "visited" by the market price at some prior time. Strictly speaking, order depth distributions are therefore conditional on the entire history of the order book and in particular on the distance to the best quote. Of course, the further away a given level is from the best quote, the closer the order depth distribution will be to Poisson.

The asymptotic behavior of the order book is clearly in contrast to real markets where order depth falls off rapidly away from the best quote. We would therefore not expect the behavior of the SFGK model deep in the order book to accurately mimic a real market. What is more important, at least for realized variance estimation, is that close to the best quote, the SFGK model *does* appear to capture some salient aspects of real markets. For example, a less-developed model in a similar style of [14] was shown in [12] to accurately predict the distribution of bid–offer spreads in the foreign exchange market. Later, [15] estimated μ, α and δ from London Stock Exchange SETS order book data, comparing average actual bid–offer spreads and volatilities against SFGK model predictions and finding agreement to be quite good. As emphasized in [13], there are inevitably some aspects of real markets that zero-intelligence models cannot mimic, most notably autocorrelation of trade signs which arises when "intelligent" traders place orders based on the shape of the order book and trade history. Nevertheless, the SFGK model generates realistic order book behavior, at least to first approximation, and therefore serves as a useful alternative to the more stylized "random walk plus i.i.d. noise" model—not least because it allows us to simulate different price processes (trade prices, mid-quotes, and micro-prices) in an internally consistent manner.

With regard to the actual implementation of the ZI market simulation, we note that it is clearly impossible to simulate order arrivals and cancelations at integer price levels extending from $-\infty$ to $+\infty$. A reasonable simplification is to only consider order arrivals and cancelations in a moving band of price levels centered around the current best quotes. The width B of this band should be chosen conservatively so as to ensure minimal edge effects. Thus, for each allowable price level within the band, the arrival rate of limit orders is α. The probability of an order inside the band being canceled is given by δ times the outstanding number of shares at a given price level. Outside the band, orders may neither arrive nor be canceled. Specifically, the simulation proceeds as follows. At each time $t = 1, \ldots, T$,

- Compute the best bid $p^b(t)$ and best offer $p^a(t)$.
- Compute the number N_b of shares on the bid side of the book from level $p^a(t) - 1$ to level $p^a(t) - B$.
- Compute the number N_a of shares on the offer side of the book from level $p^b(t) + 1$ to level $p^b(t) + B$.
- Set $\delta_b = \delta N_b$; $\delta_a = \delta N_a$.
- Draw a new event according to the relative probabilities $\{\mu/2, \mu/2, B\alpha, B\alpha, \delta_a, \delta_b\}$. These relative probabilities are respectively of a market buy, a market sell, a limit buy at one of the allowable price levels, a limit sell, a cancelation of an existing buy order, and a cancelation of an existing sell order.
- If the selected event is a limit order, draw the relative price level from $\{1, 2, \dots, B\}$.
- If the selected event is a cancelation, select randomly which order within the band to cancel.
- Update the order book and increment t.

This simulation procedure yields a time series of order book data from which we extract the relevant quantities required for our study. Specifically, we store the series of trade prices $\{p_i^*\}_{i=0}^M$ (arising from market orders) as well as the best bid and ask prices with corresponding volume that prevail prior to each trade, $\{p_i^b, p_i^a, v_i^b, v_i^a\}_{i=0}^M$. The latter series permit us to construct the mid-quote, i.e.,

$$q_i^m = \frac{p_i^b + p_i^a}{2}, \quad i = 0, \dots, M, \tag{2.1}$$

and the volume weighted mid-quote or micro-price, i.e.,

$$q_i^v = \frac{v_i^a p_i^b + v_i^b p_i^a}{v_i^a + v_i^b}, \quad i = 0, \dots, M. \tag{2.2}$$

The trade price and mid-quote series are commonly used in the literature. The micro-price, more familiar to practitioners, linearly weighs the bid and ask prices by the volume on the opposite side of the book and can thus be interpreted as the market clearing price when demand and supply curves are linear in price. A priori it is unclear which of these three price series will be most suitable for realized variance measurement, and this will be part of our investigation below.

To conclude our discussion of the SFGK model, consider the descriptive statistics of some simulated price series in Fig. 1. Here, the same model parameters are used as in the simulation study below. Panel A illustrates the three different sampling schemes considered in this paper: the bouncing of trades, staleness of mid-quote, and the more "Brownian-like" evolution of the micro-price are evident. From Panel B we see that, very much like a typical US stock, the first order autocorrelation of trade price changes is about -40% with negligible higher order autocorrelations. In contrast, the first order autocorrelation of mid-quote increments is much smaller at about -5%, reflecting a reduction of microstructure noise. The autocorrelation structure of the micro-price is similar to that of the mid-quote, albeit that higher order autocorrelations appear non-negligible, indicating some dependent noise dynamics. From the histograms of quote and transaction returns in Panel C, the discreteness of

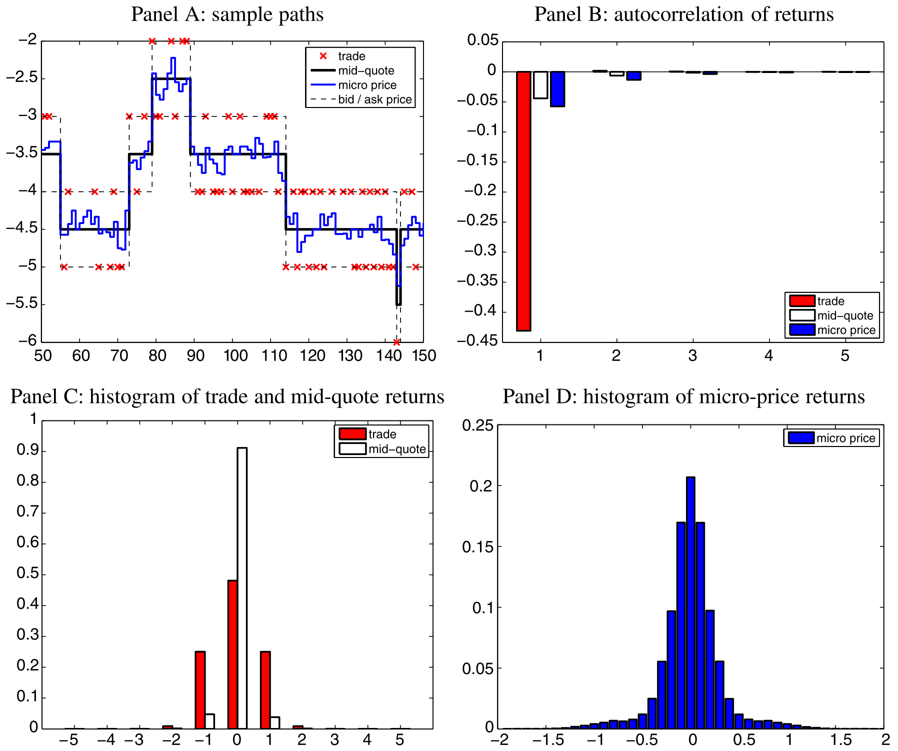

Fig. 1 Properties of the sampled price processes

the data is apparent, and we see that the spread is rarely greater than one tick with 96.5% of all market orders being filled at minimum cost. On average, the mid-quote moves only once for every 10 trades. Finally, Panel D displays the marginal distribution of micro-price returns which appears much closer to a symmetric distribution with continuous support, but with fatter tails than a Gaussian would imply. All in all, the three processes considered here have very distinct properties and, as such, constitute a good basis for a comprehensive robustness analysis of the competing realized variance measures described next.

3 Realized variance estimation

When computing realized variance type quantities, the aim is to use "high-frequency" price observations to construct an efficient ex post estimate of the "low-frequency" return variance, i.e., the variance of returns measured over a horizon that is relatively long compared to the frequency of observation and that is unaffected by the microstructure effects potentially present in the high-frequency prices. In the typical "random walk plus noise" model, the object of econometric interest is the quadratic variation of the efficient price process. In the ZI model considered here, however, there is no well-defined notion of efficient price or quadratic variation thereof. Still, with constant model parameters, the return variance free of microstructure noise can be defined as

$$\sigma^2 := \lim_{m\to\infty} E\big((p_m - p_0)^2/m\big). \tag{3.1}$$

Consequently, for a given sample size M, the object of econometric interest in the realized variance calculation is $\sigma^2 M$. While an analytical expression for σ^2 is unavailable, we can determine its value to arbitrary accuracy by simulations (details are given below).

3.1 Overview of realized variance measures

Let $\{p_i\}_{i=0}^{M}$ denote a time series of logarithmic prices (e.g. trade prices, mid-quote, or micro-prices). For ease of discussion, we think of this series as a realization of prices over a full trading day. Thus, as discussed above, when estimating the variance our interest is explicitly in the ex post daily realized variance. One estimator of this quantity is of course the daily squared return $(p_M - p_0)^2$ but, as pointed out in [4], this is a very noisy estimator and can be improved upon by using intra-day data. In particular, note the relation

$$(p_M - p_0)^2 = \sum_{i=1}^{M} r_i^2 + 2\sum_{k=1}^{M-1} \gamma(k),$$

where $r_i = p_i - p_{i-1}$ and $\gamma(k) = \sum_{i=1}^{M-k} r_i r_{i+k}$. So if returns are serially uncorrelated, then an unbiased and efficient estimate of the daily realized variance can be obtained as the sum of squared intra-day returns. With microstructure noise, however, intra-day returns sampled at the highest frequency will generally exhibit serial correlation (recall Panel B of Fig. 1), thereby invalidating the sum of squared returns as a reliable variance estimator. The estimators we discuss below are all motivated by this reasoning and aim to provide improved measures of the daily realized variance.

To facilitate exposition, we introduce some notation. Given an observed price series $\{p_i\}_{i=0}^{M}$, let

$$\gamma_{h,q}(k) = \sum_{i=1}^{m} (p_{iq+h} - p_{(i-1)q+h})(p_{(i+k)q+h} - p_{(i-1+k)q+h}),$$

where $m = \lfloor (M - h + 1)/q \rfloor - k$. As already mentioned, many of the realized variance measures are derived in a setting where observed prices are modeled as $p_{t_i} = \int_0^{t_i} \sigma(u)\,dW(u) + \omega\varepsilon_{t_i}$ with $W \perp\!\!\!\perp \varepsilon$. Their implementation requires measurements of the latent integrated variance $IV = \int_0^1 \sigma_u^2\,du$, the integrated quarticity $IQ = \int_0^1 \sigma_u^4\,du$, and the noise variance ω^2. Even though these quantities are not well defined in the ZI market model, we can of course still apply the commonly used estimators to the simulated data. Whether or not these estimates will yield useful quantities in such a scenario is explicitly part of our robustness analysis.

The realized variance estimators below naturally divide into seven different classes, where estimators within each class are distinguished by the selection of "tuning" parameters such as sampling frequency or bandwidth. We now list them in turn.

1. Realized variance,

$$RV = \frac{M/q}{\lfloor M/q \rfloor}\gamma_{0,q}(0), \tag{3.2}$$

   (a) at highest sampling frequency $q = 1$,
   (b) at ad hoc sampling frequency of 5 minutes (i.e., $q = M/78$),
   (c) at optimal sampling frequency

$$q^*_{RV} = \left(\frac{2\omega^2}{\sqrt{IQ}}\right)^{2/3} M.$$

   This choice of sampling frequency has been shown to minimize the MSE of RV with i.i.d. noise, see e.g. [2, 8], and [24] with varying assumptions on the "efficient" price process.
2. Bias-corrected RV of [28],

$$ZHOU = \frac{M}{M-q+1}\frac{1}{q}\sum_{h=0}^{q-1}\big(\gamma_{h,q}(0) + 2\gamma_{h,q}(1)\big),$$

   (a) at highest sampling frequency $q = 1$,
   (b) at ad hoc sampling frequency of 5 minutes (i.e., $q = M/78$),
   (c) at optimal sampling frequency of [28], $q^*_{Zhou} = \max\{1, 2\gamma/\sqrt{3}\}$ where $\gamma = M\omega^2/IV$.
3. Two-scale RV of [27],

$$TSRV = (1 - \overline{M}/M)^{-1}\left(\frac{1}{q}\sum_{h=0}^{q-1}\gamma_{h,q}(0) - \frac{\overline{M}}{M}\gamma_{0,1}(0)\right),$$

   where $\overline{M} = (M - q + 1)/q$,
   (a) at ad hoc subsampling frequency of $q = 5$,
   (b) at optimal subsampling frequency of [27],

$$q^*_{ZMA} = \left(\frac{12(\omega^2)^2}{IQ}\right)^{1/3} M^{2/3}.$$

4. Multi-scale RV of [26],

$$MSRV = \sum_{j=1}^{q}\frac{a_j}{j}\sum_{h=0}^{j-1}\gamma_{h,j}(0),$$

   where

$$a^*_j = \big(1 - 1/q^2\big)^{-1}\left(\frac{j}{q^2}h(j/q) - \frac{j}{2q^3}h'(j/q)\right) \quad \text{and} \quad h(x) = 12(x - 1/2),$$

(a) with ad hoc number of $q=5$ subsamples,
(b) with optimal number of subsamples of [26], $q_Z^* = c^* \sqrt{M}$ where

$$c^* = \underset{c}{\operatorname{argmin}} \left\{ 2\frac{52}{35} c IQ + \frac{48}{5} c^{-1} \omega^2 \left(IV + \omega^2/2 \right) + 48 c^{-3} \omega^4 \right\}. \tag{3.3}$$

5. Realized kernel estimator of [10],

$$KRV = \gamma_{0,1}(0) + 2\sum_{s=1}^{q} \kappa\left(\frac{s-1}{q}\right) \gamma_{0,1}(s).$$

For a given choice of kernel $\kappa(x)$, the optimal bandwidth is $q_{BHLS}^* = c^* \sqrt{M}$ where[2]

$$\begin{aligned} c^* = \underset{c}{\operatorname{argmin}} \big\{ & 4c\kappa_\bullet^{0,0} IQ - 8c^{-1} \kappa_\bullet^{0,2} \omega^2 \left(IV + \omega^2/2 \right) \\ & + 4c^{-3} \omega^4 \left(\kappa'''(0) + \kappa_\bullet^{0,4} \right) \big\}, \end{aligned} \tag{3.4}$$

and $\kappa_\bullet^{0,0} = \int_0^1 \kappa(x)^2 \, dx$, $\kappa_\bullet^{0,2} = \int_0^1 \kappa(x) \kappa''(x) \, dx$, $\kappa_\bullet^{0,4} = \int_0^1 \kappa(x) \kappa''''(x) \, dx$.
(a) Modified Tukey–Hanning kernel TH_2 with ad hoc bandwidth $q=5$,
(b) modified Tukey–Hanning kernel TH_2 with optimal bandwidth from [10], $q_{BHLS}^* = c^* \sqrt{M}$,
(c) modified Tukey–Hanning kernel TH_{16} with optimal bandwidth from [10], $q_{BHLS}^* = c^* \sqrt{M}$,
(d) cubic kernel with optimal bandwidth from [10], $q_{BHLS}^* = c^* \sqrt{M}$.

6. Fourier estimator of [20],

$$FE = \sum_{i=1}^{M} \sum_{j=1}^{M} D_q(t_i - t_j) r_i r_j. \tag{3.5}$$

With regular spacing of returns, as is the case here, we have $t_i - t_j = 2\pi(i-j)/M$. From [22], we have that for a given choice of kernel $D_q(\cdot)$, the optimal number of Fourier coefficients q can be optimally set to minimize the MSE of the estimator as

$$\begin{aligned} q_F^* = \underset{q}{\operatorname{argmin}} \bigg\{ & 2\frac{IQ}{M} + 2\frac{IQ}{2q+1} d_q + 4\omega^4 M^2 c_q + 12\omega^4 M c_q \\ & - 16\omega^4 c_q + 8\omega^2 IV + 12\omega^4 \bigg\}, \end{aligned} \tag{3.6}$$

[2] For the TH_2 kernel, i.e., $\kappa(x) = \sin^2\{\frac{\pi}{2}(1-x)^2\}$, we have $\kappa'''(0) = 6\pi^2$, $\kappa_\bullet^{0,0} = 0.218524$, $\kappa_\bullet^{0,2} = -1.71236$, $\kappa_\bullet^{0,4} = -17.4564$. For the TH_{16} kernel, i.e., $\kappa(x) = \sin^2\{\frac{\pi}{2}(1-x)^{16}\}$, we have $\kappa'''(0) = 5760\pi^2$, $\kappa_\bullet^{0,0} = 0.0317324$, $\kappa_\bullet^{0,2} = -10.26423$, $\kappa_\bullet^{0,4} = -42474.9$. For the cubic kernel, i.e., $\kappa(x) = 1 - 3x^2 + 2x^3$, we have $\kappa'''(0) = 12$, $\kappa_\bullet^{0,0} = \frac{13}{35}$, $\kappa_\bullet^{0,2} = -\frac{6}{5}$, $\kappa_\bullet^{0,4} = 0$.

where

$$c_q = 1 + D_q^2(2\pi/M) - 2D_q(2\pi/M),$$

$$d_q = \frac{2(2q+1)}{M^2 - M} \sum_{k=1}^{M-1} (M-k) D_q^2(2\pi k/M).$$ [3]

(a) Re-scaled Dirichlet (DIR) kernel with optimal number of Fourier coefficients q_F^*,

$$D_q(t) = \frac{1}{2q+1} \sum_{s=-q}^{q-1} e^{ist} = \frac{\sin(qt + t/2)}{(2q+1)\sin(t/2)},$$

(b) re-scaled Fejer (FEJ) kernel, as recommended in [21], with optimal number of Fourier coefficients q_F^*,

$$D_q(t) = \frac{1}{q+1} \sum_{s=-q}^{q} \left(1 - \frac{|s|}{q}\right) e^{ist} = \frac{1 - \cos(qt)}{2(q+1)q \sin^2(t/2)}.$$

7. Alternation estimator of [19],

$$ALT = \frac{M_c}{M_r} \gamma_{0,1}(0),$$

where M_r (M_c) are the number of reversals (continuations) in the sample. Note that $M_c = \sum_{i=2}^{M} (I_i^p I_{i-1}^p + I_{i-1}^n I_i^n)$ and $M_r = \sum_{i=2}^{M} (I_i^p I_{i-1}^n + I_{i-1}^p I_i^n)$, where $I_i^p = I(r_i > 0)$ and $I_i^n = I(r_i < 0)$. If there are zero returns in the sample, then these are first removed, i.e., the estimator is implemented using tick data.

8. Maximum likelihood estimator of [2],

$$MLRV = M\widehat{\delta}^2 (1 + \widehat{\theta})^2,$$

where $(\widehat{\eta}, \widehat{\delta^2})$ are the maximum likelihood estimates of an MA(1) model for observed returns, i.e., $r_i = \varepsilon_i + \theta\varepsilon_{i-1}$ where $-1 < \theta < 0$ and the ε_i are serially un-

[3] For computational speed, we calculate the following equivalents of (3.5) for the DIR kernel,

$$FE = \frac{1}{2q+1} \sum_{s=-q}^{q} \mathfrak{F}(r)_M(s) \mathfrak{F}(r)_M(-s),$$

and for the FEJ kernel,

$$FE = \frac{1}{q+1} \sum_{s=-q}^{q} \left(1 - \frac{|s|}{q}\right) \mathfrak{F}(r)_M(s) \mathfrak{F}(r)_M(-s),$$

where $\mathfrak{F}(r)_M(s) = \sum_{j=1}^{M} e^{-ist_{j-1}} r_j$. [22] gives an upper bound to the MSE which differs slightly from the one we use in (3.6). At least in the current setting, (3.6) appears to provide a more accurate approximation of the MSE and delivers better results. For the Dirichlet kernel, $d_q \simeq 1 - 2q/M$.

correlated with mean zero and variance δ^2. Note that within this model structure, we have that $\omega^2 = -\theta\delta^2$.

3.2 Comparison of theoretical properties

In the absence of noise, and with the price process specified as a continuous semi-martingale, it is well known that RV is a consistent and efficient estimator of IV with rate of convergence proportional to $M^{-1/2}$. With noise, however, RV is inconsistent with a bias that grows linearly with the number of sampled observations. ZHOU incorporates the first order autocovariance of returns making it an unbiased estimator of IV with i.i.d. noise. Yet, this estimator remains inconsistent implying that the best efficiency is attained at a finite sampling frequency. In other words, it is "optimal" for ZHOU not to make full use of all available data. [22] shows that for a suitably chosen number of Fourier coefficients the FE is asymptotically unbiased but, like ZHOU, also inconsistent. The TSRV overcomes this problem by cleverly combining two RV measures, one computed at the highest and one at a lower sampling frequency, yielding a consistent estimator of IV that converges at rate $M^{-1/6}$. The rate of this estimator can be improved to $M^{-1/4}$—the fastest attainable in this setting—by using multiple timescales as in the MSRV. The realized kernels provide an equally efficient alternative to the subsampling estimators with rates of convergence of $M^{-1/6}$ or $M^{-1/4}$ depending on the choice of kernel (the TH_2, TH_{16}, and cubic kernels we consider here converge at the fastest rate). Finally, both the ALT and MLRV estimators are also consistent and converge at rate $M^{-1/4}$, albeit under more restrictive (semi-)parametric assumptions. An important feature of the non-parametric RV measures TSRV, MSRV, and KRV is that they allow for stochastic volatility, leverage, and can be made robust to dependent noise. ZHOU is biased with dependent noise, ALT rules out leverage effects and requires uncorrelated noise, and although MLRV can be modified to take account of dependent noise, formally it does not allow for stochastic volatility.

So a priori, based on asymptotic properties only, we should anticipate the following ordering of estimators in terms of their performance:

(1) MSRV and KRV with fastest convergence rate $M^{-1/4}$,
(2) MLRV and ALT with convergence rate $M^{-1/4}$, but susceptible to misspecification due to their (semi-)parametric nature,
(3) TSRV with convergence rate $M^{-1/6}$,
(4) inconsistent ZHOU,
(5) inconsistent FE with finite sample bias, and
(6) RV with exploding bias.

Yet, in practice, other considerations for choosing an estimator may come into play. For instance, a faster *asymptotic* convergence rate does not necessarily translate into superior *finite sample* performance. To illustrate this, note that the best attainable asymptotic variance of the TSRV estimator—assuming constant volatility $\overline{\sigma}^2$ and regular sampling—is equal to $2(12\omega^4)^{1/3}\overline{\sigma}^{8/3}M^{-1/3}$ (see [27], p. 1403). Similarly, the lower bound on the asymptotic variance of the MSRV and KRV estimators is equal to $8\omega\overline{\sigma}^3M^{-1/2}$. Upon comparing the two, we see that when $M < \frac{256}{9}\overline{\sigma}^2/\omega^2$, the

TSRV estimator has a lower variance as implied by the asymptotic distribution despite its slower rate of convergence (e.g. in the empirically reasonable setting where $\overline{\sigma}^2 = 1000\omega^2$, the critical value for M is over 28,000!). Another important consideration when selecting a realized variance estimator is its robustness to violations of assumptions underlying its construction. For instance, in relation to the MSRV [26] writes "*... though we have assumed that the ϵ are i.i.d., our estimator is quite robust to the nature of the noise*". [3] develops this point and formally shows that the TSRV and MSRV measures can both be robustified to general types of dependent noise by controlling the timescales involved in the construction of these estimators. Similarly, [11] proposes a subsampling version of the realized kernels that "*... leads to valid inference that is robust to both time-dependent and endogenous noise*". Finally, we point out that the properties—and consequently their relative performance—of the above RV measures with event-driven random sampling as considered here are typically unknown. Thus, all in all, it would seem unwise to select an RV measure solely based on its theoretical properties, and it is precisely this consideration that motivates our simulation study.

3.3 Notes on implementation

3.3.1 Estimation of IV, IQ, and ω^2

For the calculation of the optimal sampling frequencies, subsamples, or bandwidths q^*, measurements of IV, IQ, and ω^2 are required.[4] Here we use the standard variance and quarticity estimators averaged over $\overline{q} = \lfloor M/78 \rfloor$ subsamples of "low frequency" 5-minute returns where the impact of noise is expected to be benign, i.e.,

$$\widehat{IV} = \frac{M}{M - \overline{q} + 1} \frac{1}{\overline{q}} \sum_{h=0}^{\overline{q}-1} \sum_{i=1}^{m} (p_{i\overline{q}+h} - p_{(i-1)\overline{q}+h})^2,$$

$$\widehat{IQ} = \frac{M^2}{(M - \overline{q} + 1)^2} \frac{78}{3\overline{q}} \sum_{h=0}^{\overline{q}-1} \sum_{i=1}^{m} (p_{i\overline{q}+h} - p_{(i-1)\overline{q}+h})^4,$$

where $m = \lfloor (M - h + 1)/q \rfloor$.

Regarding the estimation of the noise variance ω^2, a number of alternative estimators have been proposed. [2] develops an MLE estimator within the MA(1) model discussed above, namely

$$\widehat{\omega}^2_{AMZ} = -\widehat{\theta\delta}^2. \tag{3.7}$$

A similar method-of-moments estimator has been suggested in [24],

$$\widehat{\omega}^2_{O} = -\gamma_{0,1}(1)/M. \tag{3.8}$$

[4] We stress again that these quantities are not well defined in the ZI market setup but, as in practice, we can still apply the commonly used estimators to the available data.

Table 1 Estimation of microstructure noise variance ω^2

	$\widehat{\omega}^2_{AMZ}$	$\widehat{\omega}^2_{O}$	$\widehat{\omega}^2_{BR}$	$\widehat{\omega}^2_{BHLS}$	Correlation	$\widehat{\omega}^2_{O}$	$\widehat{\omega}^2_{BR}$	$\widehat{\omega}^2_{BHLS}$
Panel A: *trade prices*								
mean	2.974	2.993	3.473	3.015	$\widehat{\omega}^2_{AMZ}$	0.585	0.741	0.879
std	0.045	0.076	0.043	0.049	$\widehat{\omega}^2_{O}$		0.737	0.812
					$\widehat{\omega}^2_{BR}$			0.852
Panel B: *quote prices*								
mean	0.050	0.049	0.557	0.226	$\widehat{\omega}^2_{AMZ}$	0.997	0.244	0.660
std	0.014	0.014	0.021	0.011	$\widehat{\omega}^2_{O}$		0.241	0.637
					$\widehat{\omega}^2_{BR}$			0.837
Panel C: *micro-prices*								
mean	0.069	0.067	0.587	0.248	$\widehat{\omega}^2_{AMZ}$	0.994	0.018	0.622
std	0.013	0.013	0.018	0.009	$\widehat{\omega}^2_{O}$		0.023	0.586
					$\widehat{\omega}^2_{BR}$			0.712

Note This table reports the mean, standard deviation, and correlation of the four noise variance estimators listed in (3.7–3.10). The SFGK parameters are as described in Sect. 4 and the sample size is $M = 10{,}000$. All estimates are standardized by $\sigma^2 = 0.083$

[8] re-scaled RV to deliver biased but consistent estimates of the noise variance,

$$\widehat{\omega}^2_{BR} = \gamma_{0,1}(0)/(2M), \tag{3.9}$$

while [10] advocated a bias-corrected version,

$$\widehat{\omega}^2_{BHLS} = \exp\left\{\log \widehat{\omega}^2_{BR} - KRV/\gamma_{0,1}(0)\right\}, \tag{3.10}$$

where KRV denotes a suitable realized kernel (here, we use the TH_2 with $q = 5$).

To keep the number of realized variance estimators to be implemented manageable, we first investigate the performance of the above noise variance estimators and then use a single preferred estimator throughout the remainder of the paper. Table 1 reports summary statistics of the noise variance estimates when applied to the ZI trade price, mid-quote, and micro-price series used in the simulation study below. The following observations can be made. Firstly, $\widehat{\omega}^2_{AMZ}$ and $\widehat{\omega}^2_{O}$ have a very similar mean but the former appears to have better efficiency with trade data. For the quote and micro-price series the performance of the two estimators is nearly identical with a correlation exceeding 99%. Secondly, for $\widehat{\omega}^2_{BR}$ we confirm the well-documented bias, and find it is most apparent for the quote series where noise variance estimates are increased tenfold compared to $\widehat{\omega}^2_{AMZ}$ or $\widehat{\omega}^2_{O}$. The bias-corrected version $\widehat{\omega}^2_{BHLS}$ performs well for trade data, but only partially eliminates the bias for quote and micro-price series. Motivated by these findings, we use $\widehat{\omega}^2_{AMZ}$ as our preferred noise variance estimator in the analysis below.

3.3.2 Small sample bias correction

With a fixed number of returns M, it can happen that for a particular choice of sampling frequency q the quantity M/q is not an integer. This leads to a small sample bias because the return associated with the last incomplete step, i.e., $p_M - p_{q\lfloor M/q \rfloor}$, is necessarily omitted from the calculations. To correct for this, we include a scaling constant in the definition of RV in (3.2). Similarly, we apply a small sample bias correction to ZHOU and TSRV, noting that the effective number of subsamples is $M - q + 1$ rather than M. We are not aware of any suggested corrections for the MSRV, so we implement the unaltered version. The realized kernels, when calculated at the highest frequency as is done here, conveniently do not require any bias correction. The same is the case for ALT and MLRV.

3.3.3 Calculation of q^*

The optimal number of subsamples or bandwidth q^* for the multi-scale RV or realized kernels is calculated by solving the minimization problem in (3.3) and (3.4) respectively. Because the first order condition for optimality is a quadratic form in c^2, we can obtain the closed form expression

$$q^* = \left(\frac{b\omega^2(IV + \omega^2/2)}{2aIQ} + \sqrt{\frac{b^2\omega^4(IV + \omega^2/2)^2}{4a^2IQ^2} + \frac{\omega^4 d}{aIQ}} \right)^{1/2} \sqrt{M}, \tag{3.11}$$

where $a = 104/35$, $b = 48/5$, $d = 144$ for MSRV and $a = 4\kappa_\bullet^{0,0}$, $b = -8\kappa_\bullet^{0,2}$, $d = 12(\kappa'''(0) + \kappa_\bullet^{0,4})$ for KRV.

To find the optimal number of Fourier coefficients q_F^* for the Fourier estimator, we do a grid search of (3.6) for $q \in (1, M/2)$.

3.3.4 Dealing with small values of q^*

When the (estimated) magnitude of the microstructure noise ω^2 is sufficiently small, it can happen that q^* is less than 2 for the TSRV and MSRV or less than 1 for KRV. In such a case we need to decide how to round q^*. Specifically, if we decide to round q^* down to 1 for TSRV[5] and MSRV and down to 0 for KRV, then we end up computing the equivalent of RV. If, on the other hand, we decide to round q^* up to 2 for TSRV and MSRV or 1 for KRV, then we compute a quantity that is either very close or exactly equal to ZHOU. In fact, in this case we have the identities

$$\begin{aligned} ZHOU &= KRV \\ &= MSRV + r_1^2 + r_M^2 \\ &= TSRV + \frac{TSRV - RV}{M} + r_1^2 + r_M^2, \end{aligned} \tag{3.12}$$

[5] Strictly speaking, TSRV does not nest RV, suggesting that the restriction $q \geq 2$ should be imposed in any case.

where $r_i = p_i - p_{i-1}$, ZHOU and RV are computed using $q = 1$, and the KRV is implemented using a flat-top kernel. Thus, the decision of how to round q^* in essence comes down to deciding whether to compute RV or ZHOU. Recall that under the typical assumptions of the "random walk plus i.i.d. noise" model, the MSE of RV is equal to

$$\frac{IV^2}{M^2}\left(2M\lambda^{-1} + 8M\xi - 4\xi^2 + 12M\xi^2 + 4\xi^2 M^2\right),$$

and the MSE of ZHOU is

$$\frac{IV^2}{M^2}\left(6M\lambda^{-1} + 8M\xi - 6\xi^2 + 8M\xi^2\right).$$

Hence, one possible approach may be to round q^* in such a way that it leads to the lowest MSE. Using the expressions above, this would lead one to round down when

$$\xi < \sqrt{\frac{2M\lambda}{1 + 2M + 2M^2}} \approx \sqrt{\lambda/M}$$

and round up otherwise. Note, however, that even for a relatively modest sample size of $M = 1{,}000$ (the smallest we consider below) we should always compute ZHOU unless $\xi < 0.03$. Moreover, when taking into account the measurement error in estimates of ξ and the limited efficiency gains to be had in the first place, it seems more practical to set $q = \max\{2, q^*\}$ for TSRV and MSRV and $q = \max\{1, q^*\}$ for KRV. This is therefore what we do in this paper.

4 ZI market simulation results

Because the SFGK model generates integer price moves (in ticks), we can achieve any desired dollar tick size or spread by choosing the appropriate stock price which then gives the model a scale. So, to most accurately mimic a real market, we need to find a dimensionless quantity that is characteristic of the market and then choose the model parameters to match it. The quantity we focus on here is the so-called "noise ratio" of [24] which, using our notation, is defined as $\xi = M\omega^2/\sigma^2$. Intuitively, ξ measures the magnitude of the noise relative to that of the per-tick noise-free price innovation. Because the choice of timescale is arbitrary, we set the arrival rate of limit orders $\alpha = 1$ without loss of generality. Next, setting the rate of arrival of market orders $\mu = 10$ and the cancelation rate $\delta = 0.2$ generates a simulated trade price series with $\xi \approx 2.9$ which, as we see from Panel A of Fig. 2, is pretty typical for a component of the DJ30 index. To simulate the process, we also need to set the bandwidth B. Here, we use $B = 10$. Following the discussion in Sect. 2, the asymptotic book depth (far away from the best quote) is 5 shares. Simulations indicate that this is the steady state level everywhere except at the best quote where the average size is around 4.8 shares. The spread between bid and offer is a single tick 96.5% of the time.

In order to evaluate the performance of the competing realized variance estimators, we require σ^2 as defined in (3.1). Because an analytical expression for σ^2 is unavailable, we determine its value using simulations. In particular, based on 1 million simulated price paths, we estimate $E((p_m - p_0)^2/m)$ across a range of values for m. The

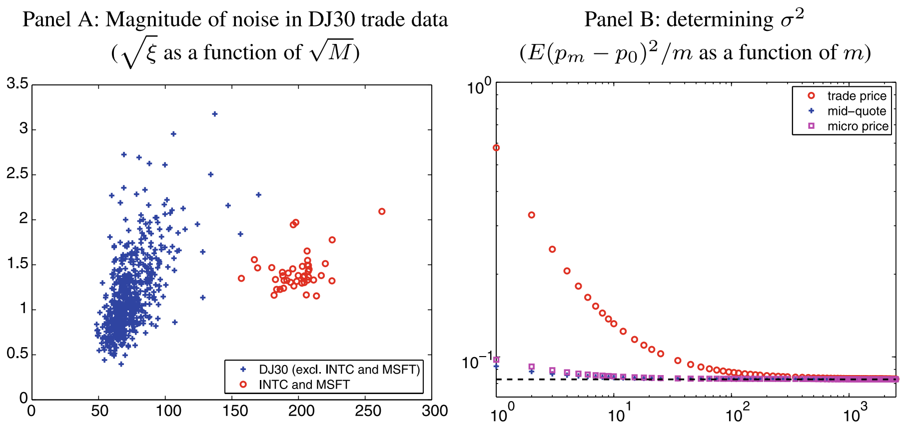

Fig. 2 Simulation of the ZI market model

results can be found in Panel B of Fig. 2. For visual aid, we use log scaling on both axes. Irrespective of the price series used, we see that the standardized return variance converges to a constant which we take to be σ^2. For the model parameters used in this paper, we find that $\sigma^2 = 0.083$, up to a negligible amount of measurement error. In reporting our results, we consider sample sizes of $M = \{1{,}000; 5{,}000; 10{,}000\}$ and standardize all realized variance estimates by $M\sigma^2$ so that an unbiased estimator should average 1.

4.1 Results for trade prices

We first consider the series of trade prices, i.e., $\{p_i^*\}_{i=1}^M$. Results can be found in Table 2. For $M = 1,000$, ALT of [19] and MLRV of [2] deliver the best performance, with the difference in MSE being statistically insignificant. Admittedly, both estimators operate in a close-to-ideal environment: As can be seen from Panels B and C in Fig. 1, the trade price process is very close to MA(1), benefitting the MLRV estimator and, at the same time, more than 95% of price moves are one-tick only and this clearly favors ALT. Interestingly, even though the MSE of both estimators is comparable, they have distinctly different properties with ALT incurring a substantial downward bias which can likely be attributed to the few large price moves of two ticks or over. Following in close second place is KRV. In fact, when the optimal bandwidth and the asymptotically most efficient TH_{16} kernel are used, the performance of KRV is statistically indistinguishable from ALT or MLRV. The TH_2 and cubic kernels incur some limited efficiency loss, and so does the use of ad hoc bandwidth selection. In third place come the subsampling based methods. The remarkably small difference in performance among the TSRV and MSRV emphasizes that asymptotic convergence rates can be a misleading guide to finite sample performance. Up to this point, the best performing measure in each class of estimators only leads to a limited deterioration in performance relative to ALT or MLRV. For instance, the TSRV with ad hoc bandwidth selection still attains a log MSE of −3.40 compared to −3.52 for MLRV. In sharp contrast, when moving to FE and ZHOU in shared fourth best place,

Table 2 Performance of alternative realized variance measures with ZI trade-price data

	$M = 1{,}000$					$M = 5{,}000$					$M = 10{,}000$				
	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*
1. Realized variance															
(a) highest ($q = 1$)	6.942	0.277	3.566	7.089	1.00	6.948	0.124	3.566	8.698	1.00	6.947	0.086	3.566	9.362	1.00
(b) ad hoc (5 mins)	1.490	0.255	−1.188	2.335	12.00	1.090	0.182	−3.185	1.946	64.00	1.044	0.171	−3.464	2.332	128.0
(c) q^*_{RV}	1.237	0.311	−1.877	1.645	24.68	1.104	0.177	−3.165	1.967	53.01	1.078	0.137	−3.694	2.103	69.27
2. Bias-corrected RV of Zhou [28]															
(a) highest ($q = 1$)	0.964	0.335	−2.174	1.349	1.00	0.961	0.147	−3.769	1.362	1.00	0.961	0.105	−4.374	1.422	1.00
(b) ad hoc (5 mins)	1.001	0.283	−2.525	0.997	12.00	0.999	0.266	−2.650	2.482	64.00	0.997	0.263	−2.674	3.122	128.0
(c) q^*_{Zhou}	0.992	0.227	−2.961	0.561	2.35	0.996	0.095	−4.708	0.424	3.20	0.996	0.067	−5.392	0.404	3.35
3. Two-scale RV of Zhang, Mykland, and Aït-Sahalia [27]															
(a) ad hoc ($q = 5$)	0.978	0.181	−3.403	0.120	5.00	0.984	0.080	−5.006	0.126	5.00	0.984	0.057	−5.645	0.152	5.00
(b) q^*_{ZMA}	0.969	0.202	−3.172	0.350	3.56	0.981	0.082	−4.949	0.183	4.47	0.982	0.058	−5.610	0.186	4.64
4. Multi-scale RV of Zhang [26]															
(a) ad hoc ($q = 5$)	0.973	0.180	−3.405	0.117	5.00	0.984	0.080	−5.016	0.116	5.00	0.985	0.057	−5.667	0.129	5.00
(b) q^*_Z	0.968	0.190	−3.297	0.225	4.10	0.983	0.080	−5.012	0.119	4.88	0.984	0.057	−5.673	0.124	5.02

Note This table reports the mean ("mean"), standard deviation ("stdev"), logarithmic MSE ("MSE"), the difference in log MSE relative to the best estimator ("loss"), and the average (sub)sampling frequency or bandwidth ("q^*") for each realized variance measure and across sample size M. Loss levels in boldface are insignificantly different from zero at a 1% bootstrapped confidence level

Table 2 (*Continued*)

	$M=1{,}000$					$M=5{,}000$					$M=10{,}000$				
	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*
5. Realized kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard [10]															
(a) TH_2, ad hoc ($q=5$)	0.985	0.187	−3.344	0.178	5.00	0.983	0.083	−4.936	0.196	5.00	0.982	0.059	−5.585	0.211	5.00
(b) TH_2, q^*_{BNLS}	0.990	0.176	−3.472	0.050	7.94	0.991	0.078	−5.089	0.043	9.47	0.991	0.055	−5.765	**0.031**	9.75
(c) TH_{16}, q^*_{BNLS}	0.989	0.173	−3.501	**0.022**	54.17	0.991	0.077	−5.111	**0.020**	64.54	0.991	0.055	−5.785	**0.011**	66.42
(d) Cubic, q^*_{BNLS}	0.992	0.180	−3.430	0.092	5.10	0.993	0.080	−5.046	0.086	6.07	0.993	0.057	−5.727	0.069	6.25
6. Fourier estimator of Malliavin and Mancino [20]															
(a) DIR kernel, q^*_F	1.102	0.204	−2.955	0.567	52.03	1.039	0.099	−4.483	0.648	165.2	1.027	0.073	−5.113	0.683	281.6
(b) FEJ kernel, q^*_F	1.081	0.194	−3.116	0.406	72.02	1.031	0.097	−4.572	0.560	226.5	1.022	0.072	−5.184	0.613	385.0
7. ALT of Large [19]	0.885	0.127	**−3.522**	★	–	0.884	0.057	−4.097	1.035	–	0.884	0.040	−4.195	1.601	–
8. MLRV of AMZ [2]	0.986	0.172	−3.517	**0.006**	–	0.988	0.076	**−5.132**	★	–	0.988	0.054	**−5.796**	★	–

we observe a considerable drop in performance. For ZHOU, when using all data (i.e., $q = 1$), the MSE increases more than threefold relative to MLRV and still nearly doubles when we use the optimal sampling frequency. Here, the high level of noise and inconsistency of the estimator lead to a poor performance. For FE, a substantial finite sample bias of around 10% is another contributing factor. As anticipated, RV comes in last place with a significant positive bias and very high MSE irrespective of the choice of sampling frequency.

Now let us consider what happens when we increase the sample size. Keep in mind here that we should not think of this as increasing the sampling frequency—after all, we should always want to use all available data—but rather view this as representing a more actively traded stock with similar noise characteristics (i.e., a horizontal move in Fig. 2, Panel A). Both with $M = 5{,}000$ and $M = 10{,}000$, MLRV remains the best estimator with the TH_{16} kernel delivering statistically indistinguishable performance. The subsampling estimators also continue to perform admirably well with limited efficiency loss relative to the MLE. In sharp contrast, ALT, which was previously in shared first place, now incurs a severe increase in MSE that can be attributed to the persistent bias that plays a more dominant role in the MSE now that the variance steadily drops with larger sample size. A similar issue appears for FE with the finite sample bias weighing down on its performance. On the other hand, ZHOU's relative performance improves with larger sample size but the efficiency loss is still in excess of 30% (compared to 50% with $M = 1{,}000$). As before, and not surprisingly, RV comes in last place.

Considering the different choices of tuning parameters, we find that for the realized kernels the theory-implied optimal bandwidth outperforms the ad hoc rule. Yet, for TSRV and MSRV we see the reverse and the ad hoc rule for subsample selection leads to the lower MSE, particularly for smaller sample sizes. Also, for KRV the optimal bandwidth varies substantially with the choice of kernel: for the cubic kernel q^* is around 5 while for TH_{16} this figure is exceeding 50 (however, the TH_{16} kernel tails off quickly and so the "effective" bandwidth is much lower). Further, while the optimal q^* for the realized kernels and subsampling estimators increases with the sample size, it does so much more slowly than by $\sqrt{M}$ as one might expect. This can be explained as follows. If we neglect the small term $\omega^2/2$ in the expression for q^* in (3.11) and use the convenient notation of the noise ratio $\xi = M\omega^2/IV$ and $\lambda = IV^2/IQ$, then we have

$$q^* = \left(\lambda \frac{b}{2a} + \frac{b}{2a}\sqrt{\lambda^2 + 4\lambda ad/b^2}\right)^{1/2} \sqrt{\xi}. \tag{4.1}$$

From this it is clear that q^* increases as λ increases and attains it maximum value when $\lambda = 1$, i.e., with constant volatility. In that scenario, we have $q_Z^* = 2.96\sqrt{\xi}$ for the multi-scale estimator and $q_{BHLS}^* = 5.75\sqrt{\xi}$ for the realized TH_2 kernel. Thus, the optimal bandwidth is nearly double the optimal number of subsamples which is exactly what we see in the results. More importantly, the above illustrates that when the noise characteristics remain unchanged as M is varied, the optimal q^* is constant. Because in our simulation setup $\xi \approx 2.9$ for all sample sizes, we see that $q_Z^* = 2.96\sqrt{2.9} \approx 5.0$ and $q_{BHLS}^* = 5.75\sqrt{2.9} \approx 9.8$. The small increase in q^* we observe when M grows is due to overestimation of IV and IQ in small sample sizes which leads to underestimation of the noise ratio ξ as well as λ.

As a final remark, we point out that the assumption of a fixed noise ratio when the sample size grows in our simulations is supported by empirical evidence. For the DJIA index components (see Fig. 2, Panel A) we find a relatively low time series correlation between $\sqrt{M}$ and $\sqrt{\xi}$ of 21.15% when averaged across the 30 DJIA components, but a much more substantial correlation in the cross section of DJIA stocks of 67.61% when averaged over our sample period (i.e., 21 days in October 2004). As a consequence, while the theory suggests that the optimal q^* should grow linearly with the square root of the sample size, we may want to think of this as a cross-sectional relation rather than a time series relationship. It is not uncommon for a given stock to experience wild fluctuations in trade intensity with relatively stable noise characteristics. If we believe that, at least for trade data, the microstructure noise can largely be attributed to the spread then this seems quite intuitive, especially for highly liquid stocks such as MSFT or INTC where the minimum tick size can be restraining and the spread would not react to intra-day or even day-to-day variation in trade intensity.

4.2 Results for quote prices

A widely used method to mitigate the level of microstructure noise in tick data—particularly the "bounce" of trade prices (e.g. [23])—is to sample the mid-quote process, i.e., $\{q_i^m\}_{i=1}^M$ as defined in (2.1). That this is an effective means of sampling for the purpose of realized variance calculations is confirmed by the heavily reduced amount of serial correlation in returns, as exhibited in Panel B of Fig. 1, and the lower noise variance estimates in Table 1. The first order autocorrelation of mid-quote returns is now less than 5%, compared to over -40% for trade data, while the noise variance estimates (standardized by σ^2) drop from around 3 to 0.05. So with the same number of observations, we expect to see a reduction in bias, a lower MSE, and a fall in optimal (sub-)sampling frequencies and bandwidths.

As an aside, although it may be obvious that sampling mid-quotes rather than trade prices reduces noise, it is not so obvious *when* to sample the mid-quotes. For example, one might suppose that sampling the mid-quote every time the book is updated might provide more information. In simulation of the zero-intelligence market, however, we find that sampling the mid-quote just prior to each trade, exactly as recommended by [7] and [13], gives by far the best results. We are unaware of a theoretical explanation for this.

The most striking finding that is immediately evident from Table 3 is that with mid-quote data, ZHOU, TSRV, MSRV, KRV, and MLRV all perform admirably well and the difference in MSE among these estimators is in most cases statistically insignificant. Unreported results indicate that the method for estimating the noise variance is relatively unimportant here because the estimates of $\widehat{\omega}^2$ are so small[6] that the optimal q^* is often set to its minimum value of 2 for TSRV and MSRV and 1 for

[6]The maximum TSRV subsampling frequency obtained over 12,500 simulation runs is 2.2 when using $\widehat{\omega}^2_{BR}$ and 0.6 when using $\widehat{\omega}^2_{AMZ}$ or $\widehat{\omega}^2_{O}$.

Table 3 Performance of alternative realized variance measures with ZI quote-price data

		$M = 1{,}000$					$M = 5{,}000$					$M = 10{,}000$				
		mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*
1.	Realized variance															
	(a) highest ($q = 1$)	1.112	0.138	−3.452	0.479	1.00	1.113	0.062	−4.097	1.360	1.00	1.113	0.043	−4.224	1.899	1.00
	(b) ad hoc (5 mins)	1.008	0.198	−3.238	0.694	12.00	1.001	0.169	−3.555	1.901	64.00	0.999	0.165	−3.609	2.513	128.0
	(c) q^*_{RV}	1.049	0.146	−3.736	0.196	1.93	1.035	0.070	−5.087	0.370	3.61	1.027	0.051	−5.689	0.434	4.63
2.	Bias-corrected RV of Zhou [28]															
	(a) highest ($q = 1$)	1.013	0.143	−3.876	0.056	1.00	1.015	0.064	−5.447	**0.009**	1.00	1.015	0.045	−6.098	**0.024**	1.00
	(b) ad hoc (5 mins)	0.995	0.278	−2.561	1.371	12.00	0.998	0.266	−2.650	2.806	64.00	0.996	0.263	−2.674	3.449	128.0
	(c) q^*_{Zhou}	1.013	0.143	−3.876	0.056	1.00	1.015	0.064	−5.447	**0.009**	1.00	1.015	0.045	−6.098	**0.024**	1.00
3.	Two-scale RV of Zhang, Mykland, and Aït-Sahalia [27]															
	(a) ad hoc ($q = 5$)	0.996	0.153	−3.752	0.179	5.00	1.002	0.068	−5.383	0.073	5.00	1.003	0.048	−6.079	0.044	5.00
	(b) q^*_{ZMA}	1.011	0.143	−3.880	0.052	2.00	1.015	0.064	−5.450	**0.006**	2.00	1.015	0.045	−6.101	**0.022**	2.00
4.	Multi-scale RV of Zhang [26]															
	(a) ad hoc ($q = 5$)	0.995	0.154	−3.741	0.191	5.00	1.002	0.068	−5.375	0.081	5.00	1.002	0.048	−6.071	0.051	5.00
	(b) q^*_Z	1.011	0.143	−3.880	0.052	2.00	1.015	0.064	−5.450	**0.006**	2.00	1.015	0.045	−6.101	**0.022**	2.00

Note This table reports the mean ("mean"), standard deviation ("stdev"), logarithmic MSE ("MSE"), the difference in log MSE relative to the best estimator ("loss"), and the average (sub)sampling frequency or bandwidth ("q^*") for each realized variance measure and across sample size M. Loss levels in boldface are insignificantly different from zero at a 1% bootstrapped confidence level

Table 3 (*Continued*)

	$M = 1{,}000$					$M = 5{,}000$					$M = 10{,}000$				
	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*
5. Realized kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard [10]															
(a) TH_2, ad hoc ($q = 5$)	1.002	0.150	−3.795	0.137	5.00	1.005	0.066	−5.425	**0.031**	5.00	1.005	0.047	**−6.123**	★	5.00
(b) TH_2, q^*_{BNLS}	1.012	0.144	−3.873	0.058	1.30	1.015	0.064	−5.444	**0.013**	1.29	1.015	0.045	−6.096	**0.027**	1.29
(c) TH_{16}, q^*_{BNLS}	1.012	0.144	−3.873	0.058	7.64	1.014	0.064	−5.449	**0.008**	8.52	1.014	0.045	−6.105	**0.018**	8.69
(d) Cubic, q^*_{BNLS}	1.013	0.144	−3.873	0.058	1.04	1.015	0.064	−5.447	**0.009**	1.01	1.015	0.045	−6.098	**0.024**	1.01
6. Fourier estimator of Malliavin and Mancino [20]															
(a) DIR kernel, q^*_F	1.039	0.143	−3.824	0.108	304.9	1.028	0.067	−5.257	0.199	924.0	1.021	0.048	−5.914	0.208	1504
(b) FEJ kernel, q^*_F	1.034	0.142	−3.855	0.077	381.6	1.025	0.066	−5.304	0.152	1293	1.019	0.047	−5.959	0.163	2121
7. ALT of Large [19]	1.002	0.240	−2.851	1.080	–	0.990	0.105	−4.506	0.950	–	0.987	0.073	−5.204	0.918	–
8. MLRV of AMZ [2]	1.007	0.140	**−3.932**	★	–	1.014	0.064	**−5.456**	★	–	1.014	0.045	−6.114	**0.009**	–

ZHOU and KRV. In such a scenario, we know from the relation in (3.12) that all these estimators are equivalent up to end-effects and so it is not surprising that they attain a similar MSE. The next best estimator is FE which, without a formal bias correction, benefits greatly from the reduced level of noise. As anticipated, ALT's performance is severely worsened because the estimator is now essentially misspecified and inapplicable. RV's performance is not good either, although the efficiency loss relative to the best performing measure is now much smaller than with trade data. This can be largely attributed to a heavily reduced bias of only 10% with mid-quote data when sampled at the highest frequency, compared to more than 500% with trade data. Finally, when we increase the sample size, the relative ranking of ZHOU, TSRV, MSRV and KRV changes around somewhat, although none of this is statistically significant.

4.3 Results for micro-prices

An alternative to using mid-quote data (which can be quite stale) is to construct a price series that incorporates the information contained in the order book depths. A popular approach is to linearly weigh the bid and ask prices by the volume on the opposite side of the book to yield a so-called volume-weighted mid-quote or micro-price series, i.e., $\{q_i^v\}_{i=1}^M$ as defined in (2.2). Particularly in scenarios where the tick size is restraining and order-queue lengths are long, the micro-price series can have desirable statistical properties in that it will fluctuate with every change in quoted volume as opposed to the mid-quote which will only change with a change in quoted price. Moreover, the micro-price has a nice interpretation: It represents the market clearing price when demand and supply curves are linear in price. See Fig. 3 for a graphical illustration of this (here price is on the horizontal axis and volume on the vertical axis). Because the ZI market model also simulates the order book depth, we can straightforwardly compute this micro-price and redo the above analysis for comparison. The results can be found in Table 4.

With $M = 1{,}000$ the results for the micro-price are qualitatively comparable to those obtained for the mid-quote data in that ZHOU, TSRV, MSRV, KRV, and MLRV all perform similarly, with MSE differences statistically insignificant as long as q^*

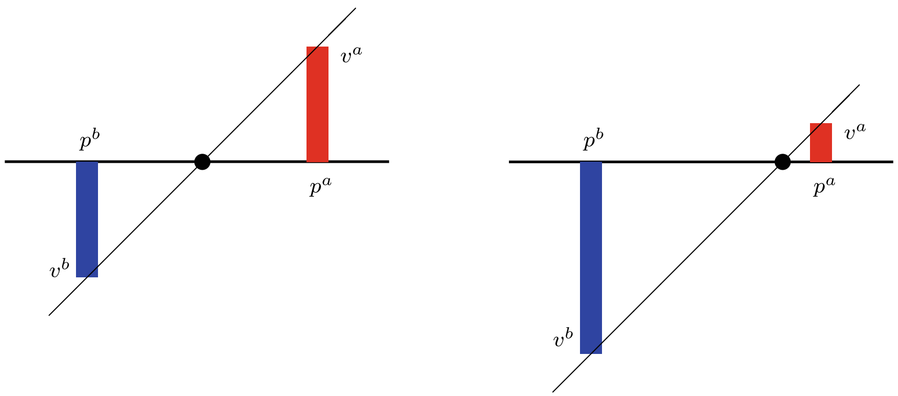

Fig. 3 Graphical illustration of micro-price construction

is chosen optimally. However, when we increase M the results change quite dramatically. The performance of ZHOU, TSRV and MSRV calculated using optimal subsampling frequencies, and even MLRV, rapidly deteriorate and incur an average efficiency loss of more than 30% relative to the best performing measures: TSRV, MSRV, and KRV with ad hoc subsampling frequency and bandwidth! Thus, the theory-implied optimal selection of q^* appears not to work well with micro-price data. To some extent, this is quite surprising given that for the micro-price, estimates of the noise ratio ξ (see Table 1) as well as the autocorrelation pattern (see Fig. 1) are broadly comparable to those for quote data. Of course, given the randomness of the market depth in this ZI market the micro-price will incorporate a little more noise than mid-quotes do—as is clear from the summary statistics—but the process is still much less noisy than trade data, and its statistical properties do not appear perverse in any way. Still, the techniques that worked well before, now perform poorly and it can therefore be argued that they lack a certain degree of robustness. As will be discussed in more detail below, a plausible explanation for this apparent failure is that in large samples the *second* order autocorrelation of micro-price returns becomes important, indicating dependent noise, and none of the procedures to optimally select q take account of this. The realized kernels or subsampling methods can of course still deal with such data: One way is to select a sufficiently high bandwidth, thereby effectively lowering the sampling frequency, and this is precisely what the ad hoc rule achieves here.

4.4 Investigating noise dependence

The properties of the microstructure noise present in high-frequency data are—to a large extent—determined by the scheme employed to sample the data.[7] Importantly, however, [2] shows that the MLRV estimator is theoretically robust to misspecification of the marginal distribution of the noise process. Thus, the observed deterioration in performance of MLRV (as well as the realized kernels and subsampling estimators with asymptotically optimal tuning parameters) when applied to micro-price data may be due to noise *dependence* instead. From Panel B of Fig. 1 we do indeed observe second-order return serial correlation that, upon closer inspection, turns out to be highly significant in large samples. To investigate this a little further, we specify an "efficient price plus MA(1) noise" model of the form

$$r_i = z_i + \eta_i - \eta_{i-1} \quad \text{where } z \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2),$$
$$\eta_i = u_i + \beta u_{i-1} \quad \text{where } u \sim \text{i.i.d. } \mathcal{N}(0, \omega^2).$$

Note that we now allow return dependence up to lag 2 with $E(r_i r_{i-1}) = -(\beta - 1)^2\omega^2$ and $E(r_i r_{i-2}) = -\beta\omega^2$. Analogously to the methodology in [2], we use maximum

[7] As already discussed in detail, the sampled trade, mid-quote, and micro-price series all have distinctly different characteristics. See also [24] and [17] for a formal study of the properties of calendar-time, business-time, transaction-time, and tick-time sampling schemes in the context of realized variance calculations.

Table 4 Performance of alternative realized variance measures with ZI micro-price data

	$M = 1{,}000$					$M = 5{,}000$					$M = 10{,}000$				
	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*
1. Realized variance															
(a) highest ($q = 1$)	1.174	0.113	−3.147	0.778	1.00	1.175	0.050	−3.406	1.990	1.00	1.175	0.035	−3.447	2.600	1.00
(b) ad hoc (5 mins)	1.016	0.196	−3.251	0.674	12.00	1.002	0.169	−3.550	1.846	64.00	1.000	0.165	−3.607	2.440	128.0
(c) q^*_{RV}	1.078	0.146	−3.593	0.332	2.46	1.049	0.073	−4.856	0.541	4.56	1.038	0.054	−5.427	0.620	5.81
2. Bias-corrected RV of Zhou [28]															
(a) highest ($q = 1$)	1.038	0.137	−3.894	**0.031**	1.00	1.040	0.061	−5.221	0.176	1.00	1.040	0.043	−5.660	0.387	1.00
(b) ad hoc (5 mins)	0.995	0.278	−2.559	1.366	12.00	0.998	0.266	−2.650	2.746	64.00	0.996	0.263	−2.674	3.373	128.0
(c) q^*_{Zhou}	1.038	0.137	−3.894	**0.031**	1.00	1.040	0.061	−5.221	0.176	1.00	1.040	0.043	−5.660	0.387	1.00
3. Two-scale RV of Zhang, Mykland, and Aït-Sahalia [27]															
(a) ad hoc ($q = 5$)	1.005	0.151	−3.777	0.147	5.00	1.011	0.067	−5.373	**0.024**	5.00	1.012	0.047	−6.040	**0.007**	5.00
(b) q^*_{ZMA}	1.036	0.137	−3.903	**0.021**	2.00	1.040	0.061	−5.227	0.169	2.00	1.040	0.043	−5.665	0.382	2.00
4. Multi-scale RV of Zhang [26]															
(a) ad hoc ($q = 5$)	1.003	0.152	−3.763	0.161	5.00	1.010	0.068	−5.366	**0.030**	5.00	1.010	0.048	−6.041	**0.006**	5.00
(b) q^*_Z	1.036	0.137	−3.903	**0.022**	2.00	1.040	0.061	−5.227	0.169	2.00	1.040	0.043	−5.665	0.382	2.00

Note This table reports the mean ("mean"), standard deviation ("stdev"), logarithmic MSE ("MSE"), the difference in log MSE relative to best estimator ("loss"), and the average (sub)sampling frequency or bandwidth ("q^*") for each realized variance measure and across sample size M. Loss levels in boldface are insignificantly different from zero at a 1% bootstrapped confidence level

Table 4 (*Continued*)

	$M = 1{,}000$					$M = 5{,}000$					$M = 10{,}000$				
	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*
5. Realized kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard [10]															
(a) TH_2, ad hoc ($q = 5$)	1.013	0.148	−3.819	0.106	5.00	1.016	0.066	**−5.396**	★	5.00	1.015	0.046	**−6.047**	★	5.00
(b) TH_2, q^*_{BNLS}	1.036	0.139	−3.883	**0.042**	1.45	1.038	0.063	−5.231	0.165	1.50	1.038	0.044	−5.690	0.357	1.52
(c) TH_{16}, q^*_{BNLS}	1.035	0.140	−3.879	0.045	9.28	1.037	0.062	−5.244	0.153	10.19	1.037	0.044	−5.712	0.335	10.33
(d) Cubic, q^*_{BNLS}	1.038	0.139	−3.882	**0.043**	1.08	1.040	0.062	−5.219	0.177	1.05	1.040	0.043	−5.660	0.387	1.04
6. Fourier estimator of Malliavin and Mancino [20]															
(a) DIR kernel, q^*_F	1.067	0.142	−3.703	0.222	257.9	1.042	0.069	−5.037	0.360	762.6	1.032	0.049	−5.670	0.377	1286
(b) FEJ kernel, q^*_F	1.059	0.141	−3.761	0.164	339.0	1.038	0.067	−5.115	0.282	1075	1.030	0.048	−5.741	0.306	1809
7. ALT of Large [19]	0.869	0.107	−3.556	0.369	–	0.869	0.047	−3.941	1.455	–	0.869	0.033	−3.997	2.050	–
8. MLRV of AMZ [2]	1.033	0.137	**−3.925**	★	–	1.037	0.062	−5.260	0.137	–	1.037	0.044	−5.733	0.314	–

Table 5 Performance of MLRV-MA(2) relative to MLRV

	$M = 1{,}000$				$M = 5{,}000$				$M = 10{,}000$			
	mean	stdev	MSE	loss	mean	stdev	MSE	loss	mean	stdev	MSE	loss
Panel A: *trade-price data*												
MLRV	0.986	0.172	**−3.517**	★	0.988	0.076	**−5.132**	★	0.988	0.054	**−5.796**	★
MLRV-MA(2)	0.996	0.185	−3.371	0.145	0.996	0.082	−5.004	0.128	0.995	0.058	−5.693	0.103
Panel B: *quote-price data*												
MLRV	1.007	0.140	**−3.932**	★	1.014	0.064	**−5.456**	★	1.014	0.045	**−6.114**	★
MLRV-MA(2)	0.983	0.143	−3.872	0.059	1.000	0.067	−5.394	0.062	1.001	0.049	−6.049	0.065
Panel C: *micro-price data*												
MLRV	1.033	0.137	**−3.925**	★	1.037	0.062	−5.260	0.072	1.037	0.044	−5.733	0.282
MLRV-MA(2)	0.998	0.145	−3.859	0.065	1.009	0.069	**−5.331**	★	1.008	0.049	**−6.014**	★

likelihood to estimate an MA(2) model for observed returns, i.e.,

$$r_i = \varepsilon_i + \theta_1 \varepsilon_{i-1} + \theta_2 \varepsilon_{i-2} \quad \text{where } \varepsilon \sim \text{i.i.d. } \mathcal{N}\big(0, \delta^2\big),$$

and then map back to the original parameterization as $\sigma^2 = \delta^2(1 + \theta_1 + \theta_2)^2$, $\omega^2 = \rho\delta^2$, and $\beta = -\theta_2/\rho$, where

$$\rho = -\theta_2 - \frac{1}{2}\theta_1(1+\theta_2) \pm \frac{1}{2}\sqrt{\theta_1(1+\theta_2)(4\theta_2 + \theta_1\theta_2 + \theta_1)}.$$

The model structure imposes for the parameters the constraints $-1 < \theta_1 < 0$ and $-1 < \theta_2 < -\theta_1/(4+\theta_1)$.

Table 5 reports the performance of this MLRV-MA(2) estimator relative to MLRV for all price series and sample sizes considered before. For trade and quote data we find that even though MLRV-MA(2) nests MLRV, it clearly underperforms. MLRV-MA(2) requires the estimation of one additional parameter which, for this data, serves no purpose and merely picks up spurious dependence, adding noise to the estimates. Importantly, however, for micro-price data the added flexibility of MLRV-MA(2) pays off and leads to superior performance when the sample size grows. Estimates for β are highly significant in this case and average at around 0.15, indicating positive noise dependence and explaining the observed negative second-order return serial correlation. Interestingly, comparison to Table 4 shows that MLRV-MA(2) is now among the best performing estimators.

The above findings support our conjecture that the observed deterioration in performance is due to noise dependence in micro-prices. MLRV is misspecified, while the realized kernels and subsampling methods select too low tuning parameters: Estimates of ω^2 underestimate the true level of noise and this is reflected in q^*.

4.5 Robustness analysis

The model parameters in the above simulations are kept constant and this may favor the parametric approaches over the non-parametric ones. To address this issue, we

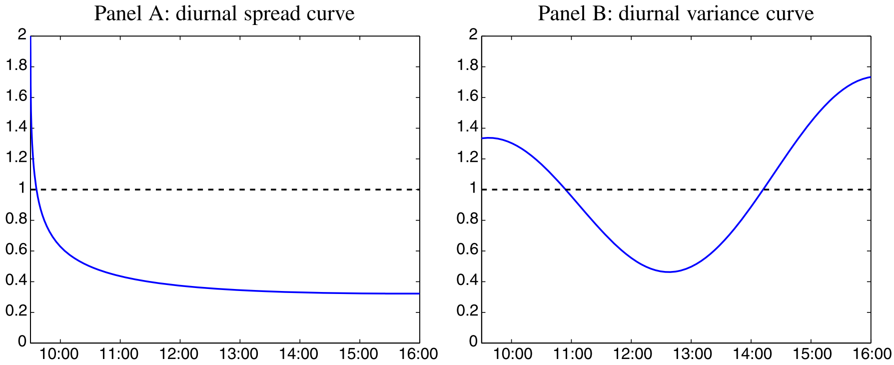

Fig. 4 Time-varying spread and volatility specification

generate a new trade-price series that incorporates a diurnal pattern in (i) the magnitude of the spread and (ii) the return volatility. In particular, based on the simulated price data used above, we construct a new series of transformed trade prices as

$$\widetilde{p}_i^* = q_i^m + s_{i/M}\left(p_i^* - q_i^m\right),$$

where

$$s_t = \frac{2}{1+8\sqrt{t}} + \frac{1}{10}t.$$

Here, s_t denotes the "spread curve" with a functional form that is motivated by the empirical observation that spreads are often wide early in the morning and tighten as the day progresses. See Panel A of Fig. 4 for an illustration. Note that while this trade-price series will have the same long-horizon variance $M\sigma^2$ as before, it now exhibits a varying degree of return serial correlation: At the start of the day when the spread is widest at $s_0 = 2$, we have the first order autocorrelation $\rho_1 \approx -47.8\%$, but then this more than halves to $\rho_1 \approx -19.5\%$ as we near the close.

A diurnal pattern in return volatility is achieved through the transformation of trade prices given by

$$\widetilde{p}_i^* - \widetilde{p}_{i-1}^* = \sqrt{v}_{i/M}\left(p_i^* - p_{i-1}^*\right),$$

where

$$v_t = \frac{4}{5} + \frac{2}{5}t + \frac{8}{15}\cos(2\pi t).$$

Here, v_t denotes the "variance curve" with a functional form that is motivated by the empirical observation that volatility is often highest at the start and end of the day while it dips around lunch time.[8] See Panel B of Fig. 4 for an illustration. Because $\int_0^1 v_t\,dt = 1$, the "true" variance of the series is again unaffected by this scaling.

[8]Time-varying volatility is much less of an issue with tick data than it is with regularly sampled data in calendar time, such as daily data. In particular, a judicious choice of sampling scheme (e.g. sampling in business or trade time) can substantially reduce heteroskedasticity in returns. For the purposes of the robustness analysis, however, we do consider this case.

Table 6 Robustness analysis with ZI trade-price data ($M = 10{,}000$)

	Time-varying spread					Time-varying volatility				
	mean	stdev	MSE	loss	q^*	mean	stdev	MSE	loss	q^*
1. Realized variance										
(a) highest ($q = 1$)	2.165	0.048	0.307	6.387	1.00	6.946	0.093	3.566	9.221	1.00
(b) ad hoc (5 mins)	1.007	0.166	−3.591	2.488	128.00	1.042	0.184	−3.332	2.324	128.00
(c) q^*_{RV}	1.044	0.084	−4.725	1.354	24.02	1.077	0.146	−3.598	2.058	66.85
2. Bias-corrected RV of Zhou [28]										
(a) highest ($q = 1$)	0.994	0.054	−5.812	0.267	1.00	0.962	0.113	−4.250	1.406	1.00
(b) ad hoc (5 mins)	0.997	0.263	−2.674	3.405	128.00	0.990	0.277	−2.563	3.093	128.00
(c) q^*_{Zhou}	0.994	0.054	−5.812	0.267	1.00	0.996	0.072	−5.254	0.402	3.38
3. Two-scale RV of Zhang, Mykland, and Aït-Sahalia [27]										
(a) ad hoc ($q = 5$)	0.994	0.049	−6.018	0.061	5.00	0.984	0.062	−5.499	0.156	5.00
(b) q^*_{ZMA}	0.992	0.054	−5.802	0.277	2.00	0.981	0.063	−5.442	0.214	4.48
4. Multi-scale RV of Zhang [26]										
(a) ad hoc ($q = 5$)	0.994	0.049	−6.011	0.068	5.00	0.985	0.061	−5.521	0.135	5.00
(b) q^*_Z	0.991	0.054	−5.818	0.261	2.27	0.983	0.061	−5.506	0.150	4.85
5. Realized kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard [10]										
(a) TH_2, ad hoc ($q = 5$)	0.996	0.048	−6.057	**0.022**	5.00	0.983	0.063	−5.451	0.205	5.00
(b) TH_2, q^*_{BNLS}	0.996	0.048	−6.065	**0.014**	4.40	0.991	0.060	−5.622	**0.034**	9.40
(c) TH_{16}, q^*_{BNLS}	0.996	0.048	−6.078	**0.001**	30.01	0.990	0.059	−5.644	**0.012**	64.13
(d) Cubic, q^*_{BNLS}	0.996	0.049	−6.041	**0.038**	2.83	0.992	0.061	−5.585	0.070	6.03
6. Fourier estimator of Malliavin and Mancino [20]										
(a) DIR kernel, q^*_F	1.019	0.058	−5.574	0.505	533.69	1.029	0.077	−4.987	0.669	288.19
(b) FEJ kernel, q^*_F	1.017	0.058	−5.626	0.453	741.80	1.023	0.076	−5.054	0.601	394.05
7. ALT of Large [19]	2.064	0.052	0.127	6.206	–	0.884	0.040	−4.192	1.463	–
8. MLRV of AMZ [2]	0.995	0.048	**−6.079**	★	–	0.988	0.058	**−5.656**	★	–
MLRV-MA(2)	0.998	0.052	−5.914	0.165	–	0.995	0.062	−5.547	0.108	–

Note This table reports the mean (“mean”), standard deviation (“stdev”), logarithmic MSE (“MSE”), the difference in log MSE relative to best estimator (“loss”), and the average (sub)sampling frequency or bandwidth (“q^*”) for each realized variance measure and across sample size M. Loss levels in boldface are insignificantly different from zero at a 1% bootstrapped confidence level

The results of the exercise can be found in Table 6 and should be compared to those in Table 2. The most striking and important finding is that the relative ranking of the estimators is qualitatively unchanged. In particular, the fully parametric MLRV remains the best performing estimator despite now being misspecified due to time-varying return volatility and serial correlation. The spread curve considered here shrinks the spread for most part of the day, and because of this the performance of all estimators improves with RV benefitting most from a reduction in noise. Also, the ad hoc implementation of KRV now performs statistically indistinguishably from MLRV. On the other hand, with time-varying volatility the performance of all estimators deteriorates slightly, but again their relative ordering is unaffected. So all in all, we find that the parametric MLRV has excellent robustness properties and remains among the top performing estimators. In a related study, [18] come to the same conclusion.

4.6 Overall comparison

In a comparison of nineteen different realized variance measures in eight distinct classes, three different types of data, and three different sample sizes, it is of interest from an applied viewpoint which estimator gives the best *all-round* performance. To facilitate discussion, consider Table 7. Here, a star (★) indicates "superior" performance where the RV measure is either the best performing or statistically indistinguishable from the best, a check-mark (✓) indicates "acceptable" performance where the RV measure is not more than 25% away from the best performing measure in terms of MSE, and all remaining RV measures are classified with a dagger (†) indicating "bad" performance.

Starting with the least desirable estimators, we find that RV is consistently among the worst performing irrespective of the sampling frequency, data type, or sample size. This finding is of course not surprising given its sensitivity to microstructure noise. Next is ALT which in certain circumstances is among the best performing, but is biased and lacks robustness to misspecification. While ZHOU constitutes an ideal estimator for mid-quote data in terms of performance and simplicity, its incurs a substantial efficiency loss with trade data and large samples of micro-price data. A similar pattern is observed for the FE. All remaining estimators perform very well.

On the one hand we have the non-parametric realized kernel and subsampling methods. These estimators are known to share some equivalence up to end-effects and their performance is really only distinguished by the choice of tuning parameters. Asymptotic convergence rates appear of secondary importance. When the tuning parameters are selected based on asymptotic criteria, this often gives good results, particularly for the realized TH_{16} kernel. However, as the micro-price results illustrate, the asymptotic bandwidth and subsample selection is susceptible to misspecification and an ad hoc rule may therefore be preferred: it is simpler and delivers an all-round good performance with a loss in MSE relative to the best performing measure of less than 10% when averaged across data types and sample sizes. On the other end of the spectrum, we have the fully parametric MLRV that ranks consistently as the best (or among the best) performing estimator. Only with the micro-price series does its performance deteriorate significantly. While the performance of the realized kernel and subsampling methods could be improved by overriding the theory-implied tuning parameters, the performance of MLRV can be improved by incorporating dependent noise into the model structure. Doing so, we find that also this extended MLRV attains a good all-round performance with an efficiency loss of less than 10% compared to the various best performing estimators, all different, across all the scenarios we considered.

Table 7 Ranking of alternative realized variance measures

	Trade prices			Quote prices			Micro-prices			loss		
$M=$	1,000	5,000	10,000	1,000	5,000	10,000	1,000	5,000	10,000	min	mean	max
1. Realized variance												
(a) highest ($q=1$)	†	†	†	†	†	†	†	†	†	0.479	3.806	9.362
(b) ad hoc (5 mins)	†	†	†	†	†	†	†	†	†	0.674	1.853	2.513
(c) q^*_{RV}	†	†	†	✓	†	†	†	†	†	0.196	0.912	2.103
2. Bias-corrected RV of Zhou [28]												
(a) highest ($q=1$)	†	†	†	✓	★	★	★	✓	†	0.009	0.535	1.422
(b) ad hoc (5 mins)	†	†	†	†	†	†	†	†	†	0.997	2.412	3.449
(c) q^*_{Zhou}	†	†	†	✓	★	★	★	✓	†	0.009	0.230	0.561
3. Two-scale RV of Zhang, Mykland, and Aït-Sahalia [27]												
(a) ad hoc ($q=5$)	✓	✓	✓	✓	✓	✓	✓	★	★	0.007	0.097	0.179
(b) q^*_{ZMA}	†	✓	✓	✓	★	★	★	✓	†	0.006	0.152	0.382
4. Multi-scale RV of Zhang [26]												
(a) ad hoc ($q=5$)	✓	✓	✓	✓	✓	✓	✓	★	★	0.006	0.098	0.191
(b) q^*_Z	✓	✓	✓	✓	★	★	★	✓	†	0.006	0.124	0.382
5. Realized kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard [10]												
(a) TH_2, ad hoc ($q=5$)	✓	✓	✓	✓	★	★	✓	★	★	0.000	0.095	0.211
(b) TH_2, q^*_{BNLS}	✓	★	★	✓	★	★	★	✓	†	0.013	0.087	0.357
(c) TH_{16}, q^*_{BNLS}	★	★	★	✓	★	★	✓	✓	†	0.008	0.074	0.335
(d) Cubic, q^*_{BNLS}	✓	✓	✓	✓	★	★	★	✓	†	0.009	0.105	0.387
6. Fourier estimator of Malliavin and Mancino [20]												
(a) DIR kernel, q^*_F	†	†	†	✓	✓	✓	✓	†	†	0.108	0.375	0.683
(b) FEJ kernel, q^*_F	†	†	†	✓	✓	✓	✓	†	†	0.077	0.303	0.613
7. ALT of Large [19]	★	†	†	†	†	†	†	†	†	0.000	1.051	2.050
8. MLRV of AMZ [2]	★	★	★	★	★	★	★	✓	†	0.000	0.052	0.314
MLRV-MA(2)	✓	✓	✓	✓	✓	✓	✓	✓	★	0.033	0.082	0.151

Note A star (★) indicates the best performing RV measure, or those statistically indistinguishable at a 1% bootstrapped confidence level, a checkmark (✓) indicates an "acceptable" RV measure with an MSE at within 25% distance from the optimum, and a dagger (†) indicates a "bad" RV measure with an MSE at more than 25% distance from the optimum

5 Conclusion

This paper compares a comprehensive set of "second generation" realized variance measures using a simulated "zero-intelligence" order book market of [25]. The emphasis in our investigation lies on statistical efficiency, implementation, and robustness. The order book model is ideally suited for this purpose as it provides a realistic microstructure setting where trade and quote price series can be studied in an internally consistent manner.

Based on the results presented here, we find that the best performing classes of estimators are the non-parametric realized kernel and subsampling methods and the parametric MLRV estimators. While implementation of the non-parametric methods requires the selection of tuning parameters, the MLRV requires appropriate model specification. Interestingly, a shared feature of both classes of estimators is that they are robust to time-varying parameters but sensitive to dependent noise.[9] Hence, our advice to the pragmatic practitioner is to sample prices at the highest available frequency and then measure realized variance using either (i) a non-parametric TSRV, MSRV, or KRV method with *ad hoc choice of tuning parameter* or (ii) a parametric MLRV that allows for moderate *noise dependence*. The performance of all these estimators is largely equivalent, and so which particular one to use would be a matter of taste. In specific circumstances, we find that a theory-driven implementation of the subsampling and realized kernel methods can deliver superior performance. Equally, under ideal conditions, MLRV performs (unsurprisingly) better than its dependent noise-extended counterpart. Yet, the micro-price results highlight that it is dangerous to rely too heavily on theory and model structure, and a better all-round performance is achieved by following the above guidelines. Doing so yields optimal or near-optimal performance and is likely to lead to substantial efficiency gains over the widely used sum of sparsely sampled returns following the "5-minute" rule prescribed in earlier literature.

In terms of data sampling we recommend the use of mid-quotes. When sampled immediately prior to a trade, we ensure the same number of observations as for the trade data but with a heavily reduced level of microstructure noise. The micro-price is also preferred over the trade data but, despite some seemingly appealing features, does not seem to improve over mid-quote data. This is likely due to the setup of our simulation where order placement is entirely random. In practice, it may well be that the micro-price, or modifications of this quantity, can lead to further efficiency improvements.

To conclude, it should be stressed that the main virtue of the ad hoc selection of subsampling frequency or bandwidth, as suggested here, lies in its simplicity and robustness together with good all-round efficiency. Still, if one is willing to put in additional effort, there can be scope for refinement of this admittedly crude approach, albeit with no guarantee of success. For instance, we know that there is significant variation in the level of noise across data types as well as in the cross section and time series of security prices, and it is only natural for this to be reflected in the selection of bandwidth. In general, the rule should be to let q grow with the level of noise, so that with little noise we compute something that is close to RV and with high levels of noise we effectively reduce the sampling frequency so as to mitigate its impact. The theory-implied optimal choice of bandwidth, as in (4.1), embodies this principle and ensures that q^* grows with the square root of the noise ratio ξ (if one decides to use this rule, we suggest to set $\lambda = 1$ which leads to conservative values of q^* and has the added advantage that the integrated variance and quarticity need not

[9] As already mentioned, modifications of the TSRV, MSRV, and KRV do exist to gain robustness to dependent and endogenous noise; see [3] and [11].

be estimated). Yet, an important drawback of this semi-parametric approach to selecting the bandwidth is a potential discrepancy between asymptotic optimality and finite sample performance, as well as misspecification of the noise process. Any alternative that seeks to improve over the ad hoc rule, should face these non-trivial challenges. In addition, in certain circumstances it may be sensible to focus on application-specific *economic* criteria rather than statistical ones (as is done here) when evaluating the performance of a realized variance measure. Recent work in this direction includes [9] that focuses on bandwidth selection in the context of option pricing and [1, 6, 16] that focuses on volatility forecasting in the presence of noise.

References

1. Aït-Sahalia, Y., Mancini, L.: Out of sample forecasts of quadratic variation. J. Econom. **147**, 17–33 (2008)
2. Aït-Sahalia, Y., Mykland, P., Zhang, L.: How often to sample a continuous-time process in the presence of market microstructure noise. Rev. Financ. Stud. **18**, 351–416 (2005)
3. Aït-Sahalia, Y., Mykland, P., Zhang, L.: Ultra-high frequency volatility estimation with dependent microstructure noise. J. Econom. (2006, forthcoming). Available at http://www.princeton.edu/~yacine/
4. Andersen, T.G., Bollerslev, T.: Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. Int. Econ. Rev. **39**, 885–905 (1998)
5. Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H.: The distribution of stock return volatility. J. Financ. Econ. **61**, 43–76 (2001)
6. Andersen, T.G., Bollerslev, T., Meddahi, N.: Market microstructure noise and realized volatility forecasting. J. Econom. (2006, forthcoming). Available at http://gremaq.univ-tlse1.fr/perso/meddahi/
7. Bandi, F.M., Russell, J.R.: Comment on JBES 2005 invited address by Peter R. Hansen and Asger Lunde. J. Bus. Econ. Stat. **24**(2), 167–173 (2006)
8. Bandi, F.M., Russell, J.R.: Separating microstructure noise from volatility. J. Financ. Econ. **79**, 655–692 (2006)
9. Bandi, F.M., Russell, J.R., Yang, C.: Realized volatility forecasting and option pricing. J. Econom. **147**, 34–46 (2008)
10. Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N.: Designing realised kernels to measure the ex post variation of equity prices in the presence of noise. Econometrica **76**, 1481–1536 (2008)
11. Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N.: Subsampling realised kernels. J. Econom. (2009, forthcoming)
12. Bollerslev, T., Domowitz, I., Wang, J.: Order flow and the bid–ask spread: An empirical probability model of screen-based trading. J. Econ. Dyn. Control **21**, 1471–1491 (1997)
13. Bouchaud, J.-P., Gefen, Y., Potters, M., Wyart, M.: Fluctuations and response in financial markets: The subtle nature of 'random' price changes. Quant. Finance **4**, 176–190 (2004)
14. Domowitz, I., Wang, J.: Auctions as algorithms: Computerized trade execution and price discovery. J. Econ. Dyn. Control **18**, 29–60 (1994)
15. Farmer, J.D., Patelli, P., Zovko, I.I.: The predictive power of zero intelligence in financial markets. Proc. Natl. Acad. Sci. USA **102**, 2254–2259 (2005)
16. Ghysels, E., Sinko, A.: Volatility forecasting and microstructure noise. J. Econom. (2006, forthcoming). Available at http://www.unc.edu/~eghysels/working_papers.html
17. Griffin, J.E., Oomen, R.C.: Sampling returns for realized variance calculations: Tick time or transaction time? Econom. Rev. **27**(1–3), 230–253 (2008)
18. Hansen, P., Large, J., Lunde, A.: Moving average-based estimators of integrated variance. Econom. Rev. **27**(1–3), 79–111 (2008)
19. Large, J.: Estimating quadratic variation when quoted prices jump by a constant increment. J. Econom. (2005, forthcoming)
20. Malliavin, P., Mancino, M.E.: Fourier series method for measurement of multivariate volatilities. Finance Stoch. **6**, 49–61 (2002)
21. Malliavin, P., Mancino, M.E.: A Fourier transform method for nonparametric estimation of multivariate volatility. Ann. Stat. **37**, 1983–2010 (2009)

22. Mancino, M.E., Sanfelici, S.: Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. Comput. Stat. Data Anal. **52**, 2966–2989 (2008)
23. Niederhoffer, V., Osborne, M.F.M.: Market making and reversal on the stock exchange. J. Am. Stat. Assoc. **61**(316), 897–916 (1966)
24. Oomen, R.C.: Properties of realized variance under alternative sampling schemes. J. Bus. Econ. Stat. **24**, 219–237 (2006)
25. Smith, E., Farmer, J.D., Gillemot, L., Krishnamurthy, S.: Statistical theory of the continuous double auction. Quant. Finance **3**, 481–514 (2003)
26. Zhang, L.: Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Bernoulli **12**, 1019–1043 (2006)
27. Zhang, L., Mykland, P.A., Aït-Sahalia, Y.: A tale of two timescales: Determining integrated volatility with noisy high frequency data. J. Am. Stat. Assoc. **100**, 1394–1411 (2005)
28. Zhou, B.: High frequency data and volatility in foreign-exchange rates. J. Bus. Econ. Stat. **14**, 45–52 (1996)