Zero-intelligence realized variance estimation

Jim Gatheral and Roel C.A. Oomen*

12 March 2007

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 view point, the choice of which particular estimator to use is not a trivial one because the study of their relative merits has primarily focussed 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 twenty 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 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 or realized kernel delivers the best overall result. We make firm practical recommendations on choosing and implementing a realized variance estimator, as well as data sampling.

*Gatheral is with Merrill Lynch, New York and the Courant Institute, NYU. Oomen is with Deutsche Bank, London and the Department of Finance, Warwick Business School, United Kingdom. E-mail: roel.oomen@wbs.ac.uk. Oomen is also a research affiliate of the Department of Quantitative Economics at the University of Amsterdam, The Netherlands. The authors wish to thank participants of the workshop on high-frequency data at Imperial College London and Yacine Aït-Sahalia for helpful comments and suggestions.

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 by Bandi, Russell, and Yang (2006), 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 Andersen, Bollerslev, Diebold, and Ebens (2001), 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 Zhang, Mykland, and Aït-Sahalia (2005), "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 dataset. Both the design and the analysis of these estimators is 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. Gaussian 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 criteria 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 twenty 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

[1] Notable exceptions include Bandi, Russell, and Yang (2006) who concentrate on the payoff of an option trading strategy based on alternative realized variance measures, and Aït-Sahalia and Mancini (2006), Andersen, Bollerslev, and Meddahi (2006), and Ghysels and Sinko (2006) who consider the forecast accuracy in the presence of noise.

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 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 of prices generated by this artificial market, 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 actually adds some noise to the sampled price series. Next, with regard to the choice of variance estimators, we find that those based on subsampling or realized kernels perform best overall, both in terms of efficiency and in terms of robustness. However, implementation of these estimators requires one to select tuning parameters such as a bandwidth for realized kernels and a sampling grid for subsampling methods. 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. Yet, we find that an *ad hoc* choice of tuning parameters gives reasonably efficient, and sometimes even superior, results. Indeed, averaged across sampling schemes and sample sizes, the realized kernels with bandwidth of 5 and the subsampling estimators with 5 subsamples, yield the best all-round performance.

The remainder of this paper is organized as follows. In Section 2, we describe the zero-intelligence market. In Section 3, we discuss the realized variance measures studied in this paper and provide details on their implementation. In Section 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 Smith, Farmer, Gillemot, and Krishnamurthy (2003, SFGK hereafter). According to the specification of this model, limit orders may 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 $b(t)$ at time t and limit buy orders at any level less than the best offer $a(t)$. 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.

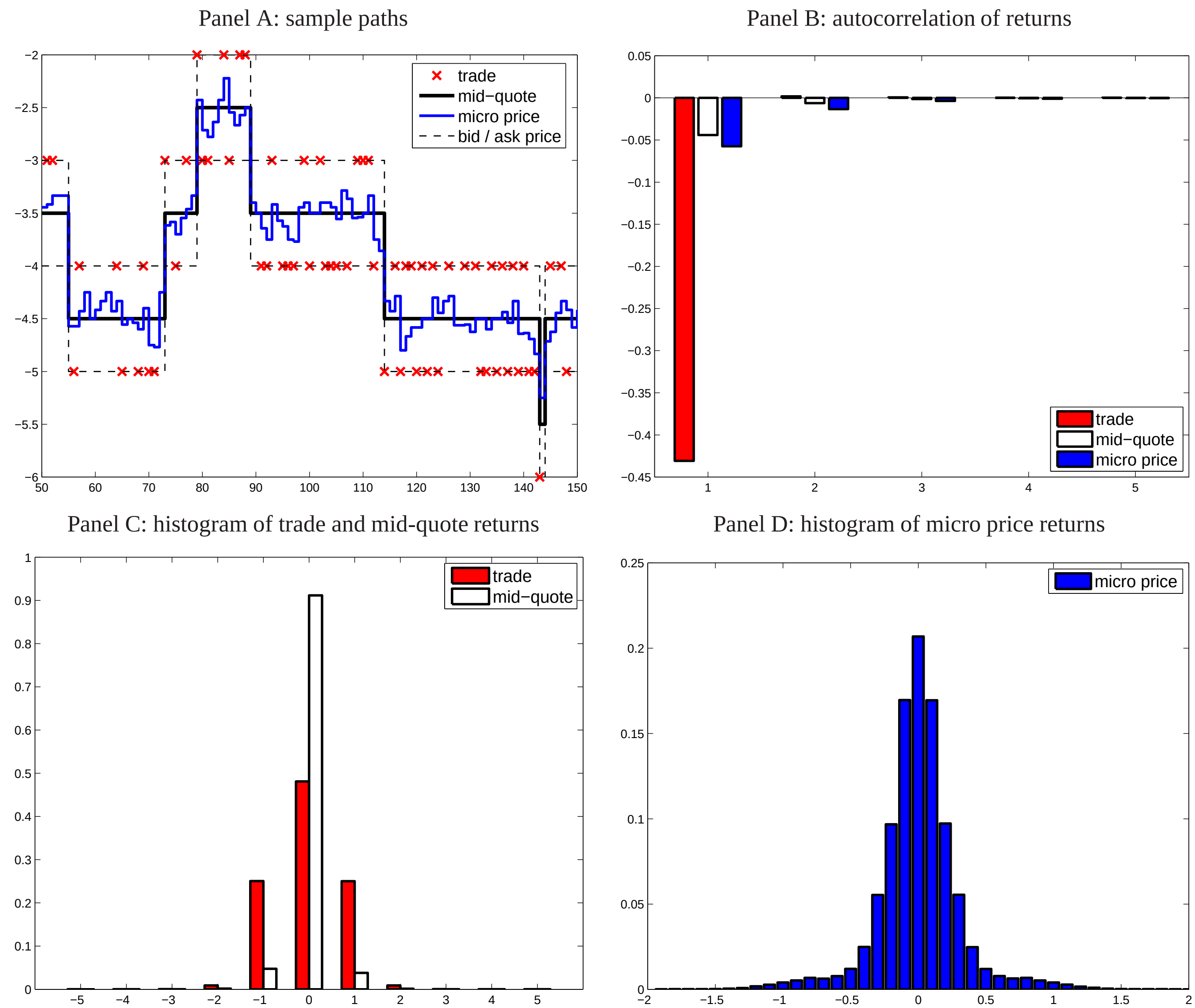
The SFGK model depends on so few parameters that we can use dimensional analysis to predict relationships between statistical properties of the model and its inputs. For example, it is easy to see that the asymptotic book depth far away from the best quote must be given by α/δ. This is because in the 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 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 $a(t)$, the best bid $b(t)$ 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 do move, and each level in the book retains some memory of having been "visited" by the market price at some prior time. Stricly 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 by Domowitz and Wang (1994) was shown by Bollerslev, Domowitz, and Wang (1997) to accurately predict the distribution of bid-offer spreads in the foreign exchange market. Later, Farmer, Patelli, and Zovko (2005) 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 by Bouchaud, Gefen, Potters, and Wyart (2004), 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. Gaussian noise" model widely used in

Figure 1: Properties of the sampled price processes



the realized variance literature.

To conclude our discussion of the SFGK model, consider the descriptive statistics of some simulated price series in Figure 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 distinct dynamics 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 higher order autocorrelations close to negligible. In contrast, the first order autocorrelation of mid-quote changes is much smaller at about -5%, reflecting a reduction of microstructure noise. The autocorrelation structure of the micro price is similar to that of 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 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. Further, the mid-quote moves only once for every 10 trades on average. 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 Overview of realized variance measures

Let $\{p_i\}_{i=0}^M$ denote a time-series of observed logarithmic prices. Without loss of generality we think of this as a realization of the trade, quote, or micro price process over a single day. 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, i.e. $(p_M - p_0)^2$ but, as pointed out by Andersen and Bollerslev (1998), this is a very noisy estimator and can be improved upon by using intra-day data. In particular, note the following 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 Figure 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

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

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 $W \perp \varepsilon$. Their implementation requires measurements of the following latent quantities: integrated variance $IV = \int_0^1 \sigma_u^2 du$, integrated quarticity $IQ = \int_0^1 \sigma_u^4 du$, and noise variance ω^2. While our ZI market data is not necessarily well characterized by such a model, the latent quantities can of course still be estimated from the simulated data, albeit that they might be misspecified. This 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

$$RV = \frac{M/q}{\lfloor M/q \rfloor}\gamma_{0,q}(0), \tag{2}$$

(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 of Bandi and Russell (2006b)

$$q^*_{BR} = M \left(\frac{IQ}{(2\omega^2)^2} \right)^{-1/3},$$

using Bandi and Russell (2006b) estimate of the noise variance, i.e.

$$\widehat{\omega}^2_{BR} = \gamma_{0,1}(0)/(2M), \tag{3}$$

(d) at optimal sampling frequency of Oomen (2006)

$$q^*_O = M \left(\frac{M}{2\xi} \right)^{-2/3},$$

using Oomen (2006) estimate of the noise variance, i.e.

$$\widehat{\omega}^2_O = -\gamma_{0,1}(1), \tag{4}$$

and corresponding noise ratio estimate $\widehat{\xi}_O = \omega^2_O/(\gamma_{0,1}(0) + 2\gamma_{0,1}(1))$.

2. Bias-corrected RV of Zhou (1996)

$$ZHOU = \frac{M}{M-q+1} \frac{1}{q} \sum_{h=0}^{q-1} (\gamma_{h,q}(0) + 2\gamma_{h,q}(1)), \tag{5}$$

(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 Zhou (1996) $q^*_{Zhou} = \max\{1, 2\xi/\sqrt{3}\}$ using $\widehat{\xi}_O$ as an estimate for the noise ratio.

3. Two-Scale RV of Zhang, Mykland, and Aït-Sahalia (2005)

$$TSRV = \left(1 - \overline{M}/M\right)^{-1} \left(\frac{1}{q} \sum_{h=0}^{q-1} \gamma_{h,q}(0) - \frac{\overline{M}}{M} \gamma_{0,1}(0) \right), \tag{6}$$

where $\overline{M} = (M - q + 1)/q$

(a) at ad-hoc sub-sampling frequency of $q = 5$,

(b) at optimal sub-sampling frequency of Zhang, Mykland, and Aït-Sahalia (2005) $q^*_{ZMA} = \left(\frac{12(\omega^2)^2}{IQ} \right)^{1/3} M^{2/3}$ and using Bandi and Russell (2006b) estimate of the noise variance $\widehat{\omega}^2_{BR}$,

(c) at optimal sub-sampling frequency of Zhang, Mykland, and Aït-Sahalia (2005) $q^*_{ZMA} = \left(\frac{12(\omega^2)^2}{IQ} \right)^{1/3} M^{2/3}$ and using Oomen (2006) estimate of the noise variance $\widehat{\omega}^2_O$,

(d) at optimal sub-sampling frequency of Bandi and Russell (2005) $q^*_{BR} = \left(\frac{3IV^2}{2IQ}\right)^{1/3} M^{1/3}$ and using Bandi and Russell (2006b) estimate of the noise variance $\widehat{\omega}^2_{BR}$.

4. Multi-Scale RV of Zhang (2006)

$$MSRV = \sum_{j=1}^{q} \frac{a_j}{j} \sum_{h=0}^{j-1} \gamma_{h,j}(0), \tag{7}$$

where

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

The optimal number of subsamples is equal to $q^*_Z = c^*\sqrt{M}$ where

$$c^* = \arg\min_c\{2\frac{52}{35}cIQ + \frac{48}{5}c^{-1}\omega^2(IV + \omega^2/2) + 48c^{-3}\omega^4\}. \tag{8}$$

(a) with ad-hoc number of $q = 5$ subsamples,

(b) with optimal number of subsamples of Zhang (2006) q^*_Z and using Bandi and Russell (2006b) estimate of the noise variance $\widehat{\omega}^2_{BR}$,

(c) with optimal number of subsamples of Zhang (2006) q^*_Z and using Oomen (2006) estimate of the noise variance $\widehat{\omega}^2_O$.

5. Realized Kernel estimator of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a)

$$KRV = \gamma_{0,1}(0) + 2\sum_{s=1}^{q} \kappa\left(\frac{s-1}{q}\right)\gamma_{0,1}(s). \tag{9}$$

For given choice of kernel $\kappa(x)$, the optimal bandwidth is equal to $q^*_{BHLS} = c^*\sqrt{M}$ where

$$c^* = \arg\min_c\left\{4c\kappa_\bullet^{0,0}IQ - 8c^{-1}\kappa_\bullet^{0,2}\omega^2(IV+\omega^2/2) + 4c^{-3}\omega^4(\kappa'''(0) + \kappa_\bullet^{0,4})\right\}, \tag{10}$$

where $\kappa_\bullet^{0,0} = \int_0^1 \kappa(x)^2\,dx$, $\kappa_\bullet^{0,2} = \int_0^1 \kappa(x)\kappa''(x)\,dx$, $\kappa_\bullet^{0,4} = \int_0^1 \kappa(x)\kappa''''(x)\,dx$. For the Modified Tukey-Hanning kernel TH$_2$, 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 Cubic kernel, i.e. $\kappa(x) = 1 - 3x^2 + 2x^3$, we have $\kappa'''(0) = 12$, $\kappa_\bullet^{0,0} = \frac{13}{35}$, $\kappa_\bullet^{0,2} = -\frac{6}{5}$, $\kappa_\bullet^{0,4} = 0$

(a) modified Tukey-Hanning kernel TH$_2$ with ad-hoc bandwidth $q = 5$,

(b) modified Tukey-Hanning[2] kernel TH$_2$ with Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a) optimal bandwidth q^*_{BHLS} and using Bandi and Russell (2006b) estimate of the noise variance $\widehat{\omega}^2_{BR}$,

(c) modified Tukey-Hanning kernel TH$_2$ with Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a) optimal bandwidth q^*_{BHLS} and using Oomen (2006) estimate of the noise variance $\widehat{\omega}^2_O$,

(d) cubic kernel with Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a) optimal bandwidth q^*_{BHLS} and using Oomen (2006) estimate of the noise variance $\widehat{\omega}^2_O$.

[2] Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006b) show that the TH$_{16}$ attains the best asymptotic efficiency with $\kappa'''(0) = 5760\pi^2$, $\kappa_\bullet^{0,0} = 0.0317324$, $\kappa_\bullet^{0,2} = -10.26423$, $\kappa_\bullet^{0,4} = -42474.9$. However, because in our simulations the optimal kernel bandwidth appears much too high we don't implement this estimator.

6. Alternation estimator of Large (2005)

$$ALT = \frac{M_c}{M_r}\gamma_{0,1}(0), \tag{11}$$

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.

7. Maximum Likelihood estimator of Aït-Sahalia, Mykland, and Zhang (2005)

$$MLRV = M\widehat{\delta}^2(1+\widehat{\eta})^2, \tag{12}$$

where $(\widehat{\eta}, \widehat{\delta^2})$ are the maximum likelihood estimates of an MA(1) model for observed returns, i.e. $r_i = \varepsilon_i + \eta\varepsilon_{i-1}$ where the $\varepsilon_i' s$ are serially uncorrelated with mean zero and variance δ^2.

3.1 Comparison of theoretical properties

In the absence of noise, 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. 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 time scales 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 and cubic kernels we consider here converges 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 it does not allow for stochastic volatility.

So a priori, based on their robustness and asymptotic properties, we would anticipate the following ordering of estimators in term of their performance:

1. MSRV and KRV with convergence rate $M^{-1/4}$

2. MLRV and ALT with convergence rate $M^{-1/4}$ but lack of robustness to misspecification

3. TSRV with convergence rate $M^{-1/6}$

4. ZHOU with inconsistency and lack of robustness to i.i.d. noise departures

5. RV with exploding bias

Of course, faster *asymptotic* convergence rates do not necessarily translate into superior *finite sample* performance.[3] Indeed, the rate of convergence may well be of second order importance in small samples. Moreover, when the sampled price path violates assumptions underlying the construction of the realized measure, its performance may be affected negatively unless there is some inherent robustness. For instance, in relation to the MSRV Zhang (2006) writes "*…though we have assumed that the ϵ are i.i.d., our estimator is quite robust to the nature of the noise*". Also, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006b) propose a subsampling version of their realized kernels that "*…leads to valid inference that is robust to both time-dependent and endogenous noise*". So 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 motivate our simulation study.

3.2 Notes on implementation

3.2.1 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 Eq. (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 implement the unaltered version. The realized kernels, when calculated at the highest frequency as is done here, conveniently don't require any bias correction. The same is the case for ALT and MLRV.

3.2.2 IV and IQ measurement

For the calculation of the optimal sampling frequencies, subsamples, or bandwidths q^*, measurements of the latent quantities IV and IQ are required. 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.

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

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

3.2.3 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 Eqs. (8) and (10) respectively. Because the first order condition for optimality is a quadratic form

[3] Suppose, for instance, we have one estimator with asymptotic variance of $M^{-1/3}$ and another with $\alpha M^{-1/2}$. Although the latter converges at a faster rate, it will only have a smaller variance when $M > \exp(6\log(\alpha))$, e.g. for $\alpha = 4$ we require $M > 4096$ and for $\alpha = 7$ we need $M > 117649$!

in c^2 we can obtain the following closed form expression:

$$q^* = \left(\frac{b\omega^2 \left(IV + \omega^2/2\right)}{2aIQ} + \sqrt{\frac{b^2\omega^4 \left(IV + \omega^2/2\right)^2}{4a^2 IQ^2} + \frac{\omega^4 d}{aIQ}} \right)^{1/2} \sqrt{M} \tag{13}$$

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.

3.2.4 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[4] 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 following identities:

$$\begin{aligned} ZHOU &= KRV \\ &= MSRV + r_1^2 + r_M^2 \\ &= TSRV + \frac{TSRV - RV}{M} + r_1^2 + r_M^2 \end{aligned} \tag{14}$$

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:

$$\frac{IV^2}{M^2} \left(2M\lambda^{-1} + 8M\xi - 4\xi^2 + 12M\xi^2 + 4\xi^2 M^2\right), \tag{15}$$

and the MSE of ZHOU is:

$$\frac{IV^2}{M^2} \left(6M\lambda^{-1} + 8M\xi - 6\xi^2 + 8M\xi^2\right). \tag{16}$$

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

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

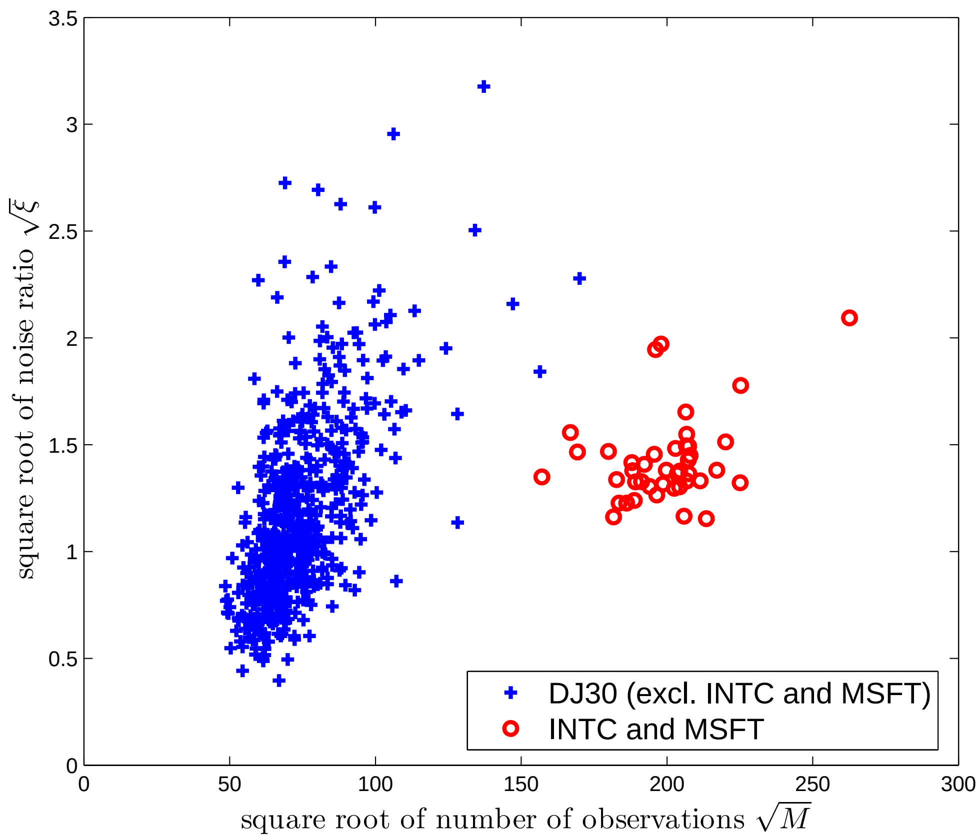
and round up otherwise. Note, however, that even for a relatively modest sample size of $M = 1,000$ (the smallest we consider below) we would always compute ZHOU unless $\xi < 0.03$. Moreover, when taking into account the measurement error in estimates of ξ and the limited efficiency gains to be had in the first place, it seem 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

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

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

Figure 2: Magnitude of microstructure noise in trade data for DJ30 components (October 2004)



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 Oomen (2006) defined as $\xi = M\omega^2/IV$. Intuitively, ξ measures the magnitude of the noise relative to that of the per-tick efficient 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 Figure 2, is pretty typical for a component of the DJ30 index. Following the discussion of Section 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.

Because it is clearly impossible to simulate order arrivals and cancelations at integer price levels extending from $-\infty$ to $+\infty$, we simplify the simulation by only considering order arrivals and cancelations in a moving band of price levels centered around the current best quotes. The width of this band is chosen conservatively so as to ensure minimal edge effects: in this particular simulation, we select a bandwidth of 10. 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 $b(t)$ and best offer $a(t)$.
- Compute the number N_b of shares on the bid side of the book from level $a(t)-1$ to level $a(t)-10$.
- Compute the number N_a of shares on the offered side of the book from level $b(t)+1$ to level $b(t)+10$.
- Set $\delta_b = \delta\, N_b;\ \delta_a = \delta\, N_a$.
- Draw a new event according to the relative probabilities $\{\mu/2, \mu/2, 10\,\alpha, 10\,\alpha, \delta_a, \delta_b\}$. 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 $\{1, 2, \ldots, 10\}$.
- If the selected event is a cancelation, select randomly which order within the band to cancel.
- Update the order book and increment t.

Note that the relative probabilities δ_b and δ_a need to be recomputed at each time step because they depend on the number of outstanding limit orders within the band.

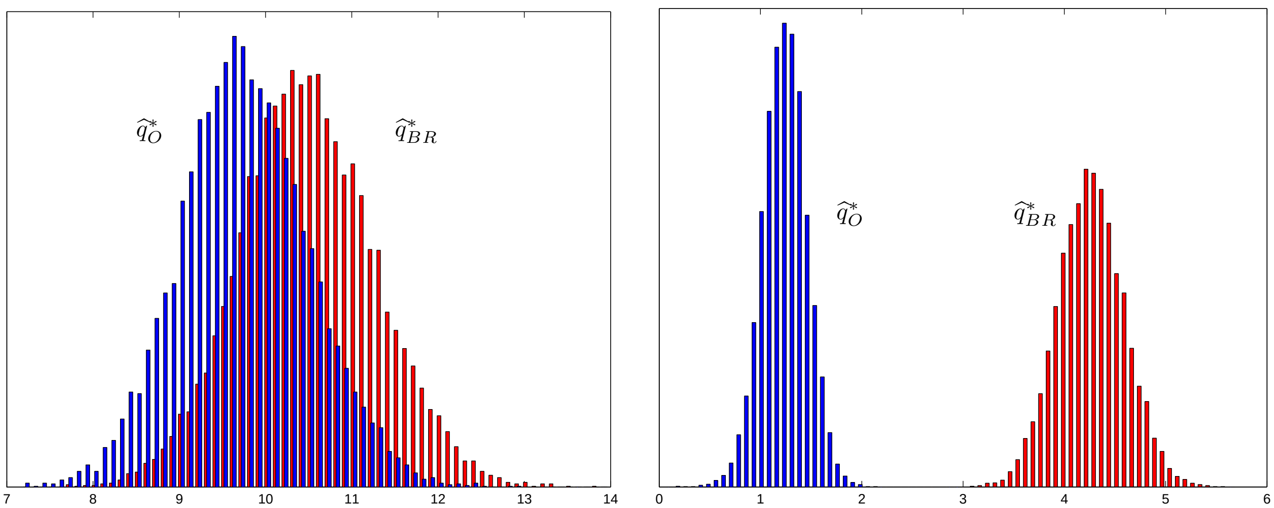
Following the above, we simulate the order book until $10,000$ trades (i.e. market orders) have been executed, storing the trade price and the mid-quote just prior each trade together with corresponding market depths at the best quotes $b(t)$ and $a(t)$. The latter quote sizes permit us to compute the micro price series. We repeat each such simulation $12,500$ times. In reporting our results, we standardize all realized variance estimates by the “true” daily return variance (which in our setup can be measured to any desirable accuracy using simulations). So an unbiased estimator should average 1.

4.1 Results for trade prices

With trade prices and $M = 1,000$, we can see from Table 2 that ALT of Large (2005) and MLRV of Aït-Sahalia, Mykland, and Zhang (2005) 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 Figure 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 provided that the Tukey-Hanning kernel with optimal bandwidth selection is used. The cubic kernel leads to some limited efficiency loss, and so does the 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. In relation to this, it is interesting to note that the Bandi-Russell “finite sample optimized” bandwidth selection procedure leads to better results than the ZMA “asymptotic” optimal bandwidth. 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 -3.4 compared to -3.5 for MLRV. In sharp contrast, when moving to ZHOU in fourth best place we observe a considerable drop in performance. When using all data (i.e. $q = 1$) the

Figure 3: Bandi-Russell and Oomen optimal TH_2 kernel bandwidth estimates for trade and mid-quote data ($M = 10,000$)

Panel A: histogram for trade data Panel B: histogram for mid-quote data



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. As anticipated, RV comes in last place with a significant positive bias and very high MSE irrespective of the choice of sampling frequency.

Considering the different choices of bandwidth, 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 leads to the lower MSE. Comparing the performance across different subsamples for TSRV we observe a clear asymmetry in loss associated with non-optimal bandwidth selection. In particular, we attain a log MSE of -3.4 at $q = 5$ which increases quite rapidly to -3.27 when we use the slightly lower (average) bandwidth of $q_{ZMA} = 3.95$ and yet only increases very slowly to -3.39 with the much higher Bandi-Russell bandwidth q_{BR} of over 11. Finally, comparing the alternative noise variance estimates, which serve as key input parameters to the optimal bandwidth selection procedure, there does not seem to be a significant difference between Oomen's and Bandi-Russell's estimate. Panel A of Figure 3 plots a histogram of $\widehat{q}_{BR}^* = M\widehat{\omega}_{BR}^2/\widehat{IV}$ and $\widehat{q}_O^* = M\widehat{\omega}_O^2/\widehat{IV}$. As expected $\widehat{q}_{BR}^*$ has a slightly higher mean (because of its well documented bias) but there is still an almost perfect correlation of 99.2% between the two.

Now let's consider what happens when we grow the sample size. Keep in mind here, that we should not think of this as increasing the sampling frequency – after all, we would 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 horizonal move in Figure 2). Both with $M = 5,000$ and $M = 10,000$ MLRV remains the best performing estimator closely followed by KRV. In fact, for large sample sizes the difference in MSE between the two is statistically indistinguishable. 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. 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

Table 1: Summary statistics of sample estimates of auxiliary variables

		$\widehat{IV}$		$\widehat{\lambda} = \widehat{IV}^2/\widehat{IQ}$		$\widehat{\xi}_{BR} = M\widehat{\omega}^2_{BR}/\widehat{IV}$		$\widehat{\xi}_O = M\widehat{\omega}^2_O/\widehat{IV}$	
data	M	mean	std	mean	std	mean	std	mean	std
trade	$1,000$	1.487	0.178	0.916	0.099	2.366	0.267	2.042	0.302
	$5,000$	1.089	0.140	0.975	0.101	3.240	0.413	2.793	0.378
	$10,000$	1.044	0.136	0.998	0.103	3.385	0.442	2.918	0.393
quote	$1,000$	1.007	0.173	0.781	0.130	0.561	0.072	0.051	0.046
	$5,000$	0.999	0.140	0.970	0.116	0.566	0.075	0.050	0.022
	$10,000$	0.999	0.136	0.998	0.112	0.567	0.075	0.050	0.016
micro	$1,000$	1.015	0.171	0.795	0.129	0.589	0.076	0.071	0.048
	$5,000$	1.001	0.140	0.970	0.116	0.597	0.079	0.069	0.023
	$10,000$	0.999	0.136	0.998	0.112	0.598	0.080	0.069	0.018

before, and not surprisingly, RV comes in last place.

We note that 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 Eq. (13) and use the convenient notation of the noise ratio $\xi = \omega^2/\left(IV/M\right)$ and $\lambda = IV^2/IQ$ then we have:

$$q^* = \left(\lambda \frac{b}{2a} + \frac{b}{2a}\sqrt{\lambda^2 + 4\lambda ad/b^2}\right)^{1/2} \sqrt{\xi}. \tag{17}$$

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 over-estimation of IV in small sample sizes which leads to under-estimation of the noise ratio ξ as well as λ (see Table 1 for detailed summary statistics on these variables). Another interesting pattern that emerges from Table 2 is that while in large samples the "asymptotically" optimal bandwidth typically outperforms the ad-hoc rule, the performance of TSRV with the "finite-sample optimized" Bandi-Russell bandwidth rapidly deteriorates when M grows. To gain some intuition here, note that for TSRV we have $q^*_{ZMA} = (12\lambda\xi^2)^{1/3}$ and $q^*_{BR} = (1.5\lambda M^2)^{1/3}$. So in line with the discussion above, q^*_{ZMA} remains roughly constant as M grows whereas q^*_{BR} does not even depend on ξ and grows at rate $M^{2/3}$ which explains why for large samples the selected bandwidth is too high.

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 Figure 2) 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

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 ($q = 1$)	6.938	0.277	3.565	7.109	1.00	6.944	0.124	3.565	8.697	1.00	6.947	0.086	3.566	9.362	1.00
(b) ad-hoc (5 mins)	1.485	0.258	-1.198	2.346	12.00	1.088	0.182	-3.202	1.930	64.00	1.044	0.171	-3.464	2.332	128.00
(c) $q^*_{BR}, \widehat{\omega}^2_{BR}$	1.210	0.317	-1.933	1.612	27.29	1.092	0.180	-3.199	1.933	58.77	1.067	0.144	-3.681	2.115	76.81
(d) $q^*_O, \widehat{\omega}^2_O$	1.126	0.499	-1.328	2.216	57.85	1.061	0.209	-3.046	2.086	90.89	1.049	0.163	-3.543	2.254	113.71
2. Bias-corrected RV of Zhou (1996)															
(a) highest ($q = 1$)	0.966	0.336	-2.169	1.375	1.00	0.963	0.150	-3.732	1.400	1.00	0.961	0.105	-4.374	1.422	1.00
(b) ad-hoc (5 mins)	1.002	0.283	-2.522	1.022	12.00	0.997	0.264	-2.662	2.470	64.00	0.997	0.263	-2.674	3.122	128.00
(c) $q^*_{Zhou}, \widehat{\omega}^2_O$	0.989	0.241	-2.844	0.700	4.51	0.997	0.095	-4.708	0.424	3.70	0.998	0.067	-5.393	0.404	3.64
3. Two-Scale RV of Zhang, Mykland, and Aït-Sahalia (2005)															
(a) ad-hoc ($q = 5$)	0.979	0.182	-3.391	0.153	5.00	0.984	0.081	-4.983	0.149	5.00	0.984	0.057	-5.645	0.152	5.00
(b) $q^*_{ZMA}, \widehat{\omega}^2_{BR}$	0.976	0.194	-3.269	0.275	3.95	0.982	0.081	-4.988	0.144	4.96	0.984	0.057	-5.658	0.138	5.14
(c) $q^*_{ZMA}, \widehat{\omega}^2_O$	0.979	0.205	-3.156	0.388	3.57	0.981	0.082	-4.948	0.184	4.48	0.983	0.058	-5.618	0.178	4.66
(d) $q^*_{BR}, \widehat{\omega}^2_{BR}$	0.980	0.183	-3.387	0.157	11.06	0.990	0.092	-4.765	0.367	19.39	0.993	0.070	-5.308	0.488	24.61
4. Multi-Scale RV of Zhang (2006)															
(a) ad-hoc ($q = 5$)	0.975	0.181	-3.399	0.146	5.00	0.984	0.080	-4.999	0.133	5.00	0.985	0.057	-5.667	0.129	5.00
(b) $q^*_Z, \widehat{\omega}^2_{BR}$	0.971	0.186	-3.341	0.204	4.41	0.983	0.080	-5.018	0.114	5.25	0.986	0.056	-5.701	0.096	5.42
(c) $q^*_Z, \widehat{\omega}^2_O$	0.974	0.191	-3.293	0.251	4.09	0.983	0.080	-5.002	0.130	4.91	0.985	0.057	-5.674	0.122	5.04
5. Realized Kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a)															
(a) TH_2, ad-hoc ($q = 5$)	0.988	0.188	-3.342	0.202	5.00	0.983	0.083	-4.936	0.195	5.00	0.982	0.059	-5.585	0.211	5.00
(b) TH_2, $q^*_{BHLS}, \widehat{\omega}^2_{BR}$	0.992	0.176	-3.476	0.068	8.56	0.991	0.078	-5.083	0.049	10.23	0.992	0.056	-5.760	**0.036**	10.53
(c) TH_2, $q^*_{BHLS}, \widehat{\omega}^2_O$	0.993	0.176	-3.470	0.074	7.95	0.991	0.078	-5.086	0.046	9.50	0.991	0.055	-5.765	**0.031**	9.78
(d) Cubic, $q^*_{BHLS}, \widehat{\omega}^2_O$	0.989	0.184	-3.378	0.166	5.38	0.986	0.079	-5.037	0.095	6.42	0.987	0.056	-5.711	0.085	6.61
6. ALT of Large	0.888	0.128	**-3.544**	★	–	0.889	0.057	-4.160	0.972	–	0.889	0.040	-4.273	1.523	–
7. MLE of AMZ	0.987	0.172	-3.520	**0.024**	–	0.988	0.076	**-5.132**	★	–	0.988	0.054	**-5.796**	★	–

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^*’) each realized variance measure and across sample size M. Loss levels in boldface are insignificantly different from zero at a 1% bootstrapped confidence level.

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 is to sample the mid-quote process rather than actual trade prices which are known to "bounce" (e.g. Niederhoffer and Osborne, 1966). 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 Figure 1. The first order autocorrelation of mid-quote returns is now less than 5%, compared to over -40% for trade data. 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 Bandi and Russell (2006a) and Bouchaud, Gefen, Potters, and Wyart (2004) 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 estimator is in most cases statistically insignificant. Interestingly, the method for estimating the noise variance is unimportant here because the estimates of $\widehat{\omega}^2$ are so small[5] that the optimal q^* is often set to its minimum value of 2 for TSRV and MSRV and 1 for ZHOU and KRV. In this case, we know from the relation in Eq. (14) that all these estimators are equivalent up to end-effects and so it is not surprising that they attain a similar MSE. 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 the 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.

As already mentioned above, with the reduced noise in mid-quote data the optimally chosen sampling frequencies or bandwidths often need to be bound from below. The only exception to this is the bandwidth for the TH_2 kernel when using Bandi-Russell noise variance estimates. If we recall that, loosely speaking, $E(RV) = IV + 2M\omega^2$ and $\widehat{\omega}^2_{BR} = RV/2M$ then it is quite clear that this estimator will be biased upwards particularly when ω^2 is small. This is exactly what happens with the mid-quote data. Note from Table 1 that $\widehat{\xi}^2_{BR}$ is about 0.56 whereas $\widehat{\xi}^2_O$ – which does not suffer from this kind of bias – is more than 10 times smaller at 0.05! Panel B of Figure 3 plots the corresponding histogram of optimal bandwidth implied by these two noise variance measures. As expected, the mean of $\widehat{q}^*_O$ is about $\sqrt{10}$ times lower than the mean of

[5] The maximum TSRV sub-sampling frequency obtained over 12,500 simulation runs is 2.2341 when using $\widehat{\omega}^2_{BR}$ and 0.5906 when using $\widehat{\omega}^2_O$

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 ($q = 1$)	1.111	0.136	-3.480	0.394	1.00	1.113	0.061	-4.108	1.367	1.00	1.113	0.043	-4.224	1.916	1.00
(b) ad-hoc (5 mins)	1.007	0.197	-3.248	0.626	12.00	0.999	0.168	-3.568	1.907	64.00	0.999	0.165	-3.609	2.530	128.00
(c) $q^*_{BR}, \widehat{\omega}^2_{BR}$	1.003	0.184	-3.387	0.487	9.89	1.001	0.100	-4.601	0.875	18.33	1.001	0.078	-5.107	1.032	23.34
(d) $q^*_O, \widehat{\omega}^2_O$	1.013	0.188	-3.339	0.535	10.19	1.006	0.098	-4.639	0.836	16.69	1.005	0.075	-5.179	0.960	21.16
2. Bias-corrected RV of Zhou (1996)															
(a) highest ($q = 1$)	1.014	0.144	-3.870	**0.005**	1.00	1.015	0.064	-5.459	**0.017**	1.00	1.015	0.045	-6.098	**0.041**	1.00
(b) ad-hoc (5 mins)	0.997	0.279	-2.555	1.319	12.00	0.996	0.264	-2.663	2.812	64.00	0.996	0.263	-2.674	3.466	128.00
(c) $q^*_{Zhou}, \widehat{\omega}^2_O$	1.014	0.144	-3.870	**0.005**	1.00	1.015	0.064	-5.459	**0.017**	1.00	1.015	0.045	-6.098	**0.041**	1.00
3. Two-Scale RV of Zhang, Mykland, and Aït-Sahalia (2005)															
(a) ad-hoc ($q = 5$)	0.997	0.152	-3.767	0.107	5.00	1.002	0.068	-5.390	0.086	5.00	1.003	0.048	-6.079	0.061	5.00
(b) $q^*_{ZMA}, \widehat{\omega}^2_{BR}$	1.012	0.144	-3.874	**0.001**	2.00	1.014	0.064	-5.462	**0.013**	2.00	1.015	0.045	-6.101	**0.038**	2.00
(c) $q^*_{ZMA}, \widehat{\omega}^2_O$	1.012	0.144	-3.874	**0.001**	2.00	1.014	0.064	-5.462	**0.013**	2.00	1.015	0.045	-6.101	**0.038**	2.00
(d) $q^*_{BR}, \widehat{\omega}^2_{BR}$	0.986	0.173	-3.506	0.368	10.51	0.994	0.090	-4.807	0.669	19.34	0.996	0.069	-5.334	0.805	24.60
4. Multi-Scale RV of Zhang (2006)															
(a) ad-hoc ($q = 5$)	0.996	0.153	-3.757	0.118	5.00	1.001	0.068	-5.380	0.095	5.00	1.002	0.048	-6.071	0.068	5.00
(b) $q^*_Z, \widehat{\omega}^2_{BR}$	1.012	0.144	-3.873	**0.002**	2.01	1.014	0.064	-5.463	**0.013**	2.04	1.014	0.045	-6.103	**0.037**	2.05
(c) $q^*_Z, \widehat{\omega}^2_O$	1.012	0.144	-3.874	**0.001**	2.00	1.014	0.064	-5.462	**0.013**	2.00	1.015	0.045	-6.101	**0.038**	2.00
5. Realized Kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a)															
(a) TH_2, ad-hoc ($q = 5$)	1.003	0.149	-3.805	0.070	5.00	1.004	0.066	-5.432	0.043	5.00	1.005	0.047	-6.123	**0.017**	5.00
(b) TH_2, $q^*_{BHLS}, \widehat{\omega}^2_{BR}$	1.004	0.147	-3.841	**0.034**	3.96	1.005	0.065	-5.453	**0.023**	4.23	1.006	0.046	**-6.139**	★	4.27
(c) TH_2, $q^*_{BHLS}, \widehat{\omega}^2_O$	1.014	0.144	-3.869	**0.005**	1.25	1.015	0.064	-5.456	**0.019**	1.19	1.015	0.045	-6.096	0.043	1.14
(d) Cubic, $q^*_{BHLS}, \widehat{\omega}^2_O$	1.011	0.144	**-3.875**	★	2.09	1.012	0.064	**-5.475**	★	2.23	1.012	0.045	-6.135	**0.004**	2.26
6. ALT of Large	1.008	0.243	-2.832	1.043	–	0.993	0.103	-4.549	0.926	–	0.992	0.073	-5.223	0.916	–
7. MLE of AMZ	1.013	0.144	-3.870	**0.004**	–	1.014	0.064	-5.466	**0.009**	–	1.014	0.045	-6.114	**0.026**	–

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^*') each realized variance measure and across sample size M. Loss levels in boldface are insignificantly different from zero at a 1% bootstrapped confidence level.

$\widehat{q}^{*}_{BR}$ and the correlation between the two is now only 58.4% (compared to 99.2% for trade data). Given these substantial differences between the two methods of calculating the optimal bandwidth, it is all the more surprising that the performance of the realized kernel is rather insensitive to all this. In fact, for $M = 10,000$ the bandwidth selection based on the heavily biased Bandi-Russell noise variance estimate leads to the lower MSE.

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 "volume-weigh" the bid and ask prices by the volume on the opposite side of the book. More formally, the volume-weighted mid-quote, or micro price, is calculated as:

$$\frac{V_{ask}P_{bid} + V_{bid}P_{ask}}{V_{ask} + V_{bid}}.$$

It is quite intuitive why this process has some appeal. If volume at the bid is much higher than volume at the ask then, with uncorrelated market order arrival, the probability of the market price breaking through the ask is higher than it breaking through the bid. The micro-price reflects this, whereas trade prices or mid-quotes do not, see Panel A of Figure 1 for an illustration. 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^* 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! The loss associated with optimal selection of q is therefore quite substantial, and can only be mitigated for the realized kernel by artificially inflating q using the biased Bandi-Russell noise variance estimate. So on the whole, these finding are indicative that the theory-implied optimal selection of q^* does not work well with micro price data. To some extent, this is quite surprising given that for the micro price estimates of λ and the noise ratio ξ (see Table 1), as well as the autocorrelation pattern (see Figure 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. A possible 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. Nevertheless, 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.

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 ($q = 1$)	1.173	0.112	-3.159	0.733	1.00	1.175	0.050	-3.413	1.997	1.00	1.175	0.035	-3.447	2.600	1.00
(b) ad-hoc (5 mins)	1.015	0.196	-3.255	0.637	12.00	1.000	0.168	-3.569	1.841	64.00	1.000	0.165	-3.607	2.440	128.00
(c) $q^*_{BR}, \widehat{\omega}^2_{BR}$	1.012	0.186	-3.359	0.533	10.28	1.005	0.102	-4.572	0.837	19.00	1.005	0.079	-5.085	0.962	24.18
(d) $q^*_O, \widehat{\omega}^2_O$	1.078	0.145	-3.612	0.280	2.69	1.050	0.072	-4.856	0.553	4.52	1.039	0.053	-5.436	0.611	5.71
2. Bias-corrected RV of Zhou (1996)															
(a) highest ($q = 1$)	1.039	0.138	-3.882	**0.010**	1.00	1.040	0.061	-5.232	0.178	1.00	1.040	0.043	-5.660	0.387	1.00
(b) ad-hoc (5 mins)	0.997	0.279	-2.556	1.337	12.00	0.996	0.264	-2.663	2.746	64.00	0.996	0.263	-2.674	3.373	128.00
(c) $q^*_{Zhou}, \widehat{\omega}^2_O$	1.039	0.138	-3.882	**0.010**	1.00	1.040	0.061	-5.232	0.178	1.00	1.040	0.043	-5.660	0.387	1.00
3. Two-Scale RV of Zhang, Mykland, and Aït-Sahalia (2005)															
(a) ad-hoc ($q = 5$)	1.006	0.150	-3.789	0.103	5.00	1.011	0.067	-5.384	**0.026**	5.00	1.012	0.047	-6.040	**0.007**	5.00
(b) $q^*_{ZMA}, \widehat{\omega}^2_{BR}$	1.037	0.138	**-3.892**	★	2.00	1.040	0.061	-5.239	0.171	2.00	1.040	0.043	-5.665	0.382	2.00
(c) $q^*_{ZMA}, \widehat{\omega}^2_O$	1.037	0.138	**-3.892**	★	2.00	1.040	0.061	-5.239	0.171	2.00	1.040	0.043	-5.665	0.382	2.00
(d) $q^*_{BR}, \widehat{\omega}^2_{BR}$	0.990	0.173	-3.511	0.381	10.57	0.996	0.090	-4.812	0.598	19.34	0.998	0.069	-5.338	0.709	24.61
4. Multi-Scale RV of Zhang (2006)															
(a) ad-hoc ($q = 5$)	1.004	0.151	-3.776	0.116	5.00	1.009	0.067	-5.378	**0.032**	5.00	1.010	0.048	-6.041	**0.006**	5.00
(b) $q^*_Z, \widehat{\omega}^2_{BR}$	1.036	0.139	-3.886	**0.006**	2.02	1.038	0.062	-5.245	0.165	2.09	1.038	0.044	-5.688	0.359	2.11
(c) $q^*_Z, \widehat{\omega}^2_O$	1.037	0.138	-3.892	**0.000**	2.00	1.040	0.061	-5.239	0.170	2.00	1.040	0.043	-5.665	0.382	2.00
5. Realized Kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a)															
(a) TH_2, ad-hoc ($q = 5$)	1.015	0.147	-3.829	0.063	5.00	1.015	0.065	**-5.410**	★	5.00	1.015	0.046	**-6.047**	★	5.00
(b) TH_2, $q^*_{BHLS}, \widehat{\omega}^2_{BR}$	1.017	0.144	-3.858	**0.035**	4.06	1.018	0.065	-5.405	**0.005**	4.36	1.018	0.046	-6.020	**0.027**	4.41
(c) TH_2, $q^*_{BHLS}, \widehat{\omega}^2_O$	1.037	0.139	-3.873	**0.019**	1.40	1.038	0.062	-5.236	0.174	1.46	1.038	0.044	-5.683	0.364	1.48
(d) Cubic, $q^*_{BHLS}, \widehat{\omega}^2_O$	1.032	0.140	-3.885	**0.007**	2.17	1.032	0.062	-5.318	0.092	2.37	1.032	0.044	-5.819	0.228	2.41
6. ALT of Large	0.874	0.105	-3.613	0.279	–	0.873	0.047	-4.007	1.403	–	0.874	0.033	-4.070	1.977	–
7. MLE of AMZ	1.035	0.140	-3.876	**0.016**	–	1.036	0.062	-5.270	0.140	–	1.037	0.044	-5.733	0.314	–

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^*') each realized variance measure and across sample size M. Loss levels in boldface are insignificantly different from zero at a 1% bootstrapped confidence level.

4.4 Overall comparison

In a comparison of twenty different realized variance measures, 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 5. Here, a blue star (★) indicates "superior" performance where the RV measure is either the best performing or statistically indistinguishable from the best, a black 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 red 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. MLRV performs very well all-round but, although not investigated here, is misspecified with time varying volatility or seasonal patterns.[6] Moreover, from a practical implementation perspective, MLRV is too computationally intensive for real-time volatility estimation. So this leaves us with the realized kernels and subsampling methods, which are known to share some equivalence up to end effects. Not surprisingly therefore, performance is comparable and is really only distinguished by the selection of tuning parameters. For TSRV, the optimal subsampling frequency of Zhang, Mykland, and Aït-Sahalia (2005) works well except for small trade series and large micro series. The MSRV works well with the optimal number of subsamples of Zhang (2006), except for large micro series. Finally, KRV works well for all sample sizes and data types when the optimal bandwidth is calculated based on Bandi-Russell's noise variance estimate. But from a robustness viewpoint, the *ad hoc* bandwidth selection for either TSRV, MSRV or KRV, always gives an all-round good performance. The loss in MSE relative to the best performing measure, averaged across data types and sample sizes, is less than 10% whereas the average improvement over the best performing RV measure is more than $1 - e^{-1} \approx 60\%$!

5 Conclusion

This paper compares a comprehensive set of "second generation" realized variance measures using a simulated "zero-intelligence" order book market of Smith, Farmer, Gillemot, and Krishnamurthy (2003). 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, our advice to the pragmatic practitioner is to sample prices at the highest available frequency and then measure realized variance using either the Two-Scale RV of Zhang, Mykland, and Aït-Sahalia (2005) with a subsampling frequency of 5, the Multi-Scale RV of Zhang (2006) with 5 subsamples, or the Realized Kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a) with a bandwidth of 5. The performance of these estimators is

[6] In a related paper Hansen, Large, and Lunde (2006) find that moving average based estimators do have some robustness to time varying volatility and that they can account for dependent noise by including higher order MA terms in the estimation.

Table 5: 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 ($q = 1$)	†	†	†	†	†	†	†	†	†	0.394	3.797	9.362
(b) ad-hoc (5 mins)	†	†	†	†	†	†	†	†	†	0.626	1.843	2.530
(c) q^*_{BR}, $\widehat{\omega}^2_{BR}$	†	†	†	†	†	†	†	†	†	0.487	1.154	2.115
(d) q^*_{O}, $\widehat{\omega}^2_{O}$	†	†	†	†	†	†	†	†	†	0.280	1.148	2.254
2. Bias-corrected RV of Zhou (1996)												
(a) highest ($q = 1$)	†	†	†	★	★	★	★	✓	†	0.005	0.537	1.422
(b) ad-hoc (5 mins)	†	†	†	†	†	†	†	†	†	1.022	2.407	3.466
(c) q^*_{Zhou}, $\widehat{\omega}^2_{O}$	†	†	†	★	★	★	★	✓	†	0.005	0.240	0.700
3. Two-Scale RV of Zhang, Mykland, and Aït-Sahalia (2005)												
(a) ad-hoc ($q = 5$)	✓	✓	✓	✓	✓	✓	✓	★	★	0.007	0.094	0.153
(b) q^*_{ZMA}, $\widehat{\omega}^2_{BR}$	†	✓	✓	★	★	★	★	✓	†	0.000	0.129	0.382
(c) q^*_{ZMA}, $\widehat{\omega}^2_{O}$	†	✓	✓	★	★	★	★	✓	†	0.000	0.151	0.388
(d) q^*_{BR}, $\widehat{\omega}^2_{BR}$	✓	†	†	†	†	†	†	†	†	0.157	0.505	0.805
4. Multi-Scale RV of Zhang (2006)												
(a) ad-hoc ($q = 5$)	✓	✓	✓	✓	✓	✓	✓	★	★	0.006	0.094	0.146
(b) q^*_{Z}, $\widehat{\omega}^2_{BR}$	✓	✓	✓	★	★	★	★	✓	†	0.002	0.110	0.359
(c) q^*_{Z}, $\widehat{\omega}^2_{O}$	†	✓	✓	★	★	★	★	✓	†	0.000	0.123	0.382
5. Realized Kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006a)												
(a) TH_2, ad-hoc ($q = 5$)	✓	✓	✓	✓	✓	★	✓	★	★	0.000	0.089	0.211
(b) TH_2, q^*_{BHLS}, $\widehat{\omega}^2_{BR}$	✓	✓	★	★	★	★	★	★	★	0.000	0.031	0.068
(c) TH_2, q^*_{BHLS}, $\widehat{\omega}^2_{O}$	✓	✓	★	★	★	✓	★	✓	†	0.005	0.086	0.364
(d) Cubic, q^*_{BHLS}, $\widehat{\omega}^2_{O}$	✓	✓	✓	★	★	★	★	✓	✓	0.000	0.075	0.228
6. ALT of Large	★	†	†	†	†	†	†	†	†	0.000	1.004	1.977
7. MLE of AMZ	★	★	★	★	★	★	★	✓	†	0.000	0.059	0.314

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

largely equivalent and their implementation equally straightforward, so which particular one to use would be a matter of taste. Still, relative to the widely used sum of sparsely sampled returns following the "5-minute" rule prescribed in earlier literature, the efficiency gain achieved with either of them is likely to be substantial.

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 Eq. (17), 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 would 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 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 Bandi, Russell, and Yang (2006) who focus on bandwidth selection in the context of option pricing and Aït-Sahalia and Mancini (2006); Andersen, Bollerslev, and Meddahi (2006); Ghysels and Sinko (2006) who focus on volatility forecasting in the presence of noise.

References

Aït-Sahalia, Y., and L. Mancini, 2006, "Out of Sample Forecasts of Quadratic Variation," manuscript Princeton University.

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

Andersen, T. G., and T. Bollerslev, 1998, "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39 (4), 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, "The Distribution of Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., T. Bollerslev, and N. Meddahi, 2006, "Realized Volatility Forecasting and Market Microstructure Noise," manuscript Tanaka Business School, Imperial College London.

Bandi, F. M., and J. R. Russell, 2005, "Market Microstructure Noise, Integrated Variance Estimators, and the Accuracy of Asymptotic Approximations," manuscript University of Chicago.

——— , 2006a, "Comment on JBES 2005 invited address by Peter R. Hansen and Asger Lunde," *Journal of Business & Economic Statistics*, 24(2), 167–173.

——— , 2006b, "Separating Microstructure Noise from Volatility," *Journal of Financial Economics*, 79, 655–692.

Bandi, F. M., J. R. Russell, and C. Yang, 2006, "Realized Volatility Forecasting and Option Pricing," manuscript GSB, The University of Chicago.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 2006a, "Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise," manuscript University of Oxford, Nuffield College.

——— , 2006b, "Subsampling Realised Kernels," manuscript University of Oxford, Nuffield College.

Bollerslev, T., I. Domowitz, and J. Wang, 1997, "Order Flow and the Bid-Ask Spread: An Empirical Probability Model of Screen-Based Trading," *Journal of Economic Dynamics and Control*, 21, 1471–1491.

Bouchaud, J.-P., Y. Gefen, M. Potters, and M. Wyart, 2004, "Fluctuations and response in financial markets: the subtle nature of 'random' price changes," *Quantitative Finance*, 4, 176–190.

Domowitz, I., and J. Wang, 1994, "Auctions as algorithms : Computerized trade execution and price discovery," *Journal of Economic Dynamics and Control*, 18(1), 29–60.

Farmer, J. D., P. Patelli, and I. I. Zovko, 2005, "The predictive power of zero intelligence in financial markets," *Proceedings of the National Academy of Sciences of the United States of America*, 102(6), 2254–2259.

Ghysels, E., and A. Sinko, 2006, "Volatility Forecasting and Microstructure Noise," manuscript University of North Carolina.

Hansen, P. R., J. Large, and A. Lunde, 2006, "Moving Average-Based Estimators of Integrated Variance," forthcoming *Econometric Reviews*.

Large, J., 2005, "Estimating Quadratic Variation When Quoted Prices Jump by a Constant Increment," manuscript University of Oxford, All Souls College.

Niederhoffer, V., and M. F. M. Osborne, 1966, "Market Making and Reversal on the Stock Exchange," *Journal of the American Statistical Association*, 61 (316), 897–916.

Oomen, R. C., 2006, "Properties of Realized Variance under Alternative Sampling Schemes," *Journal of Business & Economic Statistics*, 24 (2), 219 – 237.

Smith, E., J. D. Farmer, L. Gillemot, and S. Krishnamurthy, 2003, "Statistical Theory of the Continuous Double Auction," *Quantitative Finance*, 3, 481–514.

Zhang, L., 2006, "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12 (6), 1019–1043.

Zhang, L., P. A. Mykland, 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, B., 1996, "High Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business & Economic Statistics*, 14(1), 45–52.