# Estimating Integrated Variation With Rounding Error

Guansong Wang

This version: Dec 2013

**Abstract**

This paper studies the effect of the rounding error on the IV estimator. The rounding error cause the log-returns to have higher order time dependence which is the source of the bias in the RK estimator. We modify the estimatorto correct the bias and derive the optimal weights under the MA($q$) noise assumption. The simulation indicates the bias is severe and the correction is effective.

## 1 Introduction

The integrated variation (IV), or equally the quadratic variation, of a log price process measures the risk of the asset and is crucial in pricing theories. With access of the (ultra) high frequency intra-daily data of the process, it can be optimally estimated by the realized variation in the period of interest. However the real observations are contaminated by the micro-structure noises which introduces a bias term dominating the variation of the log-return in a small time period between samples. Instead of sampling sparsely to weaken the effect of the noise, several consistent estimators have been developed in recent years. Some achieve the optimal convergence rate ($N^{1/4}$) as is proved in Gloter and Jacod (2001a,b).

The observed prices are rounded to nearest multiples of the tick size of the market. The rounding error is the difference between the rounded price and the original price. The IV literature does not pay much attention to the rounding error and treats it as a component of the micro-structure noise which is assumed to be a martingale difference sequence. A few authors have found the rounding error bring extra difficulty in the IV estimation. Delattre and Jacod (1997) showed that the realized variation does not converge to the real IV with round-off errors unless the rounding step goes to zero fast enough. Li and Mykland (2007) discovered the Two-Scale RV developped by Zhang, Mykland, and Ait-Sahalia (2005) has a limit other than the desired IV when the other components of the micro-structure noise is too small, that is the rounding error dominates the micro-structure noise. Jacod, Li, Mykland, Podolskij, and Vetter (2009) noticed that the "pure rounding" process violates the assumption so that IV cannot be inferred. Li and Mykland (2012) proposed a bias-corrected RV to be a consistent IV estimator with a rounded process.

Failure of the supposedly consistent IV estimators implis that the rounding error has properties different from the martingale difference sequence assumption. In this paper we find the rounding error is an important even a major component of the overall micro-structure noise and has high order autocorrelations in high frequency samples but no correlation with the latent log-returns. Other authors also confirm the time dependence at higher orders in the observed log-returns. Hansen and Lunde (2006) gave warning about the time-dependent noise along with other ugly facts in high frequency samples. Recently Aït-Sahalia, Mykland, and Zhang (2011) also found the log-return process is better fitted by an ARMA process.

Based on the facts we generalize the Realized Kernel estimator designed by Jacod, Li, Mykland, Podolskij, and Vetter (2009) to the model with an MA($q$) noise process. When $q$ is greater than one the RK estimator is no longer unbiased. We modify the RK estimator to correct the bias and maintain the asymptotics. We also derive the optimal weights of the RK estimator using control variable interpretation and propose a feasible approximation. A simulation validates the performance gain of the bias correction and suggests that the rounding effect can last a longer period than the maximum lag of significant autocorrelations.

The structure of the rest of the paper is as follows. Section 2 reports the facts of the rounding error based on simulation and empirical data. In Section 3 and Section 4 we generalize the model by the MA($q$) noise process and study the bias correction and optimal weights of the RK estimator. The results of A Monte Carlo simulation are shown in Section 5 At last Section 6 concludes.

# 2 Facts of Rounding Errors

Rounding is an inevitable feature of economic and financial data and usually treated as trivial. Most high frequency data literature simply ignore the rounding error or consider it as one component of the micro-structure noise that satisfies the assumptions of the noise terms. Few raise special attention to how the rounding error affects the estimators using high frequency samples. The supposedly consistent estimators of the integrated variation are not robust with the presence of rounding, such as the Two-Scale RV discussed in Li and Mykland (2007). The simulation in Section 5 also demonstrates how the estimators severely suffer from the rounding error. Therefore the rounding error must have some unusual properties than the additive noises widely assumed as a martingale difference sequence in the literature.

In this section we find a few facts on the rounding error worth noticing from a simple model setup, simulation and empirical data. The facts are summarized below:

1. Unconditionally the rounding error is uniformly distributed.

2. Stocks with different price levels, volatilities and trading/quoting frequencies suffer from different level of round-off effect.

3. The rounding error is a significant component of the micro-structure noises and became a major component.

4. The rounding error causes the high frequency log-returns to have higher order time dependence. The order of time dependence rises as the sampling frequency becomes higher.

5. The rounding error has no or weak correlation with latent log-returns.

## 2.1 The Model

Following the high frequency IV estimator literature, the underlying log-price process $\{\log S_t = X_t\}_{t\geq 0}$ of an asset is a semi-martingale process in a probability space $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$:

$$d(\log S_t) = dX_t = \mu_t dt + \sigma_t dW_t \tag{1}$$

where the stochastic processes $\mu_t$ and $\sigma_t$ are $\mathcal{F}_t$ adapted and càdlàg. In high frequency samples the drift term becomes statistically irrelevant because $\mu_t dt$ is of lower order than the diffusion component $\sigma_t dW_t$. Formally $\mu_t$ is set to be zero throughout this paper. Our interest is to estimate the integrated variation of the process within a fixed time period such as one day $(T = 1)$:

$$IV_T = \int_0^T \sigma_t^2 dt \tag{2}$$

The prices are observed at a time grid $\{0 = t_0 < t_1 < \cdots < t_N = T\}$ when a transcation of a quote update occurs. It is well known that the realized variation (RV) of the process $X$ converges to $IV_T$ as the mesh of the partition $(\sup_i(t_i - t_{i-1}))$ diminishes to zero.

$$RV_T^N = \sum_{t_i \leq T} (X_{t_i} - X_{t_{i-1}})^2 \to_p IV_T \tag{3}$$

The convergence rate is $\sqrt{N}$ and the asymptotic distribution has been derived in Barndorff-Nielsen and Shephard (2002).

Due to the trading scheme of the market, the observed prices are embedded with micro-structure noises. Madhavan (2000) provided a comprehensive survey of how price formation and information distort the observed log-prices. An additive noise term $\varepsilon_t$ is added in the model by most literature to capture various sources of micro-structure noise. It is assumed to be a martingale difference sequence for the convenience in designing consistent estimators.

**Assumption 1.** *The noise process $\{\varepsilon_t\}$ satisfies:*

1. *Finite Moments:* $E[\varepsilon_t] = 0, V[\varepsilon_t] = E[\varepsilon_t^2] = \sigma_\varepsilon^2, E[\varepsilon_t^4] < \infty$.
2. *Non-autocorrelated:* $E[\varepsilon_t \varepsilon_s] = 0, \forall t \neq s$.
3. *Non-correlated:* $\{X_t\}$ *and* $\{\varepsilon_t\}$ *are uncorrelated.*

Formally, the observable process is $\{Y_t\}$.

$$Y_t = X_t + \varepsilon_t \tag{4}$$

## 2.2 Rounding Error

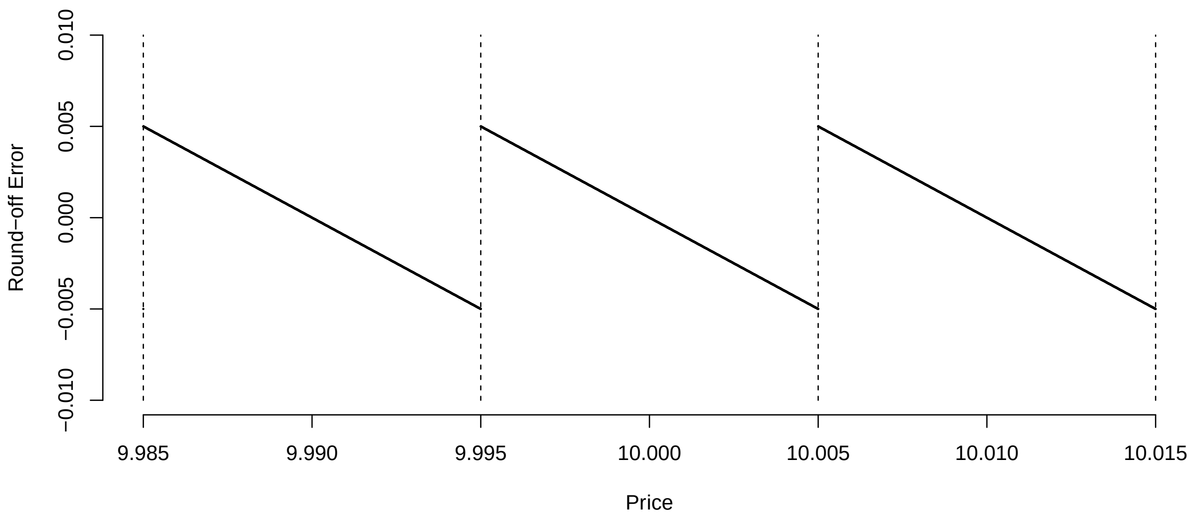
In the financial markets the quotes are usually allowed to change only by multiples of a tick $\alpha$ which may distort the quotes from their latent values to the nearest rounded discrete levels. Formally the rounding function with tick size $\alpha$ is defined as:

$$\tilde{S} = f_\alpha(S) = \begin{cases} \lfloor S/\alpha \rfloor \cdot \alpha & \text{if } S - \lfloor S/\alpha \rfloor < \alpha/2 \\ \lceil S/\alpha \rceil \cdot \alpha & \text{if } S - \lfloor S/\alpha \rfloor \geq \alpha/2 \end{cases} \tag{5}$$

where $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are the floor and ceiling functions respectively. To isolate the rounding effect, we temporarily assume that rounding is the only source of micro-structure noises. Denote $S_t$ as the latent price and $\tilde{S}_t$ as the rounded observation at time $t$. The rounding error $e_t$ is the difference between the observed price and the latent price, $e_t = \tilde{S}_t - S_t$. Since the rounding function is well-defined as in Equation (5), the rounding error $e_t$ is a deterministic function of the price $S_t$ which has a period of $\alpha$ and a range of $(-\alpha/2, \alpha/2)$. Figure 1 shows a plot of the rounding error for the prices around 10 when the tick size $\alpha$ is one cent.

The distribution of the rounding error is purely determined by the distribution of the latent raw price. The closed form distribution is difficult to derive. With the price generated

Figure 1: Rounding Error As A Function Of Raw Price




The rounding tick size ($\alpha$) is set to be 1/100 (one cent) as in present stock markets. The rounding error is a deterministic and periodic function of the raw stock price, which is ranged in $(-\alpha/2, \alpha/2)$.

by the model as in Equation (1), Gottlieb and Kalay (1985) proved that the rounding error has a limiting uniform distribution on $(-\alpha/2, \alpha/2)$ as $t \to \infty$. Intuitively the probability density of the rounding error should be evenly distributed on its range $(-\alpha/2, \alpha/2)$ as long as the stock price has large enough variation on an interval much greater than $\alpha$. In the rest of this section, we will assume that unconditionally the rounding error has a uniform distribution within its range.

The rounding occurs at the raw price level instead of the log-prices. Even though the addtive noise term $\varepsilon_t$ is capable to include the rounding error, [1] it does not fit in the model naturally. An approximation can transform $e_t$ to the additive term as in Equation (4).

$$\begin{aligned}\log \tilde{S}_t &= \log\left(S_t + e_t\right)\\ &= \log S_t + \log(1 + \frac{e_t}{S_t})\\ &\approx \log S_t + \frac{e_t}{S_t}\end{aligned} \tag{6}$$

The price $S_t$ is usually much greater than the rounding error, hence when the rounding error is the only source of the micro-structure noises the additive noise term $\varepsilon_t$ can be well

[1] See the Model 2 in Jacod, Li, Mykland, Podolskij, and Vetter (2009) as an example.

approximated by $e_t/S_t$. Consequently the higher the price level the smaller is the magnitude of the additive noise term brought by the rounding error for the rounding errors are bounded in $(-\alpha/2, \alpha/2)$ no matter what the raw prices are. Furthermore, the second moment of the additive noise term can be derived assuming that $e_t$ has a uniform distribution and the prices do not vary too much within one day ($T = 1$):

$$E[\varepsilon_t^2] \approx E[\frac{e_t^2}{\bar{S}^2}] = \frac{1}{\bar{S}^2} \int_{-\alpha/2}^{\alpha/2} e^2 \frac{1}{\alpha} de = \frac{1}{12\bar{S}^2}\alpha^2 \tag{7}$$

where $\bar{S}$ represents the price level of that trading day.

### 2.3 Empirical Study

For historical reasons the NYSE used one-eighth of a dollar as the tick size for a long period of time since its founding in 1792. It was until the summer of 1997 that the tick size was splitted into one-sixteenth of a dollar. Later in April 2001 both NYSE and Nasdaq completed decimalization, that is to set the tick size to one cent. Therefore there are three different tick sizes in the history of the U.S. stock markets, $\alpha = 1/8$, $1/16$, and $1/100$. By Equation (7), the price adjusted magnitude of the noise brought by the rounding errors with the three tick sizes are:

$$E[\varepsilon_t^2]\bar{S}^2 = \begin{cases} 1/12/8^2 \approx 1.3020 \times 10^{-3} & \text{for } \alpha = 1/8 \\ 1/12/16^2 \approx 3.2552 \times 10^{-4} & \text{for } \alpha = 1/16 \\ 1/12/100^2 \approx 8.3333 \times 10^{-6} & \text{for } \alpha = 1/100 \end{cases} \tag{8}$$

Decimalization has decreased the magnitude of the rounding error dramatically which also coincides with rising of liquitity and falling of transaction costs. It is impossible to seperate those effects from each other by analyzing the observed price processes. We compare the magnitude of the overall micro-structure noises and the magnitude of the rounding effect given in Equation (8) to measure the importance of the rounding error.

We use the intra-daily trading and quoting data of component stocks of the Dow-Jones 30 index from 1993 to 2012 excluding holidays and half trading days. [2] The data are from the TAQ database and cleaned by a procedure described in Appendix A which is based on the one suggested by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). The median price of a trading day is used as a proxy of the price level $\bar{S}$. For the estimator of the

[2] The components of the Dow-Jones 30 index has changed during that period. We choose the 30 component stocks as in Dec 2012. Also some stocks are not listed in the early years.

magnitude of the overall noises, we choose:

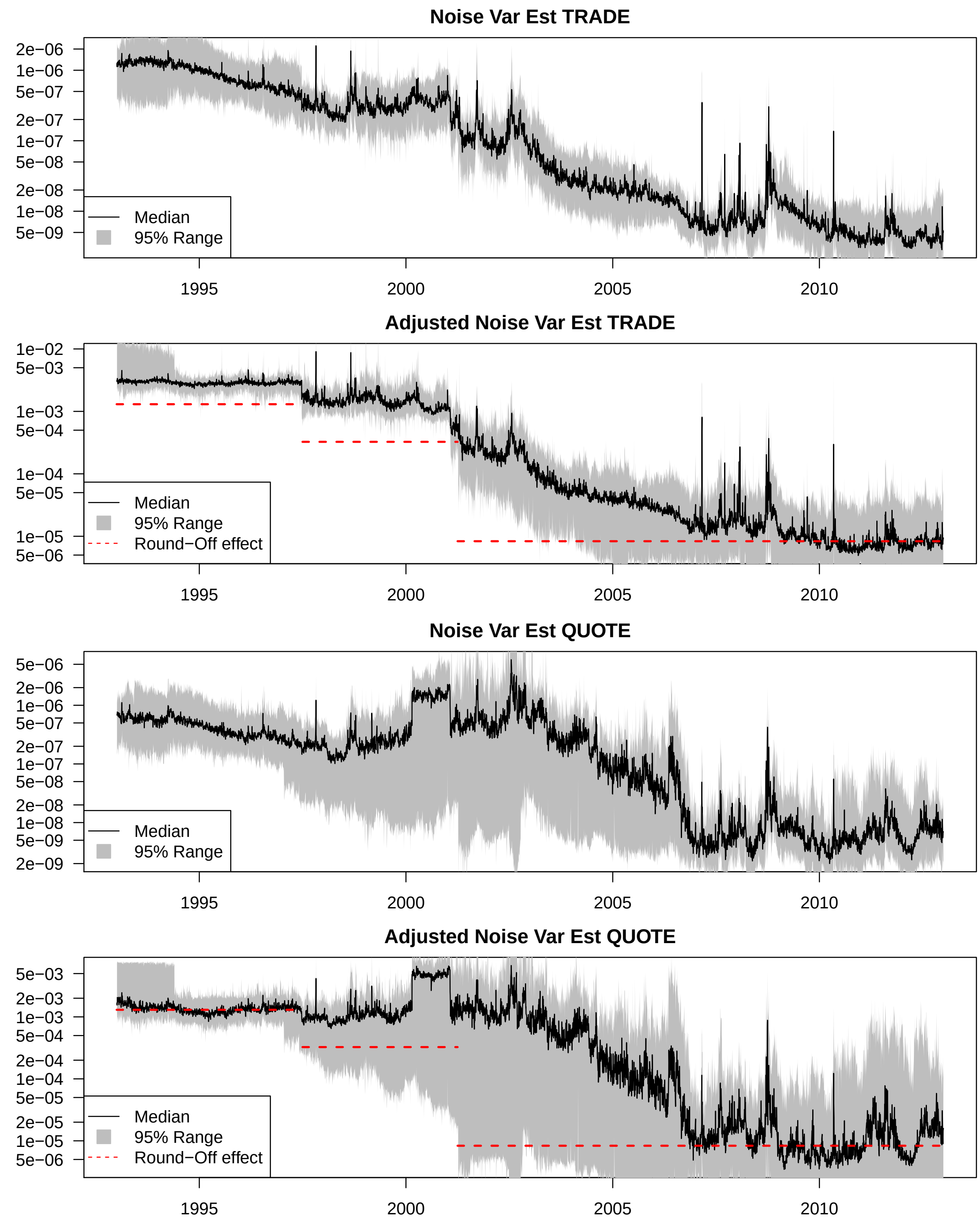
$$\widehat{E[\varepsilon_t^2]} = \frac{\sum_{i=1}^{N} \Delta Y_i^2}{2N} \tag{9}$$

where $\Delta Y_i$ are the log-returns of the process and $N$ is the sample size. It is a consistent but up-biased estimator of the variance of the additive noise. [3] There are also other estimators available yet we choose this for the following reasons. Firstly, it is easy to implement. It only requires the realized variation and sample size. Secondly, it ensures positivity which is beneficial when used as a preliminary estimate to find the optimal bandwidths of consistent IV estimators. Thirdly, the volume of intra-daily transcations and quotes are large enough to ensure accuracy. Lastly, we have used another estimator designed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) which is unbiased and ensure positivity. Both estimators produce similar time series plots. Here we only show the plots using the estimator in Equation (9).

Figure 2 shows the time series of the median and 95% interval of the estimated magnitude of the micro-structure noises. The top two plots are estimates using transaction price processes and the bottom two plots are estimates using mid-quote price processes. All the plots apply logarithm on the Y-axis. The first and third plots are the estimated variance of the additive noise term as defined in Equation (9). In both plots the variance keeps declining except during shock periods such as around 2001 and 2008, which indicates that the micro-structure noises have became smaller over years. The second and fourth plots are the price adjusted estimated variance $\widehat{E[\varepsilon_t^2]}\bar{S}^2$. For comparison, the levels of rounding effect derived in Equation (8) are marked by broken lines. Instead of the downwards trend there are two obvious sudden drops, in July 1997 and April 2001, which are just the periods of changing the tick size $\alpha$ and divide the process into three periods. In the first two periods, the downward trend and the slight increase around 2001 disappear after the price adjustment. It indicates that the fluctuation of the variance of noise is caused by the fluctuation of the price level. The levels of price adjusted variance of the overall noises in the two periods are above and of the same magnitude with the ones derived from pure rounding errors. For trading data, the rounding effect takes 44.8% and 22.6% of the overall variance of the noise in average for the first and second periods respectively. For quoting data, it takes 95.2% and 16.2% in average respectively. In the beginning of the third period, the downward trend remains after the price adjustment until around 2010. The price adjusted variance remains at a relatively stable level afterwards, of which the rounding effect contributes 99.2% and 67.9% for trading and quoting data respectively. Before the decimalization the rounding error is a significant

[3] See Zhang, Mykland, and Ait-Sahalia (2005).

Figure 2: Variance of Noise, DJ30 stocks, 1993-2012




The top two panels are for transaction prices and the bottom two panels are for mid-quote prices. The first and third panels are the time series of the estimated variance of noise defined in Equation (9) and the second and fourth panels adjust the estimators by multiplying the squared median prices. The red broken lines are the predicted levels of variance of the noise brought by the rounding errors as given in Equation (8). All plots use logarithmic scale on Y-axis.

component of the overall noise. After the decimalization the overall noise variance does not decline to the rounding error level until after 2010. In present the rounding error has became a major component in the overall noise.

## 2.4 Simulation Study

To further examine the properties of rounding error we simulate price processes based on Equation (1) and round the raw prices. The parameters are calibrated to transaction data of the "AA" [4] stock in 2012. The integrated variation is set to be constant during one day $\sigma^2 = 2.2129 \times 10^{-4}$. The sampling frequency is two ticks per second and then the sample size of log-returns for one day is $46,800$. The initial price level is 9.00. The prices $S_t = e^{X_t}$ are rounded to the nearest cent ($\alpha = 0.01$). The simulation is repeated for 100,000 times and the results are shown in Figure 3.

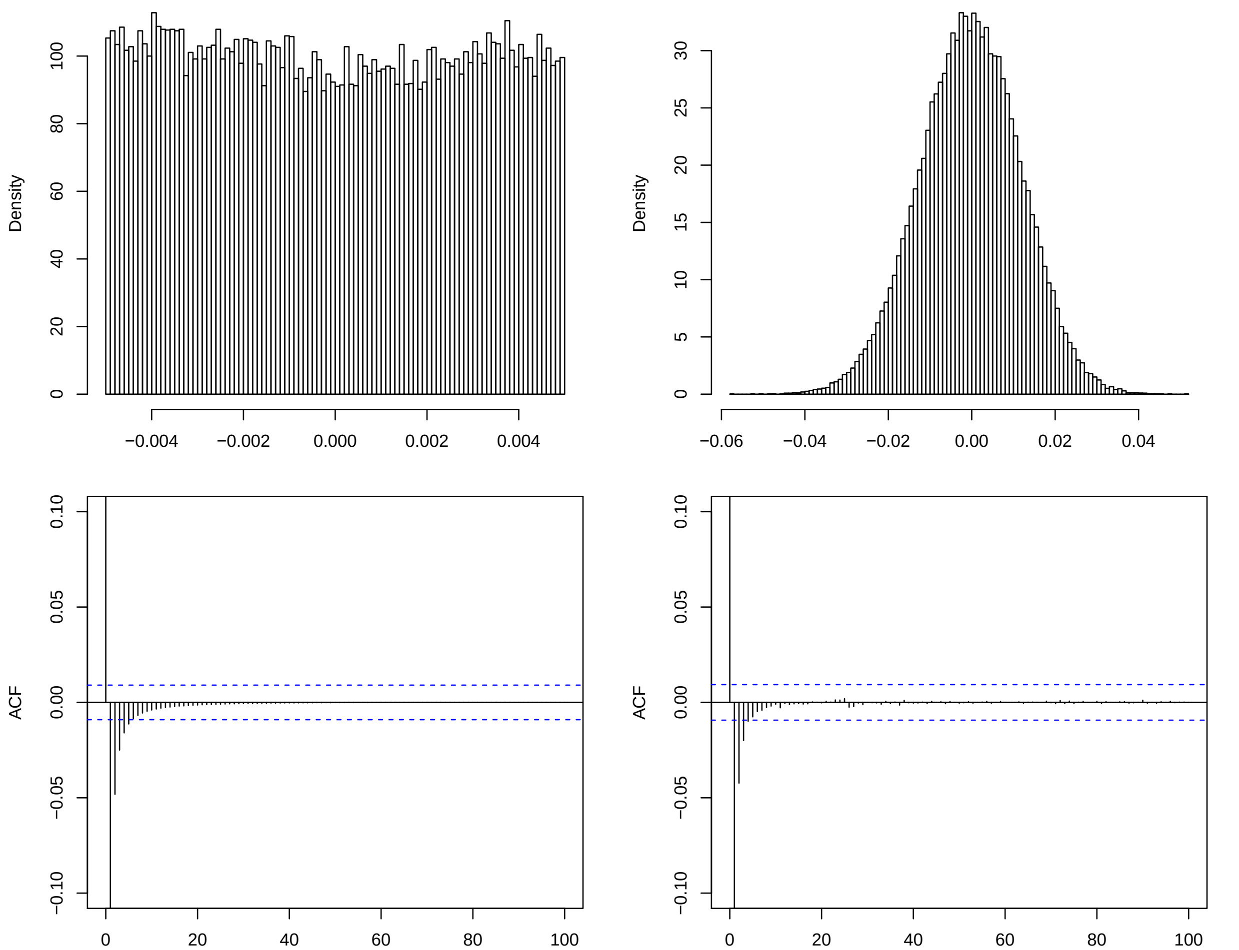
The top-left plot is the histogram of the rounding errors of one simulated price path. It resembles a uniform distribution on the interval $(-0.005, 0.005)$. The top-right plot is the histogram of the correlation between latent log-returns $\Delta X_t$ and the first differences of the additive noises $\Delta\varepsilon_t$. The distribution of the correlation is concentrated around zero with a small deviation. The term "correlation" may be misleading in this case because the rounding errors are not random when a price path is given. Another way to interpret the histogram is that the expectation $E[\Delta X_t \Delta\varepsilon_t]$ is zero when $\varepsilon_t$ is derived from rounding errors. The bottom-left plot is the autocorrelation function of the observed (rounded) log-returns averaged across simulations. The bottom-right plot is the autocorrelation function of the observed "AA" transaction log-returns averaged across days in 2012. The simulation closely replicates the time dependence structure of the empirically observed log-returns of "AA".

The higher order autocorrelations come from the additive noise term since the latent log-returns $\Delta X_t$ have no autocorrelation. Under the second assumption in Assumption 1 the observed log-returns should have only one non-zero autocorrelation at the first lag and it is the rounding error that causes autocorrelations at higher lags. When the variation between the latent log-prices is of the same magnitude as the tick size $\alpha$, which may be due to high frequency sampling or relatively small volatility, the corresponding rounding errors are likely to be close hence correlated. Consequently the order of non-zero autocorrelations increases as the sampling frequency becomes higher because the variation of the log-returns becomes smaller.

Note that the simulation is performed with one set of carlibration. Different sets of parameters can yield different shapes in the plots. A higher price or greater volatility of the

[4] Alcoa Inc. (AA) -NYSE

Figure 3: Results Of The Simulation




The data generating process is calibrated to the transaction data of "AA" in 2012. The top-left panel is the histogram of the rounding errors of one simulated price path. The top-right panel is the histogram of the correlation between the latent returns $\Delta X_t$ and the first differences of the additive noise $\Delta \varepsilon_t$. The bottom-left panel is the averaged ACF of the log-returns of the simulated rounded processes. The bottom-right panel is the averaged ACF of the log-returns of the transaction data of "AA" in 2012.

latent price process will weaken the extent of autocorrelations of higher lags.

# 3 Realized Kernel With MA Noises

The consistent IV estimators in the literature are designed under Assumption 1 and have the potential to be robust with violations against the assumptions. However the rounding error causes those estimator to fail to converge to the true value. [5] In this section we focus the effect of rounding error on the Realized Kernel (RK) estimator, find that the time dependence in the rounding error cause the RK estimator to be biased and then suggest a bias correction. We choose the the RK estimator because its design can be naturally fitted in the new assumption. Also it can have the same asymptotics as other estimators such as Two-Scale, Multi-Scale and Pre-Averaging with properly chosen kernel functions.

## 3.1 Realized Kernel Estimator

First we briefly introduce the Realized Kernel estimator of the IV by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). The basic idea is to utilize the first realized autocovariation to eliminate the bias from the noise in the realized variation and higher order realized autocovariations to control the variance of the estimator. The realized autocovariations are weighted by a kernel function. The estimators with different kernel functions have different asymptotic properties and can relate to other consistent estimators.

Formally the RK estimator is defined as:

$$RK_k(H) = \gamma_0(Y) + \sum_{h=1}^{H} k(\frac{h-1}{H})\,(\gamma_h(Y) + \gamma_{-h}(Y)) \tag{10}$$

where $k(\cdot)$ is a kernel function that equates 1 at point zero and equates 0 at point one. The realized autocovariation $\gamma_h(Y)$ is defined as:

$$\gamma_h(Y) = \sum_{i=1}^{N} \Delta Y_i \Delta Y_{i+h}, \text{ for } h = -H, \ldots, -1, 0, 1, \ldots, H \tag{11}$$

In practice where out of sample observations are unavailable, the summations in the $\gamma_h$ are taken from $i = H + 1$ to $i = N - H$ and then the estimator is scaled up accordingly.

the bandwidth $H$ has to be proportional to $N^\eta$ for the consistency where $\eta$ depends on the kernel function. The optimal bandwidth $H^*$ can be obtained by minimizing the asymptotic

[5] See Li and Mykland (2007) and Section 5.

variance, which is determined by the characteristics of the particular kernel function, the chosen bandwidth, and the parameters of the data generating process namely the noise-to-signal ratio and the measure of heteroskedasticity.

Table 1: Realized Kernel Functions

| | $k(x)$ | Rate | $H^*$ |
|---|---|---|---|
| Bartlett | $1-x$ | 1/6 | $2.28 \cdot \xi N^{2/3}$ |
| Cubic | $1-3x^2+2x^3$ | 1/4 | $3.68 \cdot \xi N^{1/2}$ |
| Tukey-Hanning$_p$ | $\sin^2(\pi/2 \cdot (1-x)^p)$ | 1/4 | $\begin{cases} 5.74 \cdot \xi N^{1/2} & p=2 \\ 39.16 \cdot \xi N^{1/2} & p=16 \end{cases}$ |
| Parzen | $\begin{cases} 1-6x^2+6x^3 & 0 \leq x \leq 1/2 \\ 2(1-x)^3 & 1/2 \leq x \leq 1 \end{cases}$ | 1/4 | $4.77 \cdot \xi N^{1/2}$ |
| Parzen (Slow) | ... | 1/5 | $3.51 \cdot \xi^{4/5} N^{3/5}$ |

The convergence rate is the $\alpha$ such that $N^\alpha(RK_k(H^*) - IV_T)$ has an asymptotic distribution. The scalars in $H^*$ are obtained by minimizing the asymptotic variance and are determined by the kernel function. The optimal bandwidth is also positively related to the noise-to-signal ratio $\xi^2$.

A few interesting kernel functions are list in Table 1. The convergence rate is the number $\alpha$ such that $N^\alpha(RK_k(H) - IV_T)$ has an asymptotic distribution when the bandwidth is taken optimally. The RK estimator with Bartlett kernel function has the slowest convergence rate and it is asymptotically equivalent to the Two Scale RV estimator of Zhang, Mykland, and Ait-Sahalia (2005). Gloter and Jacod (2001a,b) have shown that the fastest convergence rate is 1/4. It can be achieved by the kernel functions satisfying that $k'(0) = k'(1) = 0$ such as the Cubic function and the Tukey-Hanning function. The Cubic RK estimator is asymptotically equivalent to the Multi-Scale estimator of Zhang (2006). The Tukey-Hanning$_p$ RK estimator approaches the semi-parametric lower bound as $p$ increases. The Parzen kernel function can also achieve the optimal convergence rate when the bandwidth is proportional to $N^{1/2}$. When using a bandwidth proportional to $N^{3/5}$ it converges at the rate 1/5 but it guarantees positivity of the RK estimator which is recommended by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). The last column in Table 1 shows the formula for the optimal bandwidth of each kernel functions. The constant term is determined by the characteristics of the kernel function. The noise-to-signal ratio $\xi^2$ can be replaced by a preliminary estimate in practice.

## 3.2 Time Dependent Noise

According to the previous section the rounding error is an important component of the noise and it cause autocorrelations at higher lags while is uncorrelated with the latent log-returns.

Apparently the no autocorrelation in Assumption 1 is not realistic any more with rounded prices. In fact the assumption has already been questioned by a few authors. Hansen and Lunde (2006) demonstrate the "ugly" fact that the noise is time dependent with the help of the signature plot which captures the relation between an IV estimator and the sampling frequency. Aït-Sahalia, Mykland, and Zhang (2011) also show that the autocorrelogram of stock prices is better captured by an AR(1) noise structure. They also found the more liquid the stock, the more likely the noise process is autocorrelated.

We modify the model by adding a moving average structure to the additive noise term since the autocorrelations diminish quickly. The assumption of time-dependent noise is given below.

**Assumption 2.** *We assume the noise $\varepsilon$ is a MA($q-1$) process for $q \geq 1$.*

$$\varepsilon_i = \upsilon_i + \beta_1 \upsilon_{i-1} + \beta_2 \upsilon_{i-1} + \cdots + \beta_{q-1} \upsilon_{i-q+1} \tag{12}$$

*where $\upsilon_t$ is a martingale difference sequence process, $E[\upsilon^2] = \sigma_\upsilon^2$, and first differences $\Delta\upsilon_t$ are uncorrelated with the latent log-returns $\Delta X_t$.*

Note that when $q = 1$ Assumption 2 is exactly the same as Assumption 1.
The additive noise process is modeled as MA($q-1$) so that its first difference $\Delta\varepsilon_i$follows is an MA($q$) process.

$$\begin{aligned}
\Delta\varepsilon_i &= \varepsilon_i - \varepsilon_{i-1} \\
&= \upsilon_i + (\beta_1 - 1)\upsilon_{i-1} + \cdots + (\beta_{q-1} - \beta_{q-2})\upsilon_{i-q+1} - \beta_{q-1}\upsilon_{i-q} \\
&= \theta_0 \upsilon_i + \theta_1 \upsilon_{i-1} + \theta_2 \upsilon_{i-2} + \cdots + \theta_q \upsilon_{i-q}
\end{aligned} \tag{13}$$

We rewrite $\Delta\varepsilon_t$ using parameters $\{\theta_0, \ldots, \theta_q\}$ in the last step in Equation (13) for simplicity of notations. The parameters $\theta$ have to satisfy $\theta_0 = 1$ and $\sum_{\tau=0}^{q} \theta_\tau = 0$ in order to ensure $\Delta\varepsilon_i$ as a first order difference of an MA noise process. Then Assumption 2 can be equivalently expressed as:

**Assumption 3.** *We assume the first difference process of the additive noise $\varepsilon_t$ is a MA($q$) process as defined in Equation (13), where $\upsilon_t$ is the same as in Assumption 2 and the parameters $\theta$ satisfies:*

$$\theta_0 = 1 \text{ and } \sum_{\tau=0}^{q} \theta_\tau = 0 \tag{14}$$

In the rest of this paper we will use Assumption 3 and parameters $\theta$.

### 3.3 Bias Corrected Realized Kernel

Immediately from Assumption 3 the expectations of the autocovariations $\gamma_k$ can be derived.

**Theorem 3.1.** *Suppose the observed process $Y$ is defined as in Equation (4) where $X$ is defined as in Equation (1) and $\varepsilon$ satisfies Assumption 2. The expectation of $\gamma_h$ is:*

$$E[\gamma_k] = \begin{cases} IV_T + N\varphi_0\sigma_v^2 & \text{if } k = 0 \\ (N-k)\varphi_k\sigma_v^2 & \text{if } 0 < k \leq q \\ 0 & \text{if } k > q \end{cases} \quad \text{where } \varphi_k = \begin{cases} 0 & \text{if } |k| > q \\ \sum_{\tau=0}^{q-|k|}\theta_\tau\theta_{\tau+|k|} & \text{if } |k| \leq q \end{cases} \tag{15}$$

*and $\gamma_k$ is defined as $\sum_{i=1}^{N-k}\Delta Y_i \Delta Y_{i+k}$.* [6]

The proof of Theorem 3.1 is given in Appendix. It is easy to spot the bias ($N\varphi_0\sigma_v^2$) in the RV ($\gamma_0$) from the above theorem. The crucial key to eliminate the exploding bias is the prior information or assumption on the MA parameters or the $\varphi$'s.

Under Assumption 1 the additive noise process is an MA(1) process and the values of $\theta_0$ and $\theta_1$ are determined as 1 and $-1$ due to the restrictions in Equation (14). As a result, $\varphi_0$ is 2, $\varphi_1$ is $-1$ and $\varphi_k$ is 0 for any $k > 1$. With the knowledge of those parameters one can construct an unbiased IV estimator combining $\gamma_0$ and $\gamma_1$ which is suggested by Zhou (1996) at the very beginning of this literature:

$$E[\gamma_0 + \frac{2N}{N-1}\gamma_1] - IV_T = (\varphi_0 + 2\varphi_1)N\sigma_v^2 = 0 \tag{16}$$

The RK estimator defined in Equation (10) is "flat-topped" [7] for the same purpose. The non "flat-topped" RK estimator generally cannot guarantee unbiasedness and require the weight on the first autocovariation $k(1/H)$ to get close to one for consistency.

Under Assumption 3 with $q > 1$ the MA parameters can no longer be identified. The bias of $RK(H)$ becomes more involved:

$$E[RK(H)] - IV_T \approx \left(\varphi_0 + 2\sum_{h=1}^{H} k(\frac{h-1}{H})\varphi_h\right) N\sigma_v^2 \tag{17}$$

In general $\varphi_h$ is non-zero when $h \leq q$ so that without further information on $\varphi$'s it is impossible to find a kernel function that can balance out the bias. [8]

[6] The $\gamma_k$ defined above is different from in Equation (11) that it does not include out of sample log-returns. When the sample size is large the difference is negligible and the one defined above will be used in the following of this paper.

[7] It means the weight on $(\gamma_1 + \gamma_{-1})$ in Equation (10) is $k(0) = 1$ instead of $k(1/H)$.

[8] The bias can be either upwards or downwards depending on the parameters of the moving average

A sub-optimal solution to the bias problem is to sample at a lower frequency. Lower frequency sampled log-returns have greater variation which can dominate the additive noise and eliminate the higher order autocorrelations. There have been discussions on how to choose sampling frequency such as Aït-Sahalia, Mykland, and Zhang (2005) and Bandi and Russell (2008) for the RV and noise under Assumption 1. It is unclear how the accuracy of the RV with chosen best frequency is affected by the MA noise.

The second solution progressively modifies the kernel function to re-establish unbiasedness by increase the length of the "flat-top" part of the kernel to $q$. Although the bias in Equation (17) involve all the MA parameters ($\varphi_h$) the bias can be cancelled out by using the restrictions in Equation (14) and noticing that since $\sum_i \theta_i = 0$:

$$\varphi_0 + 2\varphi_1 + \cdots + 2\varphi_q = (\theta_0 + \ldots \theta_q)^2 = 0 \tag{18}$$

Thus setting the first $q$ values of the kernel to be one makes the sum inside the bracket in Equation (17) to be zero. Formally the MA($q$) bias corrected version of RK estimator is defined as:

$$RK^q(H) = \gamma_0(Y) + \sum_{h=1}^{q} \frac{2N}{N-h}\gamma_h(Y) + \sum_{h=1}^{H} k(\frac{h}{H})\frac{2N}{N-h-q}\gamma_{h+q}(Y) \tag{19}$$

where $k(\cdot)$ is the kernel function. The $\gamma$'s are the same as in Theorem 3.1 and are adjusted accordingly to fully eliminate the bias.

**Corollary 3.1.1.** *Assume $Y$ is modeled as in Theorem 3.1 and $q$ is finite. Then for any kernel function $k(\cdot)$ the bias corrected RK estimator $RK^q(H)$ defined in Equation* (19) *is unbiased and has the same asymptotic properties as the original RK estimator defined in Equation* (10).

The unbiasedness has been shown above. The modification does not affect the asymptotic of the RK estimator. The reason is that when $q$ is finite the first $q$ weights valued by the kernel function approach one in the limit as long as the bandwidth goes to infinity as the sample size does and the kernel function is flat enough near zero ($k'(0) \approx 0$). It suggests another solution to simply apply the conventional RK estimator without modification because it is still consistent. When the MA dependence is brought from the rounding error the order $q$ is likely to increase as $N$ increase, which can cause the RK estimator to be biased even in limit.

noise process, which may provide another explanation to the negative bias in signature plots other than the correlation of the latent process $X$ and the noise process as mentioned in Hansen and Lunde (2006).

The bias corrected RK estimator is clearly the best among the solutions mentioned above. In practice the order of moving average of the noise return is unknown. It may be detected either as the largest lag with significantly non-zero autocorrelation of the observed log returns or by MA model selection using AIC, BIC, or other criterion.

## 3.4 Parametric Lower Bound

Gloter and Jacod (2001a,b) derived the optimal convergence rate and asymptotic Fisher information of the $IV$ estimator for a semi-martingale process with martingale difference sequence noises. The optimal asymptotic variance of the estimator of a constant volatility is considered as the semi-parametric lower bound of any $IV_T$ estimator. We extend the result to the model with moving average noise.

First denote $\Phi$ as the $q \times 1$ vector $(\varphi_0, \varphi_1, \ldots, \varphi_q)'$, and $C^N(\sigma^2, \sigma_v^2, \Phi)$ as the variance covariance matrix of $(\Delta Y_1, \ldots, \Delta Y_N)'$. Assuming constant volatility ($\sigma_t^2 = \sigma^2, \forall t$), the $(i, j)$ entry of the matrix $C^N$ is given by the function:

$$C^N(\sigma^2, \sigma_v^2, \Phi)_{i,j} = \begin{cases} \frac{\sigma^2}{N} + \varphi_0 \sigma_v^2 & \text{if } i = j \\ \varphi_k \sigma_v^2 & \text{if } |i-j| = k \leq q \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

Applying the third case of the theory in Gloter and Jacod (2001a) to the model with MA noises gives the following theorem.

**Theorem 3.2.** *Suppose the observed process $Y$ is defined as in Equation* (4) *where $X$ is defined as in Equation* (1) *such that $\sigma_t = \sigma$ for any $t$ and $\varepsilon$ satisfies Assumption 2. Denote $\lambda_i(\sigma^2, \sigma_v^2, \Phi)$ as the eigenvalues of matrix $C^N(\sigma^2, \sigma_v^2, \Phi)$. If for any sequence $\{h_n\}$ with a limit $h$, the following two conditions are satisfied:*

$$\sup_{1 \leq i \leq N} |\delta_i^N| \to 0 \qquad\qquad \frac{1}{2h^2} \sum_{i=1}^{N} (\delta_i^N)^2 \to I(\sigma^2, \Phi) \tag{21}$$

*where $\delta_i^N = u_n h_n / N / \lambda_i(\sigma^2, \sigma_v^2, \Phi)$, $u_n = (\sigma_v^2/N)^{1/4}$, then the optimal IV estimator $\hat{\sigma}^2$ converges to $\sigma^2$:*

$$N^{\frac{1}{4}}(\hat{\sigma}^2 - \sigma^2) \to_d \mathcal{N}(0, \sigma_v I(\sigma^2, \Phi)^{-1}) \tag{22}$$

When $q$ is equal to one the model degenerates to the one with white noise assumption and the matrix $C^N$ is tri-diagonal [9] whose eigenvalues can be analytically solved (see Sun (2006)). Therefore the closed form expression of $I(\sigma^2)$ can be found as the limit in Equation (21). In that case Equation (22) becomes:

$$N^{1/4}(\hat{\sigma}^2 - \sigma^2) \rightarrow \mathcal{N}(0, 8\sigma_v\sigma^3) \tag{23}$$

The asymptotic variance $8\sigma_v\sigma^3$ has been known as the semi-parametric lower bound for the consistent IV estimators. With $q$ greater than one there is no closed form solution to the eigenvalues of matrix $C^N$. Consequently $I(\sigma^2, \Phi)$ can only be numerically approximated.

As an example Table 2 shows the approximated lower bounds of the model under Assumption 3 and $q = 2$. The dimensions of $C^N$ used in calculation are 1000, 5000, and 10000. Similarly as with the martingale difference sequence noise, the inverse of the Fisher information is proportional to $\sigma^3$ and the factor is determined by $\Phi$. Note that when $\theta_1 = -1$ or $\beta_1 = 0$ the differenced noise process is MA(1) and the approximated and the computed lower bound is 8.16 which is close to the predicted value 8 in Equation (23).

Table 2: Numerically Computed Lower Bounds of Model with MA(2) Noise

| $\beta_1$ | 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\binom{\theta_1}{\theta_2}$ | $\binom{-1}{0}$ | $\binom{-0.9}{-0.1}$ | $\binom{-0.8}{-0.2}$ | $\binom{-0.7}{-0.3}$ | $\binom{-0.6}{-0.4}$ | $\binom{-0.5}{-0.5}$ | $\binom{-0.4}{-0.6}$ | $\binom{-0.3}{-0.7}$ | $\binom{-0.2}{-0.8}$ | $\binom{-0.1}{-0.9}$ | $\binom{0}{-1}$ |
| 1,000 | 8.53 | 9.44 | 10.37 | 11.31 | 12.26 | 13.23 | 14.21 | 15.20 | 16.20 | 17.10 | 9.14 |
| 5,000 | 8.23 | 9.08 | 9.93 | 10.79 | 11.66 | 12.53 | 13.40 | 14.28 | 15.17 | 16.05 | 8.48 |
| 10,000 | 8.16 | 9.00 | 9.83 | 10.68 | 11.52 | 12.37 | 13.22 | 14.08 | 14.93 | 15.79 | 8.33 |

The $\varepsilon_t$ process is: $\varepsilon_t = v_t + \beta_1 v_{t-1}$, where the values of possible $\beta_1$ are listed in the first row. The parameters of the $\Delta\varepsilon$ process are $\theta_0 = 1$, $\theta_1$ in the first row and $\theta_2 = -1 - \theta_1$. The lower bounds are in terms of $\sigma^2\sigma_v$.

# 4 Optimal Weights With MA Noise

The RK estimator is in fact motivated by the control variable approach. Under Assumption 1 $RK(1)$ is an unbiased but inconsistent estimator which has been proposed by Zhou (1996). When $H$ is greater than 1 the higher order realized autocovariations are used as control variables to reduce the variance of the estimator. The chosen kernel function assigns weights to the control variables. The RK estimator with bandwidth $H$ utilizes $H-1$ control variables.

To briefly review the method, suppose $g(Y)$ is an unbiased estimator of the parameter of interest and there are $m$ random variables $X = (X_1, X_2, \cdots, X_m)'$ which have zero expectation. Then $g(Y) + \alpha' X$ is a new unbiased estimator where $\alpha$ is a $m \times 1$ weighting

[9] A tri-diagonal matrix has non-zero elements only on its major diagonal and the two sub-diagonals above and below.

vector. With a proper choice of $\alpha$, the new estimator has a lower variance as long as $X$'s are correlated with $g(Y)$. This is known as the control variables approach of variance reduction. The optimal choice of $\alpha$ can be solved by minimizing the variance of the new estimator:

$$\alpha^* = -V[X]^{-1}Cov[X, g(Y)]$$

Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) use the optimal weights as a comparing benchmark for other kernel functions. In this section we construct the optimal weights for the RK estimator under Assumption 3 and also propose a feasible approximation.

## 4.1 Optimal RK Weights

We use $\Gamma_H$ to denote the $(H+q+1)\times 1$ vector:

$$\Gamma_H = \begin{bmatrix} \Gamma_{H,1} \\ \Gamma_{H,2} \end{bmatrix} \text{ where } \Gamma_{H,1} = (\gamma_0, 2\gamma_1, \ldots, 2\gamma_q)', \text{ and } \Gamma_{H,2} = (2\gamma_{q+1}, \ldots, 2\gamma_{H+q}) \tag{24}$$

Each $\gamma_h$ for $h > 0$ is multiplied by 2 to make the control variable weights consistent with the scale of kernel functions. Suppose the noise follows Assumption 3, Theorem 3.1 suggests that the unbiased estimator can be rewritten as $\iota'_{q+1}\Gamma_{H,1}$, where $\iota_{q+1}$ is a $(q+1)\times 1$ vector of ones.

The RK estimator with control variable kernel is defined as:

$$RK^q_{CV}(H) = \gamma_0 + 2\sum_{h=1}^{q}\gamma_h + 2\sum_{h=1}^{H}\alpha_h\gamma_{h+q} = \iota'_{q+1}\Gamma_{H,1} + \alpha'\Gamma_{H,2} \tag{25}$$

The weight $\alpha_h$ corresponds to $k(\frac{h-1}{H})$ in Equation (10). To find the optimal weights, we need the variance covariance matrix of $\Gamma_H$:

**Theorem 4.1.** *Suppose the observed process $Y$ is modeled as in Theorem 3.1. The variance covariance matrix of $\Gamma_H$ is:*

$$\begin{aligned} \Sigma_H &= Var[\Gamma_H] \\ &= 2IQ_T/N\cdot A + 4\sigma_\upsilon^2 IV_T\cdot B + 4\sigma_\upsilon^4(N\cdot C + D) + V[\upsilon^2](N\cdot E + F) \end{aligned} \tag{26}$$

*where $IQ_T$ is the integrated quarticity $\int_0^T\sigma_\tau^4 d\tau$ and the $(H+1)\times(H+1)$ matrices $A$, $B$, $C$, $D$, $E$, and $F$ are given in Appendix B.*

The matrices all have a diagonal structure. $A$ is a diagonal matrix; $B$ has $q$ non-zero sub-diagonals on each side of the major diagonal; $C$ and $D$ have $2q$ non-zero sub-diagonals on each side of the major diagonal; while the entries of $E$ and $F$ are zero except the upper-left $(q+1)\times(q+1)$ blocks. In the case when $q$ equal to one the matrices $A$, $B$, $C$ and $D$ correspond to the matrices defined in Theorem 3 of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and are different in two ways: first, the matrices in this paper characterize the finite sample variance covariance matrix of $\Gamma_H$, while those in their paper are for asymptotic property; second, the $\gamma$'s in this paper are defined without out of sample data.

To express the optimal control variable weights we partition the matrix $\Sigma_H$ into a $2 \times 2$ block structure:

$$\Sigma_H = \begin{bmatrix} \underbrace{\Sigma_{H,11}}_{(q+1)\times(q+1)} & \bullet \\ \underbrace{\Sigma_{H,21}}_{(H-q)\times(q+1)} & \underbrace{\Sigma_{H,22}}_{(H-q)\times(H-q)} \end{bmatrix}$$

The optimal weights $\alpha_H^* = (\alpha_1^*, \ldots, \alpha_H^*)'$ are:

$$\alpha_H^* = -(\Sigma_{H,22})^{-1}(\Sigma_{H,21} \cdot \iota_{q+1}) \tag{27}$$

By definition the weights generated from the control variable approach is optimal among all the kernel functions in the RK framework when the model is correctly specified.

Note that the estimator defined in Equation (25) is almost unbiased. To completely eliminate the bias, $\gamma_h$ need to be scaled by $N/(N-h)$ for $h \leq q$. The adjustment is negligible when the sampling frequency is high. For simplicity we used unscaled $\gamma_h$ when deriving the variance-covariance matrix of $\Gamma_H$ and constructing the optimal control variable weights. The exact unbiasedness can be achieved by multiplying a scaling vector $(1, N/(N-1), \ldots, N/(N-q))$ to $\Gamma_{H,1}$ element-by-element.

## 4.2 Approximation

The exact optimal control variable weights $\alpha_H^*$ is not feasible in practice because it requires the knowledge of the data generating process to construct the matrix $\Sigma_H$ in Equation (26). Even though the values of $IV_T$ and $IQ_T$ can be replaced by consistent estimators, the volatility path $\{\sigma_i\}$, the MA parameters $\Phi$, and the variance of $\upsilon$ are unable to identify while necessary to construct the matrices. We use the following method to overcome the problem and approximate the optimal weights.

Firstly, matrices $E$ and $F$ only have non zero elements in their upper left blocks $E_{11}$ and $F_{11}$. Therefore they affect neither $\Sigma_{H,22}$ nor $\Sigma_{H,21}$ in Equation (27) so that $V[\upsilon^2]$ is irrelevant in computing $\alpha_H^*$.

Secondly, only the edge areas of the matrices depend on the volatility path $\{\sigma_i\}$. When $N$ is large we can ignore the subtle differences in the edge areas and approximate the matrices $A$, $B$ and $C$ by $\hat{A}$, $\hat{B}$ and $\hat{C}$ respectively. Matrix $\hat{A}$ is diagonal with value 2 except the first term. Matrix $\hat{B}$ has $q$ non-zero sub-diagonals and matrix $\hat{C}$ has $2q$ non-zero sub-diagonals. Matrices $\hat{B}$ and $\hat{C}$ depend purely on $\Phi$. Recall that $\Phi = (\varphi_0, \ldots, \varphi_q)'$'s are defined as in Equation (15) and furthermore we define:

$$\psi_k = \begin{cases} 0 & \text{if } k < 0 \text{ or } k > 2q \\ \sum_{j=-q}^{q-k} \varphi_{|j|}\varphi_{|j+k|} & \text{if } k = 0, 1, 2, \ldots, 2q \end{cases} \tag{28}$$

$$\hat{A} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 2 & 0 & \ldots & 0 \\ 0 & 0 & 2 & \ldots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \ldots & 0 & 2 \end{bmatrix}, \hat{B}(\Phi) = \begin{bmatrix} \varphi_0 & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ 2\varphi_1 & 2\varphi_0 & \bullet & \bullet & \bullet & \bullet & \bullet \\ \vdots & \ddots & \ddots & \bullet & \bullet & \bullet & \bullet \\ 2\varphi_q & \ddots & 2\varphi_1 & 2\varphi_0 & \bullet & \bullet & \bullet \\ 0 & \ddots & \ddots & \ddots & \ddots & \bullet & \bullet \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \bullet \\ 0 & \ldots & 0 & 2\varphi_q & \ldots & 2\varphi_1 & 2\varphi_0 \end{bmatrix}$$

$$\hat{C}(\Phi) = \begin{bmatrix} \psi_0 & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \psi_1 & \psi_0 & \bullet & \bullet & \bullet & \bullet & \bullet \\ \vdots & \ddots & \ddots & \bullet & \bullet & \bullet & \bullet \\ \psi_{2q} & \ddots & \psi_1 & \psi_0 & \bullet & \bullet & \bullet \\ 0 & \ddots & \ddots & \ddots & \ddots & \bullet & \bullet \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \bullet \\ 0 & \ldots & 0 & \psi_{2q} & \ldots & \psi_1 & \psi_0 \end{bmatrix}$$

Thirdly, even though $\sigma_v^2$ and $\varphi$'s can not be identified separately, their multiplications can be consistently estimated by:

$$\widehat{\sigma_v^2\varphi_h} = \frac{1}{N-h}\gamma_h \to_p \sigma_v^2\varphi_h \tag{29}$$

Denote $\widehat{\sigma_v^2\Phi}$ as the vector $[\gamma_0/N, \ldots, \gamma_q/(N-q)]'$. Then $\hat{B}(\widehat{\sigma_v^2\Phi})$ and $\hat{C}(\widehat{\sigma_v^2\Phi})$ consistently estimate $\sigma_v^2 B$ and $\sigma_v^4 C$ respectively.

We have left out the matrix $D$ because for a large sample size $N$ its magnitude is small compared to $N \cdot C$. However if higher accuracy is desired the matrix $\sigma_v^4(N \cdot C + D)$ can be estimated using Equation (35) and $\widehat{\sigma_v^2 \Phi}$. [10]

Finally, with the preliminary estimates of $IV_T$, $IQ_T$, and $\widehat{\sigma_v^2 \Phi}$, a consistent estimate of the optimal control variable weights is:

$$\hat{\alpha}_H^* = \left( 2\frac{\hat{IQ}_T}{N}\hat{A}_{22} + 4\hat{IV}_T\hat{B}_{22}(\widehat{\sigma_v^2\Phi}) + 4N\hat{C}_{22}(\widehat{\sigma_v^2\Phi}) \right)^{-1} \cdot \left( 2\frac{\hat{IQ}_T}{N}\hat{A}_{21} + 4\hat{IV}_T\hat{B}_{21}(\widehat{\sigma_v^2\Phi}) + 4N\hat{C}_{21}(\widehat{\sigma_v^2\Phi}) \right) \iota_{q+1} \tag{30}$$

The above is a feasible approximation of the optimal weights. The realized variation ($\gamma_0$) with sparsely sampled log-returns or the RK estimator using $\text{TH}_2$ kernel function with an arbitrary bandwidth can be used as the preliminary estimator of $IV_T$. The integrated quarticity $IQ_T$ is relatively difficult to estimate. Assuming the volatility path does not vary much during one day $\hat{IV}_T^2$ can be used as an estimate. Both Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and Jacod, Li, Mykland, Podolskij, and Vetter (2009) show how to construct a consistent estimator of $IQ_T$. Furthermore both of the brackets in the above equation are dominated by matrix $C$ when $N$ is large. When only a lower order of approximation is necessary there is no need for the preliminary estimates of $IV_T$ or $IQ_T$. The optimal bandwidth $H$ is infinity by definition of the control variable approach because the more control variables the lower is the variance. Due to the approximation and computation cost a high dimension of $H$ is undesirable. A rule of thumb is to use the same bandwidth as in $\text{TH}_p$ kernel function with $p = 2$ (or $p = 16$ if $H^*$ is not too large).

# 5 Simulation

A Monte Carlo experiment is carried out in this section to demonstrate the effect of rounding error on the existing estimators and the advantage of the bias correction.

The following estimators in the literature are considered: unbiased estmoatrs; sparsely sampling RV (10 and 30 minutes); Two-Scale RV (Zhang, Mykland, and Ait-Sahalia (2005)); Multi-Scale RV (Zhang (2006)); Pre-Average RV (Jacod, Li, Mykland, Podolskij, and Vetter

[10] The matrix $M$ in equation (35) depends only on $\Phi$. An estimate of $\sigma_v^4(N \cdot C + D)$ is obtained by the second summation in the last step while replacing $\Phi$ by $\widehat{\sigma_v^2 \Phi}$ in matrix $M$. To include $D$ in the approximation, replace the $N\hat{C}$ in Equation (30) by the estimate of $\sigma_v^4(N \cdot C + D)$.

(2009); RK estimator with Tukey-Hanning$_p$ ($p = 2, 16$) and Parzen kernels. The bias-corrected RK estimators with the above kernels and optimal weights are also computed.

The 30 minutes sampled RV and its square are served as a preliminary IV and QV estimators respectively. The preliminary estimator of the variance of the noise $\sigma_\varepsilon^2$ is defined in Equation (9). The optimal bandwidths of all the estimators are computed from preliminary estimates.

## 5.1 Data Generating Process

The data is generated to mimic the transaction data of "AA" in 2012. The log-prices are generated using Equation (1). The sampling frequency is two ticks per second ($N = 46,800$) and the volatility $\sigma_t$ is set to be constant at level $\sigma^2 = 2.2129 \times 10^{-4}$. The log-price process initiates at $\log 9$. Finally the experiment is repeated 100,000 times.

Taking exponential on $X_t$ yields the raw level prices $S_t$ and $\tilde{S}_t$ is the rounded price with tick size of one cent. To concentrate on the effect of rounding error we assume that rounding is the only source of the additive noise term.

$$\varepsilon_t = \log S_t - \log \tilde{S}_t \tag{31}$$

In each trial we simulate the latent log-price process, the rounded log-price process, and the additive noise process. For comparison we generate a new log-price process removing the autocorrelations in the noise by adding the randomly re-ordered additive noise process onto the latent log-price process. The noise in the new process satisfies Assumption 1 and has the same magnitude as the noise in rounded log-price process. All the estimators are applied on both processes.

## 5.2 Result

Table 3 summarizes the bias and RMSE of the existing estimators in the literature.

In this paper we use an MA($q$) model to accommodate the features of rounding error. A possible drawback is that it overlooks the correlation between the log-returns and the differenced noises which could be another source of bias:

$$E[\sum \Delta Y_i^2] = IV_T + 2\sum E[\Delta X_i \Delta\varepsilon_i] + NE[\Delta\varepsilon^2] \tag{32}$$

The first row in the table reports the mean and root uncentered second moment of the cross-product term normalized by the true $IV_T$. As we have mentioned in the facts of the rounding error the cross-product term does not affect the bias of the RV and RK estimator.

Table 3: Bias And RMSE of IV Estimators With Rounding Errors

| | Re-ordered Noise | | | Rounding Noise | | |
|---|---|---|---|---|---|---|
| | Bandwidth | Bias | RMSE | Bandwidth | Bias | RMSE |
| $2\sum \Delta X_i \Delta \varepsilon_i$ | - | -0.0001 | 0.0610 | - | -0.0009 | 0.0833 |
| $\mathrm{RV}_{10min}$ | - | 0.0361 | 0.2375 | - | 0.0357 | 0.2374 |
| $\mathrm{RV}_{30min}$ | - | 0.0120 | 0.3983 | - | 0.0114 | 0.3982 |
| $\mathrm{RV}_{Unbiased}$ | - | -0.0008 | 0.2911 | - | 4.3377 | 4.3470 |
| $\mathrm{RV}_{TSRV}$ | 20 | -0.0004 | 0.0491 | 9 | 2.2575 | 2.2821 |
| $\mathrm{RV}_{MSRV}$ | 17 | -0.0006 | 0.0315 | 9 | 2.1308 | 2.1493 |
| $\mathrm{RV}_{PARV}$ | 24 | 0.0171 | 0.0373 | 13 | 2.2141 | 2.2357 |
| $\mathrm{RV}_{RKTH2}$ | 29 | -0.0002 | 0.0307 | 16 | 1.9108 | 1.9265 |
| $\mathrm{RV}_{RKTH16}$ | 196 | -0.0003 | 0.0302 | 106 | 1.9434 | 1.9591 |
| $\mathrm{RV}_{RKPZ}$ | 109 | 0.0012 | 0.0518 | 67 | 0.4667 | 0.4805 |

The bias and RMSE are the average across the $100,000$ Monte Carlo trials and are normalized by the true values $\sigma^2$. The first row reports the effect of the correlation of the log-returns and differenced noises on the bias. Its bias and RMSE are computed as if its expectation is zero. The bandwidth columns are the averaged optimal bandwidth of the estimators except the sparsely sampled RV's and the unbiased estimator as in Equation (16).

When the MA time dependence in the noise is removed by randomly re-ordering the additive noises the estimators perform as it is predicted in the literature. The sparsely sampled RV's are able to remove most of the bias while has a fairly large deviation around 30%. The consistent estimators have almost zero bias and small RMSE. Among them the fast converging estimators (MSRV, PARV, RKTH) tend to have better lower RMSE than the slow conerging estimators (TSRV, RKPZ).

The accuracy of the estimators drops dramatically when the log-prices are with additive noises derived from rounding error. The 10 and 30 minutes RV's are merely affected by the rounding error. In high frequency samples all the supposedly consistent estimators have high bias and RMSE where the RMSE mainly consists of the bias. The optimal bandwidths are lower than those in the re-ordered noise case for the noise variance estimator defined in Equation (9) converges to a value other than $E[\varepsilon^2]$ when the noise is an MA process. Interestingly the RK estimator with Parzen kernel dominates other "consistent" estimators while still cannot out-perform the sparsely sampled RV's.

Table 4 summarizes the bias and RMSE of the bias-corrected RK estimators. The results of the RK estimator with Tukey-Hanning$_{16}$ kernel are close to those one with Tukey-Hanning$_2$ kernel and are not reported. The unbiased estimator is $\iota'_{q+1}\Gamma_{H,1}$ as in Equation (25)

The first column is the degree of bias-correction, that is the order of MA process considered. When $q = 1$ there is no bias-correction. The advantage of bias correction is obvious for all the estimators. For example, the bias of the RK estimator with Tukey-Hanning$_2$ kernel drops from 191.08% to 0.02% and the RMSE drops from 192.65% to 11.23% after the bias-

Table 4: Bias And RMSE of Bias-Corrected IV Estimators

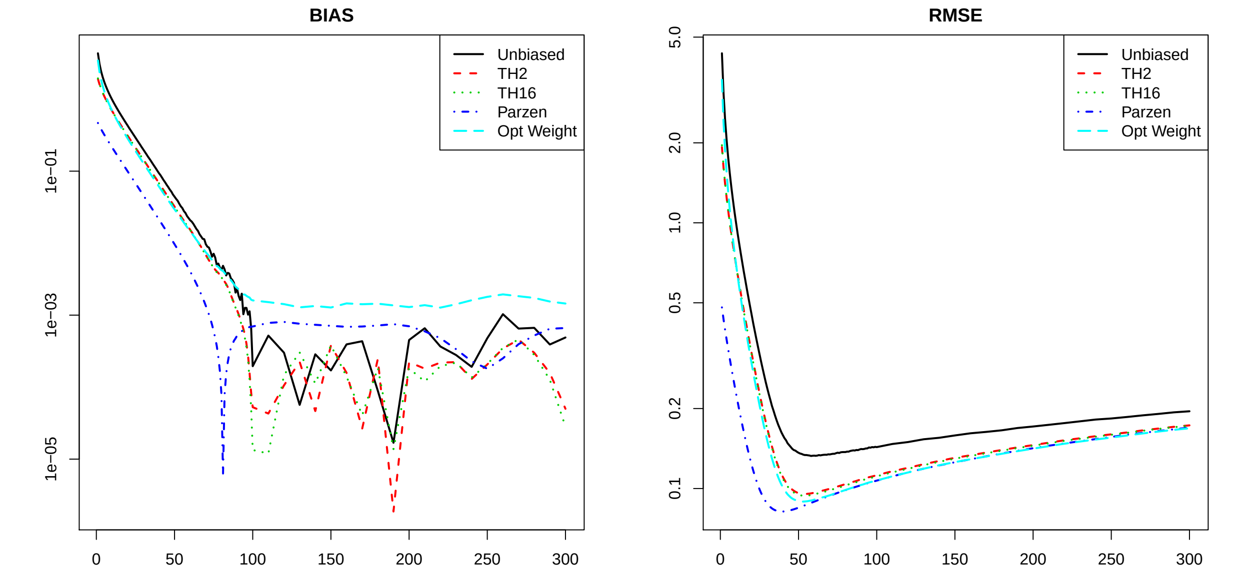
| | Bias | | | | RMSE | | | |
|---|---|---|---|---|---|---|---|---|
| $q$ | Unbiased | RKTH2 | RKPZ | RKCV | Unbiased | RKTH2 | RKPZ | RKCV |
| 1 | 4.3377 | 1.9108 | 0.4667 | 3.4494 | 4.3470 | 1.9265 | 0.4805 | 3.4562 |
| 2 | 3.0952 | 1.6095 | 0.4215 | 2.4055 | 3.1043 | 1.6213 | 0.4347 | 2.4116 |
| 3 | 2.4533 | 1.3980 | 0.3836 | 1.8751 | 2.4626 | 1.4079 | 0.3965 | 1.8810 |
| 4 | 2.0429 | 1.2356 | 0.3508 | 1.5377 | 2.0525 | 1.2445 | 0.3635 | 1.5436 |
| 5 | 1.7507 | 1.1047 | 0.3219 | 1.2992 | 1.7607 | 1.1129 | 0.3344 | 1.3051 |
| 10 | 0.9904 | 0.6896 | 0.2145 | 0.6939 | 1.0033 | 0.6973 | 0.2284 | 0.7006 |
| 15 | 0.6381 | 0.4584 | 0.1456 | 0.4319 | 0.6546 | 0.4676 | 0.1630 | 0.4402 |
| 20 | 0.4291 | 0.3110 | 0.0993 | 0.2840 | 0.4508 | 0.3230 | 0.1227 | 0.2951 |
| 50 | 0.0441 | 0.0320 | 0.0095 | 0.0294 | 0.1362 | 0.0961 | 0.0847 | 0.0896 |
| 100 | 0.0002 | 0.0002 | -0.0006 | 0.0017 | 0.1433 | 0.1123 | 0.1071 | 0.1072 |
| 150 | -0.0002 | -0.0004 | -0.0007 | 0.0003 | 0.1584 | 0.1305 | 0.1256 | 0.1298 |
| 200 | -0.0005 | -0.0002 | -0.0007 | 0.0003 | 0.1712 | 0.1457 | 0.1418 | 0.1423 |
| 250 | 0.0005 | 0.0002 | -0.0002 | 0.0008 | 0.1835 | 0.1600 | 0.1564 | 0.1558 |
| 300 | 0.0005 | 0.0001 | -0.0007 | 0.0005 | 0.1953 | 0.1729 | 0.1696 | 0.1686 |

correction with $q = 100$. The RMSE after bias correction is greater than the RMSE with re-ordered noise which reflects that the semi-parametric lower bound for the IV estimators is different with MA noise process.

Figure 4 plots the bias (in absolute value) and RMSE of the IV estimators against the degree of bias correction. A logarithmic scale is applied on Y axis. The curves of bias and RMSE are very similar. When $q$ is less than 50, the bias and RMSE almost monotonically decrease as $q$ increases. When $q$ is greater than 50, the bias fluctuates wildly around a near zero level and the RMSE shows an increasing trend. With a large $q$ the bias-corrected estimators are indifferent in RMSE sense.

The results suggests that the "real" order of MA of the addtitive noise derived by the rounding error is likely to be 50. Recall that the sampling frequency is two ticks per second hence the bandwidth of 50 implies that the MA effect brought by the rounding error last as long as 25 seconds in the current simulation. It does not reflect the autocorrelation function plot in Figure 3 whereas the autocorrelations are significant up to five lags. It raises the problem of choosing the degree of bias correction in practice. A rule of thumb is to be aggresive because a large $q$ does not harm the estimators much.

Figure 4: Bias And RMSE of Bias-Corrected Estimators




The rounding tick size ($\alpha$) is set to be 1/100 (one cent) as in present stock markets. The rounding error is a deterministic and periodic function of the raw stock price, which is ranged in $(-\alpha/2, \alpha/2)$.

# 6 Conclusion

This paper studies the effect of the rounding error on the existing IV estimators in the literature and suggests a bias correction of the RK estimator.

The rounding error is treated as a deterministic function of the raw price. After transforming the rounding error into the additive noise term in the log-price process, the magnitude of the noise depends on the raw price level. From the simulation we find the rounding error causes the observed (rounded) log-returns have non-zero autocorrelations at lags higher than 1 in high frequency samples. The pattern of the autocorrelation function resembles the one of the empirical data. Comparing the predicted magnitude of the pure rounding error noise term and the estimated variance of the overall micro-structure noises over two decades, we find the rounding error is an important component of the noise and has became a major one. Therefore the existing consistent IV estimators fail to converge to the true value when overlooking the rounding error.

To accommodate features of the rounding error we study the RK estimator under the assumption that the noise is an MA($q$) process. The conventional RK estimator is biased but consistent when $q$ is greater than 1 and finite. We suggest a bias correction of the RK estimator which re-establish the unbiasedness and does not affect the asymptotics. Furthermore we derive the optimal weights of the RK estimator with MA noise and propose a feasible approximation.

At last We perform a Monte Carlo simulation using the parameters calibrated to the transcation data of "AA" in 2012. It is the time dependence brought by the rounding error that cause the consistent estimators to suffer from large bias. The RK estimators gain great improvement from the bias correction. From the plots of bias and RMSE against the assumed $q$ the autocorrelation brought by the rounding error lasts as long as 50 ticks or 25 seconds in this particular calibration.

# Appendix

# A Data Cleaning

High frequency transaction and quotation data are obtained from the TAQ database. The cleaning procedure is essential since faulty trade or quote reports can dramatically change the IV estimators and the autocorrelations. It also demands large amount of computation because of the large volume especially for quotation data. The procedure applied in this paper is based on Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) with modification to keep as many records as possible and to reduce the computation requirement.

- General Cleaning

  1. Keep the records when the exchange is open that is within 9:30AM to 4:00PM. If a different time window is of interest one can modify this filter although the data are noisier in pre-market and post-market.
  2. Delete the records with zero bid, ask or transaction prices. The transaction prices are seldomly zero while the quotes are often zero when one side quotes are recorded.
  3. (Optional) Keep the records from a single or selected exchanges. In this paper all exchanges are included.

- Abnormal Records

  1. (For transaction data)
     Keep the records with the field "COND" ("TR_SCOND" in millisecond database) is '@', 'E', 'F', or blank.
     Keep the records which the field "CORR" ("TR_CORR" in millisecond database) is '0', '1', or '2'.

2. (For quotation data)
   Delete the records with bid greater than ask.
   Keep the records with the field "MODE" ("QU_COND" in millisecond database) is blank or '12' ('R' in millisecond database).

- Outliers The above procedure filters out most of the faulty records however some outliers remains in the series. The following outliers detecting procedure is milder and easier to computer compared to the T4 and Q4 in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009).

  1. The high frequency returns that are higher than 100% or lower than $-50\%$ are treated as faulty records. This procedure is to filter out mistakes such as a misplaced decimal point.
  2. For each point of the trade series or mid-quote series compute the rolling average distance from the 25 observations before and 25 observations after. Keep the records whose average distance is within 10 standard deviation from the median of the daily average distance.

Finally one can use the median price, bid or ask if there are multiple records for the same time stamp. Then the sampling frequency is bounded by per second and a large amount of records will be discarded. However with the milliseconds time database the price series are stamped at a sub-second frequency. Hence to utilize as many records as possible, we keep all the ticks.

# B Proof of Theorem 3.1 and 4.1

*Proof.* The observed log returns $\{\Delta Y_i\}_{i=1}^N$ are consisted of two components. The latent log returns $\{\Delta X_i\}_{i=1}^N$ are discrete sampled process of the semi-martingale defined in equation (1). The process $\{\Delta \varepsilon_i\}_{i=1}^N$ is generated by the first order difference of a noise process $\{\varepsilon_i\}$ which satisfies Assumption 2.

Denote $\Delta Y$ as the vector of observations $(\Delta Y_N, \Delta Y_{N-1}, \dots, \Delta Y_1)'$. Similarly denote $\Delta X$ and $\Delta\varepsilon$ as the $N$-by-1 vectors of the latent log returns and the MA($q$) noise differences. Then we can write $\Delta Y = \Delta X + \Delta\varepsilon$.

The process $\{\Delta\varepsilon_i\}$ is an MA($q$) process driven by an m.d.s. innovation $\upsilon$ with parameters $(\theta_0, \dots, \theta_q)$ as in equation (13). Denote the $(N+q)$-by-1 vector $(\upsilon_N, \upsilon_{N-1}, \dots, \upsilon_{-q+1})'$ as $U$.

Then we can write $\Delta\varepsilon$ as $T \cdot U$, where the $N$-by-$(N+q)$ matrix $T$ is given by:

$$T = \begin{bmatrix} \theta_0 & \theta_1 & \dots & \theta_q & 0 & \dots & 0 \\ 0 & \theta_0 & \theta_1 & \dots & \theta_q & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \theta_0 & \theta_1 & \dots & \theta_q \end{bmatrix} \tag{33}$$

To represent $\gamma_h$ by $\Delta Y$, define the $N$-by-$N$ matrix $J_k$ as a symmetric matrix with ones on both of the $k$-th subdiagonal and zeros otherwhere. For example, $J_0$ is the identity matrix and $J_1$ is:

$$J_1 = \begin{bmatrix} 0 & 1 & 0 & \\ 1 & 0 & 1 & \ddots \\ 0 & 1 & 0 & \ddots \\ & \ddots & \ddots & \ddots \end{bmatrix} \tag{34}$$

We can write $2\gamma_k$ as $\Delta Y' J_k \Delta Y$.

**Expectations**

$$\begin{aligned} E[\gamma_0] &= E[\Delta Y' J_0 \Delta Y] = E[(\Delta X + \Delta\varepsilon)' J_0 (\Delta X + \Delta\varepsilon)] \\ &= E[\Delta X' J_0 \Delta X] + E[U'T'J_0TU] = \sum E[\Delta X_i^2] + \sigma_v^2 \text{trace}(T'T) \\ &= IV_T + N\varphi_0\sigma_v^2 \end{aligned}$$

For $k > 0$:

$$\begin{aligned} E[2\gamma_k] &= E[\Delta Y' J_k \Delta Y] = E[\Delta X' J_k \Delta X] + E[U'T'J_kTU] = \sigma_v^2 \text{trace}(T'J_kT) \\ &= 2(N-k)\varphi_k\sigma_v^2 \end{aligned}$$

Note that $\varphi_k$ is zero when $k > q$ as defined in equation (15).

**Variance Covariance Matrix**

When $|k-l| > 2q$, the covariance between $\gamma_k$ and $\gamma_l$ is zero. When $|k-l| \leq 2q$,

$$\begin{aligned} Cov[\gamma_k, \gamma_l] &= Cov[\Delta Y' J_k \Delta Y, \Delta Y' J_l \Delta Y] \\ &= \underbrace{Cov[\Delta X' J_k \Delta X, \Delta X' J_l \Delta X]}_{\text{part i}} + 4\underbrace{Cov[\Delta X' J_k \Delta \varepsilon, \Delta X' J_l \Delta \varepsilon]}_{\text{part ii}} \\ &+ \underbrace{Cov[\Delta \varepsilon' J_k \Delta \varepsilon, \Delta \varepsilon' J_l \Delta \varepsilon]}_{\text{part iii}} \end{aligned}$$

Part i:

$$Cov[\Delta X' J_k \Delta X, \Delta X' J_l \Delta X] = 0 \text{ if } k \neq l$$

$$\begin{aligned} Cov[\Delta X' J_k \Delta X, \Delta X' J_k \Delta X] &= V[\Delta X' J_k \Delta X] \\ &= \begin{cases} \sum V[\Delta X_i] = 2\sum \sigma_i^4 & \text{if } k = 0 \\ 4\sum V[\Delta X_i \Delta X_{i+k}] = 4\sum \sigma_i^2 \sigma_{i+k}^2 & \text{if } k > 0 \end{cases} \end{aligned}$$

Therefore, matrix $A$ is a diagonal matrix such that:

$$A(k,l) = \begin{cases} 1 & \text{if } k = l = 1 \\ 2\sum \sigma_i^2 \sigma_{i+k}^2 / \sum \sigma_i^4 & \text{if } k = l \neq 1 \\ 0 & \text{if } k \neq l \end{cases}$$

Part ii:

$$\begin{aligned} Cov[\Delta X' J_k \Delta \varepsilon, \Delta X' J_l \Delta \varepsilon] &= E[\Delta X' J_k T E[UU'|X] T' J_l' \Delta X] \\ &= \sigma_v^2 E[\Delta X' J_k T T' J_l \Delta X] \\ &= \sigma_v^2 \sum \sigma_i^2 (J_k T T' J_l)_{i,i} \end{aligned}$$

where $(\cdot)_{i,i}$ is the $i$-th diagonal element of the matrix.

The matrix $B$ is diagonal up to the $q$-th subdiagonals, which means that when $s > q$, the

$(k, k+s)$ element of matrix $B$ is zero and when $s \leq q$, the $(k, k+s)$ element of matrix $B$ is:

$$B(k, k+s) = \frac{1}{IV_T} \left( \varphi_s (\sum_{i=1}^{k} (\sigma_i^2 + \sigma_{N+1-i}^2) + (\varphi_s + \varphi_{2k+s}) \sum_{i=k+1}^{k+s} (\sigma_i^2 + \sigma_{N+1-i}^2) \right.$$
$$\left. +2(\varphi_s + \varphi_{2k+s}) \sum_{i=k+s+1}^{N-k-s} \sigma_i^2 \right)$$

Note that $\varphi_k$ is zero when $k > q$ as defined in equation (15).

Part iii:

$$\begin{aligned} & Cov[\Delta\varepsilon' J_k \Delta\varepsilon, \Delta\varepsilon' J_l \Delta\varepsilon] = Cov[U' \overbrace{T' J_k T}^{M_k} U, U' \overbrace{T' J_l T}^{M_l} U] \\ & = Cov\left[ \sum M_k(i,i) U_i^2 + 2 \sum_{i<j} M_k(i,j) U_i U_j, \sum M_l(i,i) U_i^2 + 2 \sum_{i<j} M_l(i,j) U_i U_j \right] \\ & = V[v^2] \underbrace{\sum M_k(i,i) M_l(i,i)}_{N \cdot E + F} + 4\sigma_v^4 \underbrace{\sum_{i<j} M_k(i,j) M_l(i,j)}_{N \cdot C + D} \end{aligned} \tag{35}$$

The summations are difficult to write out explicitly. They both include a dominant term ($N$ times $C$ or $E$) and a reminder term ($D$ or $F$). The matrix $C$ and $E$ can be expressed as following:

On the edge of the $C$ matrix,

$$C(0, s) = \begin{cases} \sum_{i=-q}^{-1} \varphi_i^2 & \text{if } s = 0 \\ \sum_{i=-q}^{-1} \varphi_i (\varphi_{i+s} + \varphi_{i-s}) & \text{if } 0 < s \leq 2q \\ 0 & \text{if } s > 2q \end{cases}$$

In the interior of matrix $C$, $(k > 0)$

$$C(k, k+s) = \sum_{i=-q}^{k-1} (\varphi_i + \varphi_{i-2k})(\varphi_{i+s} + \varphi_{i-2k-s})$$

When $k > q$, the above equation is simply:

$$C(k, k+s) = \begin{cases} 0 & \text{if } s > 2q \\ \psi_s & \text{if } 0 \leq s \leq 2q \end{cases}$$

where $\psi_s$ is defined in equation (28).

The matrix $E$ has non-zero elements only in the upper left block $E_{11}$,

$$E_{11} = \begin{bmatrix} \varphi_0 & 2\varphi_1 & 2\varphi_2 & \dots & 2\varphi_q \end{bmatrix} \cdot \begin{bmatrix} \varphi_0 & 2\varphi_1 & 2\varphi_2 & \dots & 2\varphi_q \end{bmatrix}'$$

□

# References


Aït-Sahalia, Y., P. Mykland, and L. Zhang (2011): "Ultra high frequency volatility estimation with dependent microstructure noise," *Journal of Econometrics*, 160(1), 160–175.

Aït-Sahalia, Y., P. A. 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.

Bandi, F. M., and J. R. Russell (2008): "Microstructure noise, realized variance, and optimal sampling," *The Review of Economic Studies*, 75(2), 339–369.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2008): "Designing realized kernels to measure the ex post variation of equity prices in the presence of noise," *Econometrica*, 76(6), 1481–1536.

——— (2009): "Realized kernels in practice: Trades and quotes," *The Econometrics Journal*, 12(3), C1–C32.

Barndorff-Nielsen, O., and N. Shephard (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 253–280.

Delattre, S., and J. Jacod (1997): "A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors," *Bernoulli*, 3(1), 1–28.

Gloter, A., and J. Jacod (2001a): "Diffusions with measurement errors. I. Local asymptotic normality," *ESAIM: Probability and Statistics*, 5(1), 225–242.

——— (2001b): "Diffusions with measurement errors. II. Optimal estimators," *ESAIM: Probability and Statistics*, 5(1), 243–260.

Gottlieb, G., and A. Kalay (1985): "Implications of the discreteness of observed stock prices," *The Journal of Finance*, 40(1), 135–153.

Hansen, P., and A. Lunde (2006): "Realized variance and market microstructure noise," *Journal of Business and Economic Statistics*, 24(2), 127–161.

Jacod, J., Y. Li, P. Mykland, M. Podolskij, and M. Vetter (2009): "Microstructure noise in the continuous case: the pre-averaging approach," *Stochastic Processes and their Applications*, 119(7), 2249–2276.

Li, Y., and P. A. Mykland (2007): "Are volatility estimators robust with respect to modeling assumptions?," *Bernoulli*, 13(3), 601–622.

——— (2012): "Rounding Errors and Volatility Estimation," Discussion paper, Working paper.

Madhavan, A. (2000): "Market microstructure: A survey," *Journal of Financial Markets*, 3(3), 205–258.

Sun, Y. (2006): "Best quadratic unbiased estimators of integrated variance in the presence of market microstructure noise," *UCSD, Department of Economics. Manuscript.*

Zhang, L. (2006): "Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach," *BERNOULLI-LONDON-*, 12(6), 1019.

Zhang, L., P. Mykland, and Y. Ait-Sahalia (2005): "A tale of two time scales," *Journal of the American Statistical Association*, 100(472), 1394–1411.

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