Realized Kernels with Moving Average Noises and Optimal Weights

Guansong Wang

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Economics.

Chapel Hill
2013

Approved by:

Eric Renault

Abstract

GUANSONG WANG: Realized Kernels with Moving Average Noises and Optimal Weights
(Under the direction of Eric Renault)

This paper studies the effect of the time dependence in micro-structure noise on the finite sample properties of the realized kernel estimator of the integrated variation of an asset price series using high frequency sample. A bias correction is proposed to eliminate the extra bias brought from the MA noises in the RK estimator. With the control variable interpretation of the RK estimator the optimal weights and its feasible approximation can be constructed under the MA noise assumption.

1 Introduction

The integrated variation, 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 sum of returns within the period of interest. However the real observations are known to be contaminated by the micro-structure noises which introduce a bias term dominating the information of interest into the estimator. Several consistent estimators of the integrated variation using noisy observations have been developed in the literature. Some achieve the optimal convergence rate ($N^{1/4}$) as proved in Gloter and Jacod (2001a,b).

Among those methods, the Realized Kernel (RK) estimator designed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) is of great interest. First, different kernel functions can be applied to obtain different asymptotic properties. The estimator using Tukey-Hanning kernel almost achieves the semi-parametric lower bound. Second, other estimators such as Two-Scale estimator by Zhang, Mykland, and Ait-Sahalia (2005), Multiple-Scale estimator by Zhang (2006), and Pre-Average estimator by Jacod, Li, Mykland, Podolskij, and Vetter (2009) can be associated with the realized kernel estimators by applying corresponding kernel functions. Third, sacrificing the convergence rate, the RK estimators can possess other advantages. With a slower convergence rate of $N^{1/5}$, the RK estimators can ensure positivity in probability. With a convergence rate of $N^{1/6}$, its asymptotic is robust with the violation of the white noise assumption.

For the simplicity, most of the literature assumes the noise process to have no auto-correlations which may be problematic when trying to utilize tick-by-tick sample sets. Hansen and Lunde (2006) gave warning about the time-dependent noise along with other ugly facts in high frequency sample. Recently Aït-Sahalia, Mykland, and Zhang (2011) also found evidence against the simple assumption. A study of Dow Jones Industrial Average component stocks in 2011 agrees with the above finding. An example of stocks AA and IBM is shown in Section 3.

In this paper, I adopt a moving average micro-structure noise assumption and study its effect on the RK estimator. Under the new assumption the RK estimator is still consistent but biased in finite sample. The bias can be either negative or positive. Parzen kernel with slower convergence rate is robust to moving average noises due to its greater optimal bandwidth increasing at the rate of $N^{3/5}$.

To overcome the finite sample bias one can choose to sample less frequently or to ignore the bias using Parzen kernel. Alternatively a bias correction of the RK estimator is proposed. It simply put weight 1 on the first q autocovariations to fully eliminate the noises following MA(q) process. This can also be interpreted by the control variable approach. The biased corrected realized variation RV_q is used as an unbiased estimator and the autocovariations with higher orders (greater than q) are used as control variables to reduce variance. The optimal weights for the control variables, as mentioned in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), can be constructed with the knowledge of the data generating process and a feasible approximation is available using preliminary estimates of the integrated variation and autocovariations.

The simulation experiment confirms the significant improvement of the bias correction over the conventional RK estimator. The RK estimator with Parzen kernel is relatively robust to the MA noises while it still benefit from the bias correction when using highest frequency samples. The RK estimator with optimal weights performs better than other kernels with small bandwidths. Compared to the RK estimator, the bias corrected RV estimator is fairly accurate and can be a better preliminary estimator than the sparsely sampled RV.

The structure of the rest of the paper is as follows. Section 2 give a brief review of the model and the design of the Realized Kernel estimator. In Section 3 an empirical examination on the stock AA and IBM suggests the noises are autocorrelated. With the MA noise assumption the semi-parametric lower bound of the estimator can be obtained numerically. Then I show the bias in the RK estimator and propose a bias correction.

In Section 4 the RK estimator and bias correction is interpreted by the control variable approach from which the optimal weights can be constructed. The results of a Monte Carlo experiment and an empirical application are summarized in Section 5 and Section 6. At last Section 7 concludes.

2 Realized Kernel

2.1 Model

We assume the underlying efficient log-price process $\{X_t\}_{t\geq 0}$ of an asset is a semi-martingale process in a probability space $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$:

$$dX_t = \mu_t dt + \sigma_t dW_t \tag{1}$$

where the stochastic processes μ_t and σ_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 μ_t is set to be zero through out 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 observations are maken on a time grid $\{0 = t_0 < t_1 < \cdots < t_N = T\}$. There are different ways to construct observation time grids. Most of the literature assume calendar time sampling (CTS) such that each interval between two observations is of the same length such as one second, one minute or twenty minutes. Tick time sampling (TTS) is to sample every fixed number of trades or quotes. If the sampling intervals are chosen to equally divide the IV_T it is called business time sampling (BTS). Even though the BTS is not feasible, empirical results suggest that TTS may approximate BTS. For more on sampling schemes, see Hansen and Lunde (2006). The RK estimator is designed under the CTS however it is shown to be robust under stochastic sampling. Further in Section 4 the control variable interpretation of the RK estimator naturally adapts to the TTS.

It is well known that the realized variation (RV) or quadratic variation of X converges to IV_T as the mesh of the partition $(\sup_i(t_i - t_{i-1}))$ diminishes to zero when N rises to infinity:

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

The convergence is at the rate $\sqrt{N}$ and the asymptotic distribution has been derived in Barndorff-Nielsen and Shephard (2002). It suggests that one should sample as frequently as possible to efficiently utilize high frequency data sets.

Nevertheless sampling too frequently may not be a wise choice in practice when the observed log prices are embedded with micro-structure noises. Denote $\{\varepsilon_t\}_{t\geq 0}$ as the noise process, the observable process $\{Y_t\}_{t\geq 0}$ is:

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

The term ε_t summarizes various micro-structure noises due to the trading scheme of the stock market. Madhavan (2000) provided a comprehensive survey of how price formation and information distort the observed log-prices. Aït-Sahalia, Mykland, and Zhang (2011) pointed out that even though some market frictions such like discreteness of prices, bid-ask bounces, and different market trades have been reduced by the decimalization and electronic trading, some market frictions still linger such as delayed trade reports. As the sampling grid becomes finer the magnitude of the noise dominates the magnitude of the realized variation. Consequently RV_T^N with high frequency data is severely biased.

For the convenience to design consistent estimators most literature including the

Realized Kernel estimators impose the assumption that the noise process is uncorrelated with the efficient log price process and has no serial dependence. The estimators are then generalized to more realistic circumstances. Formally, the assumptions are summarized below:

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.*

Under the above assumptions, the observed log-returns $\Delta Y_i = Y_{t_i} - Y_{t_{i-1}}$ are composed of the latent log-returns ΔX_i and the first difference of the noises $\Delta \varepsilon_i$ which are an MA(1) process. The goal of RK estimators is to estimate IV_T of the log-price process X using the observed log-returns $\{\Delta Y_i\}_{i=1}^N$.

2.2 RK Estimator

This section summarizes the Realized Kernel estimator of integrated variation under the martingale difference sequence micro-structure noises. The RK estimator utilizes the first realized autocovariation to eliminate the bias from the noise in the realized variation and higher order realized autocovariations to reduce the variance of the estimator. The realized autocovariations are weighted by a kernel function. RK estimators with different kernel functions have different asymptotic properties and can relate to other consistent estimators in the literature.

Formally the RK estimator is designed as the following:

$$RK_k(H) = \gamma_0(Y) + \sum_{h=1}^{H} k(\frac{h-1}{H}) \left(\gamma_h(Y) + \gamma_{-h}(Y)\right) \tag{5}$$

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:

$$\gamma_h(Y) = \sum_{i=1}^{N} \Delta Y_i \Delta Y_{i+h}, \text{ for } h = -H, \ldots, -1, 0, 1, \ldots, H \tag{6}$$

Note that γ_0 is the realized variation of the process Y. In practice where out of sample observations are unavailable, the summations in the γ_h are taken from $i = H + 1$ to $i = N - H$ and then the estimator is scaled up accordingly.

The RK estimator defined in Equation (5) is "flat-topped" as the weight on the γ_1 and γ_{-1} is $k(0) = 1$ instead of $k(1/H)$. It is designed this way to correct the bias in the realized variation under the m.d.s. noise assumption. Conditional on the volatility path, the expectation of γ_0 is $IV_T + 2N\sigma_\varepsilon^2$, the expectation of γ_1 and γ_{-1} is $-N\sigma_\varepsilon^2$, and γ_h with $|h| > 1$ has zero expectation. Therefore the unit weight on the first realized autocovariation balances out the bias in the realized variation due to the noise. For non "flat-topped" RK estimator the weight on γ_h is $k(\frac{h}{H})$ rather than $k(\frac{h-1}{H})$ so that it generally cannot guarantee unbiasedness. In order to obtain consistency for non "flat-topped" RK estimator the weight on γ_1 and γ_{-1} has to be close enough to one, that is, the first derivative of the kernel function at point zero has to close enough to zero ($k'(0) \approx 0$).

For the consistency the bandwidth H is proportional to N^η, where η depends on the kernel function. The asymptotic variance of the RK estimator is determined by the characteristics of the particular kernel function, the chosen bandwidth, and the parameters of the data generating process such as the noise-to-signal ratio ξ^2 and the measure of heteroskedasticity ρ:

$$\xi^2 = \frac{\sigma_\varepsilon^2}{\sqrt{T\int_0^T \sigma_t^4 dt}} \text{ and } \rho = \frac{\int_0^T \sigma_t^2 dt}{\sqrt{(T\int_0^T \sigma_t^4 dt)}} \tag{7}$$

For the chosen kernel function, optimal bandwidth H^* can be obtained by minimizing the asymptotic variance.

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	$\begin{cases} 1-6x^2+6x^3 & 0 \le x \le 1/2 \\ 2(1-x)^3 & 1/2 \le x \le 1 \end{cases}$	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 α 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. For the kernel functions with 1/4 convergence rate the scalars also depend on ρ as in Equation (7) and are computed as $\rho = 1$.

A few interesting kernel functions are list in Table 1. The convergence rate is the α such that $N^{\alpha}(RK_k(H) - IV_T)$ has an asymptotic distribution as 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 lower bound as p increases. The Parzen kernel function can also achieve the optimal convergence rate when the bandwidth is proportional to $N^{1/2}$. However it is more attractive to be used as a non "flat-topped" kernel paired with a bandwidth proportional to $N^{3/5}$ and a slower convergence rate since it guarantees positivity of the RK estimator.

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 optimal bandwidth is also positively related to the noise-to-signal ratio ξ^2.

3 Time Dependent Noise

An attractive feature of RK estimator is that it is robust against scenarios violating Assumption 1 such as stochastically sampling, endogenous noise and serial dependent noise, the last of which this section is focused on. The no autocorrelation assumption has already been questioned by a few authors. In the following subsections I confirm their studies using the stock prices of AA and IBM in the year of 2011 and propose a numeric method to compute the semi-parametric lower bound of IV estimators with the serial dependent noises. A bias correction can improve the accuracy of the RK estimator in finite sample.

3.1 Empirical Evidence

Hansen and Lunde (2006) demonstrated 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. For illustration, define the RV estimators:

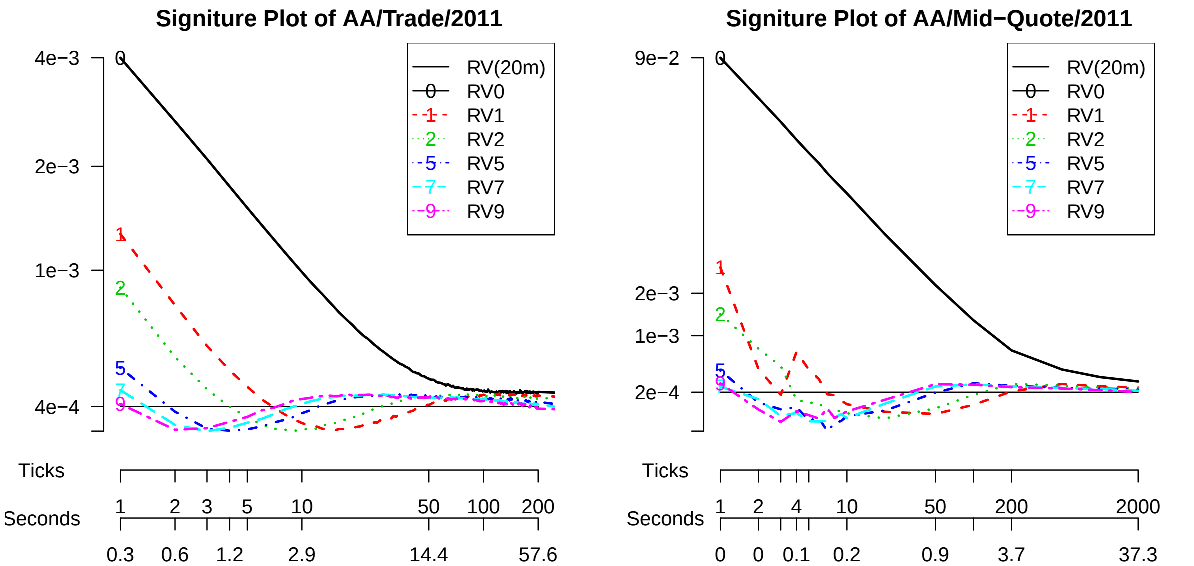
$$RV_H = \gamma_{-H}(Y) + \gamma_{-H+1}(Y) + \cdots + \gamma_{H-1}(Y) + \gamma_H(Y) \tag{8}$$

where $\gamma_h(Y)$ is the realized variation defined in Equation (6) and note that $\gamma_0(Y)$ is the realized variation. It is easy to show that under Assumption 1:

$$E[\gamma_h] = \begin{cases} IV_T + 2N\sigma_\varepsilon^2 & h = 0 \\ -N\sigma_\varepsilon^2 & |h| = 1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

As a result the realized variation RV_0 suffers from an upwards bias which increases as the observations are sampled more frequently. Therefore the signature plot of RV_0 should be a downward sloping curve converging to the real level as the dataset is sampled more sparsely. Under Assumption 1 the RV_1 or any RV_h for $h > 0$ is unbiased. The RV_1, which can be seen as the RK estimator with bandwidth 1, is indeed the estimator introduced by Zhou (1996). Therefore the signature plot of RV_h for $h > 1$ should be around the real level regardless of the sampling frequency.

Figure 1: Signature Plots of AA in 2011



The left figure is for transcation data and the right figure is for mid-quote data. The horizontal line is the level of the averaged daily sparse sampled (20 minutes) RV in the year which is at around 0.0004. Each line represents the averaged estimated integrated variation with different sampling frequencies. RV_h is the realized variation estimator defined in Equation (8). There are two X-axes: ticks and corresponding seconds of the subsampling interval in average.

Figure 1 shows the signature plots of the stock AA in the year 2011. The X-axis denotes the length of the sampling frequency by ticks and seconds and the rightside represents lower frequency. The horizontal line at the level around 0.0004 is the averaged realized variation with 20 minutes log-returns of the 251 full trading days, which is used as a proxy of the real average daily IV.

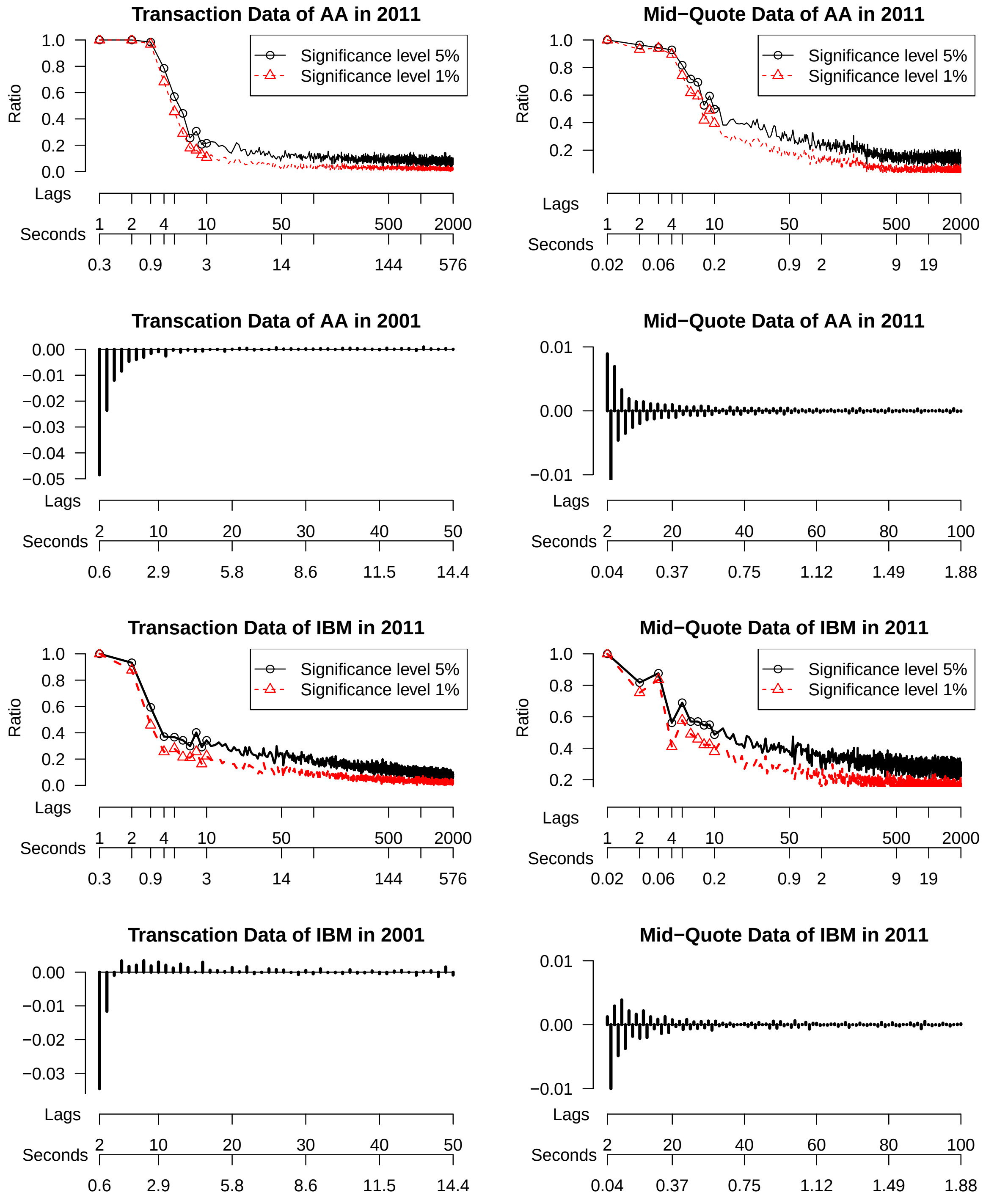
As predicted, RV_0 decreases to the sparsed sampled average level as the log-returns are sampled less frequently. The fact that the signature plots for RV_1 and RV_2 are still significantly upwards biased in high frequency samples suggests the assumption of non-autocorrelation is likely to be false. As higher orders of γ added into the RV estimator the bias is reduced. The signature plots for RV_5, RV_7 and RV_9 are close to the horizontal line at all frequencies. Hansen and Lunde (2006) found the time dependence in the stock AA lasts for 2 minutes or 30 ticks in 2000 and 2004 and Figure 1 suggests that the it lasts for 9 ticks in 2011 since RV_9 appears to be unbiased.

Figure 2 illustrates the auto-correlation functions of stock AA and IBM in 2011. The first and third rows are the ratio of days with significant autocorrelations at 5% and 1% confidence levels and the second and fourth rows are the averaged autocorrelations. The first autocorrelations are all significant without exception. The mid-quote data has more time dependence than the transaction data in the sense that the ratio of significant days is higher for most lags. For AA the time dependence seems to exist up to 10 lags which coincides with the pattern in Figure 1. The average ACF plots exclude the first autocorrelations since they are of much greater magnitude than the rest. The first autocorrelations for transaction and mid-quote data of AA are -0.34 and -0.50 and those for transaction and mid-quote data of IBM are -0.30 and -0.49. Even though with a high ratio of significance the higher order autocorrelations are of quite small magnitude. The ACF of transaction data resembles a moving average process while the ACF of mid-quote data is more likely an auto-regressive process particularly for AA. Aït-Sahalia, Mykland, and Zhang (2011) also showed 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.

3.2 Moving Average Assumption

Even though the magnitude of the autocorrelation is small, a large sample size makes its effect on the IV estimator nontrivial . We model the noise process by a moving average structure because the autocorrelation disappears quickly and the MA structure is easier to relate to existing theories and estimators. The assumption of time-dependent noise is given below.

Figure 2: Percentage of Days with Significant ACF in 2011



The top two panels are for stock "AA" and the bottom two panels are for stock "IBM". The left panels are for transaction prices and the right panels are for mid-quote series. The first the third panels are the precentage of days with (5% and 1%) significant ACF of log-returns for lags from 1 to 2000. The second and fourth panels are the yearly averaged ACF of log-returns. There are two X-axes: lags of ACF and corresponding seconds of sampling interval in average.

Assumption 2. *We assume the noise ε is a MA$(q-1)$ process.*

$$\varepsilon_i = \upsilon_i + \beta_1 \upsilon_{i-1} + \beta_2 \upsilon_{i-1} + \cdots + \beta_{q-1} \upsilon_{i-q+1} \tag{10}$$

where υ is a martingale difference sequence process, and $E[\upsilon^2] = \sigma_\upsilon^2$.

The noise process is modeled by MA$(q-1)$ so that the noise component $\Delta\varepsilon_i$ in log return ΔY_i follows a MA(q) process.

$$\begin{aligned}
\Delta\varepsilon_i &= \varepsilon_i - \varepsilon_{i-1} \\
&= \upsilon_i + (\beta_1 - 1)\upsilon_{i-1} + \cdots + (\beta_{q-1} - \beta_{q-2})\upsilon_{i-q+1} - \beta_{q-1}\upsilon_{i-q} \\
&= \theta_0 \upsilon_i + \theta_1 \upsilon_{i-1} + \theta_2 \upsilon_{i-2} + \cdots + \theta_q \upsilon_{i-q}
\end{aligned} \tag{11}$$

We rewrite $\Delta\varepsilon$ using parameters $\{\theta_0, \ldots, \theta_q\}$ in the last step in Equation (11) for simplicity of notations. The parameters θ 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.

Given that $\Delta\varepsilon$ follows an MA(q) process, the covariances of the observed log returns ΔY_i are:

$$Cov[\Delta Y_i, \Delta Y_{i+k}] = \begin{cases} \sigma_i^2 + \varphi_k \sigma_\upsilon^2 & \text{if } k = 0 \\ \varphi_k \sigma_\upsilon^2 & \text{if } k = 1, 2, \ldots \end{cases} \tag{12}$$

$$\text{where } \sigma_i^2 = \int_{t_{i-1}}^{t_i} \sigma_\tau^2 d\tau \text{ and } \varphi_k = \begin{cases} 0 & \text{if } |k| > q \\ \sum_{\tau=0}^{q-|k|} \theta_\tau \theta_{\tau+|k|} & \text{if } |k| \leq q \end{cases} \tag{13}$$

Note that when $q = 1$ Assumption 2 is exactly the same as Assumption 1.

3.3 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 the IV_T estimators. Here we will extend the result to the model with moving average noise.

First denote Φ as the $q \times 1$ vector $(\varphi_0, \varphi_1, \ldots, \varphi_q)'$, and $C^N(\sigma^2, \sigma_\upsilon^2, \Phi)$ 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:

$$C^N(\sigma^2, \sigma_\upsilon^2, \Phi)_{i,j} = \begin{cases} \frac{\sigma^2}{N} + \varphi_0 \sigma_\upsilon^2 & \text{if } i = j \\ \varphi_k \sigma_\upsilon^2 & \text{if } |i - j| = k \le q \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

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

Theorem 3.1. *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 ε satisfies Assumption 2. Denote $\lambda_i(\sigma^2, \sigma_\upsilon^2, \Phi)$ as the eigenvalues of matrix $C^N(\sigma^2, \sigma_\upsilon^2, \Phi)$. If for any sequence $\{h_n\}$ with a limit h, the following two conditions are satisfied:*

$$\sup_{1 \le i \le N} |\delta_i^N| \to 0 \qquad\qquad \frac{1}{2h^2} \sum_{i=1}^{N} (\delta_i^N)^2 \to I(\sigma^2, \Phi) \tag{15}$$

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

$$N^{\frac{1}{4}}(\hat{\sigma}^2 - \sigma^2) \to_d \mathcal{N}(0, \sigma_\upsilon I(\sigma^2, \Phi)^{-1}) \tag{16}$$

When q is equal to one , the model degenerates to the one with white noise assumption and the matrix C^N is tri-diagonal [1] 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 (15). In that case, Equation (16) becomes:

$$N^{1/4}(\hat{\sigma}^2 - \sigma^2) \to \mathcal{N}(0, 8\sigma_v \sigma^3) \tag{17}$$

The asymptotic variance $8\sigma_v\sigma^3$ has been known as the semi-parametric lower bound for the consistent IV estimators. While with q greater than one, there is no closed form solution to the eigenvalues of matrix C^N. Consequently, we are unable to write out the explicit expression of $I(\sigma^2, \Phi)$ but only to approximate the lower bounds numerically.

As an example, Table 2 shows the approximated lower bounds of the model under Assumption 2 and $q = 2$ which are calculated with the dimension of C^N as 1000, 5000, and 10000. Similarly as the case when $q = 1$, the inverse of the Fisher information is proportional to σ^3 and the factor is determined by Φ. 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 (17).

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

β_1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$N\backslash$ θ_1	-1	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0
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 parameters of the ε process are $\beta_0 = 1$ and β_1 in the first row. The parameters of the $\Delta\varepsilon$ process are $\theta_0 = 1$, θ_1 in the first row and $\theta_2 = -1 - \theta_1$. The lower bounds are in terms of $\sigma^2\sigma_v$.

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

3.4 Bias Corrected Realized Kernel

The RK estimator is initially designed under the white noise assumption the estimator still has the same optimal convergence rate with time-dependent noise but different asymptotic variance. However depending on the autocorrelations of the log-returns the RK estimator can be severely biased either upwards or downwards in rather large finite samples.

Theorem 3.2. *Suppose the observed process Y is defined as in Equation* (4) *where X is defined as in Equation* (1) *and ε satisfies Assumption 2. The expectation of γ_h is:*

$$E[\gamma_k] = \begin{cases} IV_T + N\varphi_0\sigma_v^2 & \text{if } k = 0 \\ (N-k)\varphi_k\sigma_v^2 & \text{if } 0 < k \le q \\ 0 & \text{if } k > q \end{cases} \tag{18}$$

where γ_k is defined as $\sum_{i=1}^{N-k} \Delta Y_i \Delta Y_{i+k}$ and φ's are defined in Equation (13)*.*

The γ_k defined above is different from in Equation (6) 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.

In the case of Assumption 1 where q is equal to one, there are only two realized autocovariations with non-zero expectations and $(2, -1)'$ is the only possible value of $(\varphi_0, \varphi_1)'$ because of the restriction of the assumption. Therefore a flat-top kernel function ($k(0) = 1$) guarantees unbiasedness.

$$\begin{aligned} E[RK(H)] &= E[\gamma_0] + k(0)\frac{2N}{N-1}E[\gamma_1] \\ &= IV_T + (\varphi_0 + 2\varphi_1 k(0))\sigma_v^2 = IV_T, \text{ if } k(0) = 1 \end{aligned}$$

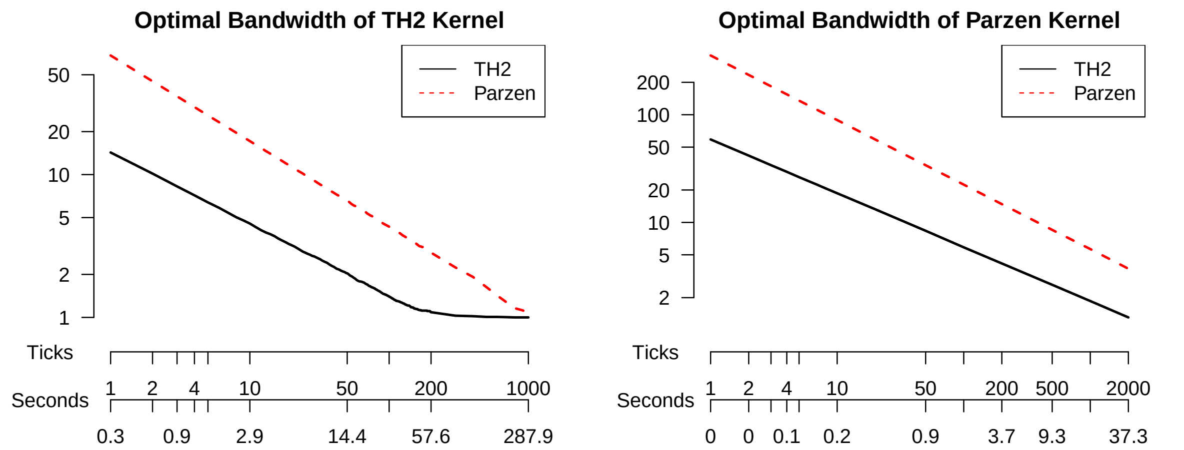
When q is greater than one, $(\varphi_0, \varphi_1)'$ has other possible values and φ_h is nonzero up to $h = q$. The expectation of $RK(H)$ becomes more involved that the flat-top kernel function cannot be certain to balance out the bias in realized variation with the first realized autocovariation. In general, the conventional realized kernel estimators are biased in finite sample with time-dependent noise. The bias can be either upwards or downwards depending on the parameters of the moving average noise process. This also provides another explanation to the negative bias in signature plots (Figure 1) other than the correlation of the latent process X and the noise process as mentioned in Hansen and Lunde (2006).

The first solution to the bias problem is not to sample at the highest frequency available. In Figure 1 the signature plots converge after 50 ticks (14 seconds) for transaction data and 200 ticks (3.7 seconds) for mid-quote data. Hence sparsely sampling the log-returns will mostly eliminate the time dependence in the noises. This solution is sub-optimal since it cannot fully benefit from the high frequency records and there lacks a standard procedure to find the suitable sampling frequency.

The second solution is to apply the conventional RK estimator without modification for the time dependence noises, because the RK estimator is still consistent and fast convergent. The key of that is the fact that the bandwidth goes to infinity as the sample size goes to infinity and the kernel function is flat enough near zero ($k'(0) \approx 0$). As long as q is finite the first q value of the kernel function will eventually be close to one enough to correct the bias as N and H go to infinity. In Figure 3 it is clear the bandwidth of Parzen kernel is significantly greater than that of Tukey-Hanning$_2$ kernel at highest sampling frequencies. It adds another reason to prefer Parzen kernel with slower convergence rate over kernel functions with optimal convergence rate such like the Tukey-Hanning kernel. This is also mentioned in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009).

Under the circumstance of a rather small noise-to-signal ratio ξ the bandwidth of the Parzen kernel may not be large enough to balance the bias. The third solution progressively modifies the kernel function to re-establish unbiasedness by increase the

Figure 3: Signature Plots for Optimal Bandwidths of AA



The signature plots are for the averaged optimal bandwidth of Tukey-Hanning$_2$ kernel and Parzen (slow convergence) kernel of stock "AA" in 2011. The left panels are for trade data and the right panels are for mid-quote data.

length of the "flat-top" part of the kernel to q. 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 $k(\cdot)$ and γ's are the same as in Equation (5) and Theorem 3.2.

Corollary 3.2.1. *Assume Y is modeled as in Theorem 3.2 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* (5).

The proof of Theorem 3.2 is shown in Appendix. The corollary is straightforward. For unbiasedness, we simply apply the theorem and notice that since $\sum_i \theta_i = 0$

$$\varphi_0 + 2\varphi_1 + \dots + 2\varphi_q = (\theta_0 + \dots \theta_q)^2 = 0 \tag{20}$$

The coefficients on the expectation brought by the noise (σ_v^2) have sum zero when the first q realized autocovariations have weight one.

In practice, the order of moving average of the noise return is unknown. It can be determined 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.

4 Optimal Weights

In this section, we will interpret the RK estimator with control variables approach and obtain the optimal weights under Assumption 2. The basic idea of control variables is reviewed below.

In general, 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 and correlated with $g(Y)$. It is possible to construct a new unbiased estimator $g(Y) + \alpha' X$, where α is a $m \times 1$ weighting vector. With a proper choice of α, 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 α can be solved by minimizing the variance of the new estimator, with the knowledge of the probability distribution of Y and X. The optimal weighting α^* is solved as:

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

The RK estimator is in fact motivated by the control variable approach. $RK(1)$ is the same unbiased but inconsistent estimator 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. Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) applied the optimal weights under Assumption 1 as a comparison to other kernel functions. In this section, I construct the optimal weights under Assumption 2 and show a feasible approximation in practice.

4.1 Optimal RK Weights

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

$$\Gamma_H = \begin{bmatrix} \Gamma_{H,1} \\ \Gamma_{H,2} \end{bmatrix} \text{ where } \Gamma_{H,1} = (\gamma_0, 2\gamma_1, \ldots, 2\gamma_q)', \text{ and } \Gamma_{H,2} = (2\gamma_{q+1}, \ldots, 2\gamma_{H+q}) \quad (21)$$

Each γ_h for $h > 0$ is multiplied by 2 to make the control variable weights consistent with the scale of other kernel functions and to simplify the notation in the proof of Theorem 4.1. Suppose the noise follows Assumption 2, Corollary 3.2.1 suggests that the unbiased estimator can be rewritten as $\iota'_{q+1}\Gamma_{H,1}$, where ι_{q+1} is a $(q+1)\times 1$ vector of ones.

The RK estimator with control variable kernel is defined as:

$$RK^q_{CV}(H) = \gamma_0 + 2\sum_{h=1}^{q}\gamma_h + 2\sum_{h=1}^{H}\alpha_h\gamma_{h+q} = \iota'_{q+1}\Gamma_{H,1} + \alpha'\Gamma_{H,2} \quad (22)$$

The weight α_h corresponds to $k(\frac{h-1}{H})$ of a kernel function. Note that the estimator defined above is almost unbiased. To completely eliminate the bias, γ_h need to be scaled by $N/(N-h)$ for $h \leq q$. However, the effect is negligible when the sampling frequency is high. For simplicity we used unscaled γ_h when deriving the variance-covariance matrix of Γ_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 $\Gamma_{H,1}$ and the covariance matrix between $\Gamma_{H,1}$ and $\Gamma_{H,2}$ when computing control variable weights.

To find the optimal weights, we need the variance covariance matrix of Γ_H:

Theorem 4.1. *Suppose the observed process Y is modeled as in Theorem 3.2. The* *variance covariance matrix of* Γ_H *is:*

$$\begin{aligned}\Sigma_H &= Var[\Gamma_H] \\ &= 2IQ_T/N \cdot A + 4\sigma_\upsilon^2 IV_T \cdot B + 4\sigma_\upsilon^4(N \cdot C + D) + V[\upsilon^2](N \cdot E + F)\end{aligned} \tag{23}$$

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 defined 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 with 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 Γ_H, while those in their paper are for asymptotic property; second, the γ's in this paper are defined without out of sample data.

To express the optimal control variable weights we partition the matrix Σ_H into a 2×2 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 $\alpha_H^* = (\alpha_1^*, \ldots, \alpha_H^*)'$ can be derived by:

$$\alpha_H^* = -(\Sigma_{H,22})^{-1}(\Sigma_{H,21} \cdot \iota_{q+1}) \tag{24}$$

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

4.2 Approximation

The exact optimal control variable weights α_H^* is not feasible in practice because it requires the knowledge of the data generating process to construct the matrix Σ_H in equation (23). Even though the values of IV_T and IQ_T can be replaced by some consistent estimators, the volatility path $\{\sigma_i\}$, the MA parameters Φ, and the variance of υ are unable to identify while necessary to construct the matrices. However we can 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 blocks of Σ_H in equation (24) so that $V[\upsilon^2]$ is irrelevant in computing α_H^*.

Secondly, only the edge areas of the matrices depend on the volatility path $\{\sigma_i\}$. When N is large we can approximate the matrices A, B and C by $\hat{A}$, $\hat{B}$ and $\hat{C}$ respectively. Remind that φ's are defined as in equation (13) and further we define:

$$\psi_k = \begin{cases} 0 & \text{if } k < 0 \text{ or } k > 2q \\ \sum_{j=-q}^{q-k} \varphi_{|j|}\varphi_{|j+k|} & \text{if } k = 0, 1, 2, \ldots, 2q \end{cases} \tag{25}$$

Matrix $\hat{A}$ is diagonal. 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 Φ.

$$\hat{A} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & 2 & \dots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 2 \end{bmatrix}$$

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

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

Thirdly, even though σ_v^2 and φ's can not be identified separately, their multiplications can be consistently estimated by:

$$\widehat{\sigma_v^2 \varphi_h} = \frac{1}{N-h}\gamma_h \to_p \sigma_v^2 \varphi_h \tag{26}$$

Denote $\widehat{\sigma_v^2 \Phi}$ as the vector $[\gamma_0/N, \dots, \gamma_q/(N-q)]'$. 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 (32) and $\widehat{\sigma_v^2 \Phi}$. [2]

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:

$$\hat{\alpha}_H^* = \left(2\frac{\hat{IQ}_T}{N}\hat{A}_{22} + 4\hat{IV}_T\hat{B}_{22}(\widehat{\sigma_v^2\Phi}) + 4N\hat{C}_{22}(\widehat{\sigma_v^2\Phi})\right)^{-1} \cdot \left(2\frac{\hat{IQ}_T}{N}\hat{A}_{21} + 4\hat{IV}_T\hat{B}_{21}(\widehat{\sigma_v^2\Phi}) + 4N\hat{C}_{21}(\widehat{\sigma_v^2\Phi})\right)\iota_{q+1} \tag{27}$$

The above equation is feasible. The realized variation (γ_0) with sparsely sampled log-returns or the RK estimator using 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 an almost constant volatility path ($\rho \approx 1$) $\hat{IV}_T^2$ can be used as an estimate. Both Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and Jacod, Li, Mykland, Podolskij, and Vetter (2009) showed 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 very large. If so preliminary estimates of IV_T and IQ_T are not required any more to construct optimal weights. The optimal bandwidth H is infinity by definition of the control variable approach because the more control variables the lower is the variance. However an overly large H is undesirable because inverting matrix is computation demanding. A rule of thumb is to use the same bandwidth as in TH_p kernel function with $p = 2$ (or $p = 16$ if H^* is not too large).

[2]The matrix M in equation (32) depends only on Φ. An estimate of $\sigma_v^4(N \cdot C + D)$ is obtained by the second summation in the last step while replacing Φ by $\widehat{\sigma_v^2\Phi}$ in matrix M. To include D in the approximation, replace the $N\hat{C}$ in equation (27) by the estimate of $\sigma_v^4(N \cdot C + D)$.

5 Simulation Study

In this section, I perform a Monte Carlo experiment to demonstrate the improvement of the bias correction and the RK estimator using optimal weights in the case that the noise process ε follows an MA(q) process.

I choose the RV estimator with 20 minutes log-returns as a preliminary IV_T estimator and the square of that RK estimator to approximately estimate IQ_T. Ignoring the MA noise structure the RK estimator applies $(\gamma_0 - \widehat{IV}_T)/(2N)$ as the estimator of the variance of noise. The RV estimator is defined as in Equation 8; the RK estimator is defined as in Equation 5. The Tukey-Hanning$_2$ kernel and the slow convergence Parzen kernel functions are used with their estimated optimal bandwidths. The optimal weights are computed with the same bandwidths of Tukey-Hanning$_2$ kernel.

The original RK estimator uses TH$_2$ kernel functions with $p = 2$ and $p = 16$. The optimal bandwidths are computed from preliminary estimates.

5.1 Data Generating Process

The data is generated to mimic the transaction data of AA in 2011. I simulate the log-prices for one day (6.5 hours) and the sample size for one days is 81,282 which is the average of AA in 2011. The volatility path σ_t is generated by the one-factor stochastic volatility model:

$$
\begin{aligned}
\sigma_t &= \sqrt{V} \cdot exp(\beta_0 + \beta_1 \tau_t) \\
d\tau_t &= \alpha \tau_t dt + dB_t
\end{aligned}
\tag{28}
$$

which is used in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). We choose the following values for the parameters to match their configuration:

$$\alpha = -0.025, \beta_1 = 0.125, \beta_0 = \beta_1^2/2\alpha, V = 4.03 \times 10^{-4}$$

V is the average of the RV of the 20 minutes log-returns. τ_0 is drawn from its stationary distribution $\mathcal{N}(0, -1/(2\alpha))$. The experiment is conditional on the volatility path, therefore there is only one volatility path in the simulation.

The noise return process $\Delta\varepsilon$ follows an MA(5) process and the parameters are fitted to mimic the autocorrelations of the stock AA in 2011. The parameters as in Equation 10 are:

$$\beta = (1.00, 0.35, 0.18, 0.09, 0.03)', \text{ and } \sigma_v^2 = 3.88 \times 10^{-8} \tag{29}$$

The latent log-returns are generated by the volatility process $\{\sigma_t\}$ and a simulated Wiener process. The latent log-prices are the cumulated sum of the log-returns with the initial value of log(16) which is the usual level of the stock AA. The observed log-prices are the sum of the noise process and the latent log-prices. The two Wiener processes generating the log-returns the stochastic volatility are uncorrelated. Finally the experiments are repeated 100,000 times.

5.2 Result

The results are summarized in Table 3. All the estimators are computed without any prior information and sampling at the highest frequency except the RV_{sparse} which is the RV estimator using 20 minutes log-returns. Different orders of bias correction q are performed for RV and RK estimators. The bias and RMSE are stated as the percentage ratio normalized by the true integrated variation of the process which is 5.03×10^{-4} in this experiment. The last column of each panel captures the ratio of squared bias in the total MSE. A high ratio implies the bias is mostly responsible for the estimation error.

The advantage of bias correction is obvious for all the estimators. Ignoring the time dependence of the noise the bias dominates the MSE. Parzen kernel has a greater optimal bandwidth than Tukey-Hanning$_2$ kernel and is much less affected by the MA noises. The

Table 3: Numerically Computed Lower Bounds of Model with MA(2) Noise

Realized Variations				RK Tukey-Hanning$_2$			
	Bias	RMSE	Bias^2/MSE	Average Bandwidth: 12.50			
RV_{sparse}	1.35	23.15	0.34	MA order	Bias	RMSE	Bias^2/MSE
RV_0	922.04	922.06	100.00				
RV_1	1086.50	1086.54	99.99	1	179.38	182.75	96.34
RV_3	228.99	229.14	99.86	3	29.72	30.75	93.42
RV_5	-0.00	8.43	0.00	5	-0.01	4.10	0.00
RV_7	-0.01	8.49	0.00	7	0.01	4.20	0.00
RV_9	-0.02	8.51	0.00	9	0.04	4.33	0.01
RK Optimal Weights				RK Parzen			
Same Bandwidth as RK TH$_2$				Average Bandwidth: 62.72			
MA order	Bias	RMSE	Bias^2/MSE	MA order	Bias	RMSE	Bias^2/MSE
1	264.36	267.59	97.60	1	6.11	7.07	74.51
3	71.45	71.64	99.48	3	0.84	3.45	5.87
5	0.09	3.17	0.07	5	-0.02	3.46	0.00
7	0.09	3.26	0.07	7	-0.02	3.60	0.00
9	0.09	3.41	0.07	9	-0.02	3.73	0.00

Bias and RMSE (root mean square error) are divided by the real value (5.0331×10^{-4}) and multiplied by 100. The ratio of Bias^2 and MSE measures the composition of MSE and is multiplied by 100. MA order is the applied order of bias correction (number of unit weights).
The top-left panel is for realized variations defined in Equation (8); the top-right panel is for RK estimators with Tukey-Hanning$_2$ kernel function; the bottom-left panel is for RK estimators with optimal weights; the bottom-right panel is for RK estimator with Parzen kernel ($N^{1/5}$ convergence rate).

optimal weight estimator has greater bias than Tukey-Hanning$_2$ kernel since it usually weights the first few γ's less than TH$_2$ kernel. When the time dependence of the noises is correctly specified ($q = 5$) the bias is eliminated for all the estimators. Overestimating the MA order of th noise does not have much affect the estimators except a slight increase of the RMSE. Therefore a safe strategy is to choose bias correction order aggressively.

After bias correction, the RK estimators have similar RMSE. The Parzen kernel with slower convergence rate has lower RMSE than the TH$_2$ kernel with optimal convergence rate. It suggests that the asymptotic of the RK estimators relies on the large bandwidth H more than the large sample size N. Nevertheless the optimal weights with the smaller bandwidth as the TH$_2$ kernel still has lower RMSE than the Parzen kernel which confirms its optimality. Considering computation cost the bias correction enables the TH_2 kernel to obtain the accuracy as the Parzen kernel with much less γ's required. With the extra cost of inverting a matrix of order $q+H^*$ the optimal weights can achieve better accuracy than the Parzen kernel.

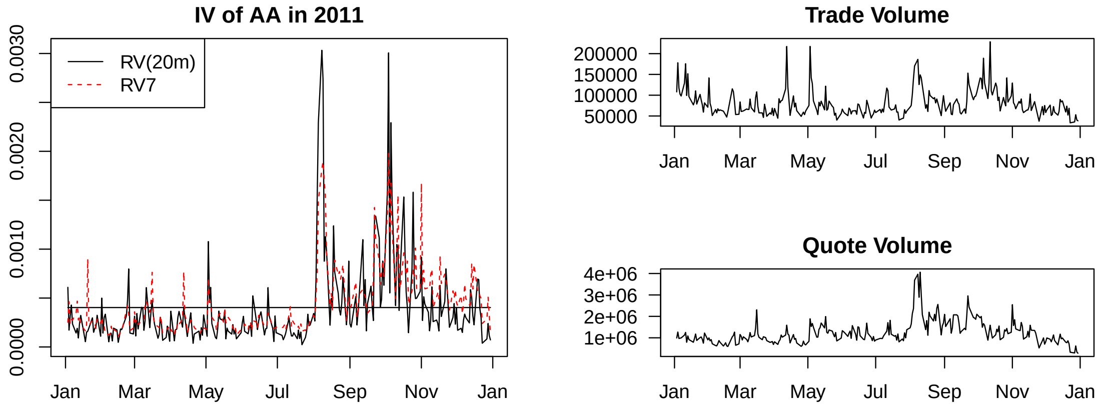
Interestingly the bias corrected RV estimator performs fairly well. It has smaller bias and lower RMSE than the sparsely sampled RV estimator which is widely used as a preliminary IV estimator. In this experiment the extra variance reduction of the RK estimator from the RV estimator is significant but mild. Therefore the biased corrected RV estimator may be a better preliminary IV estimator and a fairly accurate IV estimator which is easier to compute.

6 Empirical Application

In this section various IV estimators are applied to the high frequency transaction and mid-quote data of Alcoa Inc, which is traded with the ticker symbol "AA". It is the component of the Down Jones Industrial Average and is used as an example in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) and Hansen and Lunde (2006). The trading volume has spiked during the financial crisis and remains at a much higher level than the years analyzed in the papers mentioned above. The new milliseconds database in TAQ also makes the high frequency log-returns available in sub-seconds level.

The dataset of the year 2011 is extracted from the TAQ database in the Wharton Research Data Services. There are 251 full trading days in the year. For each trading day only the records within the open exchange time window and regular sale or quote conditions are kept. The detail of data cleaning procedure is summarized in Appendix A The time series of daily IV estimates and volumes are plotted in Figure 4. The 20 minutes RV and RV7 move similarly most of the time and the RV7 has milder jumps. Both the volume and volatility have three biggest jumps at May, August and October.

Figure 4: IV Estimate and Volume of AA in 2011



The horizontal line represents the yearly average of 20 minutes RV. The estimators use transaction data and the time series using mid-quote data will coincide with the curves. The volume are the number of records after data cleaning procedures.

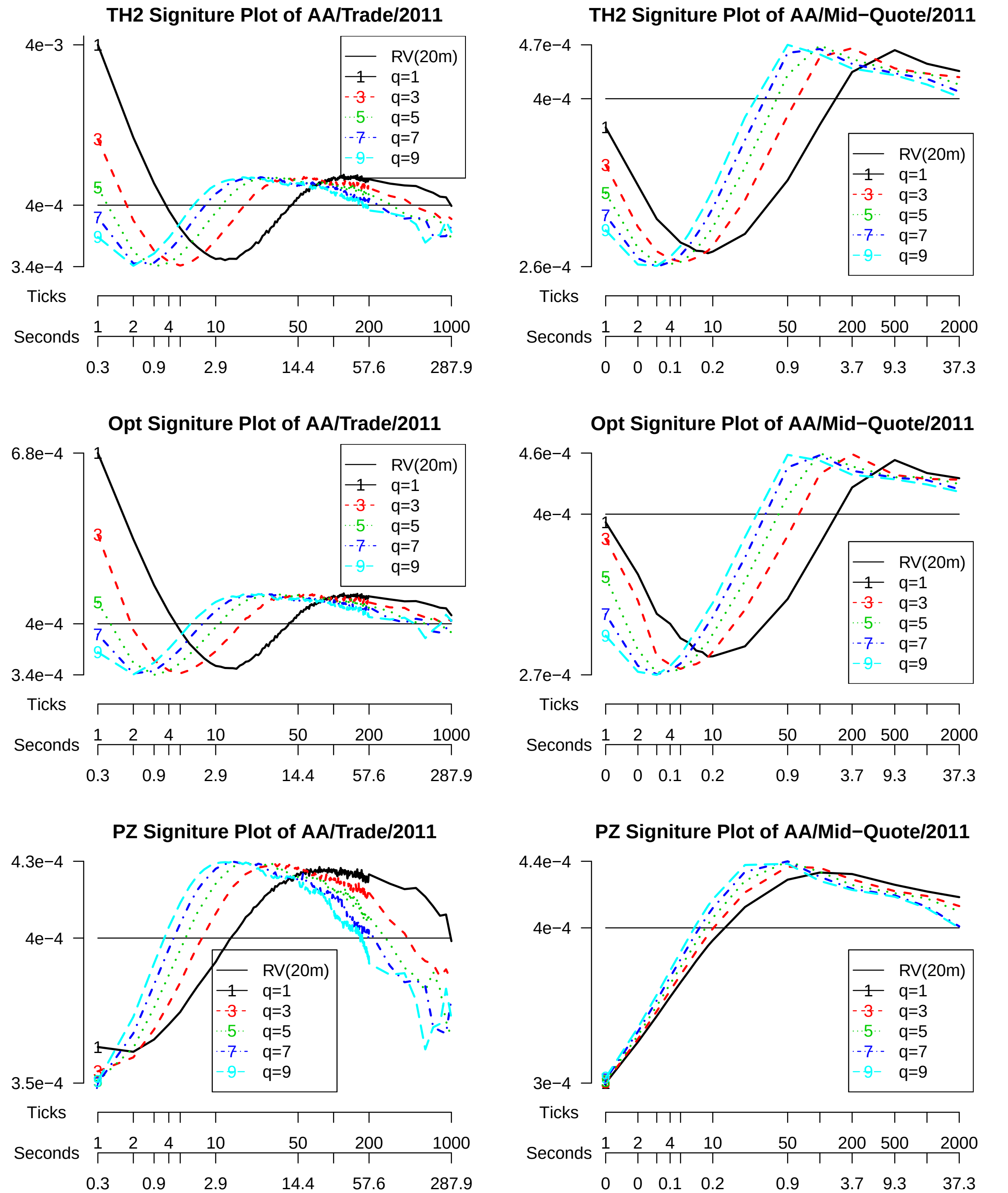
Figure 5 provides the signature plots of the RK estimators with various kernel functions and order of bias correction. The signature plots of the optimal bandwidths are given in Figure 3.

First thing to notice in the plots is the wave pattern shared by the RK estimators. As the sampling frequency decreases (right of the X-axis) the RK estimators first decline and then rise to the 20 minutes RV level. It may be caused by the fact that log-returns at different sampling frequencies have different time dependence structure and as the sampling frequency gets closer to 20 minutes the RK estimators will coincide with the RV estimator. The pattern is also shared with the bias corrected RV estimator shown in Figure 1 but of less magnitude.

Secondly, the RK estimator without bias correction can be very unreliable as seen in the transaction data. Note the Y-axis is taken logarithm the deviation is larger than it appears in the plots. This is not the case in the mid-quote data because the optimal bandwidth is large enough to reduce most of the bias.

Thirdly, the RK estimator with Parzen kernel is relatively robust to bias correction especially for mid-quote data. The reason is that the optimal bandwidth of the Parzen kernel is much greater than that of the TH_2 kernel. When applying a similarly large

Figure 5: Signature Plots for RK estimators of AA



The signature plots are for the averaged RK estimators of stock "AA" in 2011. The first row is for the Tukey-Hanning$_2$ kernel function; the second row is for the optimal weights; the third row is for the Parzen kernel function with slow convergence rate. The left panels are for trade data and the right panels are for quote data. The number q is the order of bias correction.

bandwidth the RK estimator with TH_2 and optimal weights will have similar signature plots.

And finally the RK estimators are evidently below the 20 minutes log-returns RV. According to the ACF in Figure 2 and the "aggressively estimating order of bias correction" rule the RK estimators with $q = 9$ is likely to be close to the real IV at the highest frequency. The RK estimators are around 3.6×10^{-4} for transaction data and 3.0×10^{-4} for mid-quote data, both of which are smaller than the averaged 20 minutes log-returns RV 4.0×10^{-4} for both transaction and mid-quote data. Since the discretization error should be negative in quadratic variation this can be evidence of the correlation of the latent log-prices and the noises.

7 Conclusion

This paper studies the effect of time dependent micro-structure noises on the Realized Kernel estimators. The empirical results suggest that the higher order (than one) autocorrelations in the noises are significant and small in magnitude. The RK estimator still converges to the true value in this case especially when slower convergence Parzen kernel is applied although it still suffers from finite sample bias when the optimal bandwidth is small due to the low noise-to-signal ratio ξ^2.

In order to utilize the highest frequency samples a bias correction of the RK estimator is proposed under the assumption that the noises follow a moving average process. It puts weight 1 on the first q autocovariations to fully balance out the bias from MA(q) noises. As long as q is finite the bias correction does not change the asymptotic properties of the RK estimator.

The RK estimator can be interpreted by the control variable approach. The biased corrected realized variation RV_q is an unbiased estimator of IV_T and H number of higher order autocovariations are used as control variables to reduce variance. The optimal weights with MA noises are more difficult to compute than those with m.d.s. noises mentioned in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). A feasible approximation is proposed using preliminary estimations of IV_T and autocovariations γ.

The simulation experiment shows the advantage of bias correction is obvious. With highest frequency samples the MSE of the conventional RK estimator is mostly composed of the squared bias. Overestimating the order of the MA effect in noises does not affect the unbiasedness but slightly raise the MSE of the estimator. Therefore a safe strategy is to choose the bias correction order aggressively. The slower convergence Parzen kernel is relatively robust to the MA noise while it can still benefit from the bias correction.

A comparison of different kernel functions suggests that the Parzen kernel is preferred to the Tukey-Hanning$_2$ even after the bias correction. The optimal weights estimator has lower MSE than the Parzen kernel after bias correction. The plain unbiased estimator RV_q performs fairly well compared with the RK estimators. Finally the empirical application on the stock AA suggests the RV with 20 minutes log-returns overestimates IV_T which can be an evidence of the correlation of the latent log-prices and the noises.

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 30 DJIA component stocks have an average of $75,786$ transaction records and $1,208,332$ quotation records in 2011. 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. 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 dropped. However with the milliseconds time database the price series can be sampled at a sub-second frequency. To utilize as many records as possible, I keep all the data in this paper.

Appendix B Proof of Theorem 3.2 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 ΔY as the vector of observations $(\Delta Y_N, \Delta Y_{N-1}, \ldots, \Delta Y_1)'$. Similarly denote ΔX 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 υ with parameters $(\theta_0, \ldots, \theta_q)$ as in equation (11). Denote the $(N+q)$-by-1 vector $(\upsilon_N, \upsilon_{N-1}, \ldots, \upsilon_{-q+1})'$ as U. Then we can write $\Delta\varepsilon$ as $T \cdot U$, where the N-by-$(N+q)$ matrix T is given by:

$$T = \begin{bmatrix} \theta_0 & \theta_1 & \ldots & \theta_q & 0 & \ldots & 0 \\ 0 & \theta_0 & \theta_1 & \ldots & \theta_q & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & \theta_0 & \theta_1 & \ldots & \theta_q \end{bmatrix} \tag{30}$$

To represent γ_h by ΔY, 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:

$$J_1 = \begin{bmatrix} 0 & 1 & 0 & \\ 1 & 0 & 1 & \ddots \\ 0 & 1 & 0 & \ddots \\ & \ddots & \ddots & \ddots \end{bmatrix} \tag{31}$$

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

Expectations

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

For $k > 0$:

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

Note that φ_k is zero when $k > q$ as defined in equation (13).

Variance Covariance Matrix

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

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

<u>Part i</u>:

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

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

Therefore, matrix A is a diagonal matrix such that:

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

<u>Part ii</u>:

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

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:

$$\begin{aligned} B(k, k+s) = \frac{1}{IV_T} \Bigg(& \varphi_s \Big(\sum_{i=1}^{k} (\sigma_i^2 + \sigma_{N+1-i}^2) + (\varphi_s + \varphi_{2k+s}) \sum_{i=k+1}^{k+s} (\sigma_i^2 + \sigma_{N+1-i}^2) \\ & + 2(\varphi_s + \varphi_{2k+s}) \sum_{i=k+s+1}^{N-k-s} \sigma_i^2 \Bigg) \end{aligned}$$

Note that φ_k is zero when $k > q$ as defined in equation (13).

Part iii:

$$Cov[\Delta\varepsilon' J_k \Delta\varepsilon, \Delta\varepsilon' J_l \Delta\varepsilon] = Cov[U' \overbrace{T' J_k T}^{M_k} U, U' \overbrace{T' J_l T}^{M_l} U]$$

$$= Cov\left[\sum M_k(i,i)U_i^2 + 2\sum_{i<j} M_k(i,j)U_iU_j, \sum M_l(i,i)U_i^2 + 2\sum_{i<j} M_l(i,j)U_iU_j\right]$$

$$= V[\upsilon^2] \underbrace{\sum M_k(i,i)M_l(i,i)}_{N\cdot E+F} + 4\sigma_\upsilon^4 \underbrace{\sum_{i<j} M_k(i,j)M_l(i,j)}_{N\cdot C+D} \tag{32}$$

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 ψ_s is defined in equation (25).

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

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

$\square$

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.

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.

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.

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.

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.