GCOE Discussion Paper Series

Global COE Program

Human Behavior and Socioeconomic Dynamics

Discussion Paper No.101

Bias-Corrected Realized Variance under Dependent Microstructure Noise

Kosuke Oya

December 2009

GCOE Secretariat
Graduate School of Economics
OSAKA UNIVERSITY
1-7 Machikaneyama, Toyonaka, Osaka, 560-0043, Japan

Bias-Corrected Realized Variance under Dependent Microstructure Noise *

Kosuke Oya † ‡

Abstract

The aim of this study is to develop a bias-correction method for realized variance (RV) estimation, where the equilibrium price process is contaminated with market microstructure noise, such as bid-ask bounces and price changes discreteness. Though RV constitutes the simplest estimator of daily integrated variance, it remains strongly biased and many estimators proposed in previous studies require prior knowledge about the dependence structure of microstructure noise to ensure unbiasedness and consistency. The dependence structure is unknown however and it needs to be estimated. A bias-correction method based on statistical inference from the general noise dependence structure is thus proposed. The results of Monte Carlo simulation indicate that the new approach is robust with respect to changes in the dependence of microstructure noise.

JEL Classifications: C01; C13; C51

Keywords: Realized variance; Dependent microstructure noise; Two-time scales

* The author is grateful to Naoto Kunitomo, Michael McAleer and Nabil Maghrebi for helpful comments. This work is partially supported by the Japanese Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (A), 2009, under reference number 10153313 and JST, CREST.
† Graduate School of Economics, Osaka University, Toyonaka, Osaka, Japan. oya@econ.osaka-u.ac.jp
‡ Japan Science and Technology Agency, CREST, Toyonaka , Osaka, Japan

1 Introduction

The estimation of the daily integrated variance of returns on financial assets is important for derivatives pricing and risk management purposes. While realized variance (RV) constitutes a simple but useful estimator of daily integrated variance (IV), it remains also a strongly biased estimator, where the equilibrium price process is contaminated with market microstructure noise. This microstructure noise can be induced by various market frictions such as bid-ask bounces and the discreteness of price changes, inter alia. There are three approaches to cope with noise contamination, including (i) use of returns on the appropriate interval length based on optimal sampling frequency proposed by Bandi and Russell [2], (ii) subsampling and bias correction proposed by Zhang et al. [8] and (iii) kernel estimation following Barndorff-Nielsen, et al. [4]. McAleer and Medeiros [6] provide an extensive review of the recent literature on RV estimation. It is the time-dependent noise structure that ensures the unbiasedness and consistency of IV estimators. The estimations proposed in previous studies ultimately require prior knowledge about this noise dependency, which needs to be rather estimated. The present study addresses these estimation issues and uses the consistent cross-covariance and autocovariance estimators of microstructure noise, and the tests statistics developed by Ubukata and Oya [7] to identify the noise dependence structure. The selection procedure of time scales based on Aït-Sahalia et al.[1]'s Two Scales RV (TSRV) is also provided under the general conditions of dependent noise. An alternative bias-corrected estimator of IV can be also proposed using the autocovariance of microstructure noise. The remainder of this paper is organized as follows. The price process and market microstructure noise are presented in section 2. Section 3 discusses the realized variance and related estimators are given. Section 4 provides a brief review of the autocovariance estimator of microstructure noise proposed by Ubukata and Oya [7] followed by a discussion of of the selection two-scales TSRV estimator and alternative bias-corrected RV estimator. Section 5 presents the results of Monte Carlo simulation for the finite sample properties of the proposed selection procedure and the new bias-corrected estimator under the general noise dependence. Section 6 concludes the paper.

2 Price Process and Microstructure Noise

It is assumed that the equilibrium asset price follows a continuous semi-martingale process $\mathrm{d}P^*(t) = \mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W(t)$ where $P^*(t)$ is the logarithmic equilibrium continuously compounded intra-daily price, $W(t)$ is a standard Brownian motion, and both $\mu(t)$ and $\sigma(t) > 0$ are bounded measurable functions. The diffusion term $\sigma(t)$ can be estimated according to the integrated variance over a fixed interval $[0, T]$

$$IV = \langle P^*, P^* \rangle_T = \int_0^T \sigma^2(t) \mathrm{d}t \tag{1}$$

using the observed logarithmic asset price of the asset for $t \in [0, T]$. The market closing time is denoted as T. We assume that the drift term $\mu(t)$ equals to zero since the trend term of the price process is likely to be small during the trading hours on a given day. Suppose that the asset price can be observed at the discrete time points $t_0 = 0 < t_1 < t_2 < \cdots < t_n = T$, where t_i represents the i-th transaction time. The length of the i-th interval is defined as $\Delta t_i = t_i - t_{i-1}$. It is noted that $\Delta t_i = T/n$ only under the restrictive conditions of regular sampling, and that $\Delta t_i \neq \Delta t_j$ for $i \neq j$ for non-regular sampling.

In order to examine the impact of market microstructure noise, it is also assumed that the observed price process $P(t)$ consists of the equilibrium continuously compounded intra-daily price process $P^*(t)$, which is unobservable, and the noise process $\eta(t)$

$$P(t) = P^*(t) + \eta(t). \tag{2}$$

The market microstructure noise $\eta(t)$ is also assumed to represent a serially dependent random variable. This is rather a plausible assumption given the behavior of noise determinants such as bid-ask bounces, order-flow clustering and other market imperfections. Thus, the following set of assumptions about the microstructure noise can be made.

Assumption 1 (Market microstructure noise) *Suppose (a) $\{\eta(t)\}$ is a sequence of random variables with zero mean, (b) the noise process is covariance stationary with autocovariance function, which has a finite dependence structure in the sense that:*

$$\gamma_\eta(\ell) = \mathrm{E}[\eta(t)\eta(t-\ell)] = 0, \quad \textit{for all} \ \ | \ell | > m$$

where m is a finite positive integer, (c) there exists some positive number $\beta > 1$ that satisfies $\mathrm{E}\Big|\eta(t)\eta(s)\Big|^{4\beta} < \infty$ for all t, s and (d) the noise process is independent of the equilibrium price process.

With respect to assumption (d), it is noted that as the number of observations increases, the effect of dependence is dominated by noise variation, even when the noise terms are correlated with equilibrium prices. Hansen and Lunde [5] suggest that the independence assumption (d) does not significantly affect the analysis of asset prices with high-frequency trading.

3 Realized Variance and Related Estimators

3.1 Realized Variance

The most widely used estimator of the integrated variance defined in (1) is the realized variance, which is defined as the sum of squared returns given by

$$RV = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (P(t_i) - P(t_{i-1}))^2, \tag{3}$$

where the i-th transaction price is $P(t_i)$ and the i-th intraday log return is defined as $r_i = P(t_i) - P(t_{i-1})$.

Given the time intervals $I^i = (t_{i-1}, t_i]$ for all i, it is possible to the expectation of RV conditional on the stochastic arrival times as $\mathrm{E}_I[\,\cdot\,]$. Thus, the conditional expectation of (3) can be written as

$$\mathrm{E}_I[RV] = \sum_{i=1}^{n} \mathrm{E}_I[(r_i^* + \eta(t_i) - \eta(t_{i-1}))^2] = \sum_{i=1}^{n} \mathrm{E}_I[\, r_i^{*2}] + 2n\gamma_\eta(0) - 2\sum_{i=1}^{n} \gamma_\eta(\Delta t_i) \tag{4}$$

where $r_i^* = P^*(t_i) - P^*(t_{i-1})$. It is straightforward to demonstrate that the variance $\gamma_\eta(0)$ and the sum of autocovariances $\sum_{i=1}^{n} \gamma_\eta(\Delta t_i)$ can introduce a bias in the estimation of RV according to (3). For the sake of simplicity, it is assumed hereafter that the sampling scheme is regular and that $\Delta t_i = 1$, $i = 1, \ldots, n$. Under such conditions, the total bias is represented by $2n(\gamma_\eta(0) - \gamma_\eta(1))$.

3.2 Two-Scales Realized Variance

Zhang et al. [8] proposed the Two Scales Realized Variance (TSRV) which is unbiased when the microstructure noise is independent. Denote the original grid of observation times as $\mathcal{G} = \{t_0, t_1, \ldots, t_n\}$. $\mathcal{G}$ is partitioned into K nonoverlapping subgrids, $\mathcal{G}_K^{(j)}$, $j = 1, \ldots, K$, such that $\mathcal{G} = \cup_{j=1}^{K} \mathcal{G}_K^{(j)}$, where $\mathcal{G}_K^{(j)} \cap \mathcal{G}_K^{(\ell)} = \phi$ for $j \neq \ell$. Given the assumption of regular sampling scheme, the j-th nonoverlapping subgrid can be represented as $\mathcal{G}_K^{(j)} = \{t_{j-1}, t_{j-1+K}, t_{j-1+2K}, \ldots, t_{j-1+n_jK}\}$ for $j = 1, \ldots, K$ where n_j is the integer making t_{j-1+n_jK} the last element in the subgrid $\mathcal{G}_K^{(j)}$. Then the realized variance for the subgrid $\mathcal{G}_K^{(j)}$ can be written as

$$RV_K^{(j)} = \sum_{i=1}^{n_j} (P(t_{(j-1)+iK}) - P(t_{(j-1)+(i-1)K}))^2. \tag{5}$$

Let $RV^{(all)}$ be the realized variance for the full grid $\mathcal{G}$. Then TSRV by Zhang et al. [8] can be represented as

$$RV_K = (1/K)\sum_{j=1}^{K} RV_K^{(j)} - (\bar{n}/n)RV^{(all)} \tag{6}$$

where $\bar{n} = \sum_{j=1}^{K} n_j/K = (n-K+1)/K$. The first term of (6) represents the average of $RV_K^{(j)}$ estimators for the subgrid $\mathcal{G}_K^{(j)}$, $j = 1, \ldots, K$, which is a biased estimator of IV. In case of independent noise, the second term of (6) represents the bias-correction term since the bias in the first term is $2\bar{n}\gamma_\eta(0)$ and $RV^{(all)}/(2n)$ is a consistent estimator of $\gamma_\eta(0)$. The refinement to correct for the finite sample bias in (6) following Zhang et al. [8] is conducive to the following expression

$$RV_K^{(adj)} = (1-\bar{n}/n)^{-1}\, RV_K. \tag{7}$$

Although these two-scales realized variances are unbiased and consistent estimators of IV under the independent noise assumption, these desirable features are not guaranteed when the noise terms are not independently distributed.

3.3 Extended Two-Scales Realized Variance

Aït-Sahalia et al. [1] extended the TSRV to allow for dependent market microstructure noise. The first term of (6) can be rewritten in the form of the average lag K realized variance under regular sampling intervals.

$$RV_K^{(avg)} = (1/K)\sum_{j=1}^{K} RV_K^{(j)} = (1/K)\sum_{i=0}^{n-K} (P(t_{i+K}) - P(t_i))^2. \tag{8}$$

It is also possible to use an alternative lag J instead of lag K, $(1 \le J < K \le n)$ in (8) and express using two different lags J and K, the extended TSRV as follows.

$$RV_{J,K} = RV_K^{(avg)} - (\bar{n}_K/\bar{n}_J)RV_J^{(avg)} \tag{9}$$

where $\bar{n}_J = (n-J+1)/J$, $\bar{n}_K = (n-K+1)/K$, $1 \le J < K \le n$ and $K = o(n)$.

The main difference between TSRV expressed in (6) and extended TSRV estimators is captured by the second term of (9) which represents a bias-correction term. It is easy to identify the bias term in $RV_{J,K}$ as

$$2\bar{n}_K(\gamma_\eta(t_{i+J} - t_i) - \gamma_\eta(t_{i+K} - t_i)) = 2\bar{n}_K(\gamma_\eta(J) - \gamma_\eta(K)). \tag{10}$$

This bias consists of the autocovariances of microstructure noise $\gamma_\eta(J)$ and $\gamma_\eta(K)$. It should be noted that these autocovariances become negligible when the selected lags J and K are large enough. Based on Assumption 1 that the microstructure noise process is m-dependent, it is possible to select J such that $J = m + 1$ and $K = O(n^{2/3})$ as in Aït-Sahalia et al. [1]. The finite-sample correction of the extended TSRV can be thus expressed as

$$RV_{J,K}^{(adj)} = (1 - \bar{n}_K/\bar{n}_J)^{-1}\, RV_{J,K}. \tag{11}$$

4 Lags $(J,\, K)$ Selection and Alternative Estimator

The extended TSRV is an appropriate estimator of IV when the assumption of independent noise assumption is not valid. As discussed in the previous section however, the important issue remains as to how the optimal lags J and K should be selected. Although Aït-Sahalia et al. [1] argue that the extended TSRV is robust with respect to lags (J, K) selection, the following section 5.2 demonstrates that this estimation may still be significantly sensitive to lag J selection.

The present section discusses the new methodology that allows for the selection of the appropriate lag J, based on the testing procedure proposed in Ubukata and Oya [7]. An alternative IV estimator which utilizes a different bias-correction method from the extended TSRV is also briefly introduced.

4.1 Microstructure Noise Autocovariance Estimation

Ubukata and Oya [7] proposed an unbiased and consistent estimator of the microstructure noise autocovariance $\gamma_\eta(\ell)$, and derived its asymptotic properties. The test statistic of the null hypothesis $\gamma_\eta(\ell) = 0$ is also applied to examine the significance of the microstructure noise dependence. Suppose that the threshold value of microstructure noise dependence as m, is such that $\gamma_\eta(m+1) = 0$ and $\gamma_\eta(m) \neq 0$. The threshold value m can be determined through the test statistic defined in Ubukata and Oya [7, section 3.2].

To obtain an unbiased estimator of $\gamma_\eta(\ell)$, it is possible to construct the product of returns $Z_{\ell,ij}^{(\pm)}$ for all i, j, such that $\ell = t_{j-1} - t_i$ using the selected threshold value m

$$Z_{\ell,ij}^{(\pm)} = \underline{r}_i^{(-)} \overline{r}_j^{(+)} = (P(t_i) - P(\underline{t}_{i-1}^{(-)}))(P(\overline{t}_j^{(+)}) - P(t_{j-1})) \tag{12}$$

where $\overline{t}_j^{(+)}$ is the first transaction time, which follows t_j subject to $\overline{t}_j^{(+)} - t_i > m$, and $\underline{t}_{i-1}^{(-)}$ is the last transaction time, which is followed by t_{i-1} subject to $t_{j-1} - \underline{t}_{i-1}^{(-)} > m$. Thus, for a given $\ell = t_{j-1} - t_i$, expected product of returns can be expressed as $\mathrm{E}[Z_{\ell,ij}^{(\pm)}] = -\gamma_\eta(\ell)$ for all i, j. Let $\{Z_{\ell,k}^{(\pm)}\}_{k=1}^{N_\ell}$ be the sequence that arranges $Z_{\ell,ij}^{(\pm)}$ satisfying $\ell = t_{j-1} - t_i$ in ascending order of index i. N_ℓ represents the number of observations in the sequence. It is then possible to construct the unbiased autocovariance estimator of the microstructure noise using the sample mean of $\{Z_{\ell,k}^{(\pm)}\}_{k=1}^{N_\ell}$.

Autocovariance Estimator (Ubukata and Oya [7]): *The autocovariance estimator of the microstructure noise and its asymptotic distribution are given as*

$$\hat{\gamma}_\eta(\ell) = -\frac{1}{N_\ell} \sum_{k=1}^{N_\ell} Z_{\ell,k}^{(\pm)}, \quad N_\ell^{1/2}(\hat{\gamma}_\eta(\ell) - \gamma_\eta(\ell)) \xrightarrow{a} N(0, \omega_\ell^2) \tag{13}$$

where $\omega_\ell^2 = \lim_{N_\ell \to \infty} N_\ell \mathrm{E}[(\hat{\gamma}_\eta(\ell) - \gamma_\eta(\ell))^2]$.

The test statistic to examine the significance of $\gamma_\eta(\ell)$ is also discussed in Ubukata and Oya [7, corollary 2]. Let the null hypothesis and the alternative be represented as $\gamma_\eta(\ell) = 0$ and $\gamma_\eta(\ell) \neq 0$, respectively. The test statistic of the noise autocovariance can be expressed as

$$\tau_\eta^*(\ell) = \sqrt{N_\ell}\, \hat{\gamma}_\eta(\ell)/\hat{\omega}_\ell. \tag{14}$$

This test statistic is asymptotically distributed as standard normal under the appropriate conditions, which are not further discussed herein for the sake of brevity. Reference can be made to Ubukata and Oya [7] for more details including the explicit formulation of $\hat{\omega}_\ell$.

4.2 Lags (J, K) Selection

To select the appropriate lag J, it is required first to examine whether $\gamma_\eta(1) = 0$ through the test statistic $\tau_\eta^*(\ell)$. When the null is rejected, there is a need to verify whether $\gamma_\eta(2) = 0$, and this test $\gamma_\eta(\ell) = 0$ is reiterated until the null cannot be rejected. The distance ℓ for which the null hypothesis $\gamma_\eta(\ell) = 0$ cannot be rejected for the first time is denoted as $\hat{J}$ for the extended TSRV.

The optimal choice of lag K is clearly provided in Zhang et al. [8] under the *i.i.d.* noise assumption. However, the important issue of optimal lag selection remains under less restrictive conditions of dependent microstructure noise. It is possible to devise a simple approach to address this issue of optimal lag selection, where the original grid of observation times denoted by $\mathcal{G} = \{t_0, t_1, \ldots, t_n\}$ as defined in previous section, is supplemented by a new subgrid $\mathcal{G}_{\hat{J}} = \{t_0, t_{\hat{J}}, t_{2\hat{J}}, \ldots, t_{[n/\hat{J}]}\}$ where $\hat{J}$ is the selected lag, representing the threshold value of noise dependence based on the test statistic $\tau^*_\eta(\ell)$. It is reasonable to suppose that the microstructure noise $\{\ \eta(t_{i\hat{J}})\ \}_{i=0}^{[n/\hat{J}]}$ is an uncorrelated random sequence. The optimal lag K can thus be obtained by applying the methodology proposed in Zhang et al. [8] to the sequence of the observed transaction price $\{\ P(t_{i\hat{J}})\ \}_{i=0}^{[n/\hat{J}]}$. The extended TSRV with selected lags ($\hat{J}$, $\hat{K}$) and the bias-adjusted extended TSRV are denoted hereafter as $RV_{\hat{J},\hat{K}}$ and $RV^{(adj)}_{\hat{J},\hat{K}}$, respectively.

4.3 Alternative Bias-Corrected Estimator

In order to correct the bias of RV, an alternative IV estimator $RV^{(bc)}_K$ can also be constructed using autocovariance defined in (13) as

$$RV^{(bc)}_K = RV^{(avg)}_K - 2\bar{n}_K\hat{\gamma}_\eta(0). \tag{15}$$

The unbiasedness and consistency of $RV^{(bc)}_K$ can be immediately established from the unbiasedness and consistency of the autocovariance estimator (13) and the result given in Zhang et al. [8].

5 Monte Carlo Simulation

5.1 Simulation Design

A series of Monte Carlo simulations is performed in order to examine the impact of lag (J, K) selection on the extended TSRV. This simulation exercise allows also for the assessment of the estimator (15) properties relative to the extended TSRV under dependent microstructure noise. The return-generating process is defined exactly as in Zhang et al. [8].

$$\begin{aligned} \mathrm{d}P^*(t) &= (0.05 - \nu(t)/2)\mathrm{d}t \ + \ \sigma(t)\mathrm{d}B(t) \\ \mathrm{d}\nu(t) &= 5(0.04 - \nu(t))\mathrm{d}t \ + \ 0.5\nu(t)^{1/2}\mathrm{d}W(t), \ \ \nu(t) = \sigma^2(t) \end{aligned}$$

The correlation between the two Brownian motions B and W is set to -0.5. A total of 10000 sample paths of the process by the Euler scheme are generated at time intervals $\Delta t = 1$ second. It is noted that $T = 1$ day and a day, which consists of 6.5 hours of open trading, or the equivalent of 23400 seconds. The price levels are observed discretely under the microstructure noise. In the simulation, the time interval between observations is set to be 5 seconds, i.e., $t_0 = 0, t_1 = 5, \ldots, t_{4680} = 23400$. Since the focus is made on the case of dependent microstructure noise, different dependent patterns are considered, including the following autoregressive model and moving average model for the noise process

$$\begin{aligned} \text{AR}(1): &\quad \eta(t_i) = \rho\, \eta(t_{i-1}) \ + \ \varepsilon(t_i), \\ \text{MA}(3): &\quad \eta(t_i) = \varepsilon(t_i) \ + \ \sum_{s=1}^{3} \theta_s \varepsilon(t_{i-s}) \end{aligned}$$

where $\rho =$ -0.8, -0.4, 0.0, 0.4, 0.8 for AR(1), $(\theta_1, \theta_2, \theta_3) =$ (-0.6, 0, 0), (0.6, 0, 0) for MA(1), (0.6, 0.5, 0) for MA(2) and (0.6, 0.5, 0.4) for MA(3). The variance of microstructure noise $\mathrm{E}[\eta(t)^2]$ should be carefully selected because the effect of the microstructure noise may be negligible only when $\mathrm{E}[\eta(t)^2]$ is very small. Hansen and Lunde [5] reported that the *Noise-to-Signal Ratio (NSR)* defined as $\mathrm{E}[\eta(t)^2]/IV$ for a sample of stocks listed on the NYSE and NASDAQ markets ranges from 0.0002 to 0.006. Consistent with on these previous empirical findings, the NSR for simulated paths $\mathrm{E}[\eta(t)^2]$ is set to 0.004, which is equal to the average of the above NSR reference values. The observed price is also given as $P(t_i) = P^*(t_i) + \eta(t_i)$.

Under these conditions, it is possible in the following subsections, to examine the influence of lag J selection on the extended TSRV estimators, with K = 50, 100, 200 and assess the statistical properties of selected lag J using the test statistic $\tau_\eta^*(\ell)$. Furthermore, it is important to compare the extended TSRV estimator with selected lags (J, K), to the bias-adjusted TSRV extended estimator as well as (15) proposed in this study. These distinct estimators can be obtained for each simulated sample path.

5.2 Influence of Lag Selection

The integrated variance is estimated using $RV_{J,K}$ and $RV_{J,K}^{(adj)}$ with $J = 1, \ldots, 40$ and K= 50, 100, 200. The relative bias of these estimators represented by the sample means of $(estimate - IV)/IV$ is reported in the first and third columns of Figure 1. The sample root-mean-squared-errors (RMSE) of $estimate/IV$ is also provided in the second and fourth columns. The first and third rows of Figure 1 show the bias and RMSE for $RV_{J,K}$. Similarly, the second and fourth rows refer to the bias and RMSE for $RV_{J,K}^{(adj)}$. The horizontal axis is provided for J=1, . . . , 40. The models used for the noise process are AR(1) with $\rho = -0.8$ and 0.8, *i.i.d.* and MA(1) with θ=−0.6. It is clear that the bias of $RV_{J,K}^{(adj)}$ becomes negligible after J exceeds the threshold value of noise dependence while the bias of $RV_{J,K}$ grows as J increases. Although the performance of the bias-adjusted estimator $RV_{J,K}^{(adj)}$ is relatively better, the variance seems to be sensitive to this bias adjustment process. This effect is clearly captured by the reported RMSE values.

It is important to appropriately select the lag J for both $RV_{J,K}$ and $RV_{J,K}^{(adj)}$ estimators because their RMSEs can strongly depend on the lag J selection even under more favorable conditions of independent microstructure noise. Judging also from the second and fourth columns of Figure 1, it appears that lag K selection does not exert significant influence on RMSE provided that lag J is properly selected.

Figure 1 : around here

5.3 Performance of Lag Selection Procedure

The selection of lag J is preformed by testing the null hypothesis $\gamma_\eta(\ell) = 0$ for $\ell > 0$ as described in section 4.2. Again, it is possible to examine the behavior of the test statistic $\tau_\eta^*(\ell)$ in association with lag J selection. Figure 2 shows the empirical distribution of selected lag J denoted as $\hat{J}$ for the representative cases. The mode of $\hat{J}$ is found to equal 8 and 2 for AR(1) noise dependence with $\rho =$ -0.8 and -0.4, respectively. On the other hand, mode estimates amount to just 1, 2, 3 and 4, for *i.i.d.*, MA(1), MA(2) and MA(3), respectively.

Thus, it is possible to use the test statistic $\tau_\eta^*(\ell)$ as an essential criterion in the selection of lag J. The results of lag K selection according to the procedure described in the previous section are not reported here, but the evidence suggests that variance of the empirical distribution of $\hat{K}$ rises as the noise dependence increases.

Figure 2 : around here

5.4 Comparison of IV Estimators

The analysis so far focused on important issues related to the construction of the set of three IV estimators under the assumption of dependent microstructure noise. The objective of this subsection is to examine however the statistical properties of these estimators. It is noted again that $RV_{\hat{J},\hat{K}}$ and $RV_{\hat{J},\hat{K}}^{(adj)}$ represent the extended TSRV and bias-adjusted extended

TSRV with the selected lags $(\hat{J}, \hat{K})$, while $RV_{\hat{K}}^{(bc)}$ is the estimator proposed in the previous section with the selected lag $\hat{K}$. The empirical distributions of $(estimate - IV)/IV$ for each estimator are given in Figure 3. It is easy to see that the empirical distributions of $RV_{\hat{J},\hat{K}}$ and $RV_{\hat{J},\hat{K}}^{(adj)}$ with $(\hat{J}, \hat{K})$ are found to be skewed to the right. It is noted in particular that under strong noise dependence, the empirical distribution of the proposed estimator $RV_{\hat{K}}^{(bc)}$ is rather closer to symmetry than alternative estimators.

Figure 3 : around here

Table 1 : around here

The bias and RMSE values associated with IV estimators are reported in Table 1. The bias terms in $RV_{\hat{K}}^{(bc)}$ are generally smaller than those obtained with respect to $RV_{\hat{J},\hat{K}}$ and $RV_{\hat{J},\hat{K}}^{(adj)}$. These RMSE values are rather comparable across these estimators, except in case of strong noise dependence case. These simulation results suggest that the extended TSRV and bias-adjusted extended TSRV with selected $(\hat{J},\ \hat{K})$, as well as the proposed estimator $RV_{\hat{K}}^{(bc)}$ are indeed robust to the dependence of microstructure noise.

6 Concluding Remarks

This paper proposes a new approach to the problem of appropriate lag selection for the two-scales realized variance under dependent microstructure noise. An alternative bias-adjusted estimator based on the variance of microstructure noise is also proposed, along the lines of Ubukata and Oya [7]. The evidence from Monte Carlo simulation suggests that the proposed lag selection procedure is appropriate as the the proposed estimator is associated with relatively smaller bias and RMSE values. The proposed procedure for lag selection and the new IV estimator can thus be useful for empirical studies based on transactions price data. It should be noted that kernel-type estimators may be considered to be more efficient. But as noted by Bandi and Russell [3], this asymptotic property is not necessarily satisfied in the presence of large samples typical of empirical studies based on transactions data. The comparative analysis is nevertheless important and therefore warranted. It falls beyond the scope of the present study, which introduces a new approach that may open interesting avenues for future research.

References

[1] Y. Aït-Sahalia, P. Mykland, L. Zhang, Ultra high frequency volatility estimation with dependent microstructure noise, NBER Working Paper, 11380.

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

[3] F. M. Bandi, J. R. Russell, Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations, Journal of Econometrics, forthcoming.

[4] O. E. Barndorff-Nielsen, P. Hansen, A. Lunde, N. Shephard, Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise, Econometrica 76 (2008) 1481–1536.

[5] P. R. Hansen, A. Lunde, Realized variance and market microstructure noise (with discussions), Journal of Business and Economic Statistics 24 (2006) 127–218.

[6] M. McAleer, M. C. Medeiros, Realized volatility: A review, Econometric Reviews 27 (2008) 10–45.

[7] M. Ubukata, K. Oya, Estimation and testing for dependence in market microstructure noise, Journal of Financial Econometrics 7 (2009) 106–151.

[8] L. Zhang, P. Mykland, Y. Aït-Sahalia, A tale of two time scales: Determining integrated volatility with noisy high frequency data, Journal of the American Statistical Association 100 (2005) 1394–1411.

Table 1: Relative Bias and RMSE of estimators

	Bias			RMSE		
Noise Type	$RV_{\hat{J},\hat{K}}$	$RV_{\hat{J},\hat{K}}^{(adj)}$	$RV_{\hat{K}}^{(bc)}$	$RV_{\hat{J},\hat{K}}$	$RV_{\hat{J},\hat{K}}^{(adj)}$	$RV_{\hat{K}}^{(bc)}$
AR: $\rho = -0.8$	-0.131	0.035	0.008	0.326	0.366	0.259
AR: $\rho = -0.4$	-0.064	0.009	-0.006	0.208	0.218	0.214
AR: $\rho = 0.4$	0.097	0.181	0.041	0.336	0.374	0.295
AR: $\rho = 0.8$	0.123	0.344	0.077	0.543	0.688	0.385
i.i.d.	-0.074	-0.019	-0.005	0.160	0.141	0.213
MA(1)	0.016	0.081	0.025	0.334	0.358	0.292
MA(2)	0.058	0.146	0.032	0.400	0.446	0.314
MA(3)	0.088	0.196	0.037	0.441	0.506	0.336

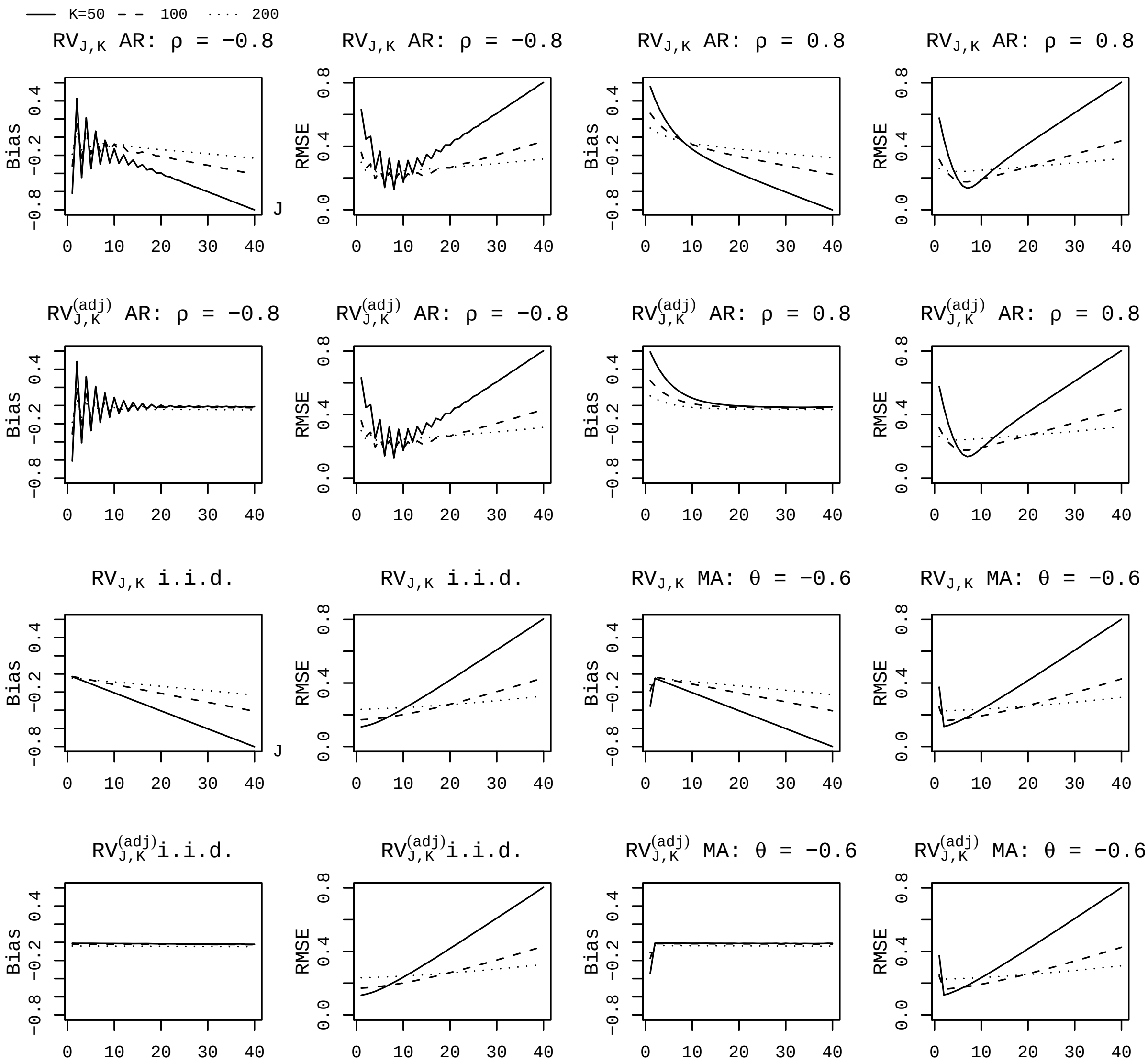

Fig. 1. Figure 1: Effect of selection lag J with different lag K values

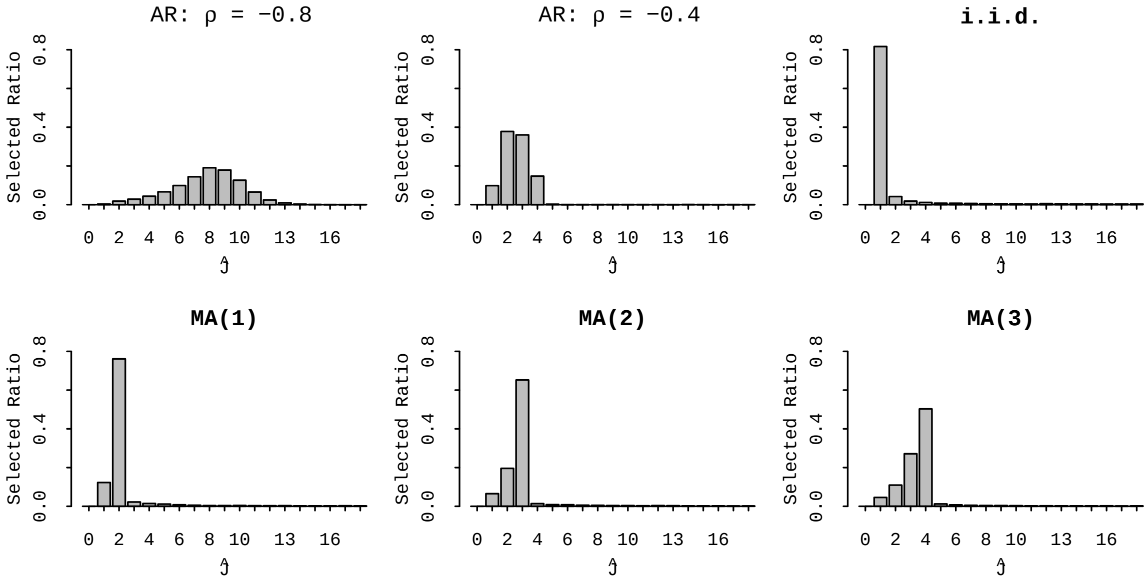

Fig. 2. Empirical distribution of selected lag J using $\tau^*_\eta(\ell)$

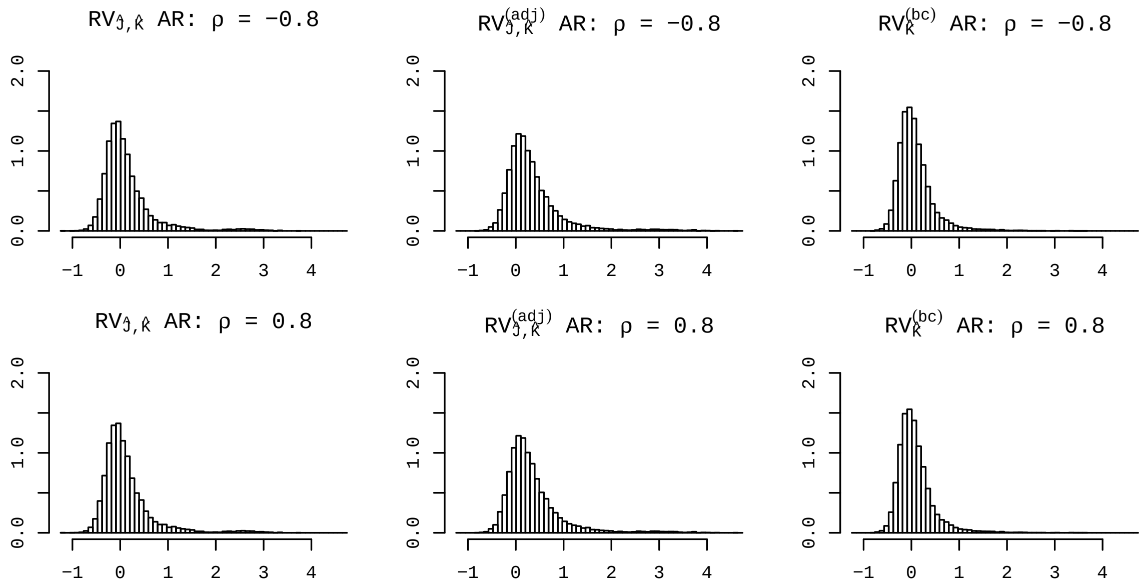

Fig. 3. Empirical distribution of $(estimate - IV)/IV$