Technical Report as Supplemental Material: Multi-scale Jump and Volatility Analysis for High-Frequency Financial Data

Jianqing Fan and Yazhen Wang

October, 2006

Abstract

The wide availability of high-frequency data for many financial instruments stimulates an upsurge interest in statistical research on the estimation of volatility. Jump-diffusion processes observed with market microstructure noise are frequently used to model high-frequency financial data. Yet, existing methods are developed for either noisy data from a continuous diffusion price model or data from a jump-diffusion price model without noise. We propose methods to cope with both jumps in the price and market microstructure noise in the observed data. They allow us to estimate both integrated volatility and jump variation from the data sampled from jump-diffusion price processes, contaminated with the market microstructure noise. Our approach is to first remove jumps from the data and then apply noise-resistant methods to estimate the integrated volatility. The asymptotic analysis and the simulation study reveal that the proposed wavelet methods can successfully remove the jumps in the price processes and the integrated volatility can be estimated as accurately as in the case with no presence of jumps in the price processes. In addition, they have outstanding statistical efficiency. The methods are illustrated by applications to two high-frequency exchange rate data sets.

Jianqing Fan is Frederick Moore'18 Professor of Finance, Department of Operation Research and Financial Engineering, Princeton University, Princeton, NJ 08544. Yazhen Wang is Professor, Department of Statistics, University of Connecticut, Storrs, CT 06269. Fan's research was partially supported by the NSF grant DMS-0532370 and Wang's research was partially supported by the NSF grant DMS-0504323. The authors thank the editor, associate editor, and two anonymous referees for stimulating comments and suggestions, which led to significant improvements in both substance and the presentation of the paper.

1 Introduction

Diffusion based stochastic models are often employed to describe complex dynamic systems where it is essential to incorporate internally or externally originating random fluctuations in the system. They have been widely applied to problems in fields such as biology, engineering, finance, physics, and psychology (Kloeden and Platen, 1999, chapter 7). As a result, there has been a great demand in developing statistical inferences for diffusion models (Prakasa Rao, 1999). This paper investigates nonparametric estimation for noisy data from a jump-diffusion model. The problem is motivated from modeling and analysis of high-frequency financial data.

In financial time series there is extensive and vibrant research on modeling and forecasting volatility of returns such as parametric models like GARCH and stochastic volatility models (Bollerslev, Chou and Kroner, 1992; Gouriéroux, 1997; Shephard, 1996; Wang, 2002), or implied volatilities from option prices in conjunction with specific option pricing models such as the Black-Scholes model (Fouque, Papanicolaou, and Sircar, 2000). These studies are for low-frequency financial data, at daily or longer time horizons. Over past decade there has been a radical improvement in the availability of intraday financial data, which are referred to as high-frequency financial data (Dacorogna, Geçay, Müller, Pictet and Olsen, 2001). Nowadays, thanks to technological innovations, high-frequency financial data are available for a host of different financial instruments on markets of all locations and at scales like individual bids to buy and sell, and the full distribution of such bids.

Historically the availability of financial data at increasingly high frequency allows us to incorporate more data in volatility modeling and to improve forecasting performance. However, because of their complex structures, it is very hard to find appropriate parametric models for high-frequency data. Volatility models at the daily level cannot readily accommodate high-frequency data, and parametric models specified directly for intradaily data generally fail to capture interdaily volatility movements (Andersen, Bollerslev, Diebold and Labys, 2003).

It is natural to use flexible nonparametric approach for high-frequency volatility analysis. One popular nonparametric method is the so-called realized volatility (RV) constructed from the summation of high-frequency intradaily squared returns (Andersen, Bollerslev, Diebold and Labys, 2003; Barndorff-Nielsen and Shephard, 2002), and another one is the realized bi-power variation (RBPV) constructed from the summation of appropriately scaled cross-products of adjacent high-frequency absolute returns (Barndorff-Nielsen and Shephard, 2006). Theoretical justifications of these nonparametric methods are based on the idealized assumption that observed high-frequency data are true underlying asset returns. Under this assumption, asymptotic theory for RV and RBPV is established by connecting them to quadratic variation

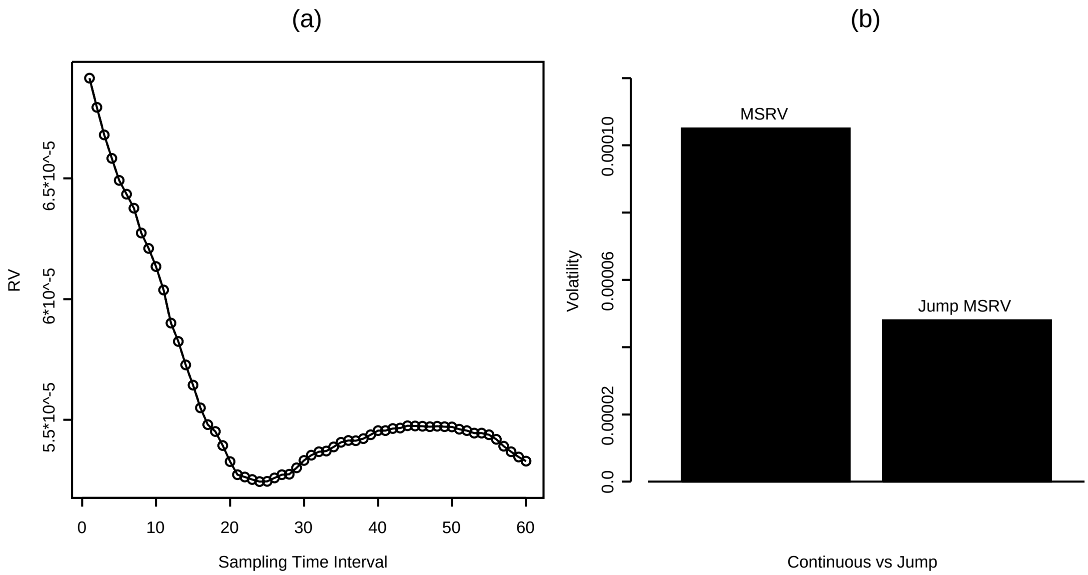

Figure 1: (a). Plot of the RV as a function of sampling time interval in minute. The horizon axis is the time interval in minute that the data are sampled from Euro-dollar exchange rates on January 7, 2004, for computing the RV. The shorter the sampling time interval is, the higher the sampling frequency. The RV shoots up as the sampling time interval gets shorter, which suggests the presence of the microstructure noise. (b). Plot of estimated integrated volatility using MSRV for Euro-dollar exchange rates on January 9, 2004. The bar on the left panel shows the estimate of integrated volatility obtained by applying MSRV to the whole data, and the right panel indicates the sum of the four estimated values obtained by applying the same MSRV procedure to each of the four pieces of the data. The left panel estimate is twice as large as the estimate in the right panel, which indicates the existence of jumps.

and bi-power variation. RV and RBPV are combined together to test for jumps in prices and to estimate integrated volatility and jump variation.

The idealized assumption is, however, severely challenged by high-frequency data. In reality, high-frequency returns are very noisy and hence do not allow for reliable inferences of RV and RBPV regarding the true underlying latent volatility. The noise is due to the imperfections of trading processes — vast array of issues collectively known as market microstructure including price discreteness, infrequent trading, and bid-ask bounce effects. The higher the frequency that prices are sampled at, the larger the realized volatility, indicating the possible presence of the microstructure noise.. These are evidenced and analyzed by, for example, Aït-Sahalia *et al.*(2005), Zhang *et al.*(2005), Zhang (2004), Bandi and Russell (2005). To provide some evidence for currency markets, Figure 1(a) plots the RV as a function of sampling time in minute for the returns of Euro-dollar exchange rates on January 7, 2004. Clearly, as the sampling interval gets smaller (or equivalently sampling frequency gets higher), the RV shoots up. For log price data following a diffusion process without noise, the RV will approach its quadratic variation, as sampling time goes to zero. If there is no noise in the data, we would expect the RV to be stabilized, as sampling frequency increases. When log price data contain market microstructure noise, the RV explodes as sampling interval approaches zero (Zhang et al., 2005). The existence of noise in the observed exchange rate data provides a reasonable explanation for the "explosion" of the RV near the origin in Figure 1(a). To mitigate the effects of microstructure noise, common practice is to sample available high-frequency data over longer time horizons and use the subsample, a small fraction of the available data, to compute RV and RBPV. The length of subsampling interval is *ad hoc*, ranging from 5 to 30 minutes. For exchange rate data, it corresponds to take a stable estimate from Figure 1(a). For the continuous price model, to better handle microstructure noise, researchers investigated the "optimal" frequency for subsampling, and prefiltering and debiasing. See Aït-Sahalia *et al.*(2005) and Bandi and Russell (2005).

Subsampling throws away most of available data. For example, at the sampling frequency of 30 minutes per data point for stock price, there are only about 16 data points per day, while sampling at frequency of 1 minute per datum, there are about 480 data points. The former results in very inefficient statistical estimates. Under the continuous diffusion price model, to improve the efficiency, Zhang *et al.*(2005) and Zhang (2006) invented two-time scale RV (TSRV) and multiple-time scale RV (MSRV) to reduce microstructure noise by averaging over RVs at various subsampling frequencies. See also Zhou (1996) and Hansen and Lunde (2006). Despite their successes, the procedures have not taken into account of the possible jumps due to the inflow of market news during the trading session. To illustrate the point, we take the Euro-dollar exchange rate data on January 9, 2004. We directly apply to the whole data set MSRV with 11 scales of returns sampled from every 5 to 15 minutes and estimate integrated volatility on that day. The estimated value is 10.5×10^{-5} labeled as 'MSRV' bar in Figure 1(b). On the other hand, the whole data are divided into 4 pieces, at three locations where the exchange rates might contain jumps. We apply the same MSRV procedure to each piece of the data to estimate integrated volatility over each subinterval, and then add up these four estimates of integrated volatilities to obtain the integrated volatility on January 9, 2004. This gives an estimated value 4.8×10^{-5}, which is labeled as 'Jump MSRV' bar in Figure 1(b). The estimated integrated volatility based on the whole data set is twice as large as the sum of the estimated integrated volatilities based on four subintervals. This gives an indication of jumps in data, since the MSRV estimate of integrated volatility based on the whole data should be close to the sum of the separate estimates for high frequency data from a continuous diffusion process. It also shows that the methods based on the continuous diffusion are not adequate for estimating integrated volatilities.

As market returns frequently contain jumps (Andersen, Bollerslev and Diebold, 2003; Barndorff-Nielsen and Shephard, 2006; Eraker, Johannes and Polson, 2003; Huang and Tauchen, 2005), it is important to have methods that handle automatically the possible jumps in the financial market. Separating variations due to jump and continuous parts is very important for asset pricing, portfolio allocation and risk management, as the former is usually less predictable than the latter due to the inflow of market news. To detect jump locations efficiently, we employ wavelets methods, which are powerful for detecting jumps as demonstrated in Wang (1995).

In this paper, we propose a wavelet based multi-scale approach to perform efficient jump and volatility analysis for high-frequency data. The proposed methods handle data with both microstructure noise and jumps in prices, providing comprehensive noise resistant estimators of integrated volatility and jump variation. The challenge of our problem is to first detect the jumps from the sample paths of diffusion processes, which are usually much rougher than those in nonparametric change point problems. Thanks to the availability of high-frequency data, jump locations and sizes can be accurately estimated. With the wavelet transformation, the information about jump locations and jump sizes is stored at high-resolution wavelet coefficients, while the low-resolution ones contain useful information for integrated volatility. With jump locations and sizes being estimated, they are removed from the observed data, resulting in jump adjusted data which are almost from a continuous process masked with microstructure noise. We demonstrate that the jump effect on the average RV for the jump adjusted data is $O_P(n^{-1/4})$, the best rate that the integrated volatility can be estimated (Gloter and Jacod, 2001 and Zhang 2004). We can then apply TSRV of Zhang *et al.*(2005) and MSRV of Zhang (2006) to the jump adjusted data and obtain estimators of integrated volatility. A wavelet based estimator of integrated volatility is also proposed and studied.

The rest of the paper is organized as follows. Section 2 specifies stochastic models for high-frequency data and statistical problems. We study the estimation of jump variation in Section 3 and the volatility estimation in Section 4 and establish convergence rates for the proposed estimators. Section 5 presents simulation results to evaluate the finite sample performance of the proposed methods and applies the methods to high-frequency exchange rate data. Section 6 features the conclusions. All technical proofs are relegated to Section 7.

2 Nonparametric volatility model

Due to market microstructure, high-frequency data are very noisy. This is convincingly demonstrated in Zhang *et al.*(2005). See also Figure 1(a). A common modeling approach is to treat microstructure noise as usual "observation error" and then to assume that the observed high-frequency data Y_t are equal to the latent, true log-price process X_t of a security plus market microstructure noise ε_t, that is,

$$Y_t = X_t + \varepsilon_t, \qquad t \in [0, 1], \tag{1}$$

where Y_t is the logarithm of the observable transaction price of the security, and is observed at times $t_i = i/n$, $i = 0, \cdots, n$, and ε_t is zero mean i.i.d. noise with variance η^2 and finite fourth moment, independent of X_t.

The true log-price X_t is generally assumed to be a semi-martingale of the form

$$X_t = \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dW_s + \sum_{\ell=1}^{N_t} L_\ell, \qquad t \in [0, 1], \tag{2}$$

where the three terms on the right hand side of (2) correspond to the drift, diffusion and jump parts of X, respectively. In the diffusion term, W_t is a standard Brownian motion, and the diffusion variance σ_t^2 is called spot volatility. For the jump part, N_t represents the number of jumps in X up to time t, and L_ℓ denotes the jump size.

The log-price process X given in (2) has quadratic variation

$$[X, X]_t = \int_0^t \sigma_s^2 \, ds + \sum_{\ell=1}^{N_t} L_\ell^2, \tag{3}$$

which consists of two parts: integrated volatility and jump variation. Denote them by

$$\Theta = \int_0^1 \sigma_s^2 \, ds, \qquad \Psi = \sum_{\ell=1}^{N_1} L_\ell^2.$$

The goal is to estimate Θ and Ψ, which will be considered in next two sections.

3 Jump analysis

To estimate jump variation Ψ, we first apply the wavelet method to the observed data and locate all jumps in the sample path of X_t and then use the estimated jump locations to estimate jump size for each estimated jump. The jump variation is estimated by the sum of squares of all estimated jump sizes.

3.1 Wavelets

Wavelets are orthonormal bases obtained by dyadically dilating and translating a pair of specially constructed functions φ and ψ which are called father wavelet and mother wavelet, respectively. The obtained wavelet basis is $\varphi(t)$, $\psi_{j,k}(t) = 2^{j/2} \, \psi(2^j \, t - k)$, $j =$

$1, 2, \cdots$, $k = 1, \cdots, 2^j$. We can expand a function over the wavelet basis. One special property of the wavelet expansion is the localization property that the coefficient of $\psi_{j,k}(t)$ reveals information content of the function at approximate location $k\, 2^{-j}$ and frequency 2^j. For example, if a function is Hölder continuous with exponent α at a point, then the wavelet coefficients of $\psi_{j,k}(t)$ with $k\, 2^{-j}$ near the point decay at order $2^{-j\,(\alpha+1/2)}$ (Daubechies, 1992, section 2.9); if the function has a jump at that point, then the wavelet coefficients of $\psi_{j,k}(t)$ with $k\, 2^{-j}$ near the given point is bounded below by $2^{-j/2}$ multiplied by a positive constant. This special feature enables us to separate jumps from the continuous part. See Daubechies (1992), Vidakovic (1999) and Wang (1995, 2006).

3.2 Jump estimation

Let $X_{j,k}$, $Y_{j,k}$, and $\varepsilon_{j,k}$ be the wavelet coefficients of X_t, Y_t, and ε_t, respectively. Then from model (1) we obtain that $Y_{j,k} = X_{j,k} + \varepsilon_{j,k}$, $k = 1, \cdots, 2^j$, $j = 1, \cdots, \log_2(n)$. The sample path corresponding to the first two terms of X_t in (2) is Hölder continuous with exponent α close to $1/2$, and the third term has a stepwise sample path. Thus, at the time point $t = k\, 2^{-j}$ where the log-price process X_t is continuous, $X_{j,k}$ is of order $2^{-j\,(\alpha+1/2)}$; whereas nearby a jump point of X, $X_{j,k}$ converges to zero at a speed no faster than $2^{-j/2}$, an order of magnitude larger than those at continuous points. As ε_t are i.i.d. noise, $\varepsilon_{j,k}$ are uncorrelated with mean zero and variance η^2/n. Thus, at high resolution levels, $X_{j,k}$ dominates $\varepsilon_{j,k}$ nearby jump locations and is negligible otherwise. From the comparison of decay order of wavelet coefficients, we easily show that, at high resolution levels j_n with $2^{j_n} \sim n/\log^2 n$, nearby jump points of the log-price process X_t, $Y_{j_n,k}$ are significantly larger than the others.

We use a threshold D_n to calibrate $|Y_{j_n,k}|$ and estimate the jump locations of the sample path of X_t by the locations of $|Y_{j_n,k}|$ that exceed D_n. That is, if $|Y_{j_n,k}| > D_n$ for some k, the corresponding jump location is estimated by $\hat{\tau} = k\, 2^{-j_n}$. One choice of threshold is the universal threshold $D_n = d\sqrt{2 \log n}$, where d is the median of $\{|Y_{j_n,k}|, k = 1, \cdots, 2^{j_n}\}$ divided by 0.6745, a robust estimate of standard deviation.

Under Conditions (A1)-(A4) stated in section 6, almost surely the sample path of X_t has finite number of jumps and otherwise is Hölder continuous with exponent close to $1/2$, and X_t and ε_t are independent. Because of the independence, by conditioning on X, the wavelet jump detection for deterministic functions (Wang, 1995; Raimondo, 1998) can be applied to the sample path of X with i.i.d. additive noise ε_t. Suppose that its sample path has $q = N_1$ jumps at $\tau_1, \cdots, \tau_q$. Denote the estimated $\hat{q}$ jump locations by $\hat{\tau}_1, \cdots, \hat{\tau}_{\hat{q}}$. Then from Wang (1995) and Raimondo (1998) we have under

Conditions (A1)-(A4),

$$\lim_{n\to\infty} P\left(\hat{q}=q, \sum_{\ell=1}^{q} |\hat{\tau}_\ell - \tau_\ell| \le n^{-1}\, \log^2\, n \,\middle|\, X\right) = 1. \tag{4}$$

3.3 Estimation of jump variation

To estimate jump variation of X_t, we need to estimate jump sizes. For each estimated jump location $\hat{\tau}_\ell$ of X_t, we choose a small neighborhood $\hat{\tau}_\ell \pm \delta_n$ for some $\delta_n > 0$. Denote by $\bar{Y}_{\hat{\tau}_\ell +}$ and $\bar{Y}_{\hat{\tau}_\ell -}$ the averages of Y_{t_i} over $[\hat{\tau}_\ell, \hat{\tau}_\ell + \delta_n]$ and $[\hat{\tau}_\ell - \delta_n, \hat{\tau}_\ell)$, respectively. We use $\hat{L}_\ell = \bar{Y}_{\hat{\tau}_\ell +} - \bar{Y}_{\hat{\tau}_\ell -}$ to estimate the true jump size $L_\ell = X_{\tau_\ell} - X_{\tau_\ell -}$. Jump variation $\Psi = \sum_{\ell=1}^{N_1} L_\ell^2$ is then estimated by the sum of squares of all the estimated jump sizes

$$\widehat{\Psi} = \sum_{\ell=1}^{\hat{q}} \hat{L}_\ell^2 = \sum_{\ell=1}^{\hat{q}} (\bar{Y}_{\hat{\tau}_\ell +} - \bar{Y}_{\hat{\tau}_\ell -})^2. \tag{5}$$

The following theorem gives its rate of convergence.

Theorem 1 *Choose $\delta_n \sim n^{-1/2}$. Then under Conditions (A1)-(A4) in Section 6, we have as $n \to \infty$,*

$$\widehat{\Psi} - \Psi = O_P(n^{-1/4}).$$

Theorem 1 shows that the proposed jump variation estimator has convergence rate $n^{-1/4}$. Thanks to large n for high-frequency data, the error is usually small. In nonparametric regression jump size can usually be estimated with much higher convergence rates. The slower rate here is due to the fact that the low degree of smoothness of the sample path of the price process and our task of separating jumps from the less-smooth sample path is much more challenging. In contrast, the task of separating jumps from smooth curves is much easier.

The $n^{-1/4}$ convergence rate matches with the optimal convergence rate for estimating integrated volatility in the continuous diffusion price model (Zhang, 2004). This enables us to show later that the integrated volatility can be estimated for the jump-diffusion price model asymptotically as well as for the continuous diffusion price model.

4 Volatility Analysis

Volatility measures the variability of the continuous part of the log-price process X_t. After removing the jump part of X_t from the noisy observation Y_t, existing methods can be used to estimate the integrated volatility $\Theta = \int_0^1 \sigma_s^2\, ds$. In this section, we first show that the estimation errors on the jump sizes and locations have negligible effects on the estimation of integrated volatility and then apply the TSRV (Zhang, Mykland and Aït-Sahalia, 2005) and MSRV (Zhang, 2004) to estimate the integrated volatility. Finally, we introduce the wavelet realized volatility estimator to the jump adjusted data and establish the connection between TSRV with Haar wavelet realized volatility.

4.1 Data adjustment

Suppose our jump estimation shows that X_t has jumps at $\hat{\tau}_\ell$, with jump size $\hat{L}_\ell$, $\ell = 1, \cdots, \hat{q}$. Then the counting process N_t and the jump part of X_t are estimated by, respectively,

$$\hat{N}_t = \sum_{\ell=1}^{\hat{q}} 1(\hat{\tau}_\ell \le t), \qquad \hat{X}_t^d = \sum_{\ell=1}^{\hat{N}_t} \hat{L}_\ell = \sum_{\hat{\tau}_\ell \le t} \hat{L}_\ell. \tag{6}$$

To remove the jump effect from the data, we adjust data Y_{t_i} by subtracting them from the estimated jump part $\hat{X}_t^d$, resulting in

$$Y_{t_i}^* = Y_{t_i} - \hat{X}_{t_i}^d = Y_{t_i} - \sum_{\hat{\tau}_\ell \le t_i} \hat{L}_\ell, \qquad i = 1, \cdots, n. \tag{7}$$

For noisy high-frequency data under the continuous diffusion price model, good estimators of Θ are based on the average of RVs for various subsampled data. As the true jump locations and jump sizes are estimated accurately, jump effect can be largely removed. Below we will show that their impact on the average realized volatility is asymptotically negligible. To demonstrate this, let us introduce some notations. For integer K, partition the whole sample into K subsamples and define

$$[Y^*, Y^*]^{(K)} = \frac{1}{K} \sum_{k=1}^{K} \sum_{j=1}^{n/K} (Y^*_{t_{k+j\,K}} - Y^*_{t_{k+(j-1)\,K}})^2 = \frac{1}{K} \sum_{i=1}^{n-K} (Y^*_{t_i+K} - Y^*_{t_i})^2$$

be the average of K subsampled realized volatilities (ASRV) for adjusted data. For comparison, denote by X^c and Y^c the continuous parts of X and Y, respectively, namely,

$$X_t^c = \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dW_s, \qquad Y_t^c = X_t^c + \varepsilon_t. \tag{8}$$

Define ASRV for $Y_{t_i}^c$,

$$[Y^c, Y^c]^{(K)} = \frac{1}{K} \sum_{k=1}^{K} \sum_{j=1}^{n/K} (Y^c_{t_{k+j\,K}} - Y^c_{t_{k+(j-1)\,K}})^2 = \frac{1}{K} \sum_{i=1}^{n-K} (Y^c_{t_i+K} - Y^c_{t_i})^2.$$

Note that Y_t^c are not observable and are treated as ideal data for the purpose of theoretical comparison.

Theorem 2 *If $K/n + \log^2 n/K \to 0$, under Conditions (A1)-(A4) in Section 6, we have*

$$[Y^*, Y^*]^{(K)} = [Y^c, Y^c]^{(K)} + O_P(n^{-1/4} + K^{-1} \log^2 n).$$

Theorem 2 shows that the effect of jumps in ASRV is of order $n^{-1/4}$, if K is chosen to be of order lower than $n^{1/4}$.

4.2 TSRV for jump adjusted data

TSRV of Zhang *et al.*(2005) is used for volatility estimation under a continuous price model. Direct application of TSRV to the noisy data Y from the jump-diffusion price model yields an inconsistent estimator of integrated volatility, as it converges in probability to the quadratic variation process given by (3). Apply TSRV to the jump adjusted data Y^* and denote by JTSRV the obtained estimator of Θ, which is given by

$$\widehat{\Theta}_K = [Y^*, Y^*]^{(K)} - \frac{1}{K} [Y^*, Y^*]^{(1)}.$$

An immediate consequence of Theorem 2 is that the same asymptotic distribution for the TSRV estimator under the continuous price model in Zhang *et al.*(2005, theorem 4) also holds for JTSRV $\widehat{\Theta}_K$ under the jump-diffusion price model.

Theorem 3 *Under Conditions (A1)-(A5) in Section 6, and $K = c\, n^{2/3}$ for some $c > 0$, we establish that the limit distribution of $n^{1/6}\, (\widehat{\Theta}_K - \Theta)/\Upsilon$ is standard normal, where*

$$\Upsilon = \left(8\, c^{-2}\, \eta^2 + 8\, c \int_0^1 \sigma_t^4\, dt/3 \right)^{1/2}, \qquad \eta^2 = Var(\varepsilon_t).$$

Thus, for estimating integrated volatility, JTSRV under the jump-diffusion price model has the same performance asymptotically as TSRV under the continuous diffusion price model.

4.3 MSRV for jump adjusted data

For the continuous diffusion price model, Zhang (2006) used ASRV over many subsampling frequencies to construct MSRV and achieve the optimal convergence rate $n^{-1/4}$ for estimating Θ (Gloter and Jacod, 2001). Like TSRV, MSRV applied to observations with jumps yields an inconsistent estimator of integrated volatility. Instead, we apply MSRV to the jump adjusted data Y^*_{ti} and denote by JMSRV the resulting estimator of Θ. The resulting JMSRV is as follows,

$$\widehat{\Theta} = \sum_{m=1}^{M} a_m [Y^*, Y^*]^{(K_m)} + \zeta \, \left([Y^*, Y^*]^{(K_1)} - [Y^*, Y^*]^{(K_M)} \right),$$

where M is an integer, $K_m = m + C$, with C the integer part of $n^{1/2}$, and

$$a_m = \frac{12\,(m+C)(m-M/2-1/2)}{M\,(M^2-1)}, \qquad \zeta = \frac{(M+C)(C+1)}{(n+1)(M-1)}.$$

We derive the convergence rate of such a procedure by using Theorems 1 and 2 and Zhang (2006). Note that, unlike Zhang (2006), we need to take the partition numbers K_m at least as big as $n^{1/2}$ in order to apply Theorem 2 and obtain $n^{-1/4}$ convergence rate. Constant C is introduced in above procedure to achieve that goal.

Theorem 4 *Under Conditions (A1)-(A5) in Section 6, and $M \sim n^{1/2}$, we have*

$$\widehat{\Theta} - \Theta = O_P(n^{-1/4}).$$

Theorem 4 shows that like the continuous diffusion price model, we can estimate the integrated volatility at the optimal convergence rate $n^{-1/4}$ for the jump-diffusion price model. Thus, jumps have no asymptotic impact on the volatility estimation.

4.4 Wavelet volatility for jump adjusted data

Before introducing our estimation method, some heuristic discussions are helpful. Let

$$y_i^* = Y_{t_i}^* - Y_{t_{i-1}}^*, \;\; x_i^c = X_{t_i}^c - X_{t_{i-1}}^c = \int_{t_{i-1}}^{t_i} \mu_s\, ds + \int_{t_{i-1}}^{t_i} \sigma_u\, dW_u.$$

Then, the realized volatility for the ideal data $\{X_{t_i}^c\}$ defined by (8) is

$$[X^c, X^c]^{(1)} = \sum_{i=1}^{n} (x_i^c)^2.$$

As $\{x_i^c\}$ are not observable, it is natural to replace $[X^c, X^c]^{(1)}$ by $\sum_{i=1}^n (y_i^*)^2$. Each y_i^* is contaminated with noise of level η, the summation cumulates the noise of order n, which dominates Θ. This is shown in Zhang et al. (2005) and also evidenced in Figure 1(a). Thus, we need to remove noise ε_t from y_i^* first before computing the realized volatility. Since X_t^c and ε_t are independent, after conditioning on X_t^c, the problem is the same as that for nonparametric regression (Vidakovic, 1999, chapter 6). Hence, the wavelet technique can be used to denoise the data.

Consider the wavelet coefficients of x_i^c. Most volatility information x_i^c is stored in relatively a small number of large wavelet coefficients at low and middle levels, whereas the information about noise is contained in wavelet coefficients at very high levels. The sum of squares contributed by the relatively small number of large wavelet coefficients at low and middle levels almost accounts for $[X^c, X^c]^{(1)}$. Similar to stationary wavelet transformation, to better suppress noise we consider data shifts and use wavelet transformation of each shifted data set to form the sum of the squared wavelet coefficients. We then take the average of all the sums as an estimator of integrated volatility.

To define the estimator specifically, we introduce some notations. Given a sequence of z_i, $i = 1, \cdots, n$, define the shift operator

$$(\mathcal{S}z)_i = z_{i+1}, \text{ for } i = 1, \cdots, n-1, \quad \text{and } (\mathcal{S}z)_i = 0, \text{ for } i \geq n.$$

Denote by $y^\ell_{j,k}$ the wavelet transformations of $(\mathcal{S}^{\ell-1}y^*)_i$, $\ell = 1, \cdots, K = 2^{J_n}$ for some integer $J_n < \log_2 n$. To estimate integrated volatility, we form the sum of squares of the wavelet coefficients $y^\ell_{j,k}$ up to the level $\log_2 n - J_n$.

Define wavelet RV (WRV)

$$\text{WRV} = \frac{1}{n} \sum_{\ell=1}^{2^{J_n}} \sum_{j=1}^{\log_2 n - J_n} \sum_{k=1}^{2^j} \left(y^\ell_{j,k}\right)^2,$$

and WRV estimator of integrated volatility Θ

$$\widehat{\Theta}_W = \text{WRV} - 2^{-J_n} [Y^*, Y^*]^{(1)}.$$

The following theorem relates $\widehat{\Theta}_W$ under Haar wavelet to JTSRV estimator $\widehat{\Theta}_K$.

Theorem 5 *When Haar wavelet is used, the WRV estimator $\widehat{\Theta}_W$ is equal to the JTSRV estimator $\widehat{\Theta}_K$ with partition number $K = 2^{J_n}$.*

Theorem 5 shows that JTSRV corresponds to the WRV estimator with Haar wavelet. Thus, they share the same asymptotic distribution. When smooth wavelets are used, the WRV estimators are different from the JTSRV estimator. It is an interesting problem to study the asymptotic behavior of the WRV estimators under smooth wavelets.

5 Numerical studies

5.1 Simulations

Simulated high-frequency data are minute by minute observations for 24 hours from model (1) and (2) with drift in log price being equal to zero, that is, 1440 equally spaced observations are simulated from the model. To obtain the 1440 observations, we first use the Euler scheme to simulate a sample path of σ_t^2 from the following Geometric Ornstein-Uhlenbeck volatility model,

$$d \log \sigma_t^2 = -(0.6802 + 0.10 \log \sigma_t^2)\, dt + 0.25\, dW_{1,t}, \qquad \text{Corr}(W_t, W_{1,t}) = -0.62.$$

Better simulation schemes (Fan, 2005) can also be used, but the difference is expected to be small. The 1440 values of $\int_0^{t_i} \sigma_s \, dW_s$, $i = 1, \cdots, 1440$, are then approximated by the discrete summations. These 1440 values form a simulated sample path of the continuous part of X_t and are denoted by $X_{t_i}^c$. For the jump-diffusion model, two jump cases are considered: X_t has one jump or three jumps over $[0, 1]$. The simulated 1440 values of the sample path of X_t with one or three jumps are obtained by adding one or three jumps to $X_{t_i}^c$, respectively, with jump locations being uniformly distributed on $[0, 1]$ and jump size following i.i.d. normal distribution with mean zero and standard deviation $1/30$. Denote the simulated values by X_{t_i}, $i = 1, \cdots, 1440$. The observed data $\{Y_{t_i}\}$ are obtained by adding i.i.d. noise $\varepsilon_{t_i} \sim N(0, \eta^2)$ to X_{t_i}. The value of η ranges from 0 to 0.001. We repeat the whole simulation process 5000 times and evaluate MSEs of the estimators based on the 5000 repetitions.

In the simulation study, Daubechies s8 wavelet was used in the calculation of wavelet coefficients for jump estimation and WRV. Although Theorem 5 shows that TSRV is equivalent to WRV with Haar wavelet, WRV with non-Haar wavelet like s8 wavelet is different from TSRV. JTSRV was evaluated with K being 5 and 15, which corresponds to 5- and 15-minute returns, respectively. We computed WRV using wavelet coefficients at the first 8 levels. JMSRV was computed with $M = 11$ and $K_m = 5, \cdots, 15$. To demonstrate the performances of the proposed estimators we compared them with RBPV, RV and TSRV estimators directly applied to the same data without jump adjustments. RBPV was evaluated for all data and subsampled data corresponding to 5- and 15-minute returns, RV was computed based on all data, and TSRV was applied with $K = 5$ and 15 partition groups corresponding to 5- and 15-minute returns, respectively.

We plotted the MSEs of the estimators against noise level for continuous price process (no jumps) in Figure 2, and for the price process with one and three jumps, respectively, in Figures 3 and 4. In each of Figures 2-4, (a), (b) and (c) plot the MSEs of RBPV, RV and TSRV, and JTSRV and WRV, respectively, along with the MSE of JMSRV for the purpose of comparison. All MSEs are multiplied by 10^4 to facilitate the presentation.

For the case of continuous price process, Figure 2 shows that, when there is no microstructure noise or noise level is very low, RV and RBPV based on all data are the best. However, they are very sensitive to noise. As the noise level increases, their performances quickly deteriorate, and they are overwhelmingly dominated by TSRV, MSRV, JTSRV, JMSRV and WRV estimators. For the subsampling based estimators, the higher the noise level is, the lower the subsampling frequency that should be used to make their MSEs smaller. Generally the MSEs of MSRV and JMSRV are between those of TSRV and JTSRV for 5- and 15-minute returns, respectively. In this case

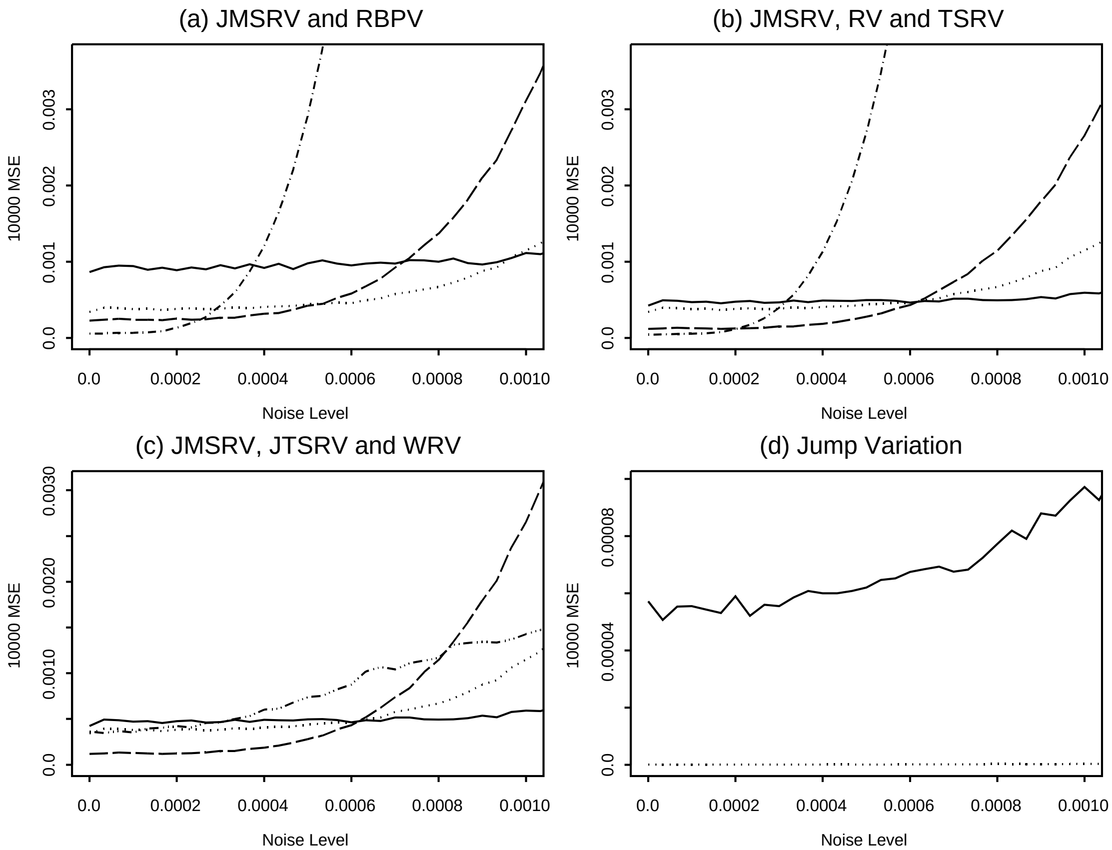

Figure 2: MSE (multiplied by 10^4) plots for continuous price process (no jumps). (a) MSE using JMSRV (dotted curve) and RBPV based on all data (dot-dash curve) and 5-minute (dash curve) and 15-minute (solid curve) subsampled data. (b) MSE using JMSRV (dotted curve), RV(dot-dashed curve) and TSRV with $K = 5$ (dashed curve) and $K = 15$ (solid curve). (c) MSE using JMSRV (dotted curve), JTSRV with $K = 5$ (dashed curve) and $K = 15$ (solid curve) and WRV (dot-dashed curve). (d) MSEs of jump variation estimation using RBPV (solid curve) and wavelets (dotted curve).

MSRV and JMSRV have almost the same MSE value. TSRV, MSRV, JTSRV, JMSRV and WRV are noise resistant and have comparable performance, but none achieve the smallest MSE uniformly over the entire range of the noise level.

For the case of price process with jumps, Figures 3 and 4 demonstrate that the proposed estimators outperform all existing estimators. RV, TSRV and MSRV are very sensitive to jumps. In this case, it is necessary to remove the jumps first before calculating the estimated integrated volatility. Regardless of noise level, one jump in the price process is enough to destroy RV, TSRV and MSRV whose MSEs are far larger than those of JTSRV, JMSRV and WRV. This is clearly demonstrated by the fact that the MSE of MSRV is between those of TSRV estimators and the rescaled

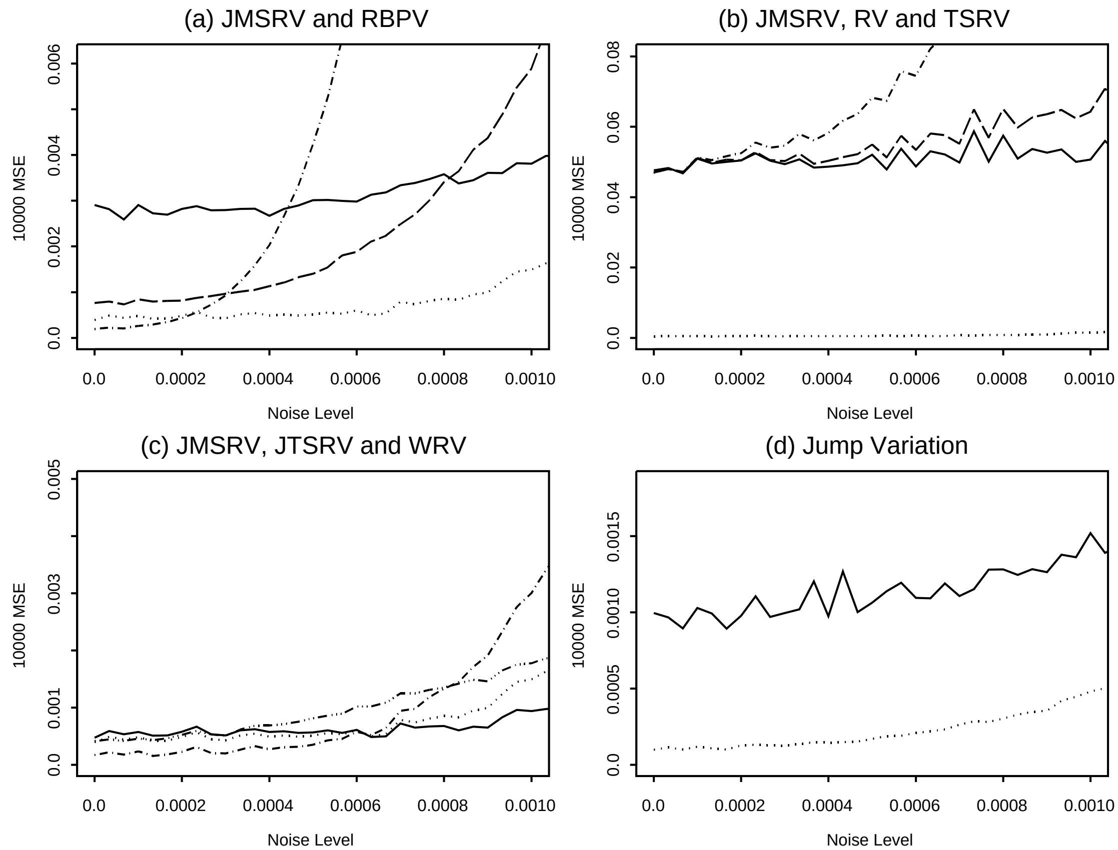

Figure 3: MSE (multiplied by 10^4) plots for price process with one jump. The same captions as those in Figure 2 are used.

MSEs of RV and TSRV in Figure 3(b) have values far above 0.04, while the rescaled MSEs of the proposed JMSRV, JTSRV and WRV in Figure 3(c) are close to 0.001. In the case of one jump, Figures 3(a) and 3(c) imply that RBPV based estimators generally have larger MSEs than the proposed JMSRV, JTSRV and WRV. RBPV is proposed to handle jumps for data without noise. For very low noise levels, RBPV for all data has slightly less MSE than the JMSRV. But as noise level increases, its MSE increases dramatically, with MSE values many times larger than that of the JMSRV. When the price process has three jumps, Figures 4(a) and 4(c) indicate that the proposed estimators obviously outperform RBPV based estimators at all noise levels, and again Figure 4(b) suggests that the behaviors of RV, TSRV and MSRV estimators are completely ruined by the jumps.

For the estimation of jump variation Ψ in the jump-diffusion price model, we have checked the performance of the proposed estimator $\hat{\Psi}$ in (5) and compared it with the existing estimator in Barndorff-Nielsen and Shephard (2006), which is the

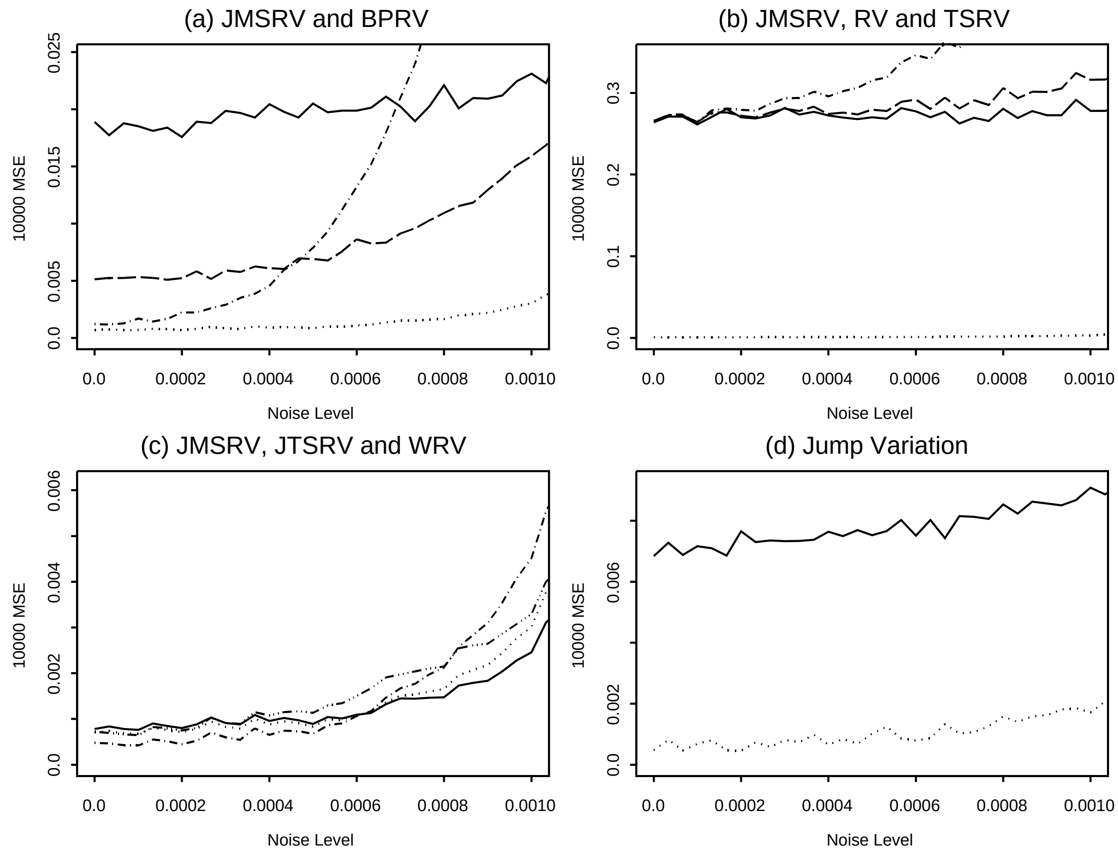

Figure 4: MSE (multiplied by 10^4) plots for price process with three jumps. The same captions as those in Figure 2 are used.

difference of RV and $\pi/2$ of RBPV of order $(1,1)$. We evaluated the existing estimator based on all and subsampled data and found that the MSEs are very close. Figures 3(d) and 4(d) plot, against noise level, the MSEs of the proposed estimator and the existing estimator based on 5-minute returns for one and three jumps, respectively. When jumps are present in the log price process, we have also checked the number of estimated jumps. The percentage of finding correct number of jumps ranges from 90% to 99% over the noise level. For the continuous price case, there is no jump and the jump variation is zero. To check the performances of the proposed estimator and the existing estimator in this case, we have plotted their MSEs in Figure 2(d). All plots show that the proposed estimators have excellent performance and are overwhelmingly better than the existing estimator.

The noise level relative to asset price found in high frequency data typically ranges from 0.001% to 0.01%. We take the ratio of η^2 and average integrated volatility as a crude proxy for the relative noise level in the simulated model and find that

$\eta \in [0, 0.001]$ in the simulation plots closely matches with the range of the relative noise level observed in real data. Indeed, since the Ornstein-Uhlenbeck process for $\log \sigma_t^2$ has a normal stationary distribution with mean -6.8020 and variance 0.3125, under this stationary distribution the expected integrated volatility $\int_0^1 E[\sigma_t^2]\, ds$ is equal to 0.0012995; The requirement of the proxy $\eta^2/0.0012995 \in [0.001\%, 0.01\%]$ is translated into $\eta \in [0.014\%, 0.14\%]$, which is close to $\eta \in [0, 0.001]$. Actually we have tried on over a much more wide range of values for η. It turns out that the plotted figures with $\eta \in [0, 0.001]$ display all MSE patterns found over the wide range of η values for the volatility estimators under the study.

5.2 Applications

We have applied the proposed methods to two exchange rate data sets: one minute Euro-dollar and Yen-dollar exchange rates for the first seven months in 2004. The data sets were obtained from Quickstars, L. L. C., Connecticut, U.S.A. Figures 5(a) and 6(a) show the plots of the two exchange rate data. We applied our jump estimation to the exchange rate data for each day and plotted the estimated number of daily jumps in Figures 5(b) and 6(b). The plots show that jumps occur very often in the exchange rates. We computed WRV and the proposed estimator of jump variation for each day and plotted the estimated daily integrated volatility and jump variation in Figures 5(c-d) and 6(c-d). For these two data sets, jump variations often make significant contribution to total variations. To quantify the contribution, we calculated the ratios of the daily estimated jump variations and the daily integrated volatilities and plotted them in Figure 7. The percentages of days with jump variation exceeding 10% and 20% of integrated volatility are, respectively, 41% and 24% for Euro-dollar exchange rates and 30% and 19% for Yen-dollar exchange rates. For several days, the estimated jump variations even surpass the estimated integrated volatilities. This shows that ignoring jump effect in volatility estimation can inflate the volatility substantially.

6 Conclusions

We develop nonparametric methods to estimate jumps and jump variations for noisy data from a jump-diffusion model. With the estimated jumps we adjust data for jumps and then propose to apply TSRV and MSRV to the jump adjusted data for the estimation of integrated volatility. We also construct WRV from the jump adjusted data to estimate integrated volatility. Asymptotic theory established for the proposed estimators shows that the integrated volatility can be estimated under the jump-diffusion model asymptotically as well as under the continuous diffusion model. Simulations demonstrate that the proposed estimators have comparable performance

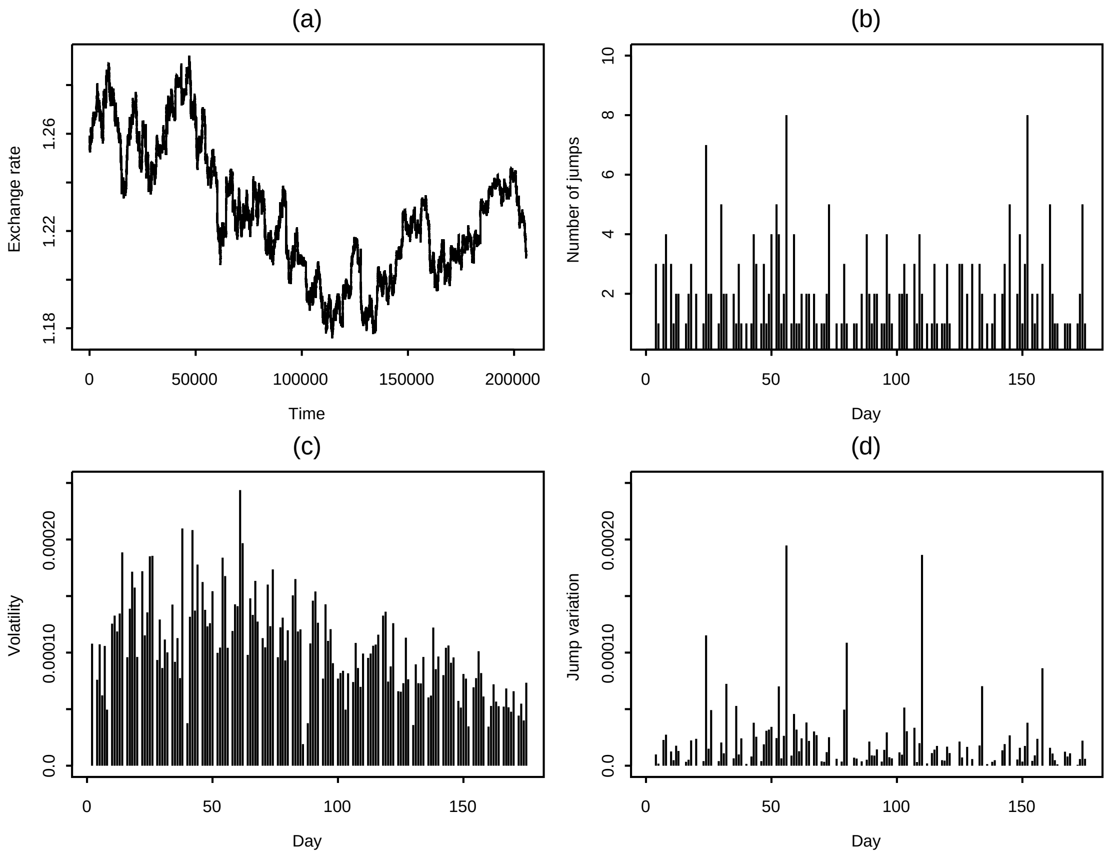

Figure 5: (a) Euro-dollar exchange rates from January to July, 2004. (b) The number of estimated jumps in each day. (c) Estimated daily integrated volatility. (d) Estimated daily jump variation.

with existing methods for either noisy data from the continuous diffusion model or noiseless data from the jump-diffusion model, but they outperform existing methods when data contain jumps. An application of the jump and volatility estimation methods to two high-frequency Euro-dollar and Yen-dollar exchange rate data reveals important information in the exchange rates.

The paper initiates a new research direction on jump and volatility estimation in the field of high-frequency financial data. The proposed methods will stimulate more research on multiscale methods. For example, the paper leaves some issues and raises interesting problems for future investigation. They include asymptotic theory of WRV under smooth wavelets, the performance of the proposed methods for data with infinitely many jumps modeled by Lévy processes, characterization of log price process whose sample path has sparse representations under wavelets and other bases, and refinement of WRV with thresholding techniques to take full advantage of bases used.

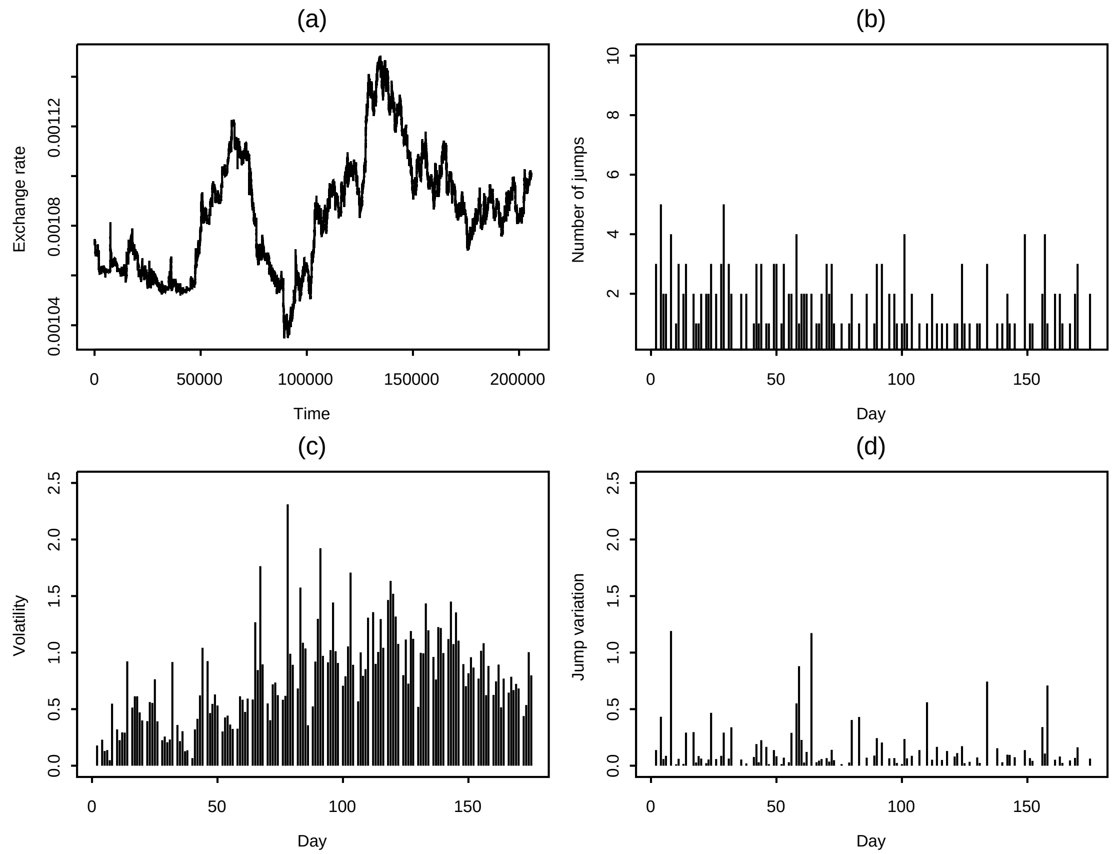

Figure 6: (a) Japanese Yen-dollar exchange rates from January to July, 2004. (b) The number of estimated jumps in each day. (c) Estimated daily integrated volatility. (d) Estimated daily jump variation.

7 Technical conditions and Proofs

We first state the technical conditions on model (1) and (2) that are needed for the technical proof and then outline the proofs of the results.

(A1). Wavelets (ψ, ϕ) used in jump estimation are differentiable.

(A2). μ_t and σ_t^2 are continuous in t almost surely and satisfy

$$E\left(\max_{0\le t\le 1} \mu_t^2\right) < \infty, \qquad E\left(\max_{0\le t\le 1} \sigma_t^2\right) < \infty.$$

(A3). For the jump part of X_t in $[0,1]$, its jump locations τ_ℓ and jump sizes L_ℓ obey

$$q = N_1 < \infty, \qquad \tau_1 < \tau_2 < \cdots < \tau_\ell < \cdots, \qquad 0 < |L_\ell| < \infty, \qquad \text{almost surely.}$$

(A4). (μ_t, σ_t^2, W_t), (N_t, L_ℓ) and ε_t are independent. ε_t are i.i.d. with $E(\varepsilon_t^4) < \infty$.

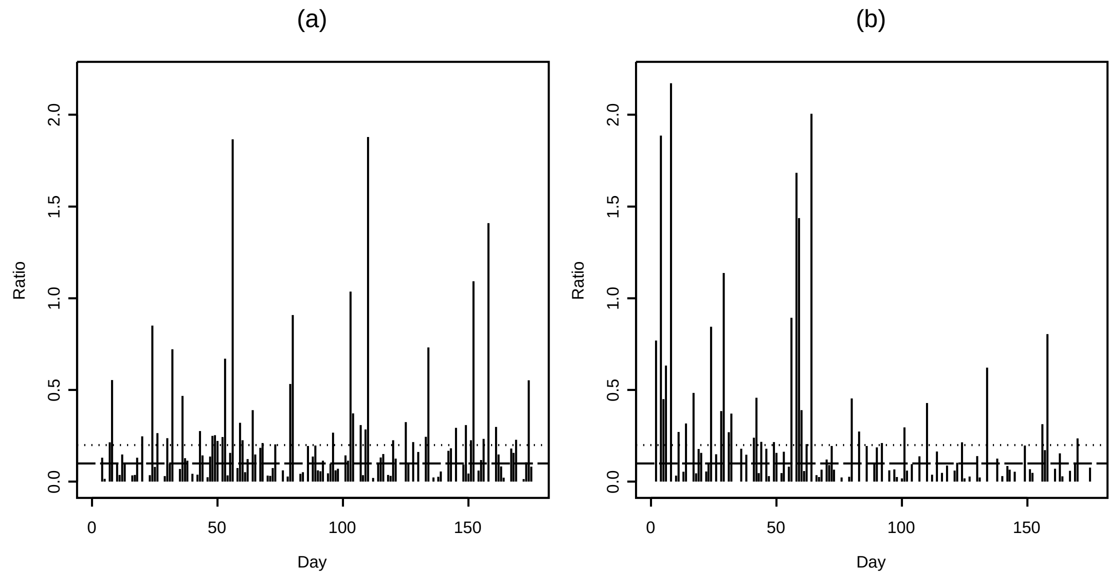

Figure 7: The ratios of estimated daily jump variations and integrated volatilities for (a) Euro-dollar and (b) Japanese Yen-dollar exchange rates. Two horizontal lines are at 10% and 20% levels.

(A5). The continuous part of X_t is an Itô process, with (μ_t, σ_t) adapted to the complete filtration generated by Brownian motion W_t.

We first prove two propositions.

Proposition 1 *Suppose that Ξ is a finite non-negative discrete random variable, and V_ℓ, $\ell = 1, 2, \cdots$, are a sequence of random variables.*

(a). If $V_\ell \neq 0$ for each fixed ℓ, then $\min\{|V_\ell|, \ell = 1, \cdots, \Xi\} > 0$.

(b). If $|V_\ell| < \infty$ for each fixed ℓ, then $\max\{|V_\ell|, \ell = 1, \cdots, \Xi\} < \infty$.

Proof. (a). The assumption implies that for each ℓ,

$$P(|V_\ell| = 0) = 0.$$

For a fixed integer m,

$$P\left(\min_{1\leq\ell\leq m} |V_\ell| = 0\right) \leq \sum_{i=1}^{m} P(|V_\ell| = 0) = 0,$$

thus $\min\{|V_\ell|, \ell = 1, \cdots, m\} > 0$. As Ξ takes non-negative integer values,

$$P\left(\min_{1\leq\ell\leq\Xi} |V_\ell| = 0\right) = \sum_{m=1}^{\infty} P\left(\min_{1\leq\ell\leq m} |V_\ell| = 0, \Xi = m\right) \leq \sum_{m=1}^{\infty} P\left(\min_{1\leq\ell\leq m} |V_\ell| = 0\right) = 0,$$

which proves the first assertion.

(b). From the assumption we have that for each ℓ,

$$\lim_{d\to\infty} P(|V_\ell| > d) = P(|V_\ell| = \infty) = 0.$$

For a fixed integer m, we have inequality

$$P\left(\max_{1\le\ell\le\Xi} |V_\ell| > d\right) \le P\left(\Xi \le m, \max_{1\le\ell\le m} |V_\ell| > d\right) + P(\Xi > m)$$
$$\le P\left(\max_{1\le\ell\le m} |V_\ell| > d\right) + P(\Xi > m)$$
$$\le \sum_{\ell=1}^{m} P(|V_\ell| > d) + P(\Xi > m).$$

We fix m, let $d \to \infty$ in above inequality and obtain

$$\lim_{d\to\infty} P\left(\max_{1\le\ell\le\Xi} |V_\ell| > d\right) \le \sum_{\ell=1}^{m} \lim_{d\to\infty} P(|V_\ell| > d) + P(\Xi > m) = P(\Xi > m).$$

As the left hand side of above equality is free of m, and Ξ is a finite discrete random variable, letting m tend to infinity we conclude

$$\lim_{d\to\infty} P\left(\max_{1\le\ell\le\Xi} |V_\ell| > d\right) \le \lim_{m\to\infty} P(\Xi > m) = 0,$$

which proves the second assertion.

Proposition 2 *Suppose that Ξ is a finite non-negative discrete random variable, $V_{n,\ell}$, $\ell = 1, 2, \cdots$, $n = 1, \cdots$, are random variables, and r_n, $n = 1, 2, \cdots$, are real numbers. Assume that as $n \to \infty$, $V_{n,\ell} = O_P(r_n)$ for each fixed ℓ. Then*

(a).

$$\max\{|V_{n,\ell}|, \ell = 1, \cdots, \Xi\} = O_P(r_n).$$

(b).

$$\sum_{\ell=1}^{\Xi} |V_{n,\ell}| = O_P(r_n).$$

Proof. (a). Note that the assumption $V_{n,\ell} = O_P(r_n)$ means that for each ℓ,

$$\lim_{d\to\infty} \lim_{n\to\infty} P(|V_{n,\ell}| > d\, r_n) = 0.$$

Similar to the proof of Proposition 1(b), the following inequality holds for fixed integer m,

$$P\left(\max_{1\le\ell\le\Xi} |V_{n,\ell}| > d\, r_n\right) \le P\left(\max_{1\le\ell\le\Xi} |V_{n,\ell}| > d\, r_n, \Xi \le m\right) + P(\Xi > m)$$
$$\le P\left(\max_{1\le\ell\le m} |V_{n,\ell}| > d\, r_n\right) + P(\Xi > m)$$
$$\le \sum_{\ell=1}^{m} P(|V_{n,\ell}| > d\, r_n) + P(\Xi > m).$$

We take limit on both sides of above inequality by letting n and then d tend to infinity sequentially and yield that for fixed m,

$$\lim_{d\to\infty}\lim_{n\to\infty} P\left(\max_{1\le\ell\le\Xi}|V_{n,\ell}|>d\,r_n\right)\le\sum_{\ell=1}^{m}\lim_{d\to\infty}\lim_{n\to\infty}P(|V_{n,\ell}|>d\,r_n)+P(\Xi>m)$$
$$=P(\Xi>m).$$

Again as the left hand side of the inequality does not depend on m, and Ξ is a finite random variable free of n, letting m tend to infinity we arrive at

$$\lim_{d\to\infty}\lim_{n\to\infty} P\left(\max_{1\le\ell\le\Xi}|V_{n,\ell}|>d\,r_n\right)\le\lim_{m\to\infty}P(\Xi>m)=0,$$

which proves the first claim.

(b). Note that

$$\sum_{\ell=1}^{\Xi}|V_{n,\ell}|\le\Xi\max_{1\le\ell\le\Xi}|V_{n,\ell}|.$$

The second claim is an immediate consequence of Part (a) and $\Xi=O_P(1)$. Indeed, the following inequality holds for any constant $d_1>0$,

$$P\left(\sum_{\ell=1}^{\Xi}|V_{n,\ell}|>d_1\,r_n\right)\le P\left(\Xi\max_{1\le\ell\le\Xi}|V_{n,\ell}|>d_1\,r_n\right)$$
$$\le P\left(\max_{1\le\ell\le\Xi}|V_{n,\ell}|>\sqrt{d_1}\,r_n\right)+P\left(\Xi>\sqrt{d_1}\right);$$

taking limit on both sides and use the proved part (a) and finiteness of Ξ, we obtain

$$\lim_{d_1\to\infty}\lim_{n\to\infty}P\left(\sum_{\ell=1}^{\Xi}|V_{n,\ell}|>d_1\,r_n\right)\le\lim_{d_1\to\infty}\lim_{n\to\infty}P\left(\max_{1\le\ell\le\Xi}|V_{n,\ell}|>\sqrt{d_1}\,r_n\right)$$
$$+\lim_{d_1\to\infty}P\left(\Xi>\sqrt{d_1}\right)=0.$$

Proof of Theorem 1. It is easy to show that the following inequality holds for any $d>0$,

$$P\left(|\hat\Psi-\Psi|>d\,n^{-1/4}\right)\le P\left(\hat q=q,|\hat\Psi-\Psi|>d\,n^{-1/4}\right)+P(\hat q\ne q).$$

From (4) we have that $\lim_{n\to\infty}P(\hat q\ne q)\to 0$. Hence,

$$\lim_{d\to\infty}\lim_{n\to\infty}P\left(|\hat\Psi-\Psi|>d\,n^{-1/4}\right)\le\lim_{d\to\infty}\lim_{n\to\infty}P\left(\hat q=q,|\hat\Psi-\Psi|>d\,n^{-1/4}\right).\qquad(9)$$

With $\hat q=q$, we have

$$\hat\Psi-\Psi=\sum_{\ell=1}^{q}\left\{\hat L_\ell^2-L_\ell^2\right\}=\sum_{\ell=1}^{q}\left\{\Pi_{n,\ell}^2+\Pi_{n,\ell}\,L_\ell\right\},\qquad(10)$$

where

$$\Pi_{n,\ell} = \hat{L}_\ell - L_\ell = (\bar{Y}_{\hat{\tau}_\ell+} - \bar{Y}_{\hat{\tau}_\ell-}) - (X_{\tau_\ell} - X_{\tau_\ell-}). \tag{11}$$

We state the following claim (whose proof will be given later) that for each ℓ,

$$\Pi_{n,\ell} = O_P(n^{-1/4}). \tag{12}$$

This claim and $L_\ell = O_P(1)$ imply that for each ℓ, $\Pi_{n,\ell}\, L_\ell = O_P(n^{-1/4})$ and $\Pi_{n,\ell}^2 = O_P(n^{-1/2})$. Applying Proposition 2 to $q = N_1$ sums of $\Pi_{n,\ell}^2$ and $|\Pi_{n,\ell}\, L_\ell|$ we obtain

$$\sum_{\ell=1}^{q} |\Pi_{n,\ell}\, L_\ell| = O_P(n^{-1/4}), \qquad \sum_{\ell=1}^{q} \Pi_{n,\ell}^2 = O_P(n^{-1/2}),$$

that is,

$$\lim_{d_1\to\infty} \lim_{n\to\infty} P\left(\sum_{\ell=1}^{q} |\Pi_{n,\ell}\, L_\ell| > d_1\, n^{-1/4}\right) = 0, \tag{13}$$

$$\lim_{d_1\to\infty} \lim_{n\to\infty} P\left(\sum_{\ell=1}^{q} \Pi_{n,\ell}^2 > d_1\, n^{-1/2}\right) = 0. \tag{14}$$

Putting together (10), (13) and (14) we deduce

$$P\left(\hat{q} = q, |\hat{\Psi} - \Psi| > d\, n^{-1/4}\right) \le P\left(\sum_{\ell=1}^{q} \left\{\Pi_{n,\ell}^2 + |\Pi_{n,\ell}\, L_\ell|\right\} > d\, n^{-1/4}\right)$$
$$\le P\left(\sum_{\ell=1}^{q} \Pi_{n,\ell}^2 > d\, n^{-1/4}/2\right) + P\left(\sum_{\ell=1}^{q} |\Pi_{n,\ell}\, L_\ell| > d\, n^{-1/4}/2\right) \to 0,$$

as n and then d sequentially go to infinity. Combining the above result and (9) we derive that $\hat{\Psi} - \Psi$ is $O_P(n^{-1/4})$.

The proof left is to show claim (12). Because of similarity, we prove the claim for $\ell = 1$. Denote by $m_\pm$ the number of t_i in $\hat{\mathcal{I}}_+ = [\hat{\tau}_1, \hat{\tau}_1 + \delta_n]$ and $\hat{\mathcal{I}}_- = [\hat{\tau}_1 - \delta_n, \hat{\tau}_1)$, respectively. Then $|m_+ - m_-| \le 2$, $m_+ \sim m_- \sim n^{1/2}$, and

$$\bar{Y}_{\hat{\tau}_1+} = \frac{1}{m_+} \sum_{t_i \in \hat{\mathcal{I}}_+} Y_{t_i}, \qquad \bar{Y}_{\hat{\tau}_1-} = \frac{1}{m_-} \sum_{t_i \in \hat{\mathcal{I}}_-} Y_{t_i}.$$

As Y_t contains noise, drift, diffusion and jump terms, denote by U_i the differences between the two averages over $\hat{\mathcal{I}}_\pm$ for the four terms, respectively. Then,

$$(\bar{Y}_{\hat{\tau}_1+} - \bar{Y}_{\hat{\tau}_1-}) - (X_{\tau_1} - X_{\tau_1-}) = \sum_{i=1}^{4} U_i - L_{N_{\tau_1}}, \tag{15}$$

and it is enough to show that U_1, U_2, U_3, and $U_4 - L_{N_{\tau_1}}$ are $O_P(n^{-1/4})$.

First of all, let us consider the noise term. U_1 is the difference of two averages of m_+ and m_- i.i.d. random variables ε_t, respectively, so U_1 has mean zero and variance $\eta^2\,(m_+^{-1}+m_-^{-1}) = O(n^{-1/2})$. Hence, $U_1 = O_P(n^{-1/4})$.

Next, consider drift term. Use Condition (A2), it is easy to see that

$$\begin{aligned} U_2 &= \frac{1}{m_+}\sum_{t_i\in\hat{\mathcal{I}}_+}\int_{\hat{\tau}_1}^{t_i}\mu_s\,ds - \frac{1}{m_-}\sum_{t_i\in\hat{\mathcal{I}}_-}\int_{t_i}^{\hat{\tau}_1}\mu_s\,ds \\ &= \frac{1}{m_+}O_P\left(\sum_{i=1}^{m_+} i/n\right) + \frac{1}{m_-}O_P\left(\sum_{i=1}^{m_-} i/n\right) = O_P(n^{-1/2}). \end{aligned}$$

Third, consider the continuous diffusion term. Note that

$$U_3 = \frac{1}{m_+}\sum_{t_i\in\hat{\mathcal{I}}_+}\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s - \frac{1}{m_-}\sum_{t_i\in\hat{\mathcal{I}}_-}\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s. \tag{16}$$

Let $\mathcal{I}_+ = [\tau_1, \tau_1+\delta_n]$ and $\mathcal{I}_- = [\tau_1-\delta_n, \tau_1)$. We will replace $\hat{\mathcal{I}}_\pm$ in above summation indices by $\mathcal{I}_\pm$, respectively, and show that the resulting difference satisfies

$$\sum_{t_i\in\hat{\mathcal{I}}_\pm}\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s - \sum_{t_i\in\mathcal{I}_\pm}\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s = O_P(n^{-1/4}\,\log^{5/2}\,n). \tag{17}$$

Because of similarity, we prove (17) for the case of $\hat{\mathcal{I}}_+$ and $\mathcal{I}_+$. The difference between $\sum_{t_i\in\hat{\mathcal{I}}_+}\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s$ and $\sum_{t_i\in\mathcal{I}_+}\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s$ involves those terms $\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s$ with t_i falling to exact one of the two intervals $\hat{\mathcal{I}}_+$ and $\mathcal{I}_+$. Since $\hat{\tau}_1-\tau_1$ is $O_P(n^{-1}\,\log^2\,n)$, the total number of such terms is $O_P(\log^2\,n)$. We get (17) by multiplying $O_P(\log^2\,n)$ with $O_P(n^{-1/4}\,\log^{1/2}\,n)$, the order in probability for $\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s$, which can be established as follows.

$$\begin{aligned} &P\left(\max\left\{\left|\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s\right|, t_i\in\hat{\mathcal{I}}_+\cup\mathcal{I}_+\right\} > n^{-1/4}\,\log^{1/2}\,n\right) \\ &\leq P\left(\max_{\tau_1-\delta_n\leq t_i\leq\tau_1+2\,\delta_n}\left|\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s\right| > n^{-1/4}\,\log^{1/2}\,n\right) + P(|\hat{\tau}_1-\tau_1|>\delta_n), \end{aligned}$$

(4) implies $P(|\hat{\tau}_1-\tau_1|>\delta_n) = o(1)$. Because of Condition (A4), (σ, W) and τ_1 are independent, and hence conditioning on τ_1, $\left\{\int_{\tau_1-2\,\delta_n}^{t}\sigma_s\,dW_s, t\geq\tau_1-2\,\delta_n\right\}$ is a martingale. We apply the maximal martingale inequality (Kloeden and Platen 1999, equation 3.6 in section 2.3) to the martingale and obtain

$$\begin{aligned} &P\left(\max_{\tau_1-\delta_n\leq t_i\leq\tau_1+2\,\delta_n}\left|\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s\right| > n^{-1/4}\,\log^{1/2}\,n\,\middle|\,\tau_1\right) \\ &\leq \frac{1}{n^{-1/2}\,\log\,n}E\left[\left(\int_{\tau_1-2\,\delta_n}^{\tau_1+2\,\delta_n}\sigma_s\,dW_s\right)^2\middle|\,\tau_1\right] \\ &= \frac{1}{n^{-1/2}\,\log\,n}E\left[\int_{\tau_1-2\,\delta_n}^{\tau_1+2\,\delta_n}\sigma_s^2\,ds\,\middle|\,\tau_1\right], \end{aligned}$$

and therefore,

$$
\begin{aligned}
&P\left(\max_{\tau_1-\delta_n\le t_i\le\tau_1+2\,\delta_n}\left|\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s\right|>n^{-1/4}\,\log^{1/2}\,n\right)\\
&=E\left[P\left(\max_{\tau_1-\delta_n\le t_i\le\tau_1+2\,\delta_n}\left|\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s\right|>n^{-1/4}\,\log^{1/2}\,n\right|\tau_1\right)\right]\\
&\le\frac{1}{n^{-1/2}\,\log\,n}E\left[\int_{\tau_1-2\,\delta_n}^{\tau_1+2\,\delta_n}\sigma_s^2\,ds\right]\\
&\le\frac{1}{n^{-1/2}\,\log\,n}\int_{-2\,\delta_n}^{2\,\delta_n}E\left(\sigma_{\tau_1+s}^2\right)\,ds\\
&\le E\left(\max_{0\le t\le 1}\sigma_t^2\right)\,O(\log^{-1}\,n)\to 0,
\end{aligned}
$$

where Condition (A2) is used in the last limit.

Substituting (17) into (16) we obtain

$$
U_3=\frac{1}{m_+}\sum_{t_i\in\mathcal{I}_+}\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s-\frac{1}{m_-}\sum_{t_i\in\mathcal{I}_-}\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s+O_P(n^{-3/4}\,\log^{5/2}\,n).\quad(18)
$$

Denote by P_τ, E_τ and Var_τ conditional probability, mean, and variance given τ_1. Note that τ_1 is independent of (σ,W). Given τ_1, for $|t_i-\tau_1|\le\delta_n$, $\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s$ has conditional mean zero and conditional covariance

$$
\mathrm{E}_\tau\left(\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s\int_{\tau_1-2\,\delta_n}^{t_j}\sigma_s\,dW_s\right)=\int_{\tau_1-2\,\delta_n}^{t_i\wedge t_j}E(\sigma_s^2)\,ds\le(t_i\wedge t_j-\tau_1+2\,\delta_n)\max_{0\le s\le 1}E\left(\sigma_s^2\right).
$$

Hence, it is easy to compute the following conditional mean and variance

$$
\frac{1}{m_+}\sum_{t_i\in\mathcal{I}_+}E_\tau\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s=\frac{1}{m_-}\sum_{t_i\in\mathcal{I}_-}E_\tau\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s=0,\qquad(19)
$$

$$
\begin{aligned}
&\mathrm{Var}_\tau\left(\frac{1}{m_+}\sum_{t_i\in\mathcal{I}_+}\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s-\frac{1}{m_-}\sum_{t_i\in\mathcal{I}_-}\int_{\tau_1-2\,\delta_n}^{t_i}\sigma_s\,dW_s\right)\\
&\le\frac{3}{m_+}\sum_{t_i\in\mathcal{I}_+}\int_{\tau_1-2\,\delta_n}^{t_i}E(\sigma_s^2)\,ds+\frac{4}{m_-}\sum_{t_i\in\mathcal{I}_-}\int_{\tau_1-2\,\delta_n}^{t_i}E(\sigma_s^2)\,ds\\
&\le 4\max_{0\le s\le 1}E(\sigma_s^2)\left(4\,\delta_n+\frac{1}{m_+}\sum_{i=1}^{m_+}i/n+\frac{1}{m_-}\sum_{i=1}^{m_-}i/n\right)=O(n^{-1/2}).\qquad(20)
\end{aligned}
$$

Equations (19) and (20) together with Tchebysheff's inequality imply that the difference of the first two terms on the right hand side of (18) is $O_P(n^{-1/4})$ under P_τ, and hence it is $O_P(n^{-1/4})$. This result together with (18) shows that $U_3=O_P(n^{-1/4})$.

Finally, we consider the jump term. Denote by $\Omega_{1,n}$ the event $\min\{\tau_\ell-\tau_{\ell-1},\ell=1,\cdots,q\}>2\,\delta_n$ $(\tau_0=0)$ and $\Omega_{2,n}$ the event $|\hat\tau_\ell-\tau_\ell|\le n^{-1}\,\log^2\,n$ for $\ell=1,\cdots,q=\hat q$, and let $\Omega_n = \Omega_{1,n} \cap \Omega_{2,n}$. Condition (A3) implies that $\min\{\tau_\ell - \tau_{\ell-1}, \ell = 1, \cdots, q\}$ is positive almost surely, and then $\lim_{n\to\infty} P(\Omega_{1,n}) = P\left(\min\{\tau_\ell - \tau_{\ell-1}, \ell = 1, \cdots, q\} > 0\right)$ $= 1$. Also (4) implies $\lim_{n\to\infty} P(\Omega_{2,n}) = 1$. Hence $\lim_{n\to\infty} P(\Omega_n) = 1$. On Ω_n, $|\hat{\tau}_1 - \tau_1| \le n^{-1} \log^2 n$ and the distances between any two jumps of X_t are great than $2\,\delta_n$, so if X_t jumps on the interval $[\hat{\tau}_1 \pm \delta_n]$, it can jump only once and at $\tau_1 \in [\hat{\tau}_1 \pm \delta_n]$. Thus, we have

$$U_4 = \frac{1}{m_+} \sum_{t_i \in \hat{\mathcal{I}}_+} \sum_{\ell=1}^{N_{t_i}} L_\ell - \frac{1}{m_-} \sum_{t_i \in \hat{\mathcal{I}}_-} \sum_{\ell=1}^{N_{t_i}} L_\ell$$

is equal to $L_{N_{\tau_1}}\,[1 - n\,(\tau_1 - \hat{\tau}_1)/m_+]$ if $\hat{\tau}_1 \le \tau_1$ and $L_{N_{\tau_1}}\,[1 - n\,(\hat{\tau}_1 - \tau_1)/m_-]$ if $\hat{\tau}_1 > \tau_1$. We conclude that on Ω_n

$$|U_4 - L_{N_{\tau_1}}| \le |L_{N_{\tau_1}}|\,(m_+^- + m_-^{-1})\,n\,|\hat{\tau}_1 - \tau_1| \le |L_{N_{\tau_1}}|\,(m_+^- + m_-^{-1}) \log^2 n.$$

By considering it on Ω_n and Ω_n^c we obtain

$$\begin{aligned}
&P(|U_4 - L_{N_{\tau_1}}| > n^{-1/2} \log^3 n) \le P(\Omega_n^c) + P\left(\Omega_n \cap \left\{|U_4 - L_{N_{\tau_1}}| > n^{-1/2} \log^3 n\right\}\right) \\
&\le P(\Omega_n^c) + P\left((m_+^- + m_-^{-1})\,|L_{N_{\tau_1}}| > n^{-1/2} \log n\right) \to 0, \quad \text{as } n \to \infty,
\end{aligned}$$

that is, $U_4 - L_{N_{\tau_1}} = O_P(n^{-1/2} \log^3 n) = o_P(n^{-1/4})$.

Proof of Theorem 2. Suppose that X has $q = N_1$ jumps at τ_ℓ, $\ell = 1, \cdots, q$. Define the jump part of X

$$X_t^d = \sum_{\ell=1}^{N_t} L_\ell = \sum_{\tau_\ell \le t} L_\ell.$$

With the continuous parts of X_t and Y_t defined in (8), we yield $X_t = X_t^c + X_t^d$ and $Y_t = Y_t^c + X_t^d$. From the definition of the data adjustment in (6) and (7), we conclude

$$Y^*_{t_{i+K}} - Y^*_{t_i} = Y^c_{t_{i+K}} - Y^c_{t_i} + \sum_{t_i < \tau_\ell \le t_{i+K}} L_\ell - \sum_{t_i < \hat{\tau}_\ell \le t_{i+K}} \hat{L}_\ell \equiv Y^c_{t_{i+K}} - Y^c_{t_i} + \xi_i,$$

With above result and the expression of ASRV given in Section 4.1 we have

$$\begin{aligned}
[Y^*, Y^*]^{(K)} - [Y^c, Y^c]^{(K)} &= \frac{1}{K} \sum_{i=1}^{n-K} \left[(Y^*_{t_{i+K}} - Y^*_{t_i})^2 - (Y^c_{t_{i+K}} - Y^c_{t_i})^2\right] \\
= \frac{1}{K} \sum_{i=1}^{n-K} \Big\{\xi_i^2 + (X^c_{t_{i+K}} - X^c_{t_i})\,\xi_i &+ (\varepsilon_{t_{i+K}} - \varepsilon_{t_i})\,\xi_i\Big\} \equiv R_1 + R_2 + R_3. \qquad (21)
\end{aligned}$$

Similar to the jump part proof of Theorem 1, denote by Ω_n the event $\min\{|\tau_\ell - \tau_{\ell-1}|, \ell = 1, \cdots, q\} > K/n$ and $|\hat{\tau}_\ell - \tau_\ell| \le n^{-1} \log^2 n$ for $\ell = 1, \cdots, q = \hat{q}$. Then Condition (A3) and (4) imply $\lim_{n\to\infty} P(\Omega_n) = 1$. Note the facts that on Ω_n, first X_t has at most one jump in each of intervals $[t_i, t_{i+K}]$ of length K/n; second, since $(K/n)/(n^{-1} \log^2 n) = K/\log^2 n \to \infty$, in comparison with intervals $[t_i, t_{i+K}]$, the interval formed by each pair τ_ℓ and $\hat{\tau}_\ell$ is tiny. Thus, on Ω_n, there are only three cases for the relationship between the two types of intervals:

(1) $t_i, t_{i+K} < \tau_\ell, \hat{\tau}_\ell$ or $t_i, t_{i+K} > \tau_\ell, \hat{\tau}_\ell$;

(2) $t_i < \tau_\ell, \hat{\tau}_\ell < t_{i+K}$, and for each $(\tau_\ell, \hat{\tau}_\ell)$ there is at most $n\,(K/n - |\tau_\ell - \hat{\tau}_\ell|) \le K$ such intervals $[t_i, t_{i+K}]$;

(3) either t_i or t_{i+K} (but not both) is between τ_ℓ and $\hat{\tau}_\ell$, and for each $(\tau_\ell, \hat{\tau}_\ell)$ there is at most $2\,n\,|\tau_\ell - \hat{\tau}_\ell| \le 2\,\log^2 n$ such intervals $[t_i, t_{i+K}]$.

Under the aforementioned three cases, ξ_i on Ω_n is equal to 0, $L_\ell - \hat{L}_\ell$, and L_ℓ or $-\hat{L}_\ell$, respectively. Classifying the $n-K$ summation terms on right hand side of (21) according to Cases (1-3), we can bound $[Y^*, Y^*]^{(K)} - [Y^c, Y^c]^{(K)}$ on Ω_n by multiplying the bounds of terms with the corresponding total numbers of associated intervals $[t_i, t_{i+K}]$ under the three cases and then dividing by K. Since the difference is zero for Case (1), we need to do multiplications only for Cases (2) and (3). Below we separately handle each R_i on the right hand side of (21).

First, ξ_i^2 is equal to $(L_\ell - \hat{L}_\ell)^2$ and L_ℓ^2 or $\hat{L}_\ell^2$ for Cases (2) and (3), respectively. Thus, on Ω_n,

$$|R_1| \le \sum_{\ell=1}^{q} |L_\ell - \hat{L}_\ell|^2 + 2\,K^{-1}\,\log^2\, n \sum_{\ell=1}^{q} (|L_\ell|^2 + |\hat{L}_\ell|^2) \equiv H_{1,n,K}. \tag{22}$$

Second, since $|X^c_{t_{i+K}} - X^c_{t_i}| \le 2\,\max_{0\le t\le 1} |X^c_t|$, with the bounds of ξ_i we obtain that on Ω_n,

$$|R_2| \le 2\,\max_{0\le t\le 1}\{|X^c_t|\}\,H_{2,n,K}, \tag{23}$$

where

$$H_{2,n,K} = \sum_{\ell=1}^{q} |L_\ell - \hat{L}_\ell| + 2\,K^{-1}\,\log^2\, n \sum_{\ell=1}^{q} (|L_\ell| + |\hat{L}_\ell|). \tag{24}$$

Third, we deal with R_3 differently under Cases (2) and (3), because $\varepsilon_{t_{i+K}} - \varepsilon_{t_i}$ are not bounded. Again on Ω_n, we have

$$R_3 = \frac{1}{K}\left(\sum_{Case(2)} + \sum_{Case(3)}\right)(\varepsilon_{t_{i+K}} - \varepsilon_{t_i})\,\xi_i, \tag{25}$$

$$\left[\sum_{Case(j)} (\varepsilon_{t_{i+K}} - \varepsilon_{t_i})\,\xi_i\right]^2 \le \sum_{Case(j)} (\varepsilon_{t_{i+K}} - \varepsilon_{t_i})^2 \sum_{Case(j)} \xi_i^2, \qquad j = 2, 3, \tag{26}$$

where the second inequality is due to Cauchy-Schwartz inequality. For $a > 0$, define

$$\Gamma_n(a) = \sum_{\ell=1}^{q}\left[\sum_{t_i \text{ or } t_{i+K}\in[\tau_\ell \pm a/n]} (\varepsilon_{t_{i+K}} - \varepsilon_{t_i})^2\right]. \tag{27}$$

Under Case (2), $t_i < \tau_\ell, \hat{\tau}_\ell < t_{i+K}$, so all these t_i and t_{i+K} must be in the interval $\tau_\ell \pm 2\,K/n$. Therefore, on Ω_n,

$$\sum_{Case(2)} (\varepsilon_{t_{i+K}} - \varepsilon_{t_i})^2 \le \Gamma_n(2\,K). \tag{28}$$

Similarly, under Case (3), t_i or t_{i+K} is between τ_ℓ and $\hat{\tau}_\ell$ whose distance is bounded by $n^{-1}\,\log^2\,n$, that is, for all these pairs (t_i, t_{i+K}), one of t_i and t_{i+K} must be in the interval $[\tau_\ell \pm n^{-1}\,\log^2\,n]$. So, on Ω_n,

$$\sum_{Case(3)} (\varepsilon_{t_{i+K}} - \varepsilon_{t_i})^2 \le \Gamma_n(\log^2\,n). \tag{29}$$

Combining (25)-(29) together with the bounds for sums of ξ_i^2 we arrive at that on Ω_n,

$$K\,|R_3| \le \left[\Gamma_n(K)\,K\,\sum_{\ell=1}^{q}(L_\ell - \hat{L}_\ell)^2\right]^{1/2} + \left[\Gamma_n(\log^2\,n)\,2\,\log^2\,n\sum_{\ell=1}^{q}(L_\ell^2 + \hat{L}_\ell^2)\right]^{1/2}. \tag{30}$$

Now we need to control $\Gamma_n(a)$. Note that the number of terms in the inner summation of $\Gamma_n(a)$ given by (27) is no more than $4\,a$, and ε_t, τ_ℓ and $q = N_1$ are independent. Conditional on whole count process N we easily derive

$$E[\Gamma_n(a)|N] \le 8\,a\,N_1\,\mathrm{Var}(\varepsilon_t).$$

For $h > 0$ denote by $A_{n,a,h}$ the event $\Gamma_n(a) \le a\,h^2$. Markov inequality immediately shows

$$\sup_{n,a} P(A_{n,a,h}^c|N) \le \sup_{n,a} \frac{E[\Gamma_n(a)|N]}{a\,h^2} \le \frac{8\,N_1\,\mathrm{Var}(\varepsilon_t)}{h^2} \to 0, \qquad \text{as } h \to \infty.$$

As $P(A_{n,a,h}^c|N) \le 1$, the dominate convergence theorem leads to

$$\lim_{h\to\infty}\sup_{n,a} P(A_{n,a,h}^c) \le \lim_{h\to\infty} E\left[\sup_{n,a} P(A_{n,a,h}^c|N)\right] = 0. \tag{31}$$

As $\Gamma_n(a)$ is bounded by $a\,h^2$ on $A_{n,a,h}$, (30) gives a bound of R_3 on $\Omega_n \cap A_{n,a,h}$,

$$|R_3| \le h\,\left[\sum_{\ell=1}^{q}(L_\ell - \hat{L}_\ell)^2\right]^{1/2} + 2\,h\,K^{-1}\,\log^2\,n\,\left[\sum_{\ell=1}^{q}(L_\ell^2 + \hat{L}_\ell^2)\right]^{1/2} \le h\,H_{2,n,K}. \tag{32}$$

Collecting (22), (23) and (32) together we bound $|R_1| + |R_2| + |R_3|$ on $\Omega_n \cap A_{n,a,h}$ by $H_{1,n,K} + 2\,(\max_{0\le t\le 1}|X_t^c| + h)\,H_{2,n,K}$, which can be re-expressed via (22) and (24) as $G_{1,n} + K^{-1}\,\log^2\,n\,G_{2,n,k}$, where

$$G_{1,n} = \sum_{\ell=1}^{q}(L_\ell - \hat{L}_\ell)^2 + 2\,(\max_{0\le t\le 1}|X_t^c| + h)\,\sum_{\ell=1}^{q}|L_\ell - \hat{L}_\ell|, \tag{33}$$

$$G_{2,n} = \sum_{\ell=1}^{q}(|L_\ell|^2 + |\hat{L}_\ell|^2) + 2\,(\max_{0\le t\le 1}|X_t^c| + h)\,\sum_{\ell=1}^{q}(|L_\ell| + |\hat{L}_\ell|). \tag{34}$$

Finally, using (21) and above bound of $|R_1|+|R_2|+|R_3|$ we consider $[Y^*,Y^*]^{(K)}-[Y^c,Y^c]^{(K)}$ on $\Omega_n\cap A_{n,a,h}$, Ω_n^c and $A_{n,a,h}^c$ and obtain

$$\begin{aligned}
&P\left(\left|[Y^*,Y^*]^{(K)}-[Y^c,Y^c]^{(K)}\right|>d\,n^{-1/4}+d\,K^{-1}\,\log^2\,n\right)+\leq P(\Omega_n^c)+P(A_{n,a,h}^c)\\
&+P\left(\Omega_n\cap A_{n,a,h}\cap\left\{|R_1+R_2+R_3|>d\,n^{-1/4}+d\,K^{-1}\,\log^2\,n\right\}\right)\\
&\leq P(\Omega_n^c)+\sup_{n,a}P(A_{n,a,h}^c)+P\left(G_{1,n}+K^{-1}\,\log^2\,n\,G_{2,n}>d\,n^{-1/4}+d\,K^{-1}\,\log^2\,n\right)\\
&\leq P(\Omega_n^c)+\sup_{n,a}P(A_{n,a,h}^c)+P\left(G_{1,n}>d\,n^{-1/4}/2\right)+P\left(G_{2,n}>d/2\right). \qquad (35)
\end{aligned}$$

Theorem 1 indicates $\sum_{\ell=1}^q|L_\ell-\hat{L}_\ell|=O_P(n^{-1/4})$, Conditions (A3) implies $\sum_{\ell=1}^q(|L_\ell|+|\hat{L}_\ell|)=O_P(1)$, and martingale property with Conditions (A2) infers $\max\{|X_t^c|,0\leq t\leq 1\}=O_P(1)$, so from (33) and (34) we derive that for each h, $G_{1,n}=O_P(n^{-1/4})$ and $G_{2,n}=O_P(1)$. The results together with (31) and $\lim_{n\to\infty}P(\Omega_n^c)=0$ immediate show that the right hand side of (35) tends to zero by sequentially letting $n\to\infty$, $d\to\infty$ and then $h\to\infty$. This completes the proof.

Proof of Theorem 3. Define

$$\hat{\Theta}_K^c=[Y^c,Y^c]^{(K)}-\frac{1}{K}[Y^c,Y^c]^{(1)},$$

the TSRV estimator of Θ for the continuous diffusion price model. The theorem 4 in Zhang *et al.*(2005) exactly shows that $n^{1/6}\,(\hat{\Theta}_K^c-\Theta)$ converges stably in law to a standard normal random variable multiplied by Υ. Convergence stably in law is stronger than convergence in law. So the theorem is proved if we show

$$\hat{\Theta}_K-\hat{\Theta}_K^c=o_p\left(n^{-1/6}\right).$$

With the relationship

$$\hat{\Theta}_K-\hat{\Theta}_K^c=[Y^*,Y^*]^{(K)}-[Y^c,Y^c]^{(K)}+\frac{1}{K}\left\{[Y^*,Y^*]^{(1)}-[Y^c,Y^c]^{(1)}\right\},$$

we need to show only that

$$[Y^*,Y^*]^{(K)}-[Y^c,Y^c]^{(K)}=o_p\left(n^{-1/6}\right), \qquad (36)$$

$$\frac{1}{K}\left([Y^*,Y^*]^{(1)}-[Y^c,Y^c]^{(1)}\right)=o_P\left(n^{-1/6}\right). \qquad (37)$$

Applying Theorem 2 with $K=c\,n^{2/3}$, we immediately obtain

$$[Y^*,Y^*]^{(K)}-[Y^c,Y^c]^{(K)}=O_P\left(n^{-1/4}\right)=o_p\left(n^{-1/6}\right),$$

which proves (36).

The rest of the proof is to show (37). Note that the unadjusted data Y_{t_i} contain $q = N_1$ jumps with jump sizes L_ℓ, and $\hat{q}$ jumps with jump sizes $\hat{L}_\ell$ are added in the adjusted data Y^*. Hence, the increments of Y^* and Y^c over $[t_i, t_{i+1}]$ satisfy

$$Y^*_{t_{i+1}} - Y^*_{t_i} = (Y^c_{t_{i+1}} - Y^c_{t_i}) + \sum_{t_i < \tau_\ell \le t_{i+1}} L_\ell - \sum_{t_i < \hat{\tau}_\ell \le t_{i+1}} \hat{L}_\ell \equiv (Y^c_{t_{i+1}} - Y^c_{t_i}) + \xi_i, \quad i = 1, \cdots, n-1.$$

Of these $(n-1)$ intervals $[t_i, t_{i+1}]$, at least $n - 1 - (\hat{q} + q)$ of them don't contain τ_ℓ or $\hat{\tau}_\ell$, and thus their corresponding ξ_i are zero. The number of non-zero ξ_i are at most $(\hat{q} + q)$, and these non-zero ξ_i are equal to $L_\ell - \hat{L}_\ell$, L_ℓ, or $-\hat{L}_\ell$, depending on both or just one of $\hat{\tau}_\ell$ and τ_ℓ falling into $[t_i, t_{i+1}]$. Therefore,

$$\begin{aligned}
|[Y^*, Y^*]^{(1)} - [Y^c, Y^c]^{(1)}| &\le 2 \sum_{\ell=1}^{q} L_\ell^2 + 2 \max_{1 \le i \le n-1} |Y^c_{t_{i+1}} - Y^c_{t_i}| \sum_{\ell=1}^{q} |L_\ell| \\
&\quad + 2 \sum_{\ell=1}^{\hat{q}} \hat{L}_\ell^2 + 2 \max_{1 \le i \le n-1} |Y^c_{t_{i+1}} - Y^c_{t_i}| \sum_{\ell=1}^{\hat{q}} |\hat{L}_\ell| \\
&\le 2 \sum_{\ell=1}^{q} L_\ell^2 + 4 \max_{1 \le i \le n} |Y^c_{t_i}| \sum_{\ell=1}^{q} |L_\ell| + 2 \sum_{\ell=1}^{\hat{q}} \hat{L}_\ell^2 + 4 \max_{1 \le i \le n} |Y^c_{t_i}| \sum_{\ell=1}^{\hat{q}} |\hat{L}_\ell|. \qquad (38)
\end{aligned}$$

Below we show that the right hand side of (38) is $O_P\left(n^{1/4} \log n\right)$ and thus,

$$[Y^*, Y^*]^{(1)} - [Y^c, Y^c]^{(1)} = O_P\left(n^{1/4} \log n\right).$$

With $K \sim n^{2/3}$,

$$\frac{[Y^*, Y^*]^{(1)} - [Y^c, Y^c]^{(1)}}{K} = O_P\left(n^{1/4 - 2/3} \log n\right) = O_P\left(n^{-5/12} \log n\right) = o_P\left(n^{-1/6}\right),$$

which proves (37).

Now we derive the order for the right hand side of (38). First, since $L_\ell = O_P(1)$ applying Proposition 2 we have that $\sum_{\ell=1}^{q} |L_\ell| = O_P(1)$ and $\sum_{\ell=1}^{q} L_\ell^2 = O_P(1)$; second,

$$\max_{1 \le i \le n} |Y^c_{t_i}| \le \max_{1 \le i \le n} \{|X^c_{t_i}| + |\varepsilon_{t_i}|\} \le \max_{0 \le t \le 1} |X^c_t| + \max_{1 \le i \le n} |\varepsilon_{t_i}|;$$

third, since X^c_t is a martingale, Doob inequality (Kloeden and Platen 1999, equation 3.7 in section 2.3) leads to

$$E\left[\max_{0 \le t \le 1} |X^c_t|^2\right] \le 2 \int_0^2 E\left(\mu_s^2\right) ds + 8\, E\left(\left|\int_0^1 \sigma_s\, dW_s\right|^2\right) = \int_0^1 E\left(2\, \mu_s^2 + 8\, \sigma_s^2\right) ds,$$

and Condition (A2) implies that $\max_{0 \le t \le 1} |X^c_t|$ has finite second moment and thus it is $O_P(1)$; fourth, the maximum of $|\varepsilon_{t_i}|$ is $O_P(n^{1/4} \log n)$, since

$$\begin{aligned}
&P\left(\max_{1 \le i \le n} |\varepsilon_{t_i}| > n^{1/4} \log n\right) = 1 - \left[1 - P\left(|\varepsilon_{t_1}| > n^{1/4} \log n\right)\right]^n \\
&\le 1 - \left[1 - \frac{E(\varepsilon_{t_1}^4)}{n \log^4 n}\right]^n \sim \frac{E(\varepsilon_{t_1}^4)}{\log^4 n} \to 0,
\end{aligned}$$

as $n \to \infty$. Therefore, the sum of the first two terms on the right hand side of (38) is $O_P(n^{1/4} \log^2 n)$. Finally, we consider the last two terms on right hand side of (38). The maximum of $|Y_{t_i}|$ has been derived above to be $O_P(n^{1/4} \log^2 n)$, so it is enough to show that $\sum_{\ell=1}^{\hat{q}} \hat{L}_\ell^2$ and $\sum_{\ell=1}^{\hat{q}} |\hat{L}_\ell|$ are $O_P(1)$. Similar to the proof of Theorem 2, we define by Ω_n the event that $\hat{q} = q$ and $|\hat{\tau} - \tau_\ell| \le n^{-1} \log^2 n$ for $\ell = 1, \cdots, q$. Then (4) implies $\lim_{n\to\infty} P(\Omega_n) = 1$ and hence we need to show that they are $O_P(1)$ on Ω_n only. Note that on Ω_n,

$$\sum_{\ell=1}^{\hat{q}} \hat{L}_\ell^2 = \sum_{\ell=1}^{q} \hat{L}_\ell^2, \qquad \sum_{\ell=1}^{\hat{q}} |\hat{L}_\ell| = \sum_{\ell=1}^{q} |\hat{L}_\ell|.$$

Since from (12) we have that for each ℓ, $\hat{L}_\ell = L_\ell + O_P(n^{-1/4}) = O_P(1)$, and again applying Proposition 2 we derive that $\sum_{\ell=1}^{q} |\hat{L}_\ell| = O_P(1)$ and $\sum_{\ell=1}^{q} \hat{L}_\ell^2 = O_P(1)$.

Remark. In fact, we can prove a much stronger result than (37),

$$[Y^*, Y^*]^{(1)} - [Y^c, Y^c]^{(1)} = O_P(\log n),$$

that is, the conclusion of Theorem 2 still holds for $K = 1$.

From the proof of Theorem 4, we see that there are at most $q + \hat{q}$ terms involved in $[Y^*, Y^*]^{(1)} - [Y^c, Y^c]^{(1)}$. Instead of the upper bound on the right hand side of (38) we can bound it better as follows,

$$\begin{aligned}
|[Y^*, Y^*]^{(1)} - [Y^c, Y^c]^{(1)}| &\le 2 \sum_{\ell=1}^{q} \left\{ L_\ell^2 + |L_\ell| \, |Y^c_{[n\,\tau_\ell+1]/n} - Y^c_{[n\,\tau_\ell]/n}| \right\} \\
&\quad + 2 \sum_{\ell=1}^{\hat{q}} \left\{ \hat{L}_\ell^2 + |\hat{L}_\ell| \, |Y^c_{[n\,\hat{\tau}_\ell+1]/n} - Y^c_{[n\,\hat{\tau}_\ell]/n}| \right\} \\
\le 2 \sum_{\ell=1}^{q} L_\ell^2 + 2 \sum_{\ell=1}^{q} |L_\ell| \, |X^c_{[n\,\tau_\ell+1]/n} &- X^c_{[n\,\tau_\ell]/n}| + 2 \sum_{\ell=1}^{q} |L_\ell| \, |\varepsilon_{[n\,\hat{\tau}_\ell+1]/n} - \varepsilon_{[n\,\hat{\tau}_\ell]/n}| \\
+ 2 \sum_{\ell=1}^{\hat{q}} \hat{L}_\ell^2 + 2 \sum_{\ell=1}^{\hat{q}} |\hat{L}_\ell| \, |X^c_{[n\,\tau_\ell+1]/n} &- X^c_{[n\,\tau_\ell]/n}| + 2 \sum_{\ell=1}^{\hat{q}} |\hat{L}_\ell| \, |\varepsilon_{[n\,\hat{\tau}_\ell+1]/n} - \varepsilon_{[n\,\hat{\tau}_\ell]/n}|, \quad (39)
\end{aligned}$$

where $[n\,\tau_\ell]$ is the integer part of $n\,\tau_1$ with τ_ℓ falling into interval $[t_{[n\,\tau_\ell]}, t_{[n\,\tau_\ell+1]})$, and it is similar to interpret $[n\,\hat{\tau}_\ell]$. Again note that

$$\max_\ell |X^c_{[n\,\hat{\tau}_\ell+1]/n} - X^c_{[n\,\hat{\tau}_\ell]/n}| \le 2 \max_{0\le t\le 1} |X^c_t|.$$

The above proof of Theorem 4 has shown that $\max_{0\le t\le 1} |X^c_t|$, $\sum_{\ell=1}^{q} L_\ell^2$, $\sum_{\ell=1}^{q} |L_\ell|$, $\sum_{\ell=1}^{\hat{q}} \hat{L}_\ell^2$ and $\sum_{\ell=1}^{\hat{q}} |\hat{L}_\ell|$ all are $O_P(1)$. So all terms on the right hand side of (39) are $O_P(1)$, except for the two terms with ε's. We will show below that of the two terms, one is $O_P(1)$ and another is $O_P(\log n)$. These together proves the claimed stronger result in the Remark.

Consider the first term with ε's on the right hand side of (39).

$$\left\{\sum_{\ell=1}^{q} |L_\ell| \, |\varepsilon_{[n\,\hat{\tau}_\ell+1]/n} - \varepsilon_{[n\,\hat{\tau}_\ell]/n}|\right\}^2 \leq \sum_{\ell=1}^{q} L_\ell^2 \sum_{\ell=1}^{q} |\varepsilon_{[n\,\hat{\tau}_\ell+1]/n} - \varepsilon_{[n\,\hat{\tau}_\ell]/n}|^2.$$

The first sum on the right hand side of above inequality has been shown to be $O_P(1)$, and for the second sum, since $(q = N_1, \tau_\ell)$ are independent of ε's, it has expectation

$$E\left[\sum_{\ell=1}^{q} |\varepsilon_{[n\,\hat{\tau}_\ell+1]/n} - \varepsilon_{[n\,\hat{\tau}_\ell]/n}|^2\right] \leq 4\, E(N_1)\, E[\varepsilon_{t_1}^2] < \infty,$$

and thus it is $O_P(1)$. Therefore, the first term with ε's on the right hand side of (39) is $O_P(1)$.

Consider the second term with ε's on the right hand side of (39).

$$\left\{\sum_{\ell=1}^{\hat{q}} |\hat{L}_\ell| \, |\varepsilon_{[n\,\hat{\tau}_\ell+1]/n} - \varepsilon_{[n\,\hat{\tau}_\ell]/n}|\right\}^2 \leq \sum_{\ell=1}^{\hat{q}} \hat{L}_\ell^2 \sum_{\ell=1}^{\hat{q}} |\varepsilon_{[n\,\hat{\tau}_\ell+1]/n} - \varepsilon_{[n\,\hat{\tau}_\ell]/n}|^2. \quad (40)$$

Again, the first sum on the right hand side of above inequality has been shown to be $O_P(1)$, and for the second sum, with Ω_n defined as in the proof of Theorem 4, and on Ω_n,

$$\begin{aligned}\sum_{\ell=1}^{\hat{q}} |\varepsilon_{[n\,\hat{\tau}_\ell+1]/n} - \varepsilon_{[n\,\hat{\tau}_\ell]/n}|^2 &= \sum_{\ell=1}^{q} |\varepsilon_{[n\,\hat{\tau}_\ell+1]/n} - \varepsilon_{[n\,\hat{\tau}_\ell]/n}|^2 \\ &\leq \sum_{\ell=1}^{q} \left[\sum_{t_i \text{ or } t_{i+1} \in [\tau_\ell \pm \log^2 n/n]} |\varepsilon_{t_{i+1}} - \varepsilon_{t_i}|^2\right],\end{aligned}$$

which has total $2\,N_1 \log^2 n$ terms in the sum and whose expectation is bounded by

$$8 \log^2 n\, E(N_1)\, E[\varepsilon_{t_1}^2] < \infty.$$

Thus, restricted on Ω_n the second sum on the right hand side of (40) has order $O_P(\log^2 n)$ and so the sum itself is $O_P(\log^2 n)$, as Ω_n has probability tending to zero. Hence, the second term with ε's on the right hand side of (39) is $O_P(\log n)$.

Proof of Theorem 4. The partition numbers $K_m = m + C$ satisfy $C \leq K_m \leq C + M$, with $C \sim n^{1/2}$ and $M \sim n^{1/2}$. Applying Theorem 2 to $[Y^*, Y^*]^{(K_m)}$ we obtain that uniformly for all $m = 1, \cdots, M$,

$$[Y^*, Y^*]^{(K_m)} - [Y^c, Y^c]^{(K_m)} = O_P\left(n^{-1/4} + K_m^{-1}\, log^2\, n\right) = O_P\left(n^{-1/4}\right).$$

So we can replace $[Y^*, Y^*]^{(K_m)}$ by $[Y^c, Y^c]^{(K_m)}$. As Y^c is from the continuous price model, the rest proof is similar to Zhang (2006) but can be much shorter and elementary, as we need to derive only convergence rate.

Proof of Theorem 5. Let $J = \log_2 n$. When processing orthogonal wavelet transformation of shifted data $(\mathcal{S}^{\ell-1} y^*)_i$, we stop the process at level $J - J_n$ and approximate $(\mathcal{S}^{\ell-1} y^*)_i$ in the Haar subspace at scale $J - J_n$. The sum of squares

$$\sum_{j=1}^{J-J_n} \sum_{k=1}^{2^{j-1}} \left(y_{j,k}^{\ell}\right)^2$$

is equal to the sum of squares of smooth wavelet coefficients corresponding to the Haar space at scale $J - J_n$. For Haar wavelet transformation of $(\mathcal{S}^{\ell-1} y^*)_i = Y^*_{t_{\ell+i}}$, the smooth wavelet coefficients at level $J - J_n$ are $2^{(J-J_n)/2}\,(Y^*_{t_{\ell+k\,2^{J_n}}} - Y^*_{t_{\ell+(k-1)\,2^{J_n}}})$, $k = 1, \cdots, 2^{J-J_n}$. Thus we have

$$\sum_{j=1}^{J-J_n} \sum_{k=1}^{2^{j-1}} \left(y_{j,k}^{\ell}\right)^2 = 2^{J-J_n} \sum_{k=1}^{2^{J-J_n}} (Y^*_{t_{\ell+k\,2^{J_n}}} - Y^*_{t_{\ell+(k-1)\,2^{J_n}}})^2.$$

Summation over ℓ leads to

$$WRV = n^{-1} \sum_{\ell=1}^{2^{J_n}} \sum_{j=1}^{J-J_n} \sum_{k=1}^{2^{j-1}} \left(y_{j,k}^{\ell}\right)^2 = 2^{-J_n} \sum_{\ell=1}^{2^{J_n}} \sum_{k=1}^{2^{J-J_n}} (Y^*_{t_{\ell+k\,2^{J_n}}} - Y^*_{t_{\ell+(k-1)\,2^{J_n}}})^2 = [Y^*, Y^*]^{\left(2^{J_n}\right)}.$$

Thus, the WRV estimator $\widehat{\Theta}_W$ agrees with JTSRV estimator $\widehat{\Theta}_K$ with $K = 2^{J_n}$.

REFERENCES

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

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2003), "Some Like It Smooth, and Some Like It Rough: Untangling Continuous and Jump Components in Measuring, Modeling, and Forecasting Asset Return Volatility," Manuscript.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), "Modeling and Forecasting Realized Volatility," Econometrica, 71, 579-625,

Bandi, F. M. and Russell, J. R. (2005), "Microstructure Noise, Realized Volatility, and Optimal Sampling," Manuscript.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2004), "Regular and Modified Kernel-Based Estimators of Integrated Variance: the Case with Independent Noise," Manuscript.

Barndorff-Nielsen, O. E. and Shephard, N. (2002), "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," Journal of Royal Statistical Society, Series B, 64, 253-280.

Barndorff-Nielsen, O. E. and Shephard, N. (2006), "Econometrics of Testing for Jumps in Financial Econometrics Using Bipower Variation," Journal of Financial Econometrics 4, 1-30.

Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992). ARCH modeling in finance. A review of the theory and empirical evidence. Journal of Econometrics **52**, 5-59.

Dacorogna, M. M., Geçay, R., Müller, U. A., Pictet, O. V. and Olsen, R. B. (2001). An Introduction to High Frequency Finance, New York: Academic Press.

Daubechies, I. (1992). Ten Lectures on Wavelets. CBMS-NSF series in applied mathematics, no. 61, SIAM, Philadelphia.

Eraker, B., Johannes, M. S. and Polson, N. (2003), "The Impact of Jumps in Volatility and Returns," Journal of Finance, 58, 1269-1300.

Fan, J.(2005), "A Selective Review of Financial Econometrics," Statistical Science, 20, 317-357.

Fouque, J. P., Papanicolaou, G. and Sircar, K. R. (2000). Derivatives in Financial Markets with Stochastic Volatility. Cambridge University Press.

Gloter, A. and Jacod, J. (2001), "Diffusions With Measurement Errors. I. Local Asymptotic Normality," ESAIM, 5, 225-242.

Gouriéroux, C. (1997). ARCH models and Financial Applications. Springer.

Hansen, P. R. and Lunde, A. (2006), "Realized Variance and Market Microstructure Noise (With Discussion)," Journal of Business and Economic Statistics, 24, 127-218.

Huang, X. and Tauchen, G. (2005), "The Relative Contribution of Jumps to Total Price Variance," Journal of Financial Econometrics, 3, 456-499.

Kloeden, P. E. and Platen, E. (1999). Numerical Solution of Stochastic Differential Equations. Springer.

Prakasa Rao, B. L. S. (1999). Statistical Inference for Diffusion Type Processes. Arnold: London.

Raimondo, M. (1998),"Minimax Estimation of Sharp Change Points," The Annals of Statistics, 26, 1379-1397.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In Time Series Models in Econometrics, Finance, and Other Fields (D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen, eds.), London: Chapman & Hall, pp. 1-67.

Vidakovic, B. (1999). Statistical Modeling by Wavelets. Wiley.

Wang, Y. (1995), "Jump and Sharp Cusp Detection via Wavelets," Biometrika, 82, 385-397.

Wang, Y. (2002), "Asymptotic Nonequivalence of ARCH Models and Diffusions," The Annals of Statistics, 30, 754-783.

Wang, Y. (2006). Selected review on wavelets. In Frontier Statistics, Festschrift for Peter Bickel (H.Koul and J.Fan, eds.), pp. 163-179.

Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," Journal of the American Statistical Association, 100, 1394-1411.

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

Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign-Exchange Rates," Journal of Business and Economic Statistics, 14, 45-52.

Detailed Proofs of Theorem 5.

Proof of Theorem 5. The main idea is similar to Zhang (2006) but the proof is shorter and elementary, because we need to derive only convergence rate. Letting $K_m = m + C$, it is easy to verify that $\zeta \sim n^{-1/2}$,

$$\bar{K} = \sum_{m=1}^{M} K_m/M = C + (M+1)/2 \sim n^{1/2},$$

$$\mathrm{Var}(K) = \frac{1}{M}\sum_{m=1}^{M}(K_m - \bar{K})^2 = (M^2-1)/12 \sim n,$$

and

$$\sum_{m=1}^{M} a_m = 1, \qquad \sum_{m=1}^{M} a_m/K_m = 0, \qquad \sum_{m=1}^{M} |a_m| \sim 2 + \frac{3\,C}{M}.$$

From Theorem 2 and Zhang *et al.*(2005, theorems 1 & 2) we have

$$\begin{aligned}
&\zeta\,([Y^*,Y^*]^{(K_1)} - [Y^*,Y^*]^{(K_M)}) = \zeta\,([Y^c,Y^c]^{(K_1)} - [Y^c,Y^c]^{(K_M)}) + O_P(\zeta\, C^{-1}\, \log^2\, n)\\
&= 2\,\zeta\, E(\varepsilon^2)[n/K_1 - n/K_M] + O_P(\zeta\, K_1^{-1}\, n^{1/2} + \zeta\, K_M^{1/2}\, n^{-1/2} + \zeta\, C^{-1}\, \log^2\, n)\\
&= \frac{2\,n}{n+1}\, E(\varepsilon^2) + O_P(n^{-1/2})\\
&= 2E(\varepsilon^2) + O_P(n^{-1/2}).
\end{aligned}$$

Appealing to Theorem 2, we have

$$\begin{aligned}
&\left|\sum_{m=1}^{M} a_m([Y^*,Y^*]^{(K_m)} - [Y^c,Y^c]^{(K_m)})\right|\\
\leq\ & \sum_{m=1}^{M} |a_m|\, \left|[Y^*,Y^*]^{(K_m)} - [Y^c,Y^c]^{(K_m)}\right|\\
\leq\ & \sum_{m=1}^{M} |a_m|\, O_P(C^{-1}\, \log^2\, n) = O_P(n^{-1/2}\, \log^2\, n).
\end{aligned}$$

Hence, combining the last two results and using $Y_t^c = X_t^c + \varepsilon_t$, we obtain

$$\begin{aligned}
\widehat{\Theta} &= \sum_{m=1}^{M} a_m[Y^c,Y^c]^{(K_m)} + 2\,E(\varepsilon^2) + O_P(n^{-1/2}\, \log^2\, n)\\
&= \sum_{m=1}^{M} a_m[X^c,X^c]^{(K_m)} + \sum_{m=1}^{M} a_m\{[\varepsilon,\varepsilon]^{(K_m)} + 2\,E(\varepsilon^2)\} + 2\,\sum_{m=1}^{M} a_m[X^c,\varepsilon]^{(K_m)}\\
&\quad + O_P(n^{-1/2}\, \log^2\, n). \qquad (41)
\end{aligned}$$

By Zhang *et al.*(2005, theorem 2), we have

$$\left|\sum_{m=1}^{M} a_m\,[X^c, X^c]^{(K_m)} - \int_0^1 \sigma_s^2\,ds\right| \le \sum_{m=1}^{M} |a_m|\,\left|[X^c, X^c]^{(K_m)} - \int_0^1 \sigma_s^2\,ds\right|$$
$$= \sum_{m=1}^{M} |a_m|\, O_P((K_M/n)^{1/2}) = O_P(n^{-1/4}).$$

Hence, it remains to show that the last two sums in (41) are of order $n^{-1/4}$.

For the second sum in (41), expanding $[\varepsilon,\varepsilon]^{(K_m)}$ and using the fact that $\sum_{m=1}^{M} a_m/K_m = 0$, we get

$$\sum_{m=1}^{M} a_m\,\{[\varepsilon,\varepsilon]^{(K_m)}+2\,E(\varepsilon^2)\} = -\sum_{m=1}^{M}\frac{a_m}{K_m}\sum_{i=1}^{K_m}\{\varepsilon_{t_i}^2+\varepsilon_{t_{n-i+1}}^2-2\,E(\varepsilon^2)\}-2\sum_{m=1}^{M}\frac{a_m}{K_m}\sum_{i=1}^{n-K_m}\varepsilon_{t_i}\varepsilon_{t_{i+K_m}}.$$

Since both sums in above expression have mean zero, we need to evaluate the orders of their variances. The variance of the last term is

$$\mathrm{Var}\left(\sum_{m=1}^{M}\frac{a_m}{K_m}\sum_{i=1}^{n-K_m}\varepsilon_{t_i}\varepsilon_{t_{i+K_m}}\right) = \sum_{m=1}^{M}\left(\frac{a_m}{K_m}\right)^2 (n-K_m+1)\,\eta^2$$

Similarly, the variance of the first term is

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{m=1}^{M}\frac{a_m}{K_m}\sum_{i=1}^{K_m}\{\varepsilon_{t_i}^2+\varepsilon_{t_{n-i+1}}^2-2\,E(\varepsilon^2)\}\right)\\
=\;& 2\,\mathrm{Var}\left(\sum_{m=1}^{M}\frac{a_m}{K_m}\sum_{i=1}^{K_m}\varepsilon_{t_i}^2\right)\\
=\;& 2\,\mathrm{Var}(\varepsilon^2)\sum_{i=1}^{K_M-C}\left(\sum_{m=K_i+1-C}^{M}\frac{a_m}{K_m}\right)^2\\
=\;& o(n^{-1/2}).
\end{aligned}$$

Hence, $\sum_{m=1}^{M} a_m\,\{[\varepsilon,\varepsilon]^{(K_m)}+2\,E(\varepsilon^2)\}$ is of order $n^{-1/4}$.

For the third sum in (41), we have

$$\begin{aligned}
\sum_{m=1}^{M} a_m\,[X^c,\varepsilon]^{(K_m)} &= \sum_{m=1}^{M}\frac{a_m}{K_m}\sum_{i=1}^{n-K_m}(X_{t_{i+K_m}}-X_{t_i})(\varepsilon_{t_{i+K_m}}-\varepsilon_{t_i})\\
&= \sum_{i=1}^{n}\varepsilon_{t_i}\sum_{m=1}^{M}\frac{a_m\,b_{i,K_m}}{K_m}, \qquad (42)
\end{aligned}$$

where

$$b_{i,K_m} = \begin{cases} (X_{t_i}-X_{t_{i-K_m}})-(X_{t_{i+K_m}}-X_{t_i}) & \text{for } i=K_m,\cdots,n-K_m,\\ -(X_{t_{i+K_m}}-X_{t_i}) & \text{for } i=1,\cdots,K_m,\\ (X_{t_i}-X_{t_{i-K_m}}) & \text{for } i=n-K_m+1,\cdots,n.\end{cases}$$

Note that

$$\sum_{i=1}^{n} |b_{i,K_m}|^2 \leq 2 \sum_{i=K_m+1}^{n} (X_{t_i} - X_{t_{i-K_m}})^2 + 2 \sum_{i=1}^{n-K_m} (X_{t_{i+K_m}} - X_{t_i})^2 \sim 4\, K_m \int_0^1 \sigma_s^2\, ds.$$

From (42) and using above inequality and expression of a_m, we obtain

$$\mathrm{Var}\left(\sum_{m=1}^{M} a_m\, [X^c, \varepsilon]^{(K_m)} \middle| X \right) = \mathrm{Var}(\varepsilon) \sum_{i=1}^{n} \left(\sum_{m=1}^{M} \frac{a_m\, b_{i,K_m}}{K_m} \right)^2.$$

By the Cauchy-Schwartz inequality, it is bounded by

$$\mathrm{Var}(\varepsilon)\, M \sum_{i=1}^{n} \sum_{m=1}^{M} \left(\frac{a_m\, b_{i,K_m}}{K_m} \right)^2 \qquad \sim 4\, \mathrm{Var}(\varepsilon) \int_0^1 \sigma_s^2\, ds \left(\frac{1}{2\, M} + \frac{C}{M^2} \right) = O_P(n^{-1/2}),$$

Thus the third sum in (41) is of order $n^{-1/4}$. This completes the proof.