

Article

# Jump Variation Estimation with Noisy High Frequency Financial Data via Wavelets

Xin Zhang, Donggyu Kim and Yazhen Wang \*

Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA;  
zhangxin@stat.wisc.edu (X.Z.); kimd@stat.wisc.edu (D.K.)

\* Correspondence: yzwang@stat.wisc.edu; Tel.: +1-608-262-6399

Academic Editor: Nikolaus Hautsch

Received: 29 February 2016; Accepted: 27 July 2016; Published: 16 August 2016

**Abstract:** This paper develops a method to improve the estimation of jump variation using high frequency data with the existence of market microstructure noises. Accurate estimation of jump variation is in high demand, as it is an important component of volatility in finance for portfolio allocation, derivative pricing and risk management. The method has a two-step procedure with detection and estimation. In Step 1, we detect the jump locations by performing wavelet transformation on the observed noisy price processes. Since wavelet coefficients are significantly larger at the jump locations than the others, we calibrate the wavelet coefficients through a threshold and declare jump points if the absolute wavelet coefficients exceed the threshold. In Step 2 we estimate the jump variation by averaging noisy price processes at each side of a declared jump point and then taking the difference between the two averages of the jump point. Specifically, for each jump location detected in Step 1, we get two averages from the observed noisy price processes, one before the detected jump location and one after it, and then take their difference to estimate the jump variation. Theoretically, we show that the two-step procedure based on average realized volatility processes can achieve a convergence rate close to  $O_p(n^{-4/9})$ , which is better than the convergence rate  $O_p(n^{-1/4})$  for the procedure based on the original noisy process, where  $n$  is the sample size. Numerically, the method based on average realized volatility processes indeed performs better than that based on the price processes. Empirically, we study the distribution of jump variation using Dow Jones Industrial Average stocks and compare the results using the original price process and the average realized volatility processes.

**Keywords:** high frequency financial data; jump variation; realized volatility; integrated volatility; microstructure noise; wavelet methods; nonparametric methods

**JEL:** C01; C14; C22; C51; C52; C58; G17

---

## 1. Introduction

### 1.1. Motivation

Volatility analysis plays an important role in finance. For example, portfolio allocation, derivative pricing and risk management all require accurate estimation of volatility. With the advance of technology, financial instruments are traded at frequencies as high as milliseconds, which produces large numbers of trading records each day and enables us to better estimate volatility. Such data are often referred to as high-frequency financial data. There are a growing number of volatility analysis studies based on high-frequency financial data. One popular approach is to use realized volatility, which is the sum of the squared difference of the log prices. See, for example, Barndorff-Nielsen and Shephard (2002) [1], Zhang et al. (2005) [2] and Zhang (2006) [3]. If the observed data are

assumed to be a continuous semi-martingale price process, realized volatility yields a consistent estimator of integrated volatility. However, high-frequency financial data are contaminated with market microstructure noise, and the contaminated price observations make the realized volatility inconsistent. If we assume that high-frequency financial data follow semi-martingale price processes with additive microstructure noises, several methods are proposed to consistently estimate integrated volatility based on high-frequency financial data. See, for example, Bandi and Russell (2008) [4], Barndorff-Nielsen et al. (2008) [5], Jacod et al. (2009) [6], Jacod et al. (2010) [7], Xiu (2010) [8] and Zhang et al. (2005) [2].

Jumps are often observed in financial markets due to events happening around the globe or the release of financial reports. See Aït-Sahalia et al. (2002) [9], Barndorff-Nielsen and Shephard (2004) [10]. Actually, a jump-diffusion model has been studied back in Merton (1976) [11]. Such jumps lead to a violation of the continuous assumption and, thus, cause the inconsistency of those established estimators. To accurately estimate and predict volatility, it is important to have methods that estimate jumps and that separate it from the continuous part. Many existing methods have focused on testing for jumps under the strong assumption that we can observe the exact price process (see Aït-Sahalia 2004 [12], Aït-Sahalia and Jacod 2009 [13], Andersen et al. 2012 [14], Barndorff-Nielsen and Shephard 2006 [15], Carr and Wu 2003 [16], Eraker et al. 2003 [17], Eraker 2004 [18], Fan and Fan 2011 [19], Huang and Tauchen 2005 [20], Jiang and Oomen 2008 [21], Lee and Hannig 2010 [22], Lee and Mykland 2008 [23], Bollerslev et al. 2009 [24] and Todorov 2009 [25]). Recently, a few tests started to take microstructure noise into account; for example, the pre-averaging method is used to take care of the microstructure noise, and a test for jumps based on power variation is employed in Aït-Sahalia et al. (2012) [26]. Furthermore, some interesting jump research going on includes, but is not limited to: estimating the degree of jump activities (see Ait-Sahalia and Jacod 2009 [27], Jing et al. 2011 [28], Jing et al. 2012 [29] and Todorov and Tauchen 2011 [30]), modeling return using pure jump process (see Jing et al. 2012 [31] and Kong et al. 2015 [32]), modeling jumps in volatility processes (see Todorov and Tauchen 2011 [33]) and forecasting volatility with the existence of jumps (see Andersen et al. 2007 [34], Andersen et al. 2011 [35] and Andersen et al. 2011 [36]).

The method proposed in this paper mainly focuses on the wavelet-based detection and estimation of jump variation based on high-frequency financial data with market microstructure noise. Wavelet analysis is a tool for decomposing signals and functions into time-frequency contents that can unfold a signal or function over the time-frequency plane to provide information on “when” such “frequency” occurs (Wang 2006 [37]). It enables us to analyze non-stationary time series and to detect local changes. With such features, wavelet methods are developed to detect sudden localized changes in Wang (1995) [38] and to study jump variation based on high-frequency financial data with microstructure noise in Fan and Wang (2007) [39]. In this paper, we are going to take advantage of the nice properties of wavelets and adopt a new approach to improve the wavelet estimation methods.

The paper is organized as follows. In the remaining part of Section 1, we will introduce the basic model and wavelet techniques. In Section 2, we analyze the statistical properties of the original log price process, realized volatility process and average realized volatility process. The simulation is in Section 3, and the empirical study of jump variation is in Section 4. Section 5 provides some concluding remarks about our results. The proofs of the main theorems are in Section 6. More detailed proofs of the lemmas are in Appendix A. Tables and figures for further illustration are in Appendix B.

### 1.2. Integrated Volatility, Realized Volatility and Jump Variation

Let  $S_t$  be the price process of a financial asset. We assume that the log price  $X_t = \log S_t$  is a semi-martingale satisfying the stochastic differential equation

$$dX_t = \mu_t dt + \sigma_t dB_t + L_t d\Lambda_t, \quad t \in [0, 1], \quad (1)$$

where we assume  $\mu_t$  and  $\sigma_t$  are both bounded and continuous,  $B_t$  is a standard Brownian motion,  $\Lambda_t$  is a counting process and the total number of jumps in time period  $[0, 1]$  is finite. The three terms on the right-hand side of Equation (1) correspond to the drift, diffusion and jump parts of  $X$ , respectively.

Integrated volatility, as an important measure of the variation of the return process during a time period  $[0, 1]$ , is defined as

$$\int_0^1 \sigma_t^2 dt. \quad (2)$$

In financial practice, the estimation of the integrated volatility is of great interest.

Suppose  $\{X(t_i), t_i \in [0, 1], i = 1, \dots, n\}$  are  $n$  observations of the log price process in time period  $[0, 1]$ . Define the realized volatility of  $X$  as

$$[X, X]_1 \equiv \sum_{t_i} \left( X(t_{i+1}) - X(t_i) \right)^2. \quad (3)$$

Stochastic analysis shows that as  $\max(t_i - t_{i-1}) \rightarrow 0$ ,

$$[X, X]_1 \rightarrow_p \int_0^1 \sigma_t^2 dt + \sum_{0 \leq t \leq 1} L_t^2, \quad (4)$$

where the first term on the right-hand side is the integrated volatility, and the second term is called the jump variation, which is the main focus of this paper. We denote it as  $\Psi$ ,

$$\Psi \equiv \sum_{0 \leq t \leq 1} L_t^2. \quad (5)$$

Equation (4) says that, as the number of observations  $n$  increases, realized volatility  $[X, X]_1$  converges in probability to integrated volatility plus jump variation  $\Psi$ . Thus, to better model and predict volatility, we need methods to estimate jump variation and to separate it from the integrated volatility.

### 1.3. Market Microstructure Noises

Microstructure noise plays a big role in the analysis of high-frequency financial data. It is often assumed that the observed data  $Y(t_i)$  are a noisy version of  $X(t_i)$  at time points  $t_i$ ,

$$Y(t_i) = X(t_i) + \epsilon(t_i), \quad i = 1, \dots, n, \quad t_i = i/n, \quad (6)$$

where  $\epsilon$ 's are the microstructure noises. We assume that  $\epsilon$ 's are i.i.d. random variables with mean zero and variance  $\eta^2$ . We also assume that  $\epsilon$  and  $X$  are independent.

With noisy observations  $Y(t_i)$ , now it becomes even more challenging. Let us define the realized volatility based on  $Y$  as

$$[Y, Y]_1 \equiv \sum_{t_i} \left( Y(t_{i+1}) - Y(t_i) \right)^2. \quad (7)$$

It can be shown (Zhang et al., 2005 [2]) that:

$$[Y, Y]_1 \approx_{\mathcal{L}} \int_0^1 \sigma_t^2 dt + \sum_{0 \leq t \leq 1} L_t^2 + 2n\eta^2 + \left[ 4nE\epsilon^4 + \frac{2}{n} \int_0^1 \sigma_t^4 dt \right]^{1/2} Z_{\text{total}}, \quad (8)$$

where  $\approx_{\mathcal{L}}$  is convergence in law and  $Z_{\text{total}}$  is a standard normal random variable. The result (8) implies that realized volatility  $[Y, Y]_1$  is an inconsistent estimator of integrated volatility.

### 1.4. Wavelet Basics

Let  $\varphi$  and  $\psi$  be the specially-constructed father wavelet and mother wavelet, respectively. Then, the wavelet basis is  $\varphi(t), \psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), j = 0, 1, 2, \dots, k = 0, \dots, 2^j - 1$ . We can expand a function  $f(t)$  over the wavelet basis as follows. Denote by  $f_{j,k}$  the wavelet coefficient of  $f(t)$  associated with location  $k2^{-j}$  and frequency  $2^j$ ,

$$f_{j,k} = \int f(t)\psi_{j,k}(t)dt. \tag{9}$$

Then, we have

$$f(t) = f_{-1,0}\varphi(t) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} f_{j,k}\psi_{j,k}(t), \tag{10}$$

where  $f_{-1,0} = \int f(t)\varphi(t)dt$ .

There are two cases regarding the order of the wavelet coefficients  $f_{j,k}$  (Daubechies, 1992 [40]):

Case 1: If  $f$  is a Hölder continuous function with exponent  $\alpha, 0 < \alpha \leq 1$ , i.e.,  $|f(x) - f(y)| \leq C|x - y|^\alpha$  for any  $x, y$ , then  $f_{j,k}$  satisfies

$$|f_{j,k}| \leq C_1 2^{-j(\alpha+1/2)}. \tag{11}$$

That means, for a Hölder continuous function  $f$ , the wavelet coefficients decrease at the order of  $2^{-j(\alpha+1/2)}$  as the frequency level  $2^j$  increases. Details are in Lemmas A1 and A2 in Appendix A.

Case 2: If  $f$  is a Hölder continuous function with exponent  $\alpha, 0 < \alpha \leq 1$ , except for a jump of size  $L$  at  $s$ . Then, for sufficiently large  $j$  with  $2^j s - k \in \mathcal{G}_d$ , there exists a constant  $C_L$  depending on  $L$ , such that

$$|f_{j,k}| \geq C_L 2^{-j/2}, \tag{12}$$

where  $\mathcal{G}_d \subset \{x : |\int_{(-\infty,x)} \psi(y)dy| \geq d\}$  is an interval around zero for some positive constant  $d$ . That means, nearby this jump point, the wavelet coefficients decrease no faster than the order of  $2^{-j/2}$  as the frequency level  $2^j$  increases. Details are in Lemma A3 in Appendix A. Examples will be shown later in Section 5 for the Daubechies wavelets used in this paper.

Note that with  $n$  discrete observations,  $j$  takes values  $0, 1, 2, \dots, J - 1$ , where  $2^J = n$ .

## 2. Statistical Analysis

### 2.1. Choosing a Frequency Level to Differentiate Jumps

In this section, we are going to study several different processes and the orders of their wavelet coefficients for both the continuous part and the jump part. For each process, we will determine a frequency level for which the order of the corresponding wavelet coefficients is significantly larger nearby jumps than the continuous part.

#### 2.1.1. Starting from $X$ and $[X, X]$

We first focus on the processes without noises. We consider  $X(t_i)$  and  $[X, X](t_i)$  at time points  $\{t_i = i/n, i = 1, \dots, n\}$ .

From Equation (1), we have

$$X(t_i) = X_0 + \int_0^{t_i} \mu_s ds + \int_0^{t_i} \sigma_s dB_s + \sum_{0 < s \leq t_i} L_s. \tag{13}$$

For simplicity, we assume that  $X(t)$  in Equation (1) is defined in  $[-1, 1]$ . Define  $[X, X](t_i)$  as:

$$[X, X](t_i) \equiv \sum_{t_{i-1} \leq t_r < t_i} \left( X(t_{r+1}) - X(t_r) \right)^2. \tag{14}$$

For  $[X, X](t_i)$ , by Lemma A7 and Corollary 1 in Appendix A, we have:

$$[X, X](t_i) \approx_{\mathcal{L}} \int_{t_{i-1}}^{t_i} \sigma_s^2 ds + \sqrt{\frac{2}{n}} \int_{t_{i-1}}^{t_i} \sigma_s^2 dB_s^{\text{discrete}} + \sum_{t_{i-1} < s \leq t_i} L_s^2, \tag{15}$$

where  $B^{\text{discrete}}$  is a standard Brownian motion whose associated diffusion term in the above equation is due to discretization, and  $\approx_{\mathcal{L}}$  is convergence in law.

We let  $X_{j,k}$  and  $[X, X]_{j,k}$  be the wavelet coefficients of  $X$  and  $[X, X]$  associated with location  $k2^{-j}$  and frequency  $2^j$ , respectively.

For  $X_{j,k}$ , from Equation (13), the drift term is Hölder continuous with exponent  $\alpha = 1$ . The diffusion term is Hölder continuous with exponent  $\alpha$  arbitrarily close to  $1/2$ , so the order of the continuous component of  $X_{j,k}$  is dominated by the diffusion term. The order of the jump component is no less than  $2^{-j/2}$ . Therefore, if we pick a frequency level at  $2^j \sim n$ , the order of the jump component is significantly larger than the other terms.

For  $[X, X]_{j,k}$ , from Equation (15), the drift term is Hölder continuous with exponent  $\alpha = 1$ . The diffusion term is Hölder continuous with exponent  $\alpha$  arbitrarily close to  $1/2$  and multiplied by an extra  $n^{-1/2}$ . Thus, again, if we pick a frequency level at  $2^j \sim n$ , the order of the jump component is significantly larger than the others.

### 2.1.2. Moving on to $Y$ and $[Y, Y]$

We now consider the noisy observations  $Y(t_i)$  from Equation (6):

$$Y(t_i) = X_0 + \int_0^{t_i} \mu_s ds + \int_0^{t_i} \sigma_s dB_s + \sum_{0 < s \leq t_i} L_s + \epsilon(t_i). \tag{16}$$

We let  $Y_{j,k}$  be the wavelet coefficients of  $Y$ , so  $Y_{j,k} = X_{j,k} + \epsilon_{j,k}$ .  $\epsilon_{j,k}$ 's are i.i.d with mean zero and variance  $n^{-1}\eta^2$ , so the order of noise component is  $O_P(n^{-1/2})$ . If we pick a frequency level at  $2^j \sim n/\log^2 n$ , the order of the jump component is significantly larger than the others (Fan and Wang, 2007 [39]).

Next, let us consider  $[Y, Y]$ . Similar to the way we rewrite  $[X, X](t_i)$  in Equation (13), we rewrite Equation (8) as follows:

$$[Y, Y](t_i) \approx_{\mathcal{L}} \int_{t_{i-1}}^{t_i} \sigma_s^2 ds + \sqrt{\frac{2}{n}} \int_{t_{i-1}}^{t_i} \sigma_s^2 dB_s^{\text{discrete}} + 2n\eta^2 + \sqrt{4nE\epsilon^4} \int_{t_{i-1}}^{t_i} dB_s^{\text{noise}} + \sum_{t_{i-1} \leq s \leq t_i} L_s^2, \tag{17}$$

where the  $B^{\text{discrete}}$  term is a diffusion term due to discretization, the  $B^{\text{noise}}$  term is a diffusion term due to noise,  $B^{\text{discrete}}$  and  $B^{\text{noise}}$  are independent Brownian motions and  $2n\eta^2$  is a constant (mean of the noise term).

We let  $[Y, Y]_{j,k}$  be the wavelet coefficients of  $[Y, Y]$ . The order for the continuous component dominated by the diffusion term due to noise is no greater than  $n^{1/2}2^{-j(\alpha+1/2)}$ , where  $\alpha$  is arbitrarily close to  $1/2$ . The order of the jump component is greater than  $2^{-j/2}$ . However, in this situation, since the order of the diffusion term due to noise is too large, we are not able to have a frequency level  $2^j \sim n$  at which we can differentiate the jumps.

### 2.1.3. Subsampling and Averaging on $[Y, Y]$

We now consider subsampling and averaging  $[Y, Y]$  to reduce the order of the diffusion term due to noise. Subsampling: Create  $M$  grids (sub-samples) on the time line so that the distance between two consecutive observations within each grid is  $M/n$  and the average size of each grid is  $\bar{n} = n/M$ . Denote by  $[Y, Y]^{(m)}, m = 1, \dots, M$ , the realized volatility of  $Y$  using grid  $m$ .

Averaging:  $[Y, Y]^{(avg)}$  denotes the average of the realized volatilities of  $Y$  using all  $M$  grids:

$$[Y, Y]^{(avg)} = \frac{1}{M} \sum_{m=1}^M [Y, Y]^{(m)}. \tag{18}$$

Rewriting the results in Zhang et al. (2005) [2], we have

$$[Y, Y](t_i)^{(avg)} \approx_{\mathcal{L}} \int_{t_{i-1}}^{t_i} \sigma_s^2 ds + \sqrt{\frac{4}{3\bar{n}}} \int_{t_{i-1}}^{t_i} \sigma_s^2 dB_s^{\text{discrete}} + 2\bar{n}\eta^2 + \sqrt{\frac{4\bar{n}}{M}} E\epsilon^4 \int_{t_{i-1}}^{t_i} dB_s^{\text{noise}} + \sum_{t_{i-1} \leq s \leq t_i} L_s^2. \tag{19}$$

Let  $[Y, Y]_{j,k}^{(avg)}$  be the wavelet coefficients of  $[Y, Y]^{(avg)}$ . We choose  $M \sim n^\gamma, 0 < \gamma \leq 2/3$ . Then, the order of the continuous component is no greater than  $n^{(1-2\gamma)/2} 2^{-j(\alpha+1/2)}$ , where  $\alpha$  is arbitrarily close to  $1/2$ . Therefore, if we pick a frequency level at  $2^{j_n} \sim n$ , the order of the jump component is larger than the others.

### 2.2. Threshold Selection and Jump Location Estimation

For each process discussed in the above section, we are able to choose a frequency level  $2^{j_n}$  to differentiate jumps, and thus, we develop a threshold on the wavelet coefficients at such a frequency level to detect those jumps. To accomplish the task, we first standardize the wavelet coefficients at chosen frequency level  $j_n$  by dividing them by their median. Given a process, if the continuous component is dominated by the diffusion term at chosen frequency level  $j_n$ , by Lemma A5, the standardized wavelet coefficients nearby a jump point are at least of the order  $2^{j_n \alpha}$ , where  $\alpha$  is arbitrarily close to  $1/2$ . As volatility process  $\sigma_t^2$  is bounded, using Lemma A6, we can show that with probability tending to one, the maximum of standardized wavelet coefficients for the continuous part can be bounded by  $c\sqrt{2 \log 2^{j_n}}/0.6745$ , where  $c \geq 1$  is some constant that may depend on  $\sigma_t^2$ , and 0.6745 is the median absolute deviation of the standard normal distribution.

For  $[Y, Y]^{(avg)}$  in (19), if we take  $\gamma < 2/3$ , the continuous part is dominated by the  $B^{\text{noise}}$  term, and by Lemma A6, we have that with probability tending to one, the maximum of standardized wavelet coefficients is bounded by  $\sqrt{2 \log 2^{j_n}}/0.6745$ . Therefore, we may use threshold

$$T_{j_n} = \frac{\sqrt{2 \log 2^{j_n}}}{0.6745}. \tag{20}$$

If any standardized wavelet coefficients exceed threshold  $T_{j_n}$ , then their associated locations are considered to be estimated jump locations. Let  $\{\tau_1, \dots, \tau_q\}$  be  $q$  true jump locations in  $X_t, t \in [0, 1]$ . Denote those estimated locations by  $\{\hat{\tau}_1, \dots, \hat{\tau}_{\hat{q}}\}$ . It is shown in Wang (1995) [38] and Raimondo (1998) [41] that at chosen frequency level  $2^{j_n}$ , we have  $P(\hat{q} = q) \rightarrow 1$  as  $n \rightarrow \infty$  and:

$$\sum_{l=1}^q |\hat{\tau}_l - \tau_l| = O_P(2^{-j_n}). \tag{21}$$

Note that for processes  $X$ ,  $[X, X]$ ,  $Y$  and  $[Y, Y]^{(avg)}$ , to differentiate jumps, we have chosen frequency levels  $2^{j_n} \sim n$ ,  $n$ ,  $n/\log^2 n$  and  $n$ , respectively. The orders of wavelet coefficients of different components and convergence rates of wavelet jump location estimators are summarized in Tables B1 and B2 in Appendix B.

### 2.3. Estimation of Jump Variation

Finally, our goal is to estimate jump variation  $\Psi$ . For each estimated jump location  $\hat{\tau}_l$ , we choose a small neighborhood  $\delta_n$  and calculate the average of the process over  $[\hat{\tau}_l - \delta_n, \hat{\tau}_l)$  and  $[\hat{\tau}_l, \hat{\tau}_l + \delta_n]$ . We estimate the jump size by taking the difference of two averages. The jump variation is estimated by the sum of squares of all jump size estimators.

#### 2.3.1. Without Microstructure Noise Assumption

To estimate jump variation without noises, we have the results using  $X$  and  $[X, X]$  in the following two theorems. Since  $[X, X]$  is smoother, its associated method can achieve a convergence rate  $O_P(n^{-1/2})$ , which is higher than convergence rate  $O_P(n^{-1/3})$  for the method based on  $X$ .

**Theorem 1.** Assume we observe  $X(t_i), i = 1, \dots, n, t_i = i/n$ . Under model (1), let  $\bar{X}_{\hat{\tau}_l+}$  be the average of  $X$  over  $[\hat{\tau}_l - \delta_n, \hat{\tau}_l)$  and  $\bar{X}_{\hat{\tau}_l-}$  the average over  $[\hat{\tau}_l, \hat{\tau}_l + \delta_n]$ . For each estimated jump location  $\hat{\tau}_l$ , let  $\hat{L}_l = \bar{X}_{\hat{\tau}_l+} - \bar{X}_{\hat{\tau}_l-}$ . Define

$$\hat{\Psi}^X \equiv \sum_{l=1}^{\hat{q}} \hat{L}_l^2. \tag{22}$$

Choose  $\delta_n \sim n^{-2/3}$ ; we have

$$\hat{\Psi}^X - \Psi = O_P(n^{-1/3}). \tag{23}$$

**Theorem 2.** Assume we observe  $X(t_i), i = 1, \dots, n, t_i = i/n$ . Under model (1), let  $\overline{[X, X]}_{\hat{\tau}_l+}$  be the average of  $[X, X]$  over  $[\hat{\tau}_l - \delta_n, \hat{\tau}_l)$  and  $\overline{[X, X]}_{\hat{\tau}_l-}$  the average of  $[X, X]$  over  $[\hat{\tau}_l, \hat{\tau}_l + \delta_n]$ . For each estimated jump location  $\hat{\tau}_l$ , let  $\hat{L}_l^2 = \overline{[X, X]}_{\hat{\tau}_l+} - \overline{[X, X]}_{\hat{\tau}_l-}$ . Define

$$\hat{\Psi}^{[X, X]} \equiv \sum_{l=1}^{\hat{q}} \hat{L}_l^2. \tag{24}$$

Choose  $\delta_n \sim n^{-1/2}$ ; we have

$$\hat{\Psi}^{[X, X]} - \Psi = O_P(n^{-1/2}). \tag{25}$$

#### 2.3.2. With the Microstructure Noise Assumption

For noisy observations, we are able to improve the convergence rate  $O_P(n^{-1/4})$  in Fan and Wang (2007) [39] by using a smoother  $[Y, Y]^{(avg)}$  to obtain a higher convergence rate  $O_P(n^{-4/9})$ .

**Theorem 3.** Under models (1) and (6), let  $\overline{[Y, Y]^{(avg)}}_{\hat{\tau}_l+}$  be the average of  $[Y, Y]^{(avg)}$  over  $[\hat{\tau}_l - \delta_n, \hat{\tau}_l)$  and  $\overline{[Y, Y]^{(avg)}}_{\hat{\tau}_l-}$  be the average of  $[Y, Y]^{(avg)}$  over  $[\hat{\tau}_l, \hat{\tau}_l + \delta_n]$ . For each estimated jump location  $\hat{\tau}_l$ , let  $\tilde{L}_l^2 = \overline{[Y, Y]^{(avg)}}_{\hat{\tau}_l+} - \overline{[Y, Y]^{(avg)}}_{\hat{\tau}_l-}$ . Define

$$\hat{\Psi}^{[Y, Y]^{(avg)}} \equiv \sum_{l=1}^{\hat{q}} \tilde{L}_l^2. \tag{26}$$

Let  $M = n^\gamma$ ,  $0 < \gamma \leq 2/3$ . If we choose  $\delta_n \sim n^{(2/3)\gamma-1}$ , then we have

$$\hat{\Psi}^{[Y, Y]^{(avg)}} - \Psi = O_P\left(n^{-(2/3)\gamma}\right). \quad (27)$$

Moreover, the convergence rate is arbitrarily close to  $O_P(n^{-4/9})$  as  $\gamma$  can get arbitrarily close to  $2/3$  for the threshold in Equation (20). For  $\gamma = 2/3$ , if we choose the threshold to be  $c\sqrt{2 \log 2^{j_n}}/0.6745$  for some constant  $c > 1$ , then the convergence rate is  $O_P(n^{-4/9})$ .

See the proof of Theorems 1–3 in Section 6. The orders of the convergence rates of the jump variation estimators are summarized in Table B3 in Appendix B.

### 3. Simulations

We have conducted a simulation study that assumes  $32,768 = 2^{15}$  number of observations for one day. Here are the simulation model and procedure.

1. A sample path of  $\sigma_t^2$  is from the geometric OU volatility model

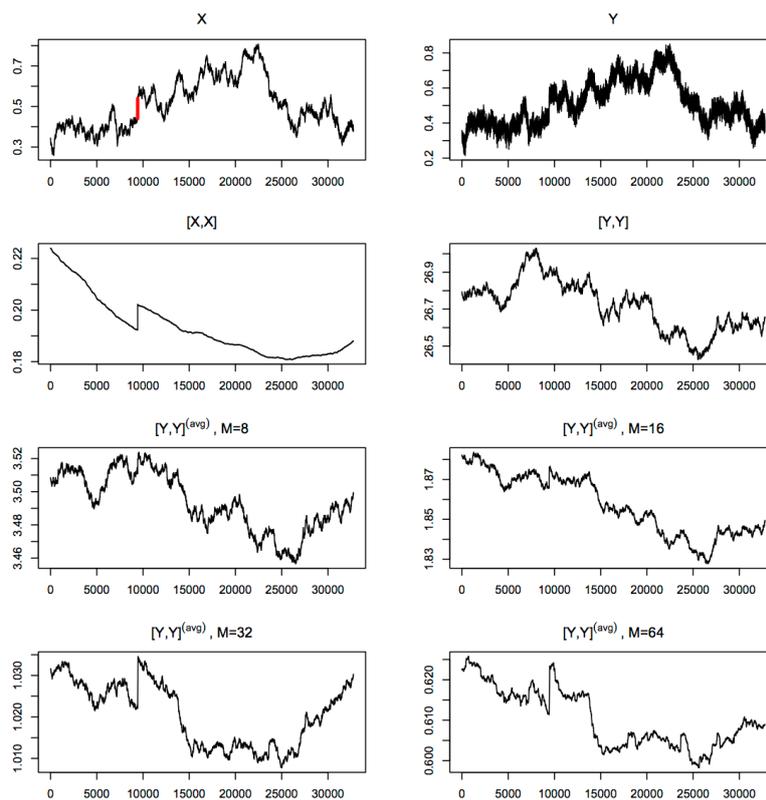
$$d \log \sigma_t^2 = -2.5(\log \sigma_t^2 - \log 0.25)dt + 0.5dW_t. \quad (28)$$

2. A sample path of  $X_t$  is from

$$dX_t = \sigma_t dB_t \quad (29)$$

with  $\text{corr}(W_t, B_t) = -0.5$ .

3. Jump locations are randomly selected, and three jumps are added to  $X_t$  with size from i.i.d.  $N(0, 0.3^2)$ .
4. Noises  $\epsilon_t$ 's are from i.i.d.  $N(0, \eta^2)$  with  $\eta$  at four different levels: 0.01, 0.02, 0.03, 0.04. Then, a sample path of  $Y_t$  is from  $Y_t = X_t + \epsilon_t$ .
5. Realized volatility processes are calculated using a moving window of 32,768 observations. We actually simulate 65,536 of records, so that we have 32,768 complete observations of all processes for calculating these realized volatility processes. There are eight of such processes:  $X$ ,  $[X, X]$ ,  $Y$ ,  $[Y, Y]$ ,  $[Y, Y]_{M=8}^{(avg)}$ ,  $[Y, Y]_{M=16}^{(avg)}$ ,  $[Y, Y]_{M=32}^{(avg)}$ ,  $[Y, Y]_{M=64}^{(avg)}$ . Figure 1 displays their sample paths as an example to show how those processes look under our scheme.
6. Discrete wavelet transformations are performed using Daubechies wavelet D20 on those realized volatility processes. We illustrate the behaviors of the wavelet coefficients at different frequency levels in Figures B1 and B2 in Appendix B using those of the  $X$  process and the  $Y$  process as an example. We use the notations in Percival and Walden (2000) [42]: W1 represents the highest frequency level, W2 the second highest, and so on and so forth. At each frequency level from W1 to W5, if any standardized wavelet coefficient exceeds the threshold  $T_{j_n}$  in (20), we declare the location associated with that coefficient as an estimated jump location. Here, we use  $T_{j_n}$ , since the results in Section 2 and the simulation study show that the method based on  $[Y, Y]^{(avg)}$  is better than others.
7. For each estimated jump location, we estimate the jump size and jump variation as described in Section 2. For  $X$  and  $Y$ , we use intervals of length 64. For  $[X, X]$ ,  $[Y, Y]$ ,  $[Y, Y]_{M=8}^{(avg)}$ ,  $[Y, Y]_{M=16}^{(avg)}$ ,  $[Y, Y]_{M=32}^{(avg)}$ ,  $[Y, Y]_{M=64}^{(avg)}$ , we use intervals of length 128.
8. The whole simulation procedure is repeated 1000 times.



**Figure 1.** Processes simulated with a true jump at location 9424, noise level  $\eta = 0.02$ .

If we do not assume the existence of market microstructure in the model, we are able to improve our estimation of jump variation using the  $[X, X]$  process from using the  $X$  process, as shown in Theorems 1 and 2. The simulation results without adding noises are given in Tables 1 and 2. Table 1 gives the average number of detected jumps and the corresponding standard deviation. We find that we are able to detect all three jumps most of the time using either  $X$  or  $[X, X]$  at any frequency level from W1–W5. Table 2 is the mean squared error (MSE) of jump variation estimation and shows a smaller MSE using  $[X, X]$  at every frequency level from W1 to W5.

**Table 1.** Mean number of detected jumps at frequency levels from W1 to W5 using  $X$  and  $[X, X]$  with the corresponding standard deviation in parenthesis.

level	$X$	$[X, X]$
W1	3.3 (0.8)	3.1 (0.6)
W2	3.2 (0.8)	3.1 (0.6)
W3	3.0 (0.9)	3.3 (0.8)
W4	2.9 (0.8)	4.0 (1.3)
W5	2.8 (0.8)	3.4 (0.9)

**Table 2.** MSE of the jump variation estimation at frequency levels from W1 to W5 using  $X$  and  $[X, X]$ .

level	$X$	$[X, X]$
W1	$3.0 \times 10^{-4}$	$1.4 \times 10^{-5}$
W2	$3.7 \times 10^{-4}$	$3.0 \times 10^{-5}$
W3	$1.7 \times 10^{-3}$	$8.6 \times 10^{-5}$
W4	$6.0 \times 10^{-3}$	$2.4 \times 10^{-4}$
W5	$2.2 \times 10^{-2}$	$8.4 \times 10^{-4}$

If we assume the existence of market microstructure in the model, we are able to improve our estimation of jump variation using  $[Y, Y]^{(avg)}$  processes from using the  $Y$  process, as shown in Theorem 3. The simulation results are summarized in Tables 3 and 4. The results are illustrated at frequency levels from W1 to W5 and noise levels  $\eta = 0.01, 0.02, 0.03, 0.04$  for processes  $Y, [Y, Y], [Y, Y]_{M=8}^{(avg)}, [Y, Y]_{M=16}^{(avg)}, [Y, Y]_{M=32}^{(avg)}$  and  $[Y, Y]_{M=64}^{(avg)}$ . Table 3 displays the average number of detected jumps and the corresponding standard deviation, with Table 4 for the MSEs of the jump variation estimation.

The results show that if we falsely detect some jumps while they are not, it does not affect too much the jump variation estimation since the estimated jump sizes are very small. However, if our detection misses some of the true jumps, it does have a certain effect on jump variation estimation.  $[X, X]$  and  $X$  are both able to detect the jumps fairly well. As the noise level increases, the  $Y$ - and  $[Y, Y]$ -based methods start to miss one or two jumps; the method based on  $[Y, Y]^{(avg)}$  can also miss detecting the jumps when  $M$  is small and noise is large, but it improves with a bigger  $M$ . In terms of the MSE of the jump variation estimation, the methods based on  $[X, X]$  and  $[Y, Y]^{(avg)}$  perform better than those based on  $X$  and  $Y$ , respectively. As the noise level increases, we need to increase  $M$  for the  $[Y, Y]^{(avg)}$  method in order to improve its performance.  $[Y, Y]$  does poorly compared to other processes.

**Table 3.** Mean number of detected jumps at frequency levels from W1 to W5 and noise levels  $\eta = 0.01, 0.02, 0.03, 0.04$  using  $Y, [Y, Y], [Y, Y]_{M=8}^{(avg)}, [Y, Y]_{M=16}^{(avg)}, [Y, Y]_{M=32}^{(avg)}$  and  $[Y, Y]_{M=64}^{(avg)}$ , with the corresponding standard deviation in parenthesis.

level	$\eta$	$Y$	$[Y, Y]$	$[Y, Y]_{M=8}^{(avg)}$	$[Y, Y]_{M=16}^{(avg)}$	$[Y, Y]_{M=32}^{(avg)}$	$[Y, Y]_{M=64}^{(avg)}$
W1	0.01	2.3 (0.9)	2.5 (0.7)	2.5 (0.8)	2.5 (0.8)	2.6 (0.9)	2.7 (0.9)
W2	0.01	2.4 (0.9)	2.7 (0.9)	2.6 (0.8)	2.7 (0.9)	3.6 (1.3)	5.1 (1.8)
W3	0.01	2.3 (0.9)	2.9 (1.1)	3.3 (1.0)	3.6 (1.2)	5.3 (1.8)	7.6 (2.6)
W4	0.01	2.6 (0.8)	3.3 (1.3)	4.2 (1.3)	6.6 (2.1)	7.8 (2.4)	9.9 (3.0)
W5	0.01	2.7 (0.8)	2.7 (1.0)	3.2 (0.9)	5.1 (1.8)	6.6 (2.1)	7.0 (2.4)
W1	0.02	1.5 (1.0)	2.0 (0.9)	2.0 (0.9)	2.1 (0.9)	2.2 (0.9)	2.2 (0.9)
W2	0.02	1.7 (0.9)	2.1 (1.0)	2.1 (0.9)	2.1 (0.9)	2.3 (0.9)	2.8 (1.1)
W3	0.02	1.7 (1.0)	2.3 (1.2)	2.7 (0.9)	2.6 (0.9)	3.0 (1.1)	4.1 (1.5)
W4	0.02	2.3 (0.9)	2.7 (1.4)	3.2 (1.0)	3.7 (1.3)	4.0 (1.4)	5.9 (2.1)
W5	0.02	2.6 (0.9)	2.0 (1.1)	2.8 (0.9)	3.3 (1.2)	4.0 (1.4)	4.7 (1.7)
W1	0.03	1.0 (0.9)	1.5 (0.9)	1.6 (0.9)	1.7 (0.9)	1.8 (0.9)	1.9 (0.9)
W2	0.03	1.1 (0.9)	1.7 (1.1)	1.7 (0.9)	1.8 (0.9)	1.9 (0.9)	2.1 (0.9)
W3	0.03	1.2 (0.9)	1.8 (1.2)	2.5 (1.0)	2.3 (0.9)	2.4 (1.0)	2.8 (1.1)
W4	0.03	1.9 (0.9)	2.2 (1.4)	2.9 (1.1)	3.2 (1.2)	3.1 (1.2)	3.7 (1.5)
W5	0.03	2.4 (0.9)	2.5 (1.1)	2.6 (1.0)	2.9 (1.1)	2.9 (1.1)	3.2 (1.4)
W1	0.04	0.6 (0.8)	1.2 (0.9)	1.2 (0.9)	1.4 (0.9)	1.5 (0.9)	1.6 (0.9)
W2	0.04	0.8 (0.8)	1.3 (1.0)	1.3 (0.9)	1.4 (0.9)	1.6 (0.9)	1.8 (0.9)
W3	0.04	0.8 (0.8)	1.4 (1.2)	2.3 (1.0)	2.0 (2.0)	2.1 (1.0)	2.2 (1.0)
W4	0.04	1.5 (0.9)	1.8 (1.4)	2.6 (1.1)	3.0 (1.2)	2.7 (1.2)	3.0 (1.3)
W5	0.04	2.2 (0.9)	1.2 (1.0)	2.3 (1.1)	2.7 (1.2)	2.5 (1.0)	2.6 (1.1)

**Table 4.** MSE of the jump variation estimation at frequency levels from W1 to W5 and noise levels  $\eta = 0.01, 0.02, 0.03, 0.04$  using  $Y, [Y, Y], [Y, Y]_{M=8}^{(avg)}, [Y, Y]_{M=16}^{(avg)}, [Y, Y]_{M=32}^{(avg)}$  and  $[Y, Y]_{M=64}^{(avg)}$ .

level	$\eta$	$Y$	$[Y, Y]$	$[Y, Y]_{M=8}^{(avg)}$	$[Y, Y]_{M=16}^{(avg)}$	$[Y, Y]_{M=32}^{(avg)}$	$[Y, Y]_{M=64}^{(avg)}$
W1	0.01	$4.3 \times 10^{-4}$	$3.0 \times 10^{-4}$	$8.9 \times 10^{-5}$	$1.2 \times 10^{-4}$	$2.1 \times 10^{-4}$	$5.8 \times 10^{-4}$
W2	0.01	$4.6 \times 10^{-4}$	$1.4 \times 10^{-3}$	$8.9 \times 10^{-5}$	$1.5 \times 10^{-4}$	$2.4 \times 10^{-4}$	$6.3 \times 10^{-4}$
W3	0.01	$1.9 \times 10^{-3}$	$1.8 \times 10^{-3}$	$1.8 \times 10^{-4}$	$1.6 \times 10^{-4}$	$2.7 \times 10^{-4}$	$6.1 \times 10^{-4}$
W4	0.01	$7.2 \times 10^{-3}$	$2.5 \times 10^{-3}$	$2.1 \times 10^{-4}$	$2.7 \times 10^{-4}$	$3.9 \times 10^{-4}$	$5.0 \times 10^{-4}$
W5	0.01	$1.2 \times 10^{-2}$	$8.0 \times 10^{-3}$	$1.2 \times 10^{-2}$	$2.8 \times 10^{-4}$	$4.2 \times 10^{-4}$	$5.1 \times 10^{-4}$
W1	0.02	$2.4 \times 10^{-3}$	$1.8 \times 10^{-3}$	$3.5 \times 10^{-4}$	$2.9 \times 10^{-4}$	$3.3 \times 10^{-4}$	$6.7 \times 10^{-4}$
W2	0.02	$2.0 \times 10^{-3}$	$3.5 \times 10^{-3}$	$3.5 \times 10^{-4}$	$2.8 \times 10^{-4}$	$3.2 \times 10^{-4}$	$6.9 \times 10^{-4}$
W3	0.02	$3.7 \times 10^{-3}$	$4.2 \times 10^{-3}$	$5.2 \times 10^{-4}$	$2.0 \times 10^{-4}$	$3.0 \times 10^{-4}$	$6.2 \times 10^{-4}$
W4	0.02	$7.5 \times 10^{-3}$	$9.9 \times 10^{-3}$	$5.1 \times 10^{-4}$	$3.6 \times 10^{-4}$	$4.6 \times 10^{-4}$	$5.3 \times 10^{-4}$
W5	0.02	$1.4 \times 10^{-2}$	$2.8 \times 10^{-2}$	$4.1 \times 10^{-2}$	$7.3 \times 10^{-4}$	$5.6 \times 10^{-4}$	$5.2 \times 10^{-4}$
W1	0.03	$1.2 \times 10^{-2}$	$6.1 \times 10^{-3}$	$1.6 \times 10^{-3}$	$9.6 \times 10^{-4}$	$7.0 \times 10^{-4}$	$7.8 \times 10^{-4}$
W2	0.03	$1.2 \times 10^{-2}$	$9.2 \times 10^{-3}$	$1.6 \times 10^{-3}$	$5.1 \times 10^{-4}$	$6.8 \times 10^{-4}$	$6.9 \times 10^{-4}$
W3	0.03	$1.3 \times 10^{-2}$	$1.3 \times 10^{-2}$	$1.0 \times 10^{-3}$	$4.9 \times 10^{-4}$	$3.5 \times 10^{-4}$	$7.2 \times 10^{-4}$
W4	0.03	$8.7 \times 10^{-3}$	$2.2 \times 10^{-2}$	$6.4 \times 10^{-4}$	$5.9 \times 10^{-4}$	$5.4 \times 10^{-4}$	$5.8 \times 10^{-4}$
W5	0.03	$1.5 \times 10^{-2}$	$4.9 \times 10^{-2}$	$5.1 \times 10^{-2}$	$1.8 \times 10^{-3}$	$6.9 \times 10^{-4}$	$6.6 \times 10^{-4}$
W1	0.04	$3.5 \times 10^{-2}$	$1.5 \times 10^{-2}$	$4.4 \times 10^{-3}$	$2.7 \times 10^{-3}$	$1.8 \times 10^{-3}$	$1.5 \times 10^{-3}$
W2	0.04	$2.6 \times 10^{-2}$	$2.1 \times 10^{-2}$	$4.7 \times 10^{-3}$	$3.2 \times 10^{-3}$	$1.6 \times 10^{-3}$	$1.2 \times 10^{-3}$
W3	0.04	$3.0 \times 10^{-2}$	$3.1 \times 10^{-2}$	$1.9 \times 10^{-3}$	$1.2 \times 10^{-3}$	$7.2 \times 10^{-4}$	$7.4 \times 10^{-4}$
W4	0.04	$1.1 \times 10^{-2}$	$4.3 \times 10^{-2}$	$1.4 \times 10^{-3}$	$1.2 \times 10^{-3}$	$5.6 \times 10^{-4}$	$6.6 \times 10^{-4}$
W5	0.04	$1.5 \times 10^{-2}$	$7.1 \times 10^{-2}$	$5.3 \times 10^{-2}$	$3.0 \times 10^{-3}$	$1.5 \times 10^{-3}$	$7.7 \times 10^{-4}$

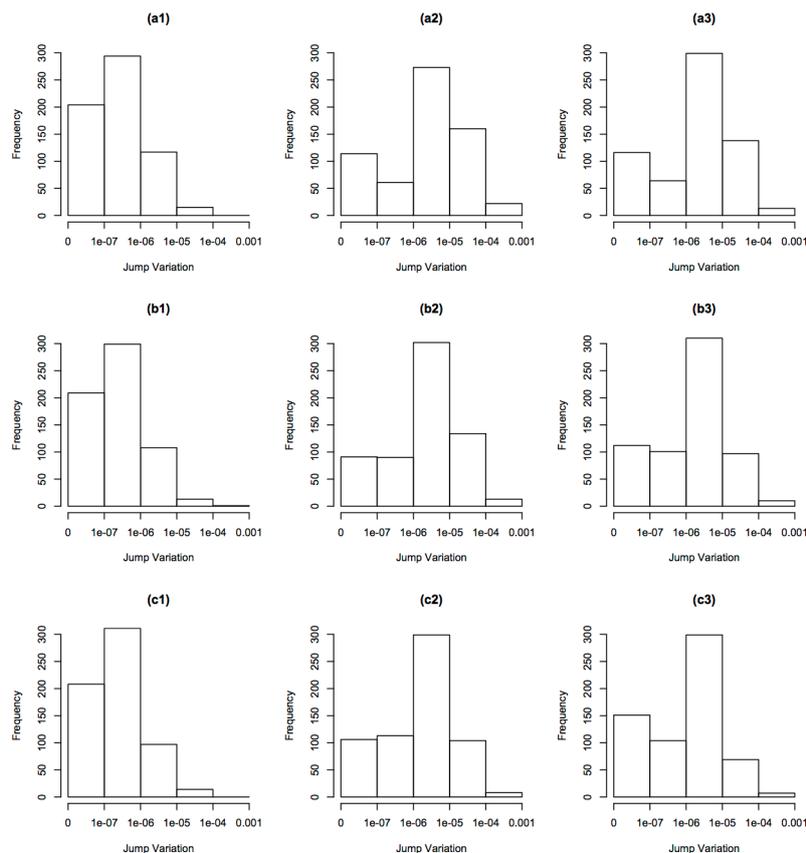
#### 4. Empirical Study

##### 4.1. Distribution of Jump Variation

We study the distribution of jump variation for returns of each Dow Jones Industrial Average stock in January 2013. We collect the tick by tick prices from 9:30 a.m.–4 p.m. for each trading day from the TAQ database. There are 21 trading days and 30 stocks, which correspond to 630 stock-days. We compare the jump variation estimation of the 630 stock-days using  $Y, [Y, Y]$  and  $[Y, Y]^{(avg)}$  and at different sampling frequencies: 1 tick, 2 ticks, 4 ticks. Here, we consider tick time equidistant, as discussed in Andersen et al., 2012 [14].

To detect the jump locations, we use Daubechies wavelet D4 at W4, W3, W2 for sampling frequencies at 1 tick, 2 ticks, 4 ticks, respectively, with the thresholds in Equation (20). We take such wavelet frequency levels to make sure of the consistency of jump detection along different sampling frequencies. To estimate the jump variation, at each estimated jump location, we take the difference between the means of intervals with width four. For  $[Y, Y]^{(avg)}$ , we use  $M = 2$ . We transform the data by the Box–Cox power transformation with power selected by the data, so that the Gaussian assumption of the threshold should be approximately followed.

The histograms of estimated jump variation for these 630 stock-days are illustrated in Figure 2. From left to right, the first column shows the result from using process  $Y$ , the second column using  $[Y, Y]$  and the third column using  $[Y, Y]^{(avg)}$ . From top to bottom, the first row is the result from sampling data at every one tick, the second row at every two ticks and the third row at every four ticks.



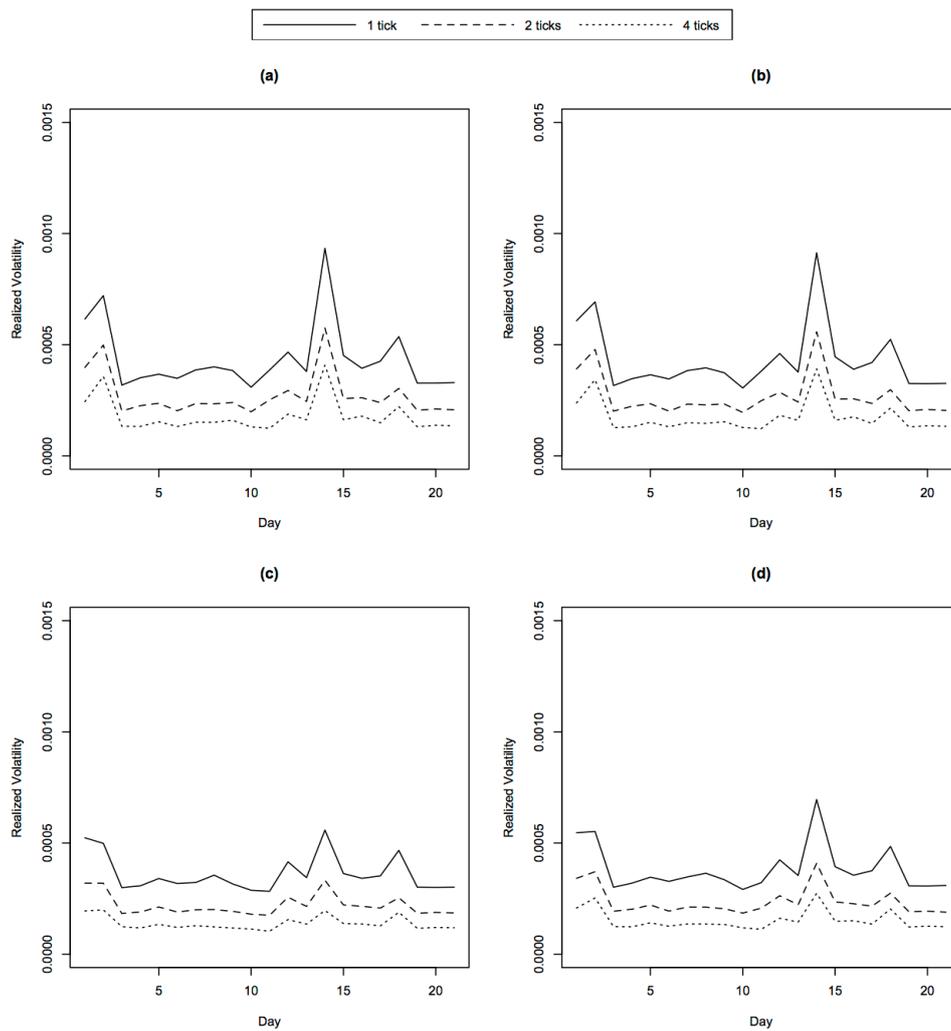
**Figure 2.** Histograms of jump variation: The first row from left to right (a1,a2,a3) shows the results using processes  $Y$ ,  $[Y, Y]$  and  $[Y, Y]^{(avg)}$ , respectively, sampled at every one tick, the second row (b1,b2,b3) sampled at every two ticks and the third row (c1,c2,c3) sampled at every four ticks.

Compared to  $Y$ , processes  $[Y, Y]$  and  $[Y, Y]^{(avg)}$  are able to pick up much larger jump variation at the same sampling frequency level. We find that around 25% of the jump variation is above  $1 \times 10^{-5}$  using  $[Y, Y]$  and  $[Y, Y]^{(avg)}$ , while using  $Y$ , we only have around 1.5% (note that a price change from 100 to 99.99 leads to a squared log return change of around  $1 \times 10^{-8}$ ). On the other hand, for the same process, the distribution of jump variation looks very similar across different sampling frequencies.

#### 4.2. Evidence of Microstructure Noises

To remove the idiosyncratic effects of each single stock, we plot the weighted average of the daily realized volatility using all stocks in Figure 3. They are before removing jump variation in (a) and after removing jump variation using processes  $Y$ ,  $[Y, Y]$  and  $[Y, Y]^{(avg)}$  in (b), (c) and (d), respectively. The solid lines are calculated from sampling at every one tick, the dashed lines at every two ticks and the dotted lines at every four ticks. The weight is decided by the mean level of the realized volatility of each stock. The larger it is, the smaller the weight is, so that each stock will contribute equally to the weighted average.

We see from Figure 3 that there is clear evidence of microstructure noises, since as the sampling frequency increases, the realized volatility increases, as well, in each case, just as we show in theory. Moreover, the effects of removing jumps are more significant using  $[Y, Y]$  and  $[Y, Y]^{(avg)}$  than  $Y$ .



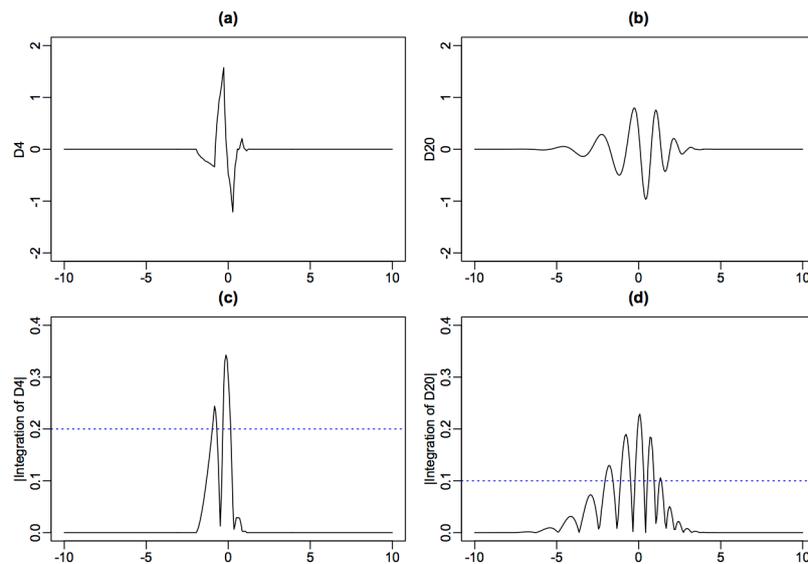
**Figure 3.** Realized volatility: (a) before removing jump variation; (b–d) after removing jump variation using processes  $Y$ ,  $[Y, Y]$  and  $[Y, Y]^{(avg)}$ , respectively. The solid lines are calculated from sampling at every one tick, the dashed lines at every two ticks and the dotted lines at every four ticks.

### 5. Discussion

We develop a nonparametric method for estimating jump variation with noisy high frequency financial data. With a better approach based on new process  $[Y, Y]^{(avg)}$  and the wavelet techniques, we are able to detect and estimate jump variation more efficiently if the variance of microstructure noise  $\eta^2$  is assumed to be a constant. Numerically, we show that the proposed  $[Y, Y]^{(avg)}$  method indeed has smaller mean square errors.

From the empirical results, we observe that the method based on  $[Y, Y]$  outperforms that based on  $Y$ . This may suggest that  $\eta^2$  is not a constant, so if we modify  $\eta^2$  to be some decreasing function of  $n$ , for example,  $\eta_n^2 \sim \log n/n$  and  $E\epsilon^4 \sim (\log n/n)^2$ , then the  $[Y, Y]$ -based method can achieve a better convergence rate than that for  $Y$ .

For the Daubechies wavelets D4 and D20 used in the paper, we display their graphs below in Figure 4, as well as the graphs of the absolute values of their integrations. Using Daubechies wavelet D4, we may pick  $\mathcal{G}_d = (-0.28, 0.14)$  for  $d = 0.2$ , while using Daubechies wavelet D20, we may pick  $\mathcal{G}_d = (-0.21, 0.28)$  for  $d = 0.1$ .



**Figure 4.** (a,b) Daubechies wavelets D4 and D20, respectively; (c,d) absolute values of integrated D4 and D20, respectively.

We leave some open problems for future study, including the extension to estimate jump size, optimal frequency level selection in practice, the irregular spaced observations, data-dependent threshold selection and its sensitivity study.

### 6. Proof of Theorems

**Theorem 1.** (Estimation of jump size/variation using  $X$ ) Let  $\hat{\tau}$  be an estimated jump location. Denote  $\bar{X}_{\hat{\tau}+}$  the average of  $X$  over  $[\hat{\tau}, \hat{\tau} + \delta_n]$  and  $\bar{X}_{\hat{\tau}-}$  the average of  $X$  over  $(\hat{\tau} - \delta_n, \hat{\tau}]$ . We estimate the jump size by

$$\hat{L} = \bar{X}_{\hat{\tau}+} - \bar{X}_{\hat{\tau}-}. \tag{30}$$

If we choose  $\delta_n \sim n^{-2/3}$ , then

$$\hat{L} - L = O_P(n^{-1/3}) \tag{31}$$

and

$$\hat{L}^2 - L^2 = O_P(n^{-1/3}). \tag{32}$$

**Proof.** Denote  $m_{\pm}$  the number of  $t_i$  in  $\hat{I}_+ = [\hat{\tau}, \hat{\tau} + \delta_n]$  and  $\hat{I}_- = (\hat{\tau} - \delta_n, \hat{\tau}]$ , respectively. Then,  $|m_+ - m_-| \leq 2$  and  $m_+ \sim m_- \sim n\delta_n$ . We decompose  $\hat{L} - L$  as

$$\hat{L} - L = U_1 + U_2 + U_3, \tag{33}$$

where  $U_1, U_2$  and  $U_3$  correspond to the drift, diffusion and jump part.

$$\begin{aligned} U_1 &= \frac{1}{m_+} \sum_{t_i \in \hat{I}_+} \int_{\hat{\tau}}^{t_i} \mu_s ds - \frac{1}{m_-} \sum_{t_i \in \hat{I}_-} \int_{t_i}^{\hat{\tau}} \mu_s ds \\ &= \frac{1}{m_+} O_P\left(\sum_{i=1}^{m_+} i/n\right) + \frac{1}{m_-} O_P\left(\sum_{i=1}^{m_-} i/n\right) = O_P(\delta_n), \end{aligned} \tag{34}$$

$$\begin{aligned}
 U_2 &= \frac{1}{m_+} \sum_{t_i \in \hat{I}_+} \int_{\hat{\tau}}^{t_i} \sigma_s dW_s - \frac{1}{m_-} \sum_{t_i \in \hat{I}_-} \int_{\hat{\tau}}^{t_i} \sigma_s dW_s \\
 &= \frac{1}{m_+} \sum_{t_i \in \hat{I}_+} \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s - \frac{1}{m_-} \sum_{t_i \in \hat{I}_-} \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s \\
 &= \left( \frac{1}{m_+} \sum_{t_i \in \hat{I}_+} \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s - \frac{1}{m_+} \sum_{t_i \in I_+} \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s \right) \\
 &\quad - \left( \frac{1}{m_-} \sum_{t_i \in \hat{I}_-} \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s - \frac{1}{m_-} \sum_{t_i \in I_-} \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s \right) \\
 &\quad + \left( \frac{1}{m_+} \sum_{t_i \in I_+} \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s - \frac{1}{m_-} \sum_{t_i \in I_-} \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s \right) \\
 &= A_1 - A_2 + A_3,
 \end{aligned} \tag{35}$$

where  $I_+ = [\tau, \tau + \delta_n]$  and  $I_- = [\tau - \delta_n, \tau)$ .

Because  $\hat{\tau} - \tau = O_p(n^{-1})$ , the total non-zero terms in  $A_1$  and  $A_2$  is  $O_p(1)$ . Furthermore, by maximal martingale inequality,

$$\begin{aligned}
 &P \left( \max \left\{ \left| \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s \right|, t_i \in \hat{I}_+ \cup I_+ \right\} > \sqrt{\delta_n \log n} \right) \\
 &\leq P \left( \max_{\tau-\delta_n \leq t_i \leq \tau+2\delta_n} \left\{ \left| \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s \right| \right\} > \sqrt{\delta_n \log n} \right) + P(|\tau - \hat{\tau}| > \delta_n) \\
 &= E \left[ P_\tau \left( \max_{\tau-\delta_n \leq t_i \leq \tau+2\delta_n} \left\{ \left| \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s \right| \right\} > \sqrt{\delta_n \log n} \right) \right] + o(1) \\
 &\leq \frac{1}{\delta_n \log n} E \left[ \int_{\tau-2\delta_n}^{\tau+2\delta_n} E(\sigma_s^2) ds \right] + o(1) \leq \max E(\sigma_s^2) O \left( \frac{1}{\log n} \right) + o(1) \rightarrow 0.
 \end{aligned} \tag{36}$$

Therefore,  $A_1 \sim A_2 \sim O_p \left( \frac{1}{n\delta_n} \sqrt{\delta_n \log n} \right)$ .

For  $A_3$ , since

$$E_\tau \left( \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s \right) = 0 \tag{37}$$

and

$$E_\tau \left( \int_{\tau-2\delta_n}^{t_i} \sigma_s dW_s \int_{\tau-2\delta_n}^{t_j} \sigma_s dW_s \right) = \int_{\tau-2\delta_n}^{t_i \wedge t_j} E(\sigma_s)^2 ds \tag{38}$$

and  $\tau$  independent of  $(\sigma, W)$ , we have

$$E(A_3) = 0 \tag{39}$$

and

$$\begin{aligned}
 &Var(A_3) = E(Var_\tau(A_3)) \\
 &\leq E \left[ \frac{3}{m_+} \sum_{t_i \in I_+} \int_{\tau-2\delta_n}^{t_i} E(\sigma_s^2) ds + \frac{4}{m_-} \sum_{t_i \in I_-} \int_{\tau-2\delta_n}^{t_i} E(\sigma_s^2) ds \right] \\
 &\leq 4 \max E(\sigma_s^2) \left( 4\delta_n + \frac{1}{m_+} \sum_{i=1}^{m_+} i/n + \frac{1}{m_-} \sum_{i=1}^{m_-} i/n \right) = O(\delta_n).
 \end{aligned} \tag{40}$$

Together, we have

$$U_2 = O_p \left( \frac{1}{n} \sqrt{\frac{\log n}{\delta_n}} + \sqrt{\delta_n} \right). \tag{41}$$

Finally,

$$\begin{aligned}
 U_3 &= \frac{1}{m_+} \sum_{t_i \in \hat{L}_+} 1_{\{t_i \geq \tau\}} L - \frac{1}{m_-} \sum_{t_i \in \hat{L}_-} 1_{\{t_i \geq \tau\}} L \\
 &= \begin{cases} \frac{1}{m_+} n(\tau - \hat{\tau})L & \text{if } \tau \geq \hat{\tau} \\ \frac{1}{m_-} n(\hat{\tau} - \tau)L & \text{if } \tau < \hat{\tau} \end{cases} \\
 &= O_P\left(\frac{1}{n\delta_n}\right).
 \end{aligned}
 \tag{42}$$

Thus, if we take  $\delta_n \sim n^{-2/3}$ , we have

$$\hat{L} - L = O_P(n^{-1/3}).
 \tag{43}$$

And

$$\hat{L}^2 - L^2 = (\hat{L} - L)(\hat{L} + L) = O_P(n^{-1/3}).
 \tag{44}$$

□

**Theorem 2.** (Estimation of jump variation using  $[X, X]$ ) Denote  $\overline{[X, X]}_{\hat{\tau}_+}$  the average of  $[X, X]$  over  $[\hat{\tau}, \hat{\tau} + \delta_n]$  and  $\overline{[X, X]}_{\hat{\tau}_-}$  the average of  $[X, X]$  over  $(\hat{\tau} - \delta_n, \hat{\tau}]$ . We estimate the jump variation by

$$\widehat{L}^2 = \overline{[X, X]}_{\hat{\tau}_+} - \overline{[X, X]}_{\hat{\tau}_-}.
 \tag{45}$$

If we choose  $\delta_n \sim n^{-1/2}$ , then

$$\widehat{L}^2 - L^2 = O_P(n^{-1/2}).
 \tag{46}$$

**Proof.** Similar to the proof of the previous theorem, we decompose the  $\widehat{L}^2 - L^2$  as

$$\widehat{L}^2 - L^2 = U_1 + U_2 + U_3,
 \tag{47}$$

where  $U_1, U_2, U_3$  correspond to the drift, diffusion and jump parts.

We still have

$$U_1 = O_P(\delta_n)
 \tag{48}$$

and

$$U_3 = O_P\left(\frac{1}{n\delta_n}\right).
 \tag{49}$$

$U_2$  has changed to

$$U_2 = O_P\left(n^{-1/2}\left(\frac{1}{n}\sqrt{\frac{\log n}{\delta_n}} + \sqrt{\delta_n}\right)\right).
 \tag{50}$$

Thus, if we take  $\delta_n \sim n^{-1/2}$ , we have

$$\widehat{L}^2 - L^2 = O_P(n^{-1/2}).
 \tag{51}$$

□

**Theorem 3.** (Estimation of jump variation using  $[Y, Y]^{(avg)}$ ) Denote  $\overline{[Y, Y]^{(avg)}}_{\hat{\tau}_+}$  the average of  $[Y, Y]^{(avg)}$  over  $[\hat{\tau}, \hat{\tau} + \delta_n]$  and  $\overline{[Y, Y]^{(avg)}}_{\hat{\tau}_-}$  the average of  $[Y, Y]^{(avg)}$  over  $(\hat{\tau} - \delta_n, \hat{\tau}]$ . We estimate the jump variation by

$$\widetilde{L}^2 = \overline{[Y, Y]^{(avg)}}_{\hat{\tau}_+} - \overline{[Y, Y]^{(avg)}}_{\hat{\tau}_-}.
 \tag{52}$$

Let  $M = n^\gamma$ ,  $0 < \gamma \leq 2/3$ . If we choose  $\delta_n \sim n^{(2/3)\gamma-1}$ , then

$$\widetilde{L}^2 - L^2 = O_P(n^{-(2/3)\gamma}). \tag{53}$$

Moreover, the convergence rate is arbitrarily close to  $O_P(n^{-4/9})$  as  $\gamma$  can get arbitrarily close to  $2/3$  for the threshold in Equation (20). For  $\gamma = 2/3$ , if we choose the threshold to be  $c\sqrt{2 \log 2^n} / 0.6745$  for some constant  $c > 1$ , then, the convergence rate is  $O_P(n^{-4/9})$ .

**Proof.** This time, we have

$$U_1 = O_P(\delta_n), \tag{54}$$

$$U_2 = O_P\left(n^{(1/2)-\gamma} \left(\frac{1}{n} \sqrt{\frac{\log n}{\delta_n}} + \sqrt{\delta_n}\right)\right), \tag{55}$$

$$U_3 = O_P\left(\frac{1}{n\delta_n}\right). \tag{56}$$

Thus, if we take  $\delta_n \sim n^{(2/3)\gamma-1}$ , we have

$$\widetilde{L}^2 - L^2 = O_P(n^{-(2/3)\gamma}).$$

□

**Acknowledgments:** Wang’s research is partly supported by NSF Grants DMS-10-5635, DMS-12-65203 and DMS-15-28375. The authors would like to thank the editor, associate editor and referees for suggestions that led to the improvement of the paper.

**Author Contributions:** Yazhen Wang proposed the project, and Xin Zhang carried out the work. Xin Zhang and Yazhen Wang contributed to the development of methodology and theory. Xin Zhang, Donggyu Kim and Yazhen Wang contributed to the theoretic proof. Xin Zhang contributed the numerical results. Xin Zhang and Yazhen Wang contributed to the writing of the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Proof of Lemmas

**Lemma A1.** (Order of wavelet coefficients in the continuous case) Suppose that  $\int |\psi(t)|(1 + |t|)dt < \infty$ . If  $f$  is Hölder continuous with exponent  $\alpha$ ,  $0 < \alpha \leq 1$ , i.e.,  $|f(x) - f(y)| \leq C|x - y|^\alpha$ . Then, its wavelet transform  $f_{j,k}$  satisfies

$$|f_{j,k}| \leq C_1 2^{-j(\alpha+1/2)}, \tag{A1}$$

where  $C_1 = C \int |y|^\alpha |\psi(y)| dy$ .

**Proof.** First, note that  $\int \psi(t)dt = 0$ , so we have

$$\begin{aligned} f_{j,k} &= 2^{j/2} \int f(t)\psi(2^j t - k)dt - 2^{j/2} \int f(2^{-j}k)\psi(2^j t - k)dt \\ &= 2^{j/2} \int (f(t) - f(2^{-j}k))\psi(2^j t - k)dt. \end{aligned} \tag{A2}$$

Taking the absolute value, we have

$$\begin{aligned} |f_{j,k}| &\leq 2^{j/2} \int C|t - 2^{-j}k|^\alpha |\psi(2^j t - k)|dt \\ &= 2^{j/2} \int C|2^{-j}y|^\alpha |\psi(y)|2^{-j}dy \\ &= 2^{-j(\alpha+1/2)} \int C|y|^\alpha |\psi(y)|dy. \end{aligned} \tag{A3}$$

Note that the integral in the last equation is finite. The result then follows.  $\square$

**Lemma A2.** (Necessary condition: order of wavelet coefficients in the continuous case) Suppose that  $\psi$  is compactly supported. Suppose also that  $f \in L^2(\mathbb{R})$  is bounded and continuous. If for some  $\alpha \in (0, 1)$ , the wavelet transform  $f_{j,k}$  of  $f$  satisfies

$$|f_{j,k}| \leq C2^{-j(\alpha+1/2)},$$

then  $f$  is Hölder continuous with exponent  $\alpha$ .

See Section 2.9 in Daubechies, 1992 [40].

**Lemma A3.** (Order of wavelet coefficients in the jump case) Suppose that  $\int |\psi(t)|(1 + |t|)dt < \infty$ . If  $g$  is Hölder continuous with exponent  $\alpha$ ,  $0 < \alpha \leq 1$ . Let

$$f(t) = \begin{cases} g(t) & \text{if } t < s \\ g(t) + L & \text{if } t \geq s. \end{cases} \tag{A4}$$

Then, for sufficiently large  $j$  with  $2^j s - k \in \mathcal{G}_d$ , there exists a constant  $C_L$  depending on  $L$ , such that

$$|f_{j,k}| \geq C_L 2^{-j/2}, \tag{A5}$$

where  $\mathcal{G}_d \subset \left\{ x : \left| \int_{(-\infty, x)} \psi(y)dy \right| \geq d \right\}$  for some positive constant  $d$ .

More specifically, for  $j \geq N$ , we have

$$|f_{j,k}| \geq 2^{-j/2} (|L|d - C_1), \tag{A6}$$

where  $N = \max \left( 0, \frac{1}{\alpha} \log_2 \frac{C_1}{|L|d} + 1 \right)$  and  $C_1$  is defined in Equation (A1).

**Proof.**

$$f_{j,k} = 2^{j/2} \int_{(-\infty, s)} (-L)\psi(2^j t - k)dt + 2^{j/2} \int g(t)\psi(2^j t - k)dt. \tag{A7}$$

For the first part, we have its absolute value

$$\begin{aligned} |2^{j/2}(-L) \int_{(-\infty, s)} \psi(2^j t - k)dt| &= 2^{-j/2}|L| \left| \int_{(-\infty, 2^j s - k)} \psi(y)dy \right| \\ &\geq 2^{-j/2}|L|d. \end{aligned} \tag{A8}$$

For the second part, by Lemma A1, we have its absolute value

$$|2^{j/2} \int g(t)\psi(2^j t - k)dt| \leq C_1 2^{-j(\alpha+1/2)}. \tag{A9}$$

Let  $N = \max \left( 0, \frac{1}{\alpha} \log_2 \frac{C_1}{|L|d} + 1 \right)$ . Putting Equations (A8) and (A9) together, we have for  $j \geq N$ ,

$$\begin{aligned} |f_{j,k}| &\geq 2^{-j/2} \left( |L|d - C_1 2^{-j\alpha} \right) \\ &\geq 2^{-j/2} \left( |L|d - C_1 2^{-N\alpha} \right) \\ &\geq 2^{-j/2} (|L|d - C_1). \end{aligned} \tag{A10}$$

The proof is complete.  $\square$

**Lemma A4.** (Robustness of the median) Let  $f_{j,k} = f'_{j,k} + f^*_{j,k}$ , where  $f'_{j,k}$  is the wavelet coefficient corresponding to the continuous part and  $f^*_{j,k}$  to the jump part. Suppose no more than  $a\%$  of  $\{f^*_{j,k}, k = 1, \dots, 2^j\}$  is nonzero. Let  $m_q$  be the  $q$ -th quantile of  $\{|f_{j,k}|, k = 0, \dots, 2^j - 1\}$  and  $m'_q$  be the  $q$ -th quantile of  $\{|f'_{j,k}|, k = 0, \dots, 2^j - 1\}$ . Then

$$m'_{0.5-a\%} - m'_{0.5} \leq m_{0.5} - m'_{0.5} \leq m'_{0.5+a\%} - m'_{0.5}. \tag{A11}$$

**Proof.** Suppose there is only one nonzero  $f^*_{j,k}$ , so  $a\% = 1/2^j$ . Let  $f^*_{j,l}$  denote this nonzero coefficient. If its corresponding  $|f'_{j,l}| \leq m'_{0.5}$ , we have the following three cases:

Case 1:  $|f_{j,l}| \leq m'_{0.5}$ . Since all other  $f_{j,k} = f'_{j,k}$ , we have  $m_{0.5} - m'_{0.5} = 0$ .

Case 2:  $m'_{0.5} < |f_{j,l}| \leq m'_{0.5+a\%}$ . Then, we have  $m_{0.5} = |f_{j,l}|$ , so  $m_{0.5} - m'_{0.5} = |f_{j,l}| - m'_{0.5} \leq m'_{0.5+a\%} - m'_{0.5}$ .

Case 3:  $|f_{j,l}| > m'_{0.5+a\%}$ . Then, we have  $m_{0.5} = m'_{0.5+a\%}$ , so  $m_{0.5} - m'_{0.5} = m'_{0.5+a\%} - m'_{0.5}$ .

Together, we have

$$m_{0.5} - m'_{0.5} \leq m'_{0.5+a\%} - m'_{0.5}.$$

If its corresponding  $|f'_{j,l}| > m'_{0.5}$ , similarly, we can prove

$$m'_{0.5-a\%} - m'_{0.5} \leq m_{0.5} - m'_{0.5}.$$

This proves the result when we have only one nonzero  $f^*_{j,k}$ . For more than one nonzero  $f^*_{j,k}$  coefficient, we can use iteration by adding one at one time. Therefore, the result holds.  $\square$

**Lemma A5.** (Order of the maximum to median ratio of wavelet coefficients in the jump case) Suppose  $f_t$  is a Hölder continuous process with exponent  $\alpha$ ,  $0 < \alpha \leq 1$ , except for a jump of size  $L$ . Then, for sufficiently large  $j$ ,

$$\frac{\max(|f_{j,k}|)}{\text{median}(|f_{j,k}|)} \geq C'_L 2^{j\alpha}, \tag{A12}$$

where  $\max(|f_{j,k}|) = \max\{|f_{j,k}|, k = 0, \dots, 2^j - 1\}$ ,  $\text{median}(|f_{j,k}|) = \text{median}\{|f_{j,k}|, k = 0, \dots, 2^j - 1\}$  and  $C'_L$  is a constant and increases as  $L$  increases.

**Proof.** First, we consider  $\max(|f_{j,k}|)$ . In a neighborhood of point  $t$  where  $f_t$  is continuous, by Lemmas A1 and A2, wavelet coefficients at such points are at order  $2^{-j(\alpha+1/2)}$  with  $\alpha$  close to  $1/2$ . On the other hand, nearby a jump point in  $f_t$ , by Lemma A3, for sufficiently large  $j$ ,  $|f_{j,k}|$  converges no faster than  $2^{-j/2}$ . Since the order  $2^{-j/2}$  is much larger than  $2^{-j(\alpha+1/2)}$ , we have  $\max(|f_{j,k}|)$  regulated by the jump part.

Second, we consider  $\text{median}(|f_{j,k}|)$ . Assume we have  $n = 2^J$  discrete observations and that the number of jumps is  $R$ . Since the wavelet functions have compact support, among the wavelet coefficients at a fine level, no more than  $bR$  (%) of them will be affected by the jumps, where  $b$  is very small. For example, if we use a wavelet of support length  $S$ , we have  $2^{J-1}$  wavelet coefficients at the finest level, and there will be no more than  $SR$  of them affected by jumps. Thus, in this case, we have  $bR = SR/2^{J-1} = (2S/n)R$ , which goes to zero as  $n$  goes to infinity. Since the number of coefficients affected by jumps is very small compared to the total number of wavelets coefficients, i.e.  $bR \ll 0.5$ , if we order the wavelet coefficients at fine levels, by Lemma A4,  $\text{median}(|f_{j,k}|)$  will be surely regulated by the continuous part.

Together, when  $j$  is sufficiently large, the ratio

$$\frac{\max(|f_{j,k}|)}{\text{median}(|f_{j,k}|)} \geq \frac{C_L 2^{-j/2}}{C 2^{-j(\alpha+1/2)}} = C'_L 2^{j\alpha}. \tag{A13}$$

□

**Lemma A6.** (Order of the maximum to median ratio of wavelet coefficients in the Brownian motion case) Let  $B_t, 0 \leq t \leq 1$  be a Brownian motion and  $B_{j,k}$  be the corresponding wavelet coefficients. Then,

$$\lim_{j \rightarrow \infty} P \left( \frac{\max(|B_{j,k}|)}{\text{median}(|B_{j,k}|)} \geq \frac{\sqrt{2 \log 2^j}}{0.6745} \right) = 0, \tag{A14}$$

where  $\max(|B_{j,k}|) = \max\{|B_{j,k}|, k = 0, \dots, 2^j - 1\}$ , and  $\text{median}(|B_{j,k}|)$  is the median of  $|B_{j,k}|$  for  $k = 0, \dots, 2^j - 1$ .

**Proof.** For fixed  $j$ ,  $B_{j,k}$  is a stationary Gaussian process with mean zero and variance function

$$\text{Var}(B_{j,k}) = 2^j \int (\psi(2^j t - k))^2 dt = \int (\psi(t))^2 dt = 1.$$

Therefore, we have

$$\text{median}(|B_{j,k}|) = 0.6745. \tag{A15}$$

We have

$$\begin{aligned} P \left( \max_k (|B_{j,k}|) \geq \sqrt{2 \log 2^j} \right) &\leq \sum_{k=0}^{2^j-1} P \left( |B_{j,k}| \geq \sqrt{2 \log 2^j} \right) \\ &\leq 2^j \frac{1}{\sqrt{2\pi} \sqrt{2 \log 2^j}} \exp \left( -\frac{2 \log 2^j}{2} \right) \\ &\leq \frac{1}{2\sqrt{\pi \log 2^j}} \end{aligned} \tag{A16}$$

where the second inequality is derived using the inequality  $P(B_{j,k} > x) \leq \frac{1}{\sqrt{2\pi}x} \exp(-\frac{x^2}{2})$ .

By (A15) and (A16), Equation (A14) follows. □

**Remark.** The covariance function of  $B_{j,k}$  is

$$\begin{aligned} E(B_{j,k} B_{j,k'}) &= E \left[ \left( 2^{j/2} \int \psi(2^j t - k) dB_t \right) \left( 2^{j/2} \int \psi(2^j t - k') dB_t \right) \right] \\ &= 2^j \int \psi(2^j t - k) \psi(2^j t - k') dt \\ &= \int \psi(t) \psi(t - (k - k')) dt, \end{aligned}$$

which is zero when  $|k - k'|$  is bigger than the range of  $\text{supp}(\psi)$ . That is,  $B_{j,k}, k = 0, 1, \dots, 2^j - 1$ , a stationary Gaussian process with the  $m$ -dependent covariance structure, and we can derive an asymptotic distribution of  $\max_k (|B_{j,k}|)$ .

**Lemma A7.** (Equivalent representation of the realized volatility of  $X$ ) Suppose  $X$  follows an Itô process satisfying the stochastic differential equation

$$dX_t = \mu_t dt + \sigma_t dB_t, \quad t \in [0, 1]. \tag{A17}$$

The realized volatility of  $X$  at  $t = 1$  is defined by

$$[X, X]_1 \equiv \sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})^2, \tag{A18}$$

where  $\{t_k, k = 1, \dots, n\}$  is an equidistant partition on  $[0, 1]$ . Then, we have

$$\sqrt{\frac{n}{2}} \left( [X, X]_1 - \int_0^1 \sigma_s^2 ds \right) \xrightarrow{\mathcal{L} \text{ stably}} \int_0^1 \sigma_s^2 dB_s^{\text{discrete}}, \tag{A19}$$

where stable convergence is defined in Rényi, 1963 [43], Aldous and Eagleson, 1978 [44], and Hall and Heyde, 1980 [45].

**Proof.** By Itô’s Lemma, we have for any  $k = 1, \dots, n$ ,

$$(X_{t_k} - X_{t_{k-1}})^2 = 2 \int_{t_{k-1}}^{t_k} (X_s - X_{t_{k-1}}) dX_s + \int_{t_{k-1}}^{t_k} \sigma_s^2 ds. \tag{A20}$$

Then, we have

$$\begin{aligned} [X, X]_1 &= \sum_{k=1}^n \left\{ 2 \int_{t_{k-1}}^{t_k} (X_s - X_{t_{k-1}}) dX_s + \int_{t_{k-1}}^{t_k} \sigma_s^2 ds \right\} \\ &= 2 \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (X_s - X_{t_{k-1}}) dX_s + \int_0^1 \sigma_s^2 ds. \end{aligned} \tag{A21}$$

Now, it is enough to show

$$\sqrt{2n} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (X_s - X_{t_{k-1}}) dX_s \xrightarrow{\mathcal{L} \text{ stably}} \int_0^1 \sigma_t^2 dB_t^{\text{discrete}}. \tag{A22}$$

Let  $X_t = X_t^D + X_t^M$ , where  $X_t^D = \int_0^t \mu_s ds$  and  $X_t^M = \int_0^t \sigma_s dB_s$ . Simple algebraic manipulations show

$$\begin{aligned} &\sum_{k=1}^n \int_{t_{k-1}}^{t_k} (X_s - X_{t_{k-1}}) dX_s \\ &= \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (X_s^D - X_{t_{k-1}}^D) dX_s^D + \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (X_s^M - X_{t_{k-1}}^M) dX_s^D \\ &\quad + \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (X_s^D - X_{t_{k-1}}^D) dX_s^M + \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (X_s^M - X_{t_{k-1}}^M) dX_s^M \\ &= T_1 + T_2 + T_3 + T_4. \end{aligned}$$

For  $T_1$ , since  $\mu_t$  is bounded, we have

$$\begin{aligned} \max_k \max_{s \in [t_{k-1}, t_k]} |X_s^D - X_{t_{k-1}}^D| &\leq \max_{t \in [0,1]} |\mu_t| \max_k \max_{s \in [t_{k-1}, t_k]} |s - t_{k-1}| \\ &\leq O_p(n^{-1}) \text{ a.s.} \end{aligned} \tag{A23}$$

Then, we have

$$\begin{aligned} |T_1| &\leq O_p(n^{-1}) \sum_{k=1}^n \int_{t_{k-1}}^{t_k} |\mu_s| ds \\ &\leq O_p(n^{-1}) \max_{t \in [0,1]} |\mu_t| \sum_{k=1}^n (t_k - t_{k-1}) \\ &= O_p(n^{-1}) \text{ a.s.,} \end{aligned} \tag{A24}$$

where the first and last lines are due to (A23) and the boundedness of  $\mu_t$ , respectively.

For  $T_2$ , we have

$$\begin{aligned} T_2 &= \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (X_s^M - X_{t_{k-1}}^M)(\mu_s - \mu_{t_{k-1}})ds + \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (X_s^M - X_{t_{k-1}}^M)\mu_{t_{k-1}}ds \\ &= T_{21} + T_{22}. \end{aligned}$$

First consider that  $T_{21}$ .  $\mu_t$  is continuous, and so, we have  $|\mu_s - \mu_{\lfloor sn \rfloor/n}| \rightarrow 0$  a.s. The boundedness of  $\mu_t^2$  and dominated convergence theorem imply  $E \left[ \int_0^1 (\mu_s - \mu_{\lfloor ns \rfloor/n})^2 ds \right] = o(1)$ . Thus, we have

$$\begin{aligned} E [|T_{21}|] &= E \left[ \left| \int_0^1 (X_s^M - X_{\lfloor ns \rfloor/n}^M)(\mu_s - \mu_{\lfloor ns \rfloor/n}) ds \right| \right] \\ &\leq \left( E \left[ \int_0^1 (X_s^M - X_{\lfloor ns \rfloor/n}^M)^2 ds \right] \right)^{1/2} \left( E \left[ \int_0^1 (\mu_s - \mu_{\lfloor ns \rfloor/n})^2 ds \right] \right)^{1/2} \\ &\leq O(n^{-1/2}) E \left[ \int_0^1 (\mu_s - \mu_{\lfloor ns \rfloor/n})^2 ds \right]^{1/2} \\ &= o(n^{-1/2}), \end{aligned} \tag{A25}$$

where the first inequality is due to Hölder’s inequality, and the second inequality is by the fact that

$$\begin{aligned} E \left[ (X_s^M - X_{t_{k-1}}^M)^2 \right] &= E \left[ \int_{t_{k-1}}^s \sigma_x^2 dx \right] \\ &= O(n^{-1}), \end{aligned} \tag{A26}$$

where the first and second equalities are due to the Itô isometry and the boundedness of  $\sigma_t$ , respectively. For  $T_{22}$ , since  $\int_{t_{k-1}}^{t_k} (X_s^M - X_{t_{k-1}}^M)\mu_{t_{k-1}} ds, k = 1, \dots, n$ , are uncorrelated, simple algebraic manipulations show

$$\begin{aligned} E [T_{22}^2] &= \sum_{k=1}^n E \left[ \left( \int_{t_{k-1}}^{t_k} (X_s^M - X_{t_{k-1}}^M)\mu_{t_{k-1}} ds \right)^2 \right] \\ &\quad + \sum_{k \neq k'}^n E \left[ \int_{t_{k-1}}^{t_k} (X_s^M - X_{t_{k-1}}^M)\mu_{t_{k-1}} ds \int_{t_{k'-1}}^{t_{k'}} (X_s^M - X_{t_{k'-1}}^M)\mu_{t_{k'-1}} ds \right] \\ &= \sum_{k=1}^n E \left[ \left( \int_{t_{k-1}}^{t_k} (X_s^M - X_{t_{k-1}}^M)\mu_{t_{k-1}} ds \right)^2 \right] \\ &\leq \sum_{k=1}^n E \left[ \int_{t_{k-1}}^{t_k} (X_s^M - X_{t_{k-1}}^M)^2 ds \int_{t_{k-1}}^{t_k} \mu_{t_{k-1}}^2 ds \right] \\ &\leq O(n^{-1}) \sum_{k=1}^n \int_{t_{k-1}}^{t_k} E \left[ (X_s^M - X_{t_{k-1}}^M)^2 \right] ds \\ &= O(n^{-2}), \end{aligned} \tag{A27}$$

where the first inequality is due to Hölder’s inequality and the fifth and sixth lines are due to the boundedness of  $\mu_t$  and (A26), respectively. By (A25) and (A27), we have

$$T_2 = o_p(n^{-1/2}). \tag{A28}$$

For  $T_3$ , the sequence,  $\int_{t_{k-1}}^{t_k} (X_s^D - X_{t_{k-1}}^D) dX_s^M, k = 1, \dots, n$ , is a martingale difference sequence, and so, we have

$$\begin{aligned} & E \left[ \left( \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (X_s^D - X_{t_{k-1}}^D) dX_s^M \right)^2 \right] \\ &= \sum_{k=1}^n E \left[ \left( \int_{t_{k-1}}^{t_k} (X_s^D - X_{t_{k-1}}^D) dX_s^M \right)^2 \right] \\ &= \sum_{k=1}^n E \left[ \int_{t_{k-1}}^{t_k} (X_s^D - X_{t_{k-1}}^D)^2 \sigma_s^2 ds \right] \\ &= O(n^{-2}), \end{aligned}$$

where the second equality is due to the Itô isometry and the last equality is from (A23) and the boundedness of  $\sigma_t$ . Thus,

$$T_3 = O_p(n^{-1}). \tag{A29}$$

For  $T_4$ , since  $\sigma_t$  is bounded and integrable,  $\int_0^1 \sigma_t^4 dt < \infty$ . Thus, an application of Theorem 5.1 in Jacod and Protter (1998) [46] leads to:

$$\sqrt{n} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (X_s^M - X_{t_{k-1}}^M) dX_s^M \xrightarrow{\mathcal{L}} \frac{1}{\sqrt{2}} \int_0^1 \sigma_t^2 dB_t^{discrete}. \tag{A30}$$

Collecting (A24), (A28)–(A30), we have (A22).  $\square$

**Corollary 1.** Suppose  $X$  follows (1). Then,  $[X, X]_1$  defined by Equation (A18) follows

$$[X, X]_1 \approx_{\mathcal{L}} \int_0^1 \sigma_t^2 dt + \sum_{0 < s \leq 1} L_s^2 + \sqrt{\frac{2}{n}} \int_0^1 \sigma_t^2 dB_t^{discrete}. \tag{A31}$$

**Appendix B. Tables and Figures**

Note that in the following tables, we take  $\alpha$  close to 1/2 and  $0 < \gamma \leq 2/3$ .

**Table B1.** Summary of the order of the wavelet coefficients of different components.

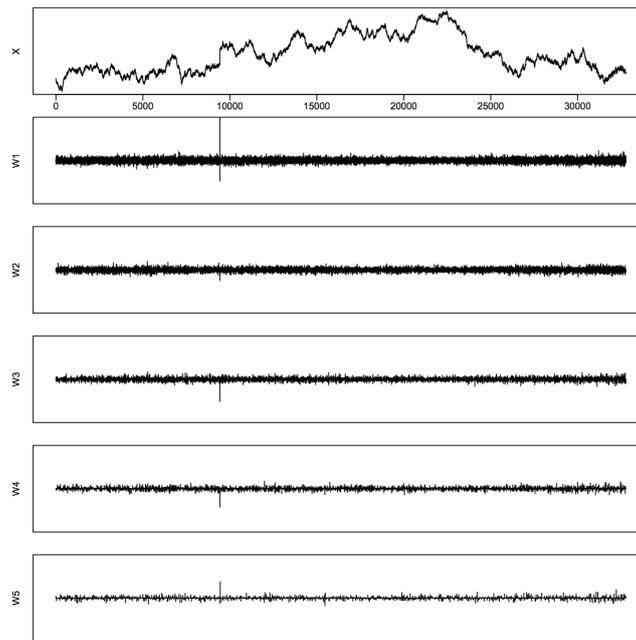
Component	$X$	$[X, X]$	$Y$	$[Y, Y]$	$[Y, Y]^{(avg)}$
cont. drift ( $\leq$ )	$2^{-j(3/2)}$	$2^{-j(3/2)}$	$2^{-j(3/2)}$	$2^{-j(3/2)}$	$2^{-j(3/2)}$
cont. diffusion ( $\leq$ )	$2^{-j(1/2+\alpha)}$	$n^{-1/2} 2^{-j(1/2+\alpha)}$	$2^{-j(1/2+\alpha)}$	$n^{1/2} 2^{-j(1/2+\alpha)}$	$n^{(1-2\gamma)/2} 2^{-j(1/2+\alpha)}$
jump ( $\geq$ )	$2^{-j/2}$	$2^{-j/2}$	$2^{-j/2}$	$2^{-j/2}$	$2^{-j/2}$
noise ( $O_p$ )	0	$\sim 0$	$n^{-1/2}$	$\sim 0$	$\sim 0$

**Table B2.** Summary of the convergence rate of jump location estimation.

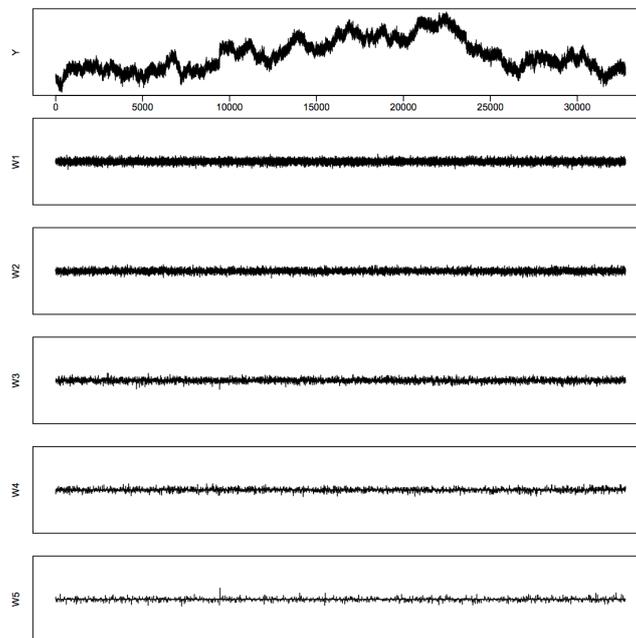
$X$	$[X, X]$	$Y$	$[Y, Y]$	$[Y, Y]^{(avg)}$
$O_p(n^{-1})$	$O_p(n^{-1})$	$O_p(n^{-1} \log^2 n)$	NA	$O_p(n^{-1})$

**Table B3.** Summary of the convergence rate of jump variation estimation.

$X$	$[X, X]$	$Y$	$[Y, Y]$	$[Y, Y]^{(avg)}$
$O_p(n^{-1/3})$	$O_p(n^{-1/2})$	$O_p(n^{-1/4})$	NA	$O_p(n^{-(2/3)\gamma})$



**Figure B1.** Wavelet coefficients of  $X$  at levels  $W1$ – $W5$ , with a jump at location 9424.



**Figure B2.** Wavelet coefficients of  $Y$  at levels  $W1$ – $W5$ , with a jump at location 9424.

## References

1. Barndorff-Nielsen, O.E.; Shephard, N. Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models. *J. R. Stat. Soc. Ser. B* **2002**, *64*, 253–280
2. Zhang, L.; Mykland, P.A.; Ait-Sahalia, Y. A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. *J. Am. Stat. Assoc.* **2005**, *100*, 1394–1411.
3. Zhang, L. Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach. *Bernoulli* **2006**, *12*, 1019–1043.
4. Bandi, F.M.; Russell, J.R. Microstructure Noise, Realized Variance, and Optimal Sampling. *Rev. Econ. Stud.* **2008**, *75*, 339–369.

5. Barndorff-Nielsen, O.E.; Hansen, P.R.; Lunde, A.; Shephard, N. Designing Realized Kernels to Measure Ex-post Variation of Equity Prices in the Presence of Noise. *Econometrica* **2008**, *76*, 1481–1536.
6. Jacod, J.; Li, Y.; Mykland, P.A.; Podolskij, M.; Vetter, M. Microstructure Noise in the Continuous Case: The Pre-averaging Approach. *Stoch. Process. Their Appl.* **2009**, *119*, 2249–2276.
7. Jacod, J.; Podolskij, M.; Vetter, M. Limit Theorems for Moving Averages of Discretized Processes Plus Noise. *Ann. Stat.* **2010**, *38*, 1478–1545.
8. Xiu, D. Quasi-maximum Likelihood Estimation of Volatility with High Frequency Data. *J. Econ.* **2010**, *159*, 235–250.
9. Aït-Sahalia, Y. Telling from Discrete Data Whether the Underlying Continuous-Time Model Is a Diffusion. *J. Finance* **2002**, *57*, 2075–2121.
10. Barndorff-Nielsen, O.E.; Shephard, N. Power and Bipower Variation with Stochastic Volatility and Jumps (with discussion). *J. Financ. Econ.* **2004**, *2*, 1–48.
11. Merton, R.C. Option Pricing When Underlying Stock Returns Are Discontinuous. *J. Financ. Econ.* **1976**, *3*, 125–144.
12. Aït-Sahalia, Y. Disentangling Volatility from Jumps. *J. Financ. Econ.* **2004**, *74*, 487–528.
13. Aït-Sahalia, Y.; Jacod, J. Testing for Jumps in a Discretely Observed Process. *Ann. Stat.* **2009**, *37*, 184–222.
14. Andersen, T.G.; Dobrev, D.; Schaumburg, E. Jump-robust Volatility Estimation Using Nearest Neighbor Truncation. *J. Econ.* **2012**, *169*, 75–93.
15. Barndorff-Nielsen, O.E.; Shephard, N. Econometrics of Testing for Jumps in Financial Economics using Bipower Variation. *J. Financ. Econ.* **2006**, *4*, 1–30.
16. Carr, P.; Wu, L. What Type of Process Underlies Options? A Simple Robust Test. *J. Finance* **2003**, *58*, 2581–2610.
17. Eraker, B.; Johannes, M.; Polson, N. The Impact of Jumps in Returns and Volatility. *J. Finance* **2003**, *58*, 1269–1300.
18. Eraker, B. Do Stock Prices and Volatility Jump? Reconciling Evidence From Spot and Option Prices. *J. Finance* **2004**, *59*, 1367–1404.
19. Fan, Y.; Fan, J. Testing and Detecting Jumps Based on a Discretely Observed Process. *J. Econom.* **2011**, *164*, 331–344.
20. Huang, X.; Tauchen, G.T. The Relative Contribution of Jumps to Total Price Variance. *J. Financ. Econ.* **2005**, *4*, 456–499.
21. Jiang, G.J.; Oomen, R.C. Testing for Jumps When Asset Prices Are Observed with Noise—A “Swap Variance” Approach. *J. Econ.* **2008**, *144*, 352–370.
22. Lee, S.S.; Hannig, J. Detecting Jumps From Lévy Jump Diffusion Processes. *J. Financ. Econ.* **2010**, *96*, 271–290.
23. Lee, S.S.; Mykland, P.A. Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. *Rev. Financ. Stud.* **2008**, *21*, 2535–2563.
24. Bollerslev, T.; Kretschmer, U.; Pigorsch, C.; Tauchen, G. A Discrete-time Model for daily S & P500 Returns and Realized Variations: Jumps and Leverage Effects. *J. Econ.* **2009**, *150*, 151–166.
25. Todorov, V. Estimation of Continuous-Time Stochastic Volatility Models with Jumps Using High-Frequency Data. *J. Econ.* **2009**, *148*, 131–148.
26. Aït-Sahalia, Y.; Jacod, J.; Li, J. Testing for Jumps in Noisy High Frequency Data. *J. Econ.* **2012**, *168*, 207–222.
27. Aït-Sahalia, Y.; Jacod, J. Estimating the Degree of Activity of Jumps in High Frequency Data. *Ann. Stat.* **2009**, *37*, 2202–2244.
28. Jing, B.Y.; Kong, X.B.; Liu, Z. Estimating the Jump Activity Index of Lévy Processes Under Noisy Observations Using High Frequency Data. *J. Am. Stat. Assoc.* **2011**, *106*, 558–568.
29. Jing, B.Y.; Kong, X.B.; Liu, Z.; Mykland, P.A. On the Jump Activity Index for Semimartingales. *J. Econ.* **2012**, *166*, 213–223.
30. Todorov, V.; Tauchen, G. Limit Theorems for Power Variations of Pure-Jump Processes with Application to Activity Estimation. *Ann. Appl. Probabil.* **2011**, *21*, 546–588.
31. Jing, B.Y.; Kong, X.B.; Liu, Z. Modeling High Frequency Data by Pure Jump Processes. *Ann. Stat.* **2012**, *40*, 759–784.
32. Kong, X.B.; Liu, Z.; Jing, B.Y. Testing For Pure-jump Processes For High-Frequency Data. *Ann. Stat.* **2015**, *43*, 847–877.
33. Todorov, V.; Tauchen, G. Volatility Jumps. *J. Bus. Econ. Stat.* **2011**, *29*, 356–371.

34. Andersen, T.G.; Bollerslev, T.; Diebold, F.X. Roughing it Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *Rev. Econ. Stat.* **2007**, *89*, 701–720.
35. Andersen, T.G.; Bollerslev, T.; Huang, X. A Reduced Form Framework for Modeling and Forecasting Jumps and Volatility in Speculative Prices. *J. Econ.* **2011**, *160*, 176–189.
36. Andersen, T.G.; Bollerslev, T.; Meddahi, N. Realized Volatility Forecasting and Market Microstructure Noise. *J. Econ.* **2011**, *160*, 220–234.
37. Wang, Y. Selected Review on Wavelets. In *Frontier Statistics, a Festschrift for Peter Bickel*; Koul, H., Fan, J., Eds.; Imperial College Press: London, UK, 2006; pp. 163–179.
38. Wang, Y. Jump and Sharp Cusp Detection by Wavelets. *Biometrika* **1995**, *82*, 385–397.
39. Fan, J.; Wang, Y. Multi-scale Jump and Volatility Analysis for High-Frequency Financial Data. *J. Am. Stat. Assoc.* **2007**, *102*, 1349–1362.
40. Daubechies, I. *Ten Lectures on Wavelets*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1992.
41. Raimondo, M. Minimax Estimation of Sharp Change Points. *Ann. Stat.* **1998**, *26*, 1379–1397.
42. Percival, D.B.; Walden, A.T. *Wavelet Methods for Time Series Analysis*; Cambridge University Press: Cambridge, UK, 2000.
43. Rényi, A. On Stable Sequences of Events. *Sankhyā Ser. A* **1963**, *25*, 293–302.
44. Aldous, D.J.; Eagleson, G.K. On Mixing and Stability of Limit Theorems. *Ann. Probab.* **1978**, *6*, 325–331.
45. Hall, P.; Heyde, C.C. *Martingale Limit Theory and Its Application*; Academic Press: Boston, MA, USA, 1980.
46. Jacod, J.; Protter, P. Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations. *Ann. Probab.* **1998**, *26*, 267–307.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).