*Quantitative Finance*, 2014
Vol. 14, No. 8, 1427–1444, http://dx.doi.org/10.1080/14697688.2013.830320

Routledge
Taylor & Francis Group

# Jump detection with wavelets for high-frequency financial time series

YI XUE†, RAMAZAN GENÇAY‡§* and STEPHEN FAGAN‡¶

†Department of Finance, School of International Trade and Economics, University of International Business and Economics, #10 Huixin Dongjie, Chaoyang District, Beijing, China
‡Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada
§Rimini Center for Economic Analysis, Rimini, Italy

(*Received 6 December 2011; accepted 25 July 2013*)

This paper introduces a new nonparametric test to identify jump arrival times in high frequency financial time series data. The asymptotic distribution of the test is derived. We demonstrate that the test is robust for different specifications of price processes and the presence of the microstructure noise. A Monte Carlo simulation is conducted which shows that the test has good size and power. Further, we examine the multi-scale jump dynamics in US equity markets. The main findings are as follows. First, the jump dynamics of equities are sensitive to data sampling frequency with significant underestimation of jump intensities at lower frequencies. Second, although arrival densities of positive jumps and negative jumps are symmetric across different time scales, the magnitude of jumps is distributed asymmetrically at high frequencies. Third, only 20% of jumps occur in the trading session from 9:30 AM to 4:00 PM, suggesting that illiquidity during after-hours trading is a strong determinant of jumps.

*Keywords*: Jump detection; Wavelets; Directional jumps; Positive jumps; Negative jumps

*JEL Classification*: G10, G11, D43, D82



## 1. Introduction

In the last two decades, a substantial amount of empirical evidence has been found to support the existence of surprise elements, commonly called jumps, in financial time series. It is well understood that compared to continuous price changes, jumps have distinctly different modeling, inference, and testing requirements. These differences impact how we perform many important financial tasks such as valuing derivative securities, inferring extreme risks, and determining optimal portfolio allocations. Thus, understanding what drives jumps in underlying securities, how to characterize jumps both theoretically and empirically, and having efficient tests available for jumps that are sufficiently robust to withstand mis-specification and small sample bias is imperative.

This paper proposes a new method to estimate a jump's location and size, as well as the number of jumps in a given time interval from high frequency data. In addition to providing a practical jump detection criterion, we derive the asymptotic distribution of the test statistic and demonstrate that the test has good size and power properties. Our methodology is based on a wavelet decomposition of the data and we show that the test of Lee and Mykland (2008) is a special case of a wavelet-type test. Additionally, we show that when a wavelet filter with less leakage is used, the performance of the test improves.‖ This improvement originates from the fact that the wavelet coefficients with less leakage are better able to isolate the high frequency components of the price process where jumps, when they occur, will dominate other components such as continuous changes and microstructure effects. Thus, we can avoid identifying spurious jumps while preserving power in detecting true jumps.††

A substantial amount of research has been dedicated to detecting jumps in asset prices. Among previous studies, Andersen *et al.* (2003) proposed a method using a jump-robust estimator of realized volatility. Barndorff-Nielsen and Shephard (2004, 2006) proposed a bi-power variation (BPV) measure to separate the jump variance and diffusive variance. Lee and Mykland (2008) developed a rolling-based nonparametric test for jumps and a method for estimating jump sizes

*Corresponding author. Email: rgencay@sfu.ca
¶Current address: BlackRock, Inc., New York, NY, USA.

‖A filter with less leakage is one that is closer to being an ideal band-pass filter.

††Spurious jumps are observations where no jump actually occurred, but the test incorrectly identified one to have occurred due to large movements originating from the presence of microstructure effects or large spot volatility.

and jump arrival times. Jiang and Oomen (2008) proposed a jump test based on the idea of a 'variance swap' and explicitly accounted for market microstructure effects. Johannes (2004) and Dungey *et al.* (2009) found significant evidence for jumps in US treasury bond prices and rates. Piazzesi (2003) demonstrated that jump modeling leads to improved bond pricing in the US treasury market. Andersen *et al.* (2007) showed that incorporation of jump components could improve the forecasting of return volatility. Fan and Wang (2007) showed that when market returns contain jumps, separating the variation in returns into jump and diffusion components is important for efficient estimation of realized volatility.

Wavelets can be a powerful tool for detecting jumps for two reasons. First, the ability of wavelets to decompose noisy time series data into separate time-scale components helps us to distinguish jumps from continuous price changes and microstructure effects. The high-frequency wavelet coefficients at jump locations are larger than other wavelet coefficients due to the fact that wavelet coefficients decay at a different rate for continuous processes and jump processes. In a given small time interval, changes in continuous price processes are close to zero while jumps are not.† Such information is contained in wavelet coefficients at the jump locations (see Wang (1995) and Fan and Wang (2007)).‡ Second, estimation of jump size is closely related to the estimation of integrated volatility. As Fan and Wang (2007) demonstrated, wavelets allow superior estimation of the integrated volatility, which can improve the efficiency of the estimation of jump size.

This paper implements our wavelet jump test to examine the jump dynamics of three US equities and finds that their estimated jump dynamics are sensitive to the data sampling frequency. This emphasizes the importance of an optimal sampling frequency to extract full jump dynamics. Our results suggest that the popular choice of 5-min sampling frequency may neglect a large proportion of jump dynamics embedded in transaction prices. Additionally, although jump arrival densities of positive jumps and negative jumps are symmetric across time scales, the magnitude of jumps is asymmetrically distributed at high frequencies. This suggests that a skewed distribution for the magnitude of jumps should be employed in risk management and asset pricing practices for high frequency trading. Finally, only 20% of jumps occur in the trading session from 9:30 AM to 4:00 PM, suggesting that illiquidity during after-hours trading is a strong determinant of jumps.

Overall, the main contributions of the paper are as follows. First, a nonparametric jump detection test based on wavelet transformations is proposed and the asymptotic distribution of the test statistic is established. This test is shown to improve jump detection by avoiding the identification of spurious jumps and providing improved estimation of jump locations. Second, we show that the jump detection test is robust in the presence of microstructure noise. Finally, the empirical implementation of the wavelet test in US equity markets demonstrates a dramatic change of jump dynamics across time scales, the asymmetric distribution of jump magnitudes at high frequencies and the occurrence of the majority of jumps outside of the primary trading session.

In Section 2, we provide the theoretical framework for jump detection in the absence of microstructure effects and introduce the wavelet-based jump test statistic. The asymptotic distribution of the test statistic follows a scaled normal distribution and the scalar is determined by the properties of the wavelet filter. Section 3 extends the framework to investigate the performance of the test in the presence of microstructure effects. We show that the asymptotic distribution of the test statistic under the null hypothesis remains the same. Section 4 presents Monte Carlo simulation results which demonstrate the finite sample behavior of the proposed test. We show that the test has desirable size and power in small samples. Section 5 discusses the empirical examination of jump dynamics in equity markets.

## 2. The jump test without microstructure noise

Wavelets have been shown to be useful for jump detection when prices are following a jump-diffusion process. Fan and Wang (2007) used the localization property of wavelet expansion to detect jumps. This property entails that if a function contains a jump but is otherwise Holder-continuous,§ then the wavelet coefficients of the high-pass filter close to the jump point decay at order $2^{\frac{1}{2}j}$, where $j$ is a scale of the wavelet decomposition. If there is no jump, the wavelet coefficients of the high pass filter decay at order $2^{\frac{3}{2}j}$. This special feature was used to separate jumps from the continuous and microstructure noise components of the process (Wang 1995).

Although Fan and Wang (2007) showed the effectiveness of the wavelets method for jump detection with a diffusion process convincingly, the distributional properties of the jump detection test statistic were not formally defined. Additionally, because a discrete wavelet transformation (DWT) was applied to the data, the jump location could only be estimated with an informational loss in time domain.

This paper proposes a framework for jump detection using wavelets and formally develops the test statistic. Before we mathematically define this jump detection statistic $J_W$, we describe the basic intuition behind the proposed detection technique. Imagine that asset prices evolve continuously over time, but due to an announcement or some other informational shock, a jump in prices occurs at time $t$. Given the additive nature of the jump, we expect to see the mean level of the price process abruptly shift. The large return at this point could be due to any combination of the drift, diffusion, or jump components of the price process, but as our sampling frequency gets higher, the contribution of the drift and diffusion components gets smaller. The contribution of an instantaneous jump however will not get smaller. Thus, we should expect to see a relatively large absolute value for the return $r_t$ at the jump location.

†Continuous price processes include cusps which are nondifferentiable kinks connecting two differentiable intervals. Differences in a price process at a heavy cusp converge to zero much more slowly than at a differentiable point. In the limit cusps are jumps.

‡Wang (1995) and Fan and Wang (2007) proposed a wavelets-based procedure using the different convergence rates of wavelet coefficients with or without jumps. In addition to jump size, Fan and Wang (2007) could estimate the number of jumps and an estimated interval of jump location.

§A function $g : R^d \to R$ is said to be Holder-continuous if there exist constants $C$ and $0 \leq E \leq 1$ such that for all $u$ and $v$ in $R^d$ : $|g(u) - g(v)| \leq C||u - v||^E$.

This intuitive description of a jump test transfers easily to one using wavelet coefficients rather than returns. Like returns, wavelet coefficients contain information on differenced prices. Unlike returns however, wavelet coefficients break the price series down into different components which correspond to different frequencies. A technical introduction to wavelet transformations is presented in Appendix A1. Since jumps are unexpected and instantaneous, their contribution will be captured in the wavelet coefficients corresponding to the highest frequency. Using wavelet coefficients from this scale should therefore provide a cleaner jump test statistic since low-frequency components have been filtered out. Thus, our jump test is based on the expectation that we should see a relatively large wavelet coefficient at jump locations.

If prices were continuously observed, jumps would be easy to identify. However, with discrete price observations, we can observe a large price change even without a jump if spot volatility is large enough. Thus, it is difficult to distinguish whether the observed large movement in prices is due to a jump in price process or a volatility of large magnitude. High frequency price data reduces the severity of this problem, but since spot volatility may vary over time, we normalize the absolute value of wavelet coefficients by the estimated spot volatility. To get a consistent estimate of the spot volatility in the presence of jumps, we apply the BPV estimator suggested by Barndorff-Nielsen and Shephard (2004).

Furthermore, we use the maximum overlap discrete wavelet transformation (MODWT) instead of DWT. MODWT generates an equal number of wavelet coefficients (high pass filter) as the original data series. Combined with zero phase correction, the locations of the wavelet coefficients naturally reveal information concerning the original data in the time domain. Thus, the jump location detection is reduced to a jump detection problem.

### *2.1. Definition of the test*

We employ a one-dimensional asset return process. Let the logarithm of the market price of the asset be $P_t = \log S_t$ where $S_t$ is the asset price at time $t$. When there are no jumps in the market price, $P_t$ is represented as

$$P_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s \tag{1}$$

where the two terms correspond to the drift and diffusion components of $P_t$. In the diffusion term, $W_t$ is a standard Brownian motion, and the square root of the diffusion variance $\sigma_t^2$ is called spot volatility. Equivalently, $P_t$ can be characterized as

$$dP_t = \mu_t dt + \sigma_t dW_t. \tag{2}$$

When there are jumps, $P_t$ is given by

$$P_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \sum_{l=1}^{N_t} L_l \tag{3}$$

where $N_t$ represents the number of jumps in $P_t$ up to time $t$ and $L_l$ denotes the jump size. Equivalently, $P_t$ can be modeled as

$$dP_t = \mu_t dt + \sigma_t dW_t + L_t dN_t \tag{4}$$

where $N_t$ is a counting process that is left unmodeled. We assume jump sizes $L_l$ are independently and identically distributed. They are also independent of other random components $W_t$ and $N_t$.

Observations of $P_t$, the log price, are only available at discrete times $0 = t_0 < t_1 < t_2 < \ldots < t_n = T$. For simplicity, we assume observations are equally spaced: $\Delta t = t_i - t_{i-1}$. Following Lee and Mykland (2008), we impose the following assumptions on price processes throughout the paper: For any small $\epsilon > 0$,

$$\text{A1} \quad \sup_i \quad \sup_{t_i \le u \le t_{i+1}} |\mu_u - \mu_{t_i}| = O_p\left(\Delta t^{\frac{1}{2}-\epsilon}\right)$$

$$\text{A2} \quad \sup_i \quad \sup_{t_i \le u \le t_{i+1}} |\sigma_u - \sigma_{t_i}| = O_p\left(\Delta t^{\frac{1}{2}-\epsilon}\right)$$

The assumption $A1$ and $A2$ can be interpreted as the drift and diffusion coefficients not changing dramatically over a short time interval.† Formally, this states that the maximum change in mean and spot volatility in a given time interval is bounded above. The assumptions $A1$ and $A2$ guarantee that the available discrete data are reasonably well-behaved such that the data are a good approximation of the continuous process of the underlying asset. The availability of high frequency financial data allows us to improve the approximation of the continuous underlying asset process using discrete data.

Now, let $P_{1,t}$ be the wavelet coefficients of $P_t = \log(S_t)$ for a first scale MODWT. We define the test statistic as $J_W$, which tests whether a jump occurs at time $t_i$ for $i = 3, \ldots, T$ as

$$J_W(i) = \frac{P_{1,t_i}}{\widehat{\sigma}_{t_{i-1}}} \tag{5}$$

where $\widehat{\sigma}^2_{t_{i-1}} = \frac{1}{i-2} \sum_{k=2}^{i-1} |P_{1,t_k}||P_{1,t_{k-1}}|$.

Following Lee and Mykland (2008), we employ a BPV method to estimate the integrated volatility of the underlying process. However, there are two major differences between our approach and theirs. First, we apply BPV on wavelet transformed data (wavelet coefficients calculated from the raw data) rather on the raw data. Wavelets have the ability to decompose the raw data into high frequency components and low frequency components. Since jumps are unexpected and instantaneous, we believe that jump dynamics should be contained in the high frequency components of prices. By using wavelet coefficients rather than returns, we rely on the ability of wavelet filter to extract the return dynamics to estimate the high-frequency component of instantaneous volatility. Second, we use all of the data points available up the jump location rather than a choice of window size $K$ as in Lee and Mykland (2008). The choice of $K$ is to minimize the impact of the jumps on the estimation of the instantaneous volatility within the window. As argued in Fan and Wang (2007), wavelets have the ability to estimate the instantaneous volatility with the presence of the jumps. We conjecture that using more sophisticated wavelet filters, e.g. $S8$ filter rather $Haar$ filter, can help us to detect jumps in the stochastic volatility framework. That is, we choose a wavelet filter rather window size to minimize

†Following Pollard (2002) and Lee and Mykland (2008), we use $O_p$ notation throughout this article to mean that, for random vectors $\{X_n\}$, and non-negative random variable $\{d_n\}$, $X_n = O_p(d_n)$, if for each $\delta > 0$, there exists a finite constant $M_\delta$ such that $P(|X_n| > M_\delta d_n) < \delta$.

the impact of the jumps on the estimation of the instantaneous volatility.

There are alternative methods for estimating the integrated volatility including the two scale realized volatility estimators (TSRV) (Zhang *et al.* 2005) and the multi-scale realized volatility estimators (MSRV) (Zhang 2006). However, Barndorff-Nielsen and Shephard (2004) demonstrated that the presence of jumps will change the asymptotic behavior of these tests. Additionally, they showed that a BPV estimator is robust in the presence of the jumps.

### 2.2. A general filter case

In this section, we demonstrate the test statistics with a general wavelet filter, i.e. we $L$ wavelet filter coefficients, $h_0, \ldots, h_{L-1}$, such that $\sum_{l=0}^{L-1} h_l = 0$. Under the null hypothesis that no jump occurs at time $t_i$, as $\Delta t$ goes to zero, $J_W(i)$ should converge to a normal distribution under the assumptions of $A1$ and $A2$. Theorem 1 demonstrates the distribution of the proposed statistic in the case of MODWT with a general filter. As argued in Lee and Mykland (2008), when working with high-frequency data, the drift (of order $dt$) is mathematically negligible compared to the diffusion (of order $\sqrt{dt}$) and the jump component (of order 1). In fact, the drift estimates have higher standard errors, so that they cause the precision of variance estimates to decrease if included in the variance estimation. Therefore, we assume zero drift in this study.†

THEOREM 1 *Assuming zero drift and constant spot volatility in the underlying price process and assuming A1 and A2 are satisfied, if there is no jump in* $(t_{i-1}, t_i)$, *as* $\Delta t \to 0$

$$sup_i \ |J_W(i) - \widehat{J}_W(i)| = O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right),$$
$$\text{where } \delta \text{ satisfies } 0 < \delta < 3/2 \text{ and } \widehat{J}_W(i) = \frac{U_i}{d}. \quad (6)$$

*where* $W_t$ *follows a Brownian motion process. And* $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, *a standard normal variable, and a constant* $d = \frac{\sqrt{E[BPV_G]}}{\sqrt{\sum_{l=0}^{L-2}(\sum_{j=0}^{l} h_j)^2}}$, *where*

$$E[BPV_G] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} E[BPV_G|x_1, \ldots, x_{L-2}]\phi(x_1) \ldots \phi(x_{L-2})dx_1 \ldots dx_{L-2},$$

$$E[BPV_G|\widetilde{r}_1, \ldots, \widetilde{r}_{L-2}] = 2\sigma^2 f\left(\frac{-\sum_{l=1}^{L-2}\left(\sum_{j=0}^{l} h_j\right)(\widetilde{r}_{L-l-l})}{h_0}\right) \times f\left(\frac{-\sum_{l=0}^{L-3}\left(\sum_{j=0}^{l} h_j\right)(\widetilde{r}_{L-l-l})}{\left(\sum_{j=0}^{L-2} h_j\right)}\right),$$

*where* $f(a) = 2\phi(a) + 2a\Phi(a) - a$, *and* $\widetilde{r}_t$*s are independently and identically distributed (iid) standard normal variables.*

Comments:

(1) The distribution of the test statistic is asymptotically normally distributed.
(2) Since $J_W(i)$ is asymptotically independent and normally distributed over time, one can easily use a joint normal distribution of the test statistics to detect multiple jumps.
(3) In addition, the scaling factor $d$ only depends on the wavelet filter coefficients. That is, when a particular wavelet filter is used, $d$ can be computed based on the values of $h_l$s.

Next, we demonstrate the implementation of the proposed test using two popular wavelets filters, i.e. *Haar* filter (proposed by Haar (1910)) and $D4$ filter (proposed by Daubechies (1988)).‡ If we implement the jump detection test using a *Haar* filter, i.e. $h_1 = 1/2$, $h_2 = -1/2$. In this case, we have the following proposition which is a direct application of Theorem 1.

PROPOSITION 2 (Haar filter case) *Assuming zero drift and A1 and A2 are satisfied. If there is no jump in* $(t_{i-1}, t_i)$, *as* $\Delta t \to 0$,

$$sup_i \ |J_W(i) - \widehat{J}_W(i)| = O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right),$$
$$\text{where } \delta \text{ satisfies } 0 < \delta < 3/2 \text{ and } \widehat{J}_W(i) = \frac{U_i}{d}. \quad (7)$$

*where* $W_t$ *follows a Brownian motion process. Here* $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, *a standard normal variable, and* $d = E[|U_i|] = \frac{\sqrt{2}}{\sqrt{\pi}}$ *is a constant.*

It is worth noting that the test statistics $J_W(i)$ in Proposition 2, i.e. the test statistics in the case of *Haar* filter, is equivalent to test statistics proposed in Lee and Mykland (2008).

Similarly, in the case of $D(4)$ filter , i.e. $h_1 = \frac{1}{8}(1 - \sqrt{3})$, $h_2 = \frac{1}{8}(\sqrt{3} - 3)$, $h_3 = \frac{1}{8}(3 + \sqrt{3})$, $h_4 = \frac{1}{8}(-1 - \sqrt{3})$, we can show that

PROPOSITION 3 ($D4$ filter) *Assuming zero drift and constant spot volatility in the underlying price process and assuming A1 and A2 are satisfied, if there is no jump in* $(t_{i-1}, t_i)$, *as* $\Delta t \to 0$

$$sup_i \ |J_W(i) - \widehat{J}_W(i)| = O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right),$$
$$\text{where } \delta \text{ satisfies } 0 < \delta < 3/2 \text{ and } \widehat{J}_W(i) = \frac{U_i}{d}. \quad (8)$$

*where* $W_t$ *follows a Brownian motion process. Here* $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, *a standard normal variable, and a constant* $d = \frac{\sqrt{2}}{\sqrt{3}}$.

## 3. The jump test with microstructure noise

At high sampling frequencies, market microstructure effects introduce noise into financial data. We accommodate such effects by assuming that the observed high-frequency prices

†We do study the statistics under non-zero drift case by simulations and demonstrate that the performance of the test is not changed under non-zero drift case.

‡See Gençay *et al.* (2001) for a good discussion on various wavelets filters.

$P_t^*$ are equal to the latent, true price process $P_t$ plus market microstructure noise $\epsilon_t$, thus:

$$P_t^* = P_t + \epsilon_t. \tag{9}$$

We consider a simple specification of microstructure noise where this noise is independently and identically distributed. More specifically, let $\epsilon_t$ be an $iid$ normal random variable with zero mean and variance $\eta^2$ that is independent of $P_t$. When there are no jumps in the market price, $P_t^*$ is represented as

$$P_t^* = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_{1,s} + \epsilon_t \tag{10}$$

where the three terms correspond to the drift and diffusion components of $P_t$ and the $iid$ market microstructure noise. When there are jumps, $P_t^*$ is given by

$$P_t^* = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_{1,s} + \epsilon_t + \sum_{l=1}^{N_t} L_l \tag{11}$$

where $N_t$ represents the number of jumps in $P_t^*$ up to time $t$ and $L_l$ denotes the jump size. Note that $N_t$ is a counting process that is left un-modeled. We assume jump sizes $L_l$ are independently and identically distributed and also independent of other random components $W_t$ and $N_t$.

Generally, for $iid$ microstructure noise, the denominator of $J_W(i)$ will be biased estimator of integrated volatility. This was demonstrated by Jiang and Oomen (2008), and they propose a method for correcting this bias. Let the spot volatility of the price process be constant, $\sigma_t = \sigma$, and let $BPV^*$ and $BPV$ be the BPV estimator based on noisy data and noise-free data, $P_t^*$ and $P_t$. With $iid$ microstructure noise and constant spot volatility,

$$E[BPV^*] = \frac{2}{\pi}(1 + c(\gamma))E[BPV]$$

where

$$c(\gamma) = (1+\gamma)\sqrt{\frac{1+\gamma}{1+3\gamma}} + \gamma\frac{\pi}{2} - 1 + 2\frac{\gamma}{(1+\lambda)\sqrt{2\lambda+1}} + 2\gamma\pi k(\lambda)$$

with $\gamma = T\eta^2/\sigma$, $\lambda = \frac{\gamma}{1+\gamma}$, $k(\lambda) = \int_{-\infty}^{\infty} x^2\Phi(x\sqrt{\lambda})(\Phi(x\sqrt{\lambda}) - 1)\phi(x)dx$, and $\Phi(.)$ and $\phi(.)$ are the CDF and PDF of the standard normal, respectively. The numerator of the test statistic $\widehat{J}_W(i)$ using MODWT with a $Haar$ filter is the wavelet coefficient at scale level 1, which is given by

$$P_{1,t}^* = \frac{1}{2}(P_t^* - P_{t-1}^*) = \frac{1}{2}(\log S_t - \log S_{t-1} + \epsilon_t - \epsilon_{t-1}).$$

Thus, it is a linear combination of two independent random variables which are normally distributed, with the mean 0 and variance $\sigma^2\Delta t + 4\eta^2$. Thus, the distribution of $J_W(i)$ is asymptotically normal with mean 0 and variance $\frac{\sigma^2\Delta t + 4\eta^2}{c^2\sigma^2}$. Formally,

THEOREM 4 *With $iid$ microstructure noise and constant spot volatility, i.e. $\sigma_t = \sigma$, assuming zero drift in the underlying process and that assumption A2 is satisfied, if there is no jump in $(t_{i-1}, t_i)$, as $\Delta t \to 0$,*

$$\sup_i |J_W(i) - \widehat{J}_W(i)| = O_p(\Delta t^{\frac{3}{2}-\epsilon}) \quad \text{where} \quad \widehat{J}_W(i) = dU_i.$$

*Here $U_i = \frac{1}{\sqrt{\Delta t}}\left(W_{t_i} - W_{t_{i-1}}\right)$ is a standard normal variable and $d = \frac{2\eta\sqrt{\pi}}{(1+c(\gamma))\sigma\sqrt{2}}$ is a constant, where*

$$c(\gamma) = (1+\gamma)\sqrt{\frac{1+\gamma}{1+3\gamma}} + \gamma\frac{\pi}{2} - 1 + 2\frac{\gamma}{(1+\lambda)\sqrt{2\lambda+1}} + 2\gamma\pi k(\lambda)$$

*with $\gamma = T\eta^2/\sigma$, $\lambda = \frac{\gamma}{1+\gamma}$, $k(\lambda) = \int_{-\infty}^{\infty} x^2\Phi(x\sqrt{\lambda})(\Phi(x\sqrt{\lambda}) - 1)\phi(x)dx$, and $\Phi(.)$ and $\phi(.)$ are the CDF and PDF of the standard normal, respectively.*

Given the above considerations, the proof proceeds along similar lines as the earlier theorems. Additionally, since the noise ratio $\gamma$ is not observed, we need to estimate it from the data. Hansen and Lunde (2006) proposed the estimator

$$\widehat{\gamma} = \frac{\overline{\eta}^2}{\overline{IV}}$$

where $\overline{\eta}^2$ is the daily average estimate of variance of microstructure noise, and $\overline{IV}$ is the daily average estimate of integrate volatility. They suggest as an estimate for $\eta$, $\frac{RV^{(m)} - RV^{(30)}}{2(m-30)}$, where $m$ is a sampling frequency and $RV^{(30)}$ is based on intraday returns that span about 30 min each.† Therefore, when microstructure noise is $iid$, the asymptotic distribution of the test statistic is still normal but with a correction of the standard deviation.

Jiang and Oomen (2008) noted that when sampling frequency is not sufficiently high and microstructure noise is relatively small, i.e. $\widehat{\gamma} \approx 0$, the correction for the standard deviation of the test statistic is approximately zero. Therefore, the distribution of test statistic is close to that in no-microstructure noise case, i.e.

$$\sup_i |J_W(i) - \widehat{J}_W(i)| = O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right) \text{ where } \widehat{J}_W(i) = \frac{U_i}{d}.$$

It is challenging to analytically derive the distribution of the test statistics with filters other than the $Haar$ filter, and so we leave it for future research. Given that a zero-phase filter may separate continuous and noise components more effectively as argued in Fan and Wang (2007), we conjecture that the zero-phase filter may perform better in the presence of the microstructure noise.

Issues related to microstructure noise are important considerations in many areas of the finance literature, including the estimation of integrated volatility. Microstructure noise might be due to the imperfections in trading processes, including price discreteness, infrequent trading, and bid-ask bounce effects. It is well-known that higher price sampling frequencies are linked to a larger the impact of microstructure noise. Zhang *et al.* (2005) demonstrated that estimation of integrated volatility via a realized volatility method using very high-frequency data is severely contaminated by microstructure noise. Fan and Wang (2007) assumes a very small noise ratio in detecting jumps. The distribution of the test proposed in Ait-Sahalia and Jacod (2009) is different in the presence of microstructure noise.

†For example, for foreign exchange market, there are 24 trading hours per day. If sampling frequency is 1 min, $m = 24 * 60 = 1440$.

Lee and Mykland (2008) chose a rather low sampling frequency (15 min) in an empirical study in order to avoid the impact of microstructure noise. Because the asymptotic distribution of this statistic is robust in the presence of microstructure noise, we do not need to decrease the sampling frequency.

In practice, the microstructure noise can have a more general form than the $iid$ case. It is also very challenging to derive the distribution of the statistics when the microstructure noise is of other general form. But the basic intuition is that given the fact that in the high frequency data, the statistic in the unit root case and $iid$ case are approximately identical, it might be the case that a more general form of microstructure noise will not affect the performance of the test significantly.

To see this, let us assume the microstructure noise is unit root process, i.e.

$$dP_t^* = \mu_t dt + \sigma_t dW_{1,t} + \eta dW_{2,t} + L_t dN_t, \tag{12}$$

where the first three terms correspond to the drift and diffusion parts of $P_t$ and the market microstructure noise, and $N_t$ is a counting process that is left un-modeled. We assume jump sizes $L_l$ are independently and identically distributed and also independent of other random components $W_t$ and $N_t$.

It is easy to see that with such persistent microstructure process, the above price diffusion can be rewritten as $dP_t^* = dP_t^* = \mu_t dt + \sigma_t^* dW_t$, where $\sigma_t^*$ is a function of spot volatility $\sigma_t$ and the volatility of microstructure noise $\eta$. Thus, it is equivalent to observe a larger spot of volatility. If we want to estimate the integrated volatility i.e. disentangling the spot volatility of the underlying process ($\sigma_t$) from the microstructure noise ($\eta$) is a challenging task, since $\sigma_t^*$ is the only information available (See Zhang *et al.* (2005) and Zhang (2006)). However, in the case of the jump detection, it is equivalent to have a larger spot volatility. Hence, we only need to estimate the 'new spot volatility' of the underlying process, $\sigma_t^*$. Note that the new underlying process is well-defined so that the BPV estimator is a consistent estimator of $\sigma_t^*$ (or $\sigma_t + \eta$). Therefore, the asymptotic distribution of the test is not changed in the presence of unit root microstructure noise.

Notice that our proposed jump detection method utilizes the ability of wavelets to extract the high frequency component, i.e. the jump component, from the data. The outcome of the 'extracting the high frequency component' from the data is the wavelets coefficients after a particular wavelets filter is applied to the data and the MODWT is done. That is the numerator of the test statistics. To disentangle the true 'jump' from the large movement due to the volatility effect discussed earlier, we need to use the estimated volatility which is robust to the presence of the jumps to standardize the numerator. In our paper, BPV method is applied not to the data, but to the wavelet coefficients. Our method to estimate the volatility is consistent with the 'pre-averaging' method proposed and developed by Podolskij and Vetter (2009) and Jacod *et al.* (2009). It is shown in Podolskij and Vetter (2009) that the volatility estimates in such method is robust to the presence of both microstructure noise and jumps. We use such estimated volatility in the denominator to standardize the high frequency component in order to find the 'true jump'. Thus, the efficiency of 'finding a true jump' relies on how efficient the wavelets filters can do to extract the high frequency component. As argued in Gençay *et al.* (2001), the less leakage wavelet filters can extract the high frequency component more efficiently from the data than other filters. We expect to see a better performance in terms of the jump detection of wavelet filter of less leakage, e.g. $S8$ filter. In the following section, we will conduct the Monte Carlo simulation to illustrate the size and power performance of the tests using different wavelet filters.

## 4. Monte Carlo simulations

In this section, we examine the effectiveness of the wavelet-based jump test using Monte Carlo simulations. The performance of the test statistic is examined at different sampling frequencies. An Euler method is used to generate continuous diffusion processes and the burn-in period observations is discarded to avoid the effects of the initial value.

### *4.1. Under the null of no jumps*

This subsection illustrates the simulated test statistic under the null hypothesis of no jumps within a given period of time. Formally, we consider

$$dP_t = \mu_t dt + \sigma_t dW_t. \tag{13}$$

where $\mu_t$ is the drift in price process and $\sigma_t$ is the spot volatility of the price process. We consider four scenarios: $\mu_t = 0$ and $\sigma_t = \sigma$ (zero drift and constant volatility as the benchmark), $\mu_t \neq 0$ and $\sigma_t = \sigma$ (non-zero drift and constant volatility), $\mu_t = 0$ and $\sigma_t \neq \sigma$ (zero drift and stochastic volatility), and $\mu_t \neq 0$ and $\sigma_t \neq \sigma$ (non-zero drift and stochastic volatility). Specifically, we assume $\mu_t = 1$ or $\mu_t = 0$ for non-zero and zero drift cases, respectively.† We employ an Ornstein-Uhlenbeck process as the volatility model for the stochastic volatility case:

$$\begin{aligned} dP_t &= \mu_t dt + \sigma_t dW_{1,t} \\ d\log\sigma_t^2 &= k(\log\overline{\sigma}^2 - \log\sigma_t^2) + \delta dW_{2,t} \end{aligned} \tag{14}$$

where $k$ measures the recovery rate of volatility to the mean, $\log\overline{\sigma}^2$ is the long-run mean of variance, and $\delta$ is the diffusion parameter for the volatility process.‡ Following Fan and Wang (2007), we assume that the correlation between $W_{1,t}$ and $W_{2,t}$ is $\rho$, which is negative and captures the asymmetric impact of the innovation in price process. We set $k = -0.1$, $\log\overline{\sigma}^2 = -6.802$, and $\delta = 0.25$. Figure 1 shows the density plot of the statistic for 1 million observations when a $Haar$ filter is used. The top left panel shows the zero mean and constant volatility case; the top right panel depicts the non-zero mean and constant volatility case; the bottom left panel shows the zero mean and stochastic volatility case; and the bottom right panel depicts the non-zero mean and stochastic volatility case.

Figure 2 shows the density plot of the statistic for one million observations when an $S8$ filter is used, and figure 3 shows the

†We also investigated various specifications with drift and the drift part has negligible impact on the main results.

‡We also investigated other specifications of the process and found that different specifications of volatility processes did not qualitatively changed the main results.

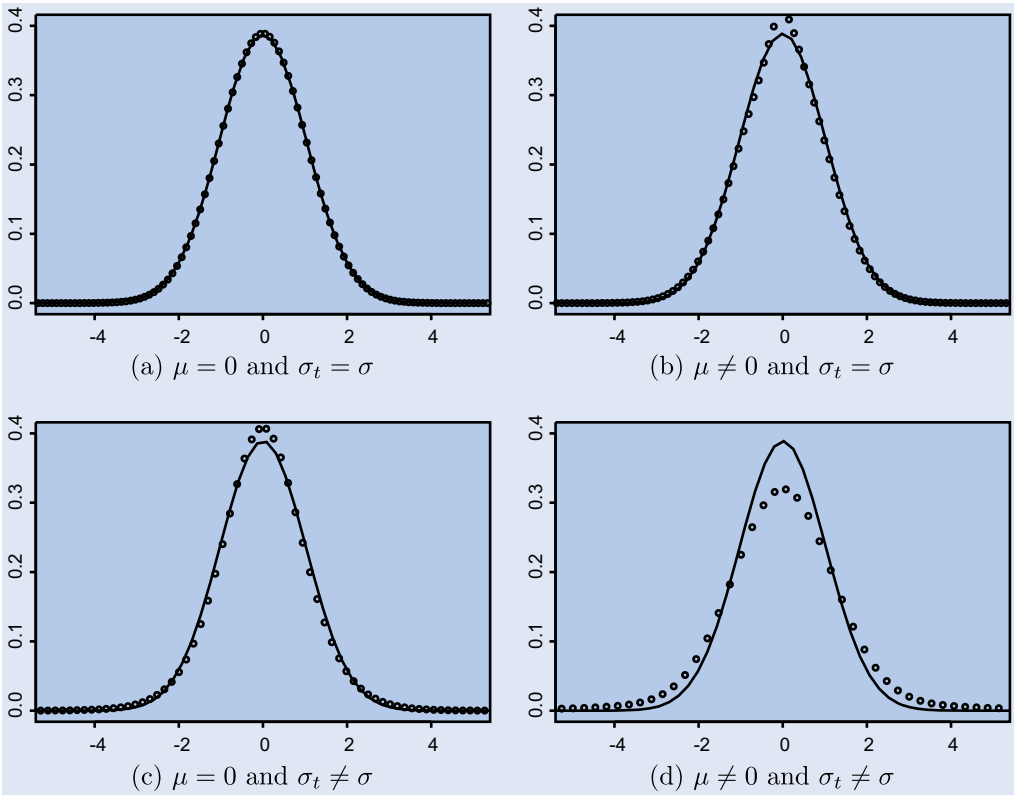


Figure 1. Density plot of the simulated statistic under the null hypothesis with the *Haar* filter. (a) Density plot of the simulated statistic with zero mean and constant volatility. (b) Density plot of the simulated statistic with non-zero mean and constant volatility. (c) Density plot of the simulated statistic with zero mean and stochastic volatility. (d) Density plot of the simulated statistic with non-zero mean and stochastic volatility. For each plot, a standard normal density function is imposed. The solid line is a standard normal density function and the line with circles is an empirical null distribution of the jump statistic from one million simulations.

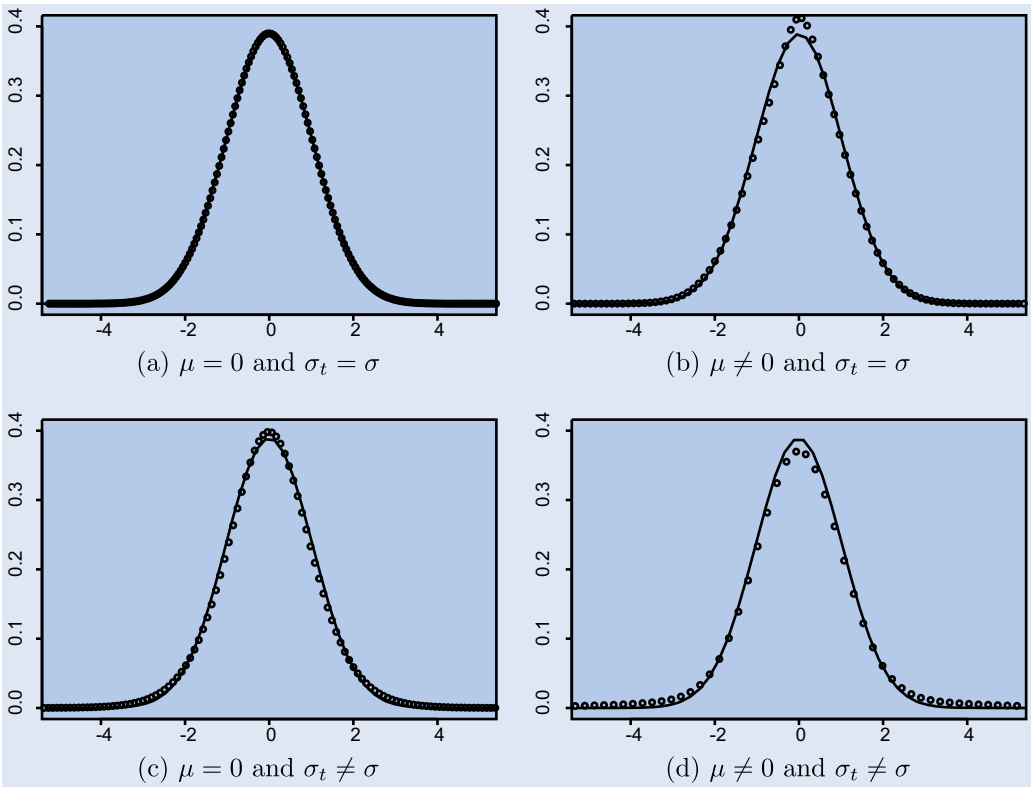


Figure 2. Density plot of the simulated statistic under the null hypothesis with *S*8 filter. (a) Density plot of the simulated statistic with zero mean and constant volatility. (b) Density plot of the simulated statistic with non-zero mean and constant volatility. (c) Density plot of the simulated statistic with zero mean and stochastic volatility. (d) Density plot of the simulated statistic with non-zero mean and stochastic volatility. For each plot, a standard normal density function is imposed. The solid line is a standard normal density function and the line with circles is an empirical null distribution of the jump statistic from one million simulations.

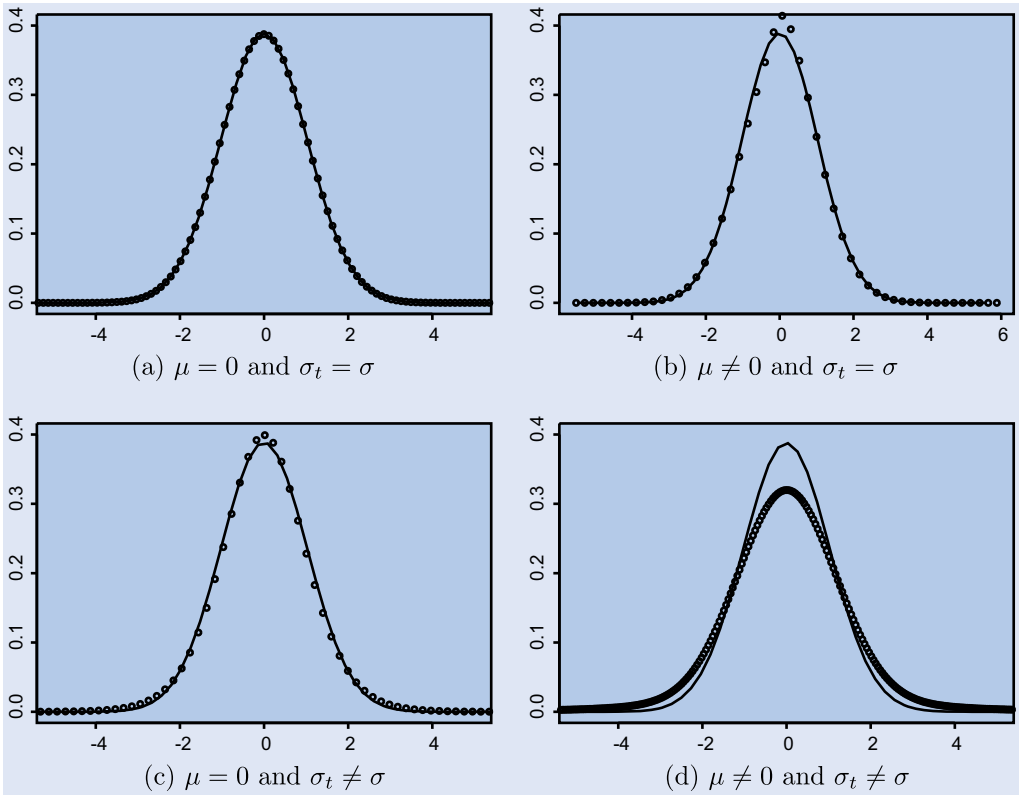


Figure 3. Density plot of the simulated statistic under the null hypothesis with the $D4$ filter. (a) Density plot of the simulated statistic with zero mean and constant volatility. (b) Density plot of the simulated statistic with non-zero mean and constant volatility. (c) Density plot of the simulated statistic with zero mean and stochastic volatility. (d) Density plot of the simulated statistic with non-zero mean and stochastic volatility. For each plot, a standard normal density function is imposed. The solid line is a standard normal density function and the line with circles is an empirical null distribution of the jump statistic from one million simulations.



same plots for the $D4$ filter.† The $S8$ filter may be preferred to the $D4$ filter since the $S8$ belongs to the least asymmetric filter class which permits nearly zero phase distortion. This property may be helpful in conveying information about the jump location in the time domain.

Additionally, figures 1–3 show that the test statistic follows a standard normal distribution when volatility is constant. When volatility is stochastic, the test statistic has a few extreme values. We found that when the frequency increases, i.e. $\Delta t \to 0$, such extreme values diminish. The extreme values of the test statistics originates from two sources: the simultaneous estimation of a drift and stochastic volatility, as well as the choice of wavelet filter. The inefficiency of non-zero mean may decrease the performance of BPV as noticed by Lee and Mykland (2008). We found this problem is most severe when we both have a drift and stochastic volatility. When we need to estimate only the drift or the stochastic volatility, the performance of the test increases significantly. Additionally, simulations demonstrate that the wavelet filter with less leakage is more effective in approximating the null empirical distribution. This is mainly due to the fact that wavelet filters with less leakage utilize information better both at the estimated location and in the neighborhood of the estimated location. If we compare figure 2(d) to figures 1(d) and 3(d), we can see that the extreme test statistics are diminished with the $S8$ filter.

### *4.2. Size and power properties*

First, we present the size of the test statistic in table 1.‡ As shown in the table, the test statistic has good size. The probability of false jump detection is quite close to their nominal values at the 1% and 5% levels.

Subsequently, we study the power of the test statistic. For the purposes of illustration, we allow only one jump per simulation. In particular, we study a sample of 1024 observations and a jump occurring at the 819th observation. The jump size could be large, e.g. $3\sigma$ (three standard deviation of return volatility) or could small, e.g, $0.1\sigma$ (10% standard deviation of return volatility).§ We also examine the performance of the test statistic at the different time scales. In the simulations, the number of repetitions is 1000.

We employ two measures of performance to characterize the test. One is the *power* of the test, which is the probability of detecting the actual jump at the time $i$ when the jump

†There is no close-form expression for $S8$ filter coefficients $h_l$. The coefficients approximately are $h_0 = 0.02278517$, $h_1 = 0.00891235$, $h_2 = -0.07015881$, $h_3 = -0.21061727$, $h_4 = 0.56832912$, $h_5 = -0.35186953$, $h_6 = -0.02095548$, $h_7 = 0.05357445$.

‡We use 99% and 95% critical values from the standard normal distribution multiplied by a scalar given in the Theorem 1, for example $\frac{\sqrt{\pi}}{\sqrt{2}}$ in the $Haar$ filter case, 1.22 in the case of $D4$ filter and 1.18 in the case of $S8$ filter.

§Fan and Wang (2007) found that the ratio of jump variation to the total return variation is 1 to 1.5 in the foreign exchange market.

Table 1. Size of test statistic at 1% and 5% levels.

| *Haar* filter | *D*4 filter | *S*8 filter | *Haar* filter | *D*4 filter | *S*8 filter |
|---|---|---|---|---|---|
| | 1% Level | | | 5% Level | |
| | | $\mu = 0$ and $\sigma_t = \sigma$ | | | |
| 0.013 | 0.011 | 0.009 | 0.058 | 0.053 | 0.048 |
| | | $\mu \neq 0$ and $\sigma_t = \sigma$ | | | |
| 0.013 | 0.016 | 0.010 | 0.057 | 0.054 | 0.050 |
| | | $\mu = 0$ and $\sigma_t \neq \sigma$ | | | |
| 0.018 | 0.016 | 0.014 | 0.057 | 0.053 | 0.049 |
| | | $\mu \neq 0$ and $\sigma_t \neq \sigma$ | | | |
| 0.006 | 0.014 | 0.012 | 0.025 | 0.046 | 0.042 |

Size is defined as the detection of spurious jumps under the null hypothesis of no jumps. The number of Monte Carlo simulations is 1000. Four scenarios are investigated; specifically, the case of zero mean and constant variance, the case of non-zero mean and constant variance, the case of zero mean and non-constant variance, and the case of non-zero mean and non-constant variance.



occurs†. Note that when we apply the test to the simulated data, we let it detect the location of jump itself without using the information about the actual jump location. The power of the test is calculated as the number of cases where the test detects a jump at $819th$ observation divided by the number of repetitions. We also consider another measure, a success rate measure, which is consistent with the probability of spurious detection of jumps as in Lee and Mykland (2008). Specifically, if the test detects the true jump without detecting any other spurious jumps, we call it a success. Recall that there is only one jump in the true data, i.e. the test should detect one and only one jump at the 819th observation. Therefore, the second measure is the success rate defined as the number of successes divided by the number of repetitions.

When the test statistic exceeds a threshold, the null hypothesis of no jumps is rejected. There are two threshold levels used in the paper: 1% and 5% levels of the null distribution. We define a jump as being detected when the test statistic exceeds the threshold level. We also investigate the performance of the test statistic with different time scales. We let the time step used to generate the continuous process of the underlying asset be 252 times the number of observations per day. For instance, we choose the time step between observations to be $252 * 1$ for daily observations. In the simulations, we choose the observations per day to be 24 or 12. When we choose the observations per day to be 24, the sample size of 1024 corresponds to two weeks (10 trading days) of hourly data assuming there are 24 trading hours per day. Alternatively, when we choose the observations per day to be 12, the sample size of 1024 corresponds to one month (20 trading days) of bi-hourly data.‡ The findings suggest that the power of the test is improved when sampling frequency is higher. This is consistent with the findings of Lee and Mykland (2008).

Table 2 shows the power comparisons for the *S*8 filter case, *D*4 filter case, *Haar* filter case (LM), linear test of Barndorff-Nielsen and Shephard (2006) (BNS), and difference test of Jiang and Oomen (2008) (JO). Our wavelet-based test has good power and outperforms other tests at both the 1% and 5% levels. all the *S*8, *D*4 and *Haar* filters can detect the actual jump without failure. For jump size $0.25\sigma$ when the *S*8 filter is used, the test can detect the actual jump with 83% probability at 5% level and 62% probability at 1% level. Similarly, for the $0.1\sigma$ case, the *S*8 filter detects jumps with 21% and 8% probability, respectively. The *Haar* filter (*D*4 filter) detects the actual jumps with a probability 97% (91%) and 91% (85%) at 5% level and 1% level, respectively, for jump size $0.25\sigma$. For jump size $0.1\sigma$, the *Haar* (*D*4) filter detects the actual jump with the probabilities 34% (27%) and 14% (11%) accordingly. Recall that wavelet coefficients of the *Haar* filter are simple differences of prices, whereas wavelet coefficients of *S*8 and *D*4 filter are weighted averages of returns in the neighborhood of a given location. Thus, it is easier for a *Haar* filter to detect large movement in returns and, not surprisingly, the *Haar* filter marginally outperforms the *S*8 filter when the magnitude of jump is very small. All the *Haar* filter, the *S*8 filter and *D*4 filter outperform BNS and JO tests using at the 1% or 5% levels. This suggests that wavelet-based methods can be more effective for detecting large movements in the presence of jumps. Note that a large variation in microstructure noise is equivalent to decrease the relative magnitude of jump size to the spot volatility of underlying process. From our simulations, wavelet-based methods indeed demonstrate superior performance for detecting jumps even when the jump size relative to spot volatility is small. This is important for detection of jumps in high frequency financial time series data, where microstructure noise inevitably exists.

As a performance measure, power only considers the ability of a test to detect the true jumps while ignoring the possibility of spurious jumps detection. Intuitively, if we set a threshold low enough to reject the null hypothesis at every point of time, the actual jumps can be detected without failure if they exist. However, too many spurious jumps may be detected. Thus, *success rate* is employed as an additional and more stringent measure. Recall that success rate measures the ability of a test to detect the actual jumps only. That is, if and only if the test only detects the true jumps, it is regarded as successful. This measure is more appropriate for describing the performance of jump detection tests. Table 3 presents success rate comparisons for the *S*8 filter, *D*4 filter, *Haar* filter (LM), linear test of Barndorff-Nielsen and Shephard (2006) (BNS), and difference test of Jiang and Oomen (2008) (JO). At 5% level, the success rate is zeros for all the *S*8, *D*4 and *Haar* filter. In contrast, BNS

†This test may also detect spurious jumps in addition to an actual jump.

‡For instance, in foreign exchange market, there are 24 trading hours per a trading day.

Table 2. Power comparison with other jump tests.

| Jump size | *S*8 filter | *D*4 filter | *Haar* filter (LM) | Linear test (BNS) | Difference test (JO) |
|---|---|---|---|---|---|
| | | | $Frequency = 2\,h$, 1% Level | | |
| $3.00\sigma$ | 1.000 | 1.000 | 1.000 | 0.855 | 0.843 |
| $1.00\sigma$ | 1.000 | 1.000 | 1.000 | 0.811 | 0.769 |
| $0.50\sigma$ | 1.000 | 1.000 | 1.000 | 0.711 | 0.647 |
| $0.25\sigma$ | 0.626 | 0.852 | 0.908 | 0.533 | 0.440 |
| $0.10\sigma$ | 0.082 | 0.112 | 0.138 | 0.111 | 0.099 |
| | | | $Frequency = 1\,h$, 1% Level | | |
| $3.00\sigma$ | 1.000 | 1.000 | 1.000 | 0.874 | 0.861 |
| $1.00\sigma$ | 1.000 | 1.000 | 1.000 | 0.846 | 0.821 |
| $0.50\sigma$ | 1.000 | 1.000 | 1.000 | 0.803 | 0.755 |
| $0.25\sigma$ | 0.700 | 0.874 | 0.912 | 0.638 | 0.554 |
| $0.10\sigma$ | 0.102 | 0.132 | 0.186 | 0.109 | 0.098 |
| | | | $Frequency = 2\,h$, 5% Level | | |
| $3.00\sigma$ | 1.000 | 1.000 | 1.000 | 0.855 | 0.843 |
| $1.00\sigma$ | 1.000 | 1.000 | 1.000 | 0.811 | 0.769 |
| $0.50\sigma$ | 1.000 | 1.000 | 1.000 | 0.711 | 0.647 |
| $0.25\sigma$ | 0.830 | 0.911 | 0.972 | 0.533 | 0.440 |
| $0.10\sigma$ | 0.216 | 0.276 | 0.336 | 0.111 | 0.099 |
| | | | $Frequency = 1\,h$, 5% Level | | |
| $3.00\sigma$ | 1.000 | 1.000 | 1.000 | 0.874 | 0.861 |
| $1.00\sigma$ | 1.000 | 1.000 | 1.000 | 0.846 | 0.821 |
| $0.50\sigma$ | 1.000 | 1.000 | 1.000 | 0.803 | 0.755 |
| $0.25\sigma$ | 0.930 | 0.955 | 0.976 | 0.638 | 0.554 |
| $0.10\sigma$ | 0.218 | 0.336 | 0.380 | 0.109 | 0.098 |

The simulation only allows one jump and assumes constant volatility and non-zero drift part in price process. The number of repetitions is 1000. The power of the test is defined as the probability that the test will detect the actual jump (even when the test also detects spurious jumps). The table shows the power of our test (*S*8 filter, *D*4 filter, *Haar* filter corresponding to Lee and Mykland (2008) (LM)), linear test of Barndorff-Nielsen and Shephard (2006) (BNS), and difference test of Jiang and Oomen (2008) (JO). The time interval for integration of the linear (BNS) and difference tests (JO) is one day. The jump sizes are 10, 25, 50, 100, 300% of the spot volatility.



provides 16% to 25% success rate. The JO test provides 5% success rate. Combined with the power results, the zero success rate suggests that all the *Haar*, *D*4 and *S*8 filters always detect at least one spurious jump in addition to the actual jump. The BNS and JO tests show a superior performance for not detecting spurious jumps compared to *Haar* filter and *S*8 filter at 5% level. However, at 1% level, the *S*8 filter outperforms BNS, LM, and JO tests at all jump size levels. This shows that when a more stricter threshold level is imposed, i.e. at 1% level, the *S*8 filter outperforms the BNS and JO tests in both power and success rate.

Additionally, the *S*8 filter is recommended rather than both the *Haar* and *D*4 filter First, at 1% level, the success rate of *S*8 filter is satisfactory compared to both the *Haar* filter and *D*4 filter and other alternative candidate tests. Second, the *S*8 filter can still provide reasonable power compared to both *D*4 and the *Haar* filter when the magnitude of jump is in a range from $0.25\sigma$ to $3\sigma$. Thus, when a filter with less leakage is used, the test performance improves.

Overall, the proposed test demonstrates satisfactory power for detecting jumps. Based on simulations, we recommend the *S*8 filter or other filters with less leakage and the use of a stringent threshold level (at 1% level) for empirical implementation of jump detection. The filter with less leakage possesses similar power to the both *Haar* filter (equivalent to the Lee and Mykland test) and *D*4 filter, but the success rate is higher. This improvement occurs because the wavelet filter with less leakage averages the noise embedded in the underlying process to facilitate the to differentiation of the jump from the noise and volatility effect.

## 5. Empirical analysis for US equity markets

We now apply the jump detection test to three major US individual equities to determine their jump dynamics. To characterize jump dynamics, the frequency of arrivals and the magnitude of jumps need to be characterized. Using a jump detection method, the jump locations can be estimated in a given time interval.† This allows us to calculate the arrival density of jumps per trading day. To characterize the magnitude of the jumps, we employ a method that follows Fan and Wang (2007). Formally, for each estimated jump location $t_i$, we choose a small neighborhood, $t_i \mp \gamma$, for a small $\gamma > 0$. Then, we calculate the average of the prices over $[t_i - \gamma, t_i]$ and $[t_i, t_i + \gamma]$. Let $P_{t_i-}$ and $P_{t_i+}$ to denote the averages accordingly. When a jump is detected, its jump size is estimated by $L_{t_i} = P_{t_i+} - P_{t_i-}$. Fan and Wang (2007) shows that this estimator of jump size is consistent when the neighborhood $\gamma$ is chosen such that $\gamma \sim T^{-1/2}$, where $T$ is the sample size.

†In this empirical study, the time interval is chosen to be one trading day. That is, we detect jump locations for each trading day. In particular, we examine the trading hours from 9:30 AM to 4:00 PM plus off-trading hours from 8:00 AM to 9:30 AM and from 4:00 PM to 6:00 PM.

Table 3. Success rate comparison with other jump tests.

| Jump size | $S8$ filter | $D4$ filter | $Haar$ filter (LM) | Linear test (BNS) | Difference test (JO) |
|---|---|---|---|---|---|
| | | | $Frequency = 2\,h$, 1% Level | | |
| $3.00\sigma$ | 0.802 | 0.232 | 0.040 | 0.251 | 0.065 |
| $1.00\sigma$ | 0.198 | 0.110 | 0.010 | 0.243 | 0.048 |
| $0.50\sigma$ | 0.050 | 0.032 | 0.008 | 0.119 | 0.001 |
| $0.25\sigma$ | 0.024 | 0.011 | 0.004 | 0.084 | 0.000 |
| $0.10\sigma$ | 0.008 | 0.003 | 0.000 | 0.051 | 0.000 |
| | | | $Frequency = 1\,h$, 1% Level | | |
| $3.00\sigma$ | 0.240 | 0.117 | 0.112 | 0.188 | 0.009 |
| $1.00\sigma$ | 0.212 | 0.104 | 0.100 | 0.169 | 0.008 |
| $0.50\sigma$ | 0.156 | 0.032 | 0.019 | 0.091 | 0.001 |
| $0.25\sigma$ | 0.112 | 0.022 | 0.012 | 0.000 | 0.000 |
| $0.10\sigma$ | 0.004 | 0.005 | 0.004 | 0.000 | 0.000 |
| | | | $Frequency = 2\,h$, 5% Level | | |
| $3.00\sigma$ | 0.000 | 0.000 | 0.000 | 0.251 | 0.065 |
| $1.00\sigma$ | 0.000 | 0.000 | 0.000 | 0.243 | 0.048 |
| $0.50\sigma$ | 0.000 | 0.000 | 0.000 | 0.119 | 0.001 |
| $0.25\sigma$ | 0.000 | 0.000 | 0.000 | 0.084 | 0.000 |
| $0.10\sigma$ | 0.000 | 0.000 | 0.000 | 0.051 | 0.000 |
| | | | $Frequency = 1\,h$, 5% Level | | |
| $3.00\sigma$ | 0.000 | 0.000 | 0.000 | 0.188 | 0.009 |
| $1.00\sigma$ | 0.000 | 0.000 | 0.000 | 0.169 | 0.008 |
| $0.50\sigma$ | 0.000 | 0.000 | 0.000 | 0.091 | 0.001 |
| $0.25\sigma$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $0.10\sigma$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

The simulation only allows one jump and assumes constant volatility and non-zero drift part in the price process. The number of repetitions is 1000. Success rate is defined as the probability that the test will detect only true jumps. The table shows the success rate of our test ($S8$ filter, $D4$ filter, $Haar$ filter corresponding to Lee and Mykland (2008) (LM)), Linear test of Barndorff-Nielsen and Shephard (2006) (BNS), and Difference test of Jiang and Oomen (2008) (JO). The time interval for integration of the linear (BNS) and difference tests (JO) is one day. The jump sizes are 10, 25, 50, 100, 300% of the spot volatility.



### *5.1. Multi-scale jump dynamics*

We use ultra-high frequency tick data from transactions on the New York Stock Exchange (NYSE) collected from the Trade and Quote (TAQ) database. We apply the jump detection test to the log transaction prices.†The time span is three months from January 1st to March 31st, 2008. The three equities we examine are Wal-Mart (WMT), IBM (IBM) and General Electric (GE). Recall that the test is robust to the presence of microstructure noise, so we employ 1 min data to improve the efficiency of estimation.

For comparison purposes, we also report jump dynamics at different time scales. Specifically, jump dynamics are reported using 1-min, 5-min, and 15-min data.‡ This comparison offers useful insights for both the theoretical framework and the implementation of jump detection. Due to the robustness of this test in the presence of microstructure noise, the difference in jump dynamics detected at different time scales should account for the difference in jump arrival densities at different time scales. Intuitively, if two jumps of similar magnitudes occur in opposite directions, i.e. a positive jump and negative jump, a large difference in price level should not be observed, and no jumps will be detected in this time interval. Specifically, low frequency data might ignore jump dynamics that could not be captured due to the sampling frequency. In order to examine the jump dynamics at a pre-chosen sampling frequency, we implicitly assume that the smallest duration between two jumps should be larger than the time interval implied by the sampling frequency. For example, with 5-min data, it may be difficult to detect two distinct jumps. For this reason, when using 5-min data, we implicitly assume that only one jump can occur every 5 min. If this is true, significant differences should not be observed in jump dynamics across different time scales. Otherwise, a higher sampling frequency should be used to extract full jump dynamics.

Table 4 reports the dynamics of jump arrivals for these three equities. Jump dynamics for the same equity are substantially different at different time scales. The jump arrival densities for GE are 11.7, 3.7, and 1.5 jumps per day at 1-min, 5-min, and 15-min sampling frequency, respectively. Similar patterns are found for IBM and WMT. The jump arrival densities for IBM are 11.3, 3.6, and 1.2 jumps per day at 1-min, 5-min, 15-min sampling frequency, respectively. The jump arrival densities for WMT are 11, 3.6, and 1.3 at 1-min, 5-min, 15-min, respectively. Thus, the jump arrival densities at 1-min sampling frequency per day for all three equities are higher than jump arrival densities at 5-min and 15-min. As stated earlier, if the sampling frequency is high enough to observe full jump dynamics, significant differences should not be observed in jump arrival densities across time scales. Therefore, sampling data at 15-min ignores a significant portion of jump dynamics that occurs at higher frequencies than the sampling frequency. Thus, sampling data at 15-min even 5-min

†We report the results using $S8$ filter with 1% significant level to detect jumps and the results based on both $Haar$ filter and $D4$ filter are not qualitatively changed.

‡Lee and Mykland (2008) used 15-min transaction data.

Table 4. Jump dynamics of individual equities.

| Sampling frequency | $N_T$ | $N_{T,+}$ | $N_{T,-}$ | $N_t$ | $Size_{N_t}$ | $Size_{N_{t,+}}$ | $Size_{N_{t,-}}$ |
|---|---|---|---|---|---|---|---|
| | | | GE | | | | |
| 1-min | 11.7 | 6.10 | 5.60 | 2.70 | −0.04% | 0.01% | −0.04% |
| 5-min | 3.70 | 1.80 | 1.90 | 0.90 | −0.09% | 0.01% | −0.09% |
| 15-min | 1.50 | 0.75 | 0.75 | 0.70 | −0.03% | 0.04% | −0.04% |
| | | | IBM | | | | |
| 1-min | 11.3 | 5.90 | 5.40 | 2.70 | −0.001% | 0.10% | −0.02% |
| 5-min | 3.60 | 1.80 | 1.80 | 0.70 | −0.05% | 0.02% | −0.05% |
| 15-min | 1.20 | 0.60 | 0.60 | 0.40 | −0.05% | 0.03% | −0.07% |
| | | | WMT | | | | |
| 1-min | 11.0 | 6.00 | 5.00 | 3.50 | −0.01% | 0.08% | −0.01% |
| 5-min | 3.60 | 1.90 | 1.70 | 1.00 | −0.20% | 0.05% | −0.05% |
| 15-min | 1.30 | 0.80 | 0.50 | 0.40 | −0.02% | 0.01% | −0.02% |

This table contains the jump dynamics of three US individual equities: GE, IBM, and WMT based on transaction prices from the NYSE during three months from January 1st to March 31st, 2008. $N_T$ is the average number of total jumps estimated in each day. $N_{T,+}$ is the average of the number of positive jumps estimated for each day. $N_{T,-}$ is the average number of negative jumps estimated for each day. $N_t$ is the average number of trading session jumps estimated for each day. $Size_{N_t}$ is the average magnitude of the trading-session jumps estimated for each day. $Size_{N_{t,+}}$ is the average magnitude of positive trading-session jumps estimated for each day. $Size_{N_{t,-}}$ is the average magnitude of the negative trading-session jumps estimated for each day.



might not be appropriate for the purpose of risk management or dynamic hedging which requires continuous adjustments of positions. If the jump arrival density is estimated at an incorrect frequency, the impact of jumps will be underestimated. Another suggestion is that the averaging of jumps with opposite directions should be considered for dynamic hedging. It should also be noted that the jump arrival densities are similar for all three equities at the same scale. This suggests that macroeconomic news might play an important role in the formation of jumps that accounts for a 'common trend.'

### *5.2. Positive jumps vs. negative jumps*

We further investigated the jump dynamics for jumps of different directions, i.e. positive jumps and negative jumps. Table 4 reports the arrival densities for jumps of different directions. Jump arrivals are symmetric for all equities. Thus, the arrival densities of positive jumps and negative jumps are similar across different time scales. For instance, the arrival densities of positive and negative jumps for GE at the 1-min sampling frequency are 6.1 and 5.6 per day, respectively. The arrival densities of positive and negative jumps for GE at 5-min sampling are 1.8 and 1.9 per day, respectively. The arrival densities of positive and negative jumps for GE at 15-min are both 0.75 per day. This pattern also holds for IBM and WMT.

While the arrival densities are symmetric for positive and negative jumps, the magnitudes of jumps are asymmetric at high frequencies for all equities. At the 1-min sampling frequency, the mean size of negative jumps for GE is −0.04% while the mean size of the positive jumps is 0.01%. In the case of IBM, the mean size of jumps are −0.02% and 0.10%, respectively. This suggests that the magnitudes of jumps are not symmetric at high sampling frequencies and so symmetric distribution of jump sizes for derivative security pricing should not be assumed in practice. This asymmetry in jump magnitude decreases at lower sampling frequencies. Therefore, the normality assumption of jump sizes might not be a good approximation for high frequency data. A skewed distribution might be required for modeling of jump size when conducting high frequency trading.

### *5.3. Trading session vs. off-trading session*

Next, we decompose the jump dynamics of these equities into two sessions: the day trading session (9:30 AM to 4:00 PM for NYSE) and the off-trading session (8:00 AM to 9:30 AM and 4:00 PM to 6:00 PM). Table 4 reports the jump dynamics in the day trading session. They demonstrate that the majority of jumps occurs in off-trading sessions. Only 20% of jumps occurs in the day trading session. The average number of jumps that occur in day trading time session at 1-min frequency is 2 to 3 per day which is comparable to the findings of Fan and Wang (2007). This suggests that the low trading volume during after-hours trading is a strong determinant of jumps.

## 6. Conclusions

This paper introduces a new nonparametric test based on wavelet transformations to detect jump arrival times in high frequency financial time series data. This test is motivated by the ability of the wavelets method to decompose data into different time scales. This localization property of wavelets is shown to be superior for jump detection. We show that the distribution of the test under the null hypothesis of no jumps is asymptotically normal. We demonstrate that the test is robust across different price processes and to the presence of market microstructure noise. A Monte Carlo simulation is conducted to demonstrate the test has good size and power properties. We also demonstrate that the use of wavelet filters with less leakage improves the success rate of the test, that is, the ability of the test to only detect true jumps.

An empirical implementation is then conducted for US equity markets, and jump dynamics are found to change dramatically across time scales. This suggests that choosing a

proper sampling frequency is very important for investigating jump dynamics. Additionally, the arrival densities of positive jumps and negative jumps are similar, but the magnitudes of the jumps are asymmetrically distributed at high frequencies. Finally, the majority of jumps occur outside of the day trading session with only 20% of jumps occuring within these sessions.

**Acknowledgements**

Yi Xue gratefully acknowledge financial support from the National Natural Science Foundation of China (NSFC), project No. 71101031. Ramazan Gençay gratefully acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada. This paper is supported by Program for Innovative Research Team in UIBE. We would like to thank two referees for their helpful comments and valuable suggestions which have improved the paper greatly. The remaining errors are ours.



**References**

Ait-Sahalia, Y. and Jacod, J., Testing for jumps in a discretely observed process. *Ann. Stat*., 2009, **37**, 184–222.

Andersen, T., Bollerslev, T., Diebold, F. and Labys, P., Modeling and forecasting realized volatility. *Econometrica*, 2003, **71**, 579–625.

Andersen, T., Bollerslev, T. and Diebold, F., Roughing it up: Including jump components in the measurement modeling and forecasting of return volatility. *Rev. Econ. Stat*., 2007, **89**, 701–720.

Barndorff-Nielsen, O.E. and Shephard, N., Power and bipower variation with stochastic volatility and jumps. *J. Financ. Econom*., 2004, **2**, 1–37.

Barndorff-Nielsen, O.E. and Shephard, N., Econometrics of testing for jumps in financial economics using bipower variation. *J. Financ. Econom*., 2006, **4**, 1–30.

Daubechies, I., Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math*., 1988, **41**, 909–996.

Dungey, M., McKenzie, M. and Smith, V., Empirical evidence on jumps in the term structure of the US treasury market. *J. Empir. Finance*, 2009, **16**, 430–445.

Fan, J. and Wang, Y., Multi-scale jump and volatility analysis for high-frequency financial data. *J. Am. Stat. Assoc*., 2007, **102**, 1349–1362.

Gençay, R., Selçuk, F. and Whitcher, B., *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, 2001 (Academic Press: New York).

Haar, A., Zur theorie der orthogonalen funktionensysteme. *Math. Ann*., 1910, **69**, 331–371.

Hansen, P.R. and Lunde, A., Realized variance and market microstructure noise. *J. Bus. Econ. Stat*., 2006, **24**, 127–161.

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M. and Vetter, M., Microstructure noise in the continuous case: The pre-averaging approach. *Stoch. Process. Appl*., 2009, **119**, 2249–2276.

Jiang, G.J. and Oomen, R.C., Testing for jumps when asset prices are observed with noise – a swap variance approach. *J. Econom*., 2008, **144**, 352–370.

Johannes, M., The statistical and economic role of jumps in continuous-time interest rate models. *J. Finance*, 2004, **59**, 227–260.

Lee, S.S. and Mykland, P.A., Jumps in financial markets: A new nonparametric test and jump dynamics. *Rev. Financ. Stud*., 2008, **21**, 2535–2563.

Leone, F.C., Nelson, L.S. and Nottingham, R.B., The folded normal distribution. *Technometrics*, 1961, **3**, 543–550.

Piazzesi, M., Bond yields and the federal reserve. *J. Polit. Econ*., 2003, **113**, 311–344.

Podolskij, M. and Vetter, M., Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 2009, **315**, 634–658.

Pollard, D., *A User's Guide of Measure Theoretic Probability*, 2002 (Cambridge University Press: Cambridge).

Protter, P., *Stochastic Integration and Differential Equations: A New Approach*, 2004 (Springer: New York).

Wang, Y., Jump and sharp cusp detection by wavelets. *Biometrika*, 1995, **82**, 385–397.

Zhang, L., Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Working Paper, Carnegie Mellon University, 2006.

Zhang, L., Ait-Sahalia, Y. and Mykland, A., A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Am. Stat. Assoc*., 2005, **100**, 1394–1411.

## Appendix A1. Wavelet Transformations†

A wavelet is a small wave which grows and decays in a limited time period.‡ To formalize the notion of a wavelet, let $\psi(.)$ be a real valued function such that its integral is zero, $\int_{-\infty}^{\infty} \psi(t)\,dt = 0$, and its square integrates to unity, $\int_{-\infty}^{\infty} \psi(t)^2\,dt = 1$. Thus, although $\psi(.)$ has to make some excursions away from zero, any excursions it makes above zero must cancel out excursions below zero, i.e. $\psi(.)$ is a small wave, or a wavelet.

Fundamental properties of the continuous wavelet functions (filters), such as integration to zero and unit energy, have discrete counterparts. Let $h = (h_0, \ldots, h_{L-1})$ be a finite length discrete wavelet (or high pass) filter such that it integrates (sums) to zero, $\sum_{l=0}^{L-1} h_l = 0$, and has unit energy, $\sum_{l=0}^{L-1} h_l^2 = 1$. In addition, the wavelet filter $h$ is orthogonal to its even shifts; that is,

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0, \text{ for all nonzero integers } n. \tag{15}$$

The natural object to complement a high-pass filter is a low-pass (scaling) filter $g$. We will denote a low-pass filter as $g = (g_0, \ldots, g_{L-1})$. The low-pass filter coefficients are determined by the *quadrature mirror relationship*§

$$g_l = (-1)^{l+1} h_{L-1-l} \quad \text{for} \quad l = 0, \ldots, L-1 \tag{16}$$

and the inverse relationship is given by $h_l = (-1)^l g_{L-1-l}$. The basic properties of the scaling filter are: $\sum_{l=0}^{L-1} g_l = \sqrt{2}$, $\sum_{l=0}^{L-1} g_l^2 = 1$,

$$\sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0, \tag{17}$$

for all nonzero integers $n$, and

$$\sum_{l=0}^{L-1} g_l h_{l+2n} = \sum_{l=-\infty}^{\infty} g_l h_{l+2n} = 0 \tag{18}$$

†This appendix offers a brief introduction to Wavelet transformations. Interested readers can consult Gençay *et al.* (2001) for more details.

‡This section closely follows Gençay *et al.* (2001). The contrasting notion is a big wave such as the sine function which keeps oscillating indefinitely.

§Quadrature mirror filters (QMFs) are often used in the engineering literature because of their ability for perfect reconstruction of a signal without aliasing effects. Aliasing occurs when a continuous signal is sampled to obtain a discrete time series.

for all integers $n$. Thus, scaling filters are average filters and their coefficients satisfy the orthonormality property that they possess unit energy and are orthogonal to even shifts. By applying both $h$ and $g$ to an observed time series, we can separate high-frequency oscillations from low-frequency ones. In the following sections, we will briefly describe DWT and MODWT.

### *A1.1 Discrete wavelet transformation*

With both wavelet filter coefficients and scaling filter coefficients, we can decompose the data using the (discrete) wavelet transformation (DWT). Formally, let us introduce the DWT through a simple matrix operation. Let $\mathbf{y}$ to be the dyadic length vector ($T = 2^J$) of observations. The length $T$ vector of discrete wavelet coefficients $\mathbf{w}$ is obtained via

$$\mathbf{w} = \mathcal{W}\mathbf{y}$$

where $\mathcal{W}$ is an $T \times T$ orthonormal matrix defining the DWT. The vector of wavelet coefficients can be organized into $J+1$ vectors, $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_J, \mathbf{v}_J]^{'}$, where $\mathbf{w}_j$ is a length $T/2^j$ vector of wavelet coefficients associated with changes on a scale of length $\lambda_j = 2^{j-1}$, and $\mathbf{v}_J$ is a length $T/2^J$ vector of scaling coefficients associated with averages on a scale of length $2^J = 2\lambda_J$.

The matrix $\mathcal{W}$ is composed of the wavelet and scaling filter coefficients arranged on a row-by-row basis. Let

$$\mathbf{h}_1 = [h_{1,N-1}, h_{1,N-2}, \ldots, h_{1,1}, h_{1,0}]^{'}$$

be the vector of zero-padded unit scale wavelet filter coefficients in reverse order. Thus, the coefficients $h_{1,0}, \ldots, h_{1,L-1}$ are taken from an appropriate ortho-normal wavelet family of length $L$, and all values $L < t < T$ are defined to be zero. Now circularly shift $\mathbf{h}_1$ by factors of two so that

$$\mathbf{h}_1^{(2)} = [h_{1,1}, h_{1,0}, h_{1,N-1}, h_{1,N-2} \ldots, h_{1,3}, h_{1,2}]^{'}$$
$$\mathbf{h}_1^{(4)} = [h_{1,3}, h_{1,2}, h_{1,1}, h_{1,0} \ldots, h_{1,5}, h_{1,4}]^{'}$$

and so on. Define the $T/2 \times T$ dimensional matrix $\mathcal{W}_1$ to be the collection of $T/2$ circularly shifted versions of $\mathbf{h}_1$. Hence,

$$\mathcal{W}_1 = [\mathbf{h}_1^{(2)}, \mathbf{h}_1^{(4)}, \ldots, \mathbf{h}_1^{(T/2-1)}, \mathbf{h}_1]^{'}$$

Let $\mathbf{h}_2$ be the vector of zero-padded scale 2 wavelet filter coefficients defined similarly to $\mathbf{h}_1$. $\mathcal{W}_2$ is constructed by circularly shifting the vector $\mathbf{h}_2$ by factor of four. Repeat this to construct $\mathcal{W}_j$ by circularly shifting the vector $\mathbf{h}_j$ (the vector of zero-padded scale $j$ wavelet filter coefficients) by $2^j$. The matrix $\mathcal{V}_J$ is simply a column vector whose elements are all equal to $1/\sqrt{T}$. Then, the $T \times T$ dimensional matrix $\mathcal{W}$ is $\mathcal{W} = [\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_J, \mathcal{V}_J]^{'}$.

When we are provided with a dyadic length time series, it is not necessary to implement the DWT down to level $J = \log_2(T)$. A *partial* DWT may be performed instead that terminates at level $J_p < J$. The resulting vector of wavelet coefficients will now contain $T - T/2^{J_p}$ wavelet coefficients and $T/2^{J_p}$ scaling coefficients.

The orthonormality of the matrix $\mathcal{W}$ implies that the DWT is a variance preserving transformation:

$$\|\mathbf{w}\|^2 = \sum_{t=1}^{T/2^J} v_{t,J}^2 + \sum_{j=1}^{J}\left(\sum_{t=1}^{T/2^j} w_{t,j}^2\right) = \sum_{t=1}^{T} y_t^2 = \|\mathbf{y}\|^2 .$$

This can be easily proven through basic matrix manipulation via

$$\|\mathbf{y}\|^2 = \mathbf{y}^{'}\mathbf{y} = (\mathcal{W}^{-1}\mathbf{w})^{'}\mathcal{W}^{-1}\mathbf{w} = \mathbf{w}^{'}(\mathcal{W}\mathcal{W}^{'})^{-1}\mathbf{w}$$
$$= \mathbf{w}^{'}\mathbf{w} = \|\mathbf{w}\|^2 .$$

The first equality holds because of definition of $\|\mathbf{y}\|^2$. The second equality holds because that given $\mathbf{w} = \mathcal{W}\mathbf{y}$, we have $\mathbf{y} = \mathcal{W}^{-1}\mathbf{w}$. Therefore, substituting $\mathbf{y} = \mathcal{W}^{-1}\mathbf{w}$ into $\mathbf{y}^{'}\mathbf{y}$ leads to the second equality. The third equality holds because of the properties of matrix inversion and transposition. The fourth equality holds the property of orthonormal matrix. Notice that since $\mathcal{W}$ is orthonormal matrix, we have $\mathcal{W}^{-1} = \mathcal{W}^{'}$. Thus, $\mathcal{W}\mathcal{W}^{'} = \mathcal{W}\mathcal{W}^{-1} = I$, where $I$ is identity matrix.

Given the structure of the wavelet coefficients, $\|\mathbf{y}\|^2$ is decomposed on a scale-by-scale basis via

$$\|\mathbf{y}\|^2 = \sum_{j=1}^{J} \|w_j\|^2 + \|v_J\|^2 \tag{19}$$

where $\|w_j\|^2 = \sum_{t=1}^{T/2^j} w_{t,j}^2$ is the sum of squared variation of $y$ due to changes at scale $\lambda_j$ and $\|v_J\|^2 = \sum_{t=1}^{T/2^J} v_{t,J}^2$ is the information due to changes at scales $\lambda_J$ and higher.

### *A1.2 Maximum overlap discrete wavelet transformation*

An alternative wavelet transform is MODWT which is computed by *not* subsampling the filtered output. Let $\mathbf{y}$ be an arbitrary length $T$ vector of observations. The length $(J+1)T$ vector of MODWT coefficients $\widetilde{\mathbf{w}}$ is obtained via

$$\widetilde{\mathbf{w}} = \widetilde{\mathcal{W}}\mathbf{y},$$

where $\widetilde{\mathcal{W}}$ is a $(J+1)T \times T$ matrix defining the MODWT. The vector of MODWT coefficients may be organized into $J+1$ vectors via

$$\widetilde{\mathbf{w}} = [\widetilde{\mathbf{w}}_1, \widetilde{\mathbf{w}}_2, \ldots, \widetilde{\mathbf{w}}_J, \widetilde{\mathbf{v}}_J]^T, \tag{20}$$

where $\widetilde{\mathbf{w}}_j$ is a length $T$ vector of wavelet coefficients associated with changes on a scale of length $\lambda_j = 2^{j-1}$ and $\widetilde{\mathbf{v}}_J$ is a length $T$ vector of scaling coefficients associated with averages on a scale of length $2^J = 2\lambda_J$, just as with the DWT.

Similar to the orthonormal matrix defining the DWT, the matrix $\widetilde{\mathcal{W}}$ is also made up of $J+1$ submatrices, each of them $T \times T$, and may be expressed as

$$\widetilde{\mathcal{W}} = \begin{bmatrix} \widetilde{\mathcal{W}}_1 \\ \widetilde{\mathcal{W}}_2 \\ \vdots \\ \widetilde{\mathcal{W}}_J \\ \widetilde{\mathcal{V}}_J \end{bmatrix}.$$

The MODWT utilizes the rescaled filters ($j = 1, \ldots, J$)

$$\tilde{\mathbf{h}}_j = \mathbf{h}_j/2^{j/2} \quad \text{and} \quad \tilde{\mathbf{g}}_J = \mathbf{g}_J/2^{J/2}.$$

To construct the $T \times T$ dimensional submatrix $\widetilde{\mathcal{W}}_1$, we circularly shift the rescaled wavelet filter vector $\tilde{\mathbf{h}}_1$ by integer units to the right so that

$$\widetilde{\mathcal{W}}_1 = \left[\tilde{\mathbf{h}}_1^{(1)}, \tilde{\mathbf{h}}_1^{(2)}, \tilde{\mathbf{h}}_1^{(3)}, \ldots, \tilde{\mathbf{h}}_1^{(N-2)}, \tilde{\mathbf{h}}_1^{(N-1)}, \tilde{\mathbf{h}}_1\right]^T. \quad (21)$$

This matrix may be interpreted as the interweaving of the DWT submatrix $\mathcal{W}_1$ with a circularly shifted (to the right by one unit) version of itself. The remaining submatrices $\widetilde{\mathcal{W}}_2, \ldots, \widetilde{\mathcal{W}}_J$ are formed similarly to equation (21), only replace $\tilde{\mathbf{h}}_1$ by $\tilde{\mathbf{h}}_j$.

In practice, a pyramid algorithm is utilized similar to that of the DWT to compute the MODWT. Starting with the data $x_t$ (no longer restricted to be a dyadic length), filter it using $\widetilde{\mathbf{h}}_1$ and $\widetilde{\mathbf{g}}_1$ to obtain the length $T$ vectors of wavelet and scaling coefficients $\widetilde{\mathbf{w}}_1$ and $\widetilde{\mathbf{v}}_1$, respectively.

For each iteration of the MODWT pyramid algorithm, we require three objects: the data vector $\mathbf{x}$, the wavelet filter $\tilde{h}_l$ and the scaling filter $\tilde{g}_l$. The first iteration of the pyramid algorithm begins by filtering (convolving) the data with each filter to obtain the following wavelet and scaling coefficients:

$$\widetilde{w}_{1,t} = \sum_{l=0}^{L-1} \tilde{h}_l y_{t-l \bmod T} \quad \text{and} \quad \widetilde{v}_{1,t} = \sum_{l=0}^{L-1} \tilde{g}_l y_{t-l \bmod T},$$

where $t = 1, \ldots, T$. The length $T$ vector of observations has been high- and low-pass filtered to obtain $T$ coefficients associated with this information. The second step of the MODWT pyramid algorithm starts by defining the data to be the scaling coefficients $\widetilde{\mathbf{v}}_1$ from the first iteration and apply the filtering operations as above to obtain the second level of wavelet and scaling coefficients

$$\widetilde{w}_{2,t} = \sum_{l=0}^{L-1} \tilde{h}_l \widetilde{v}_{1,t-2l \bmod T} \quad \text{and} \quad \widetilde{v}_{2,t} = \sum_{l=0}^{L-1} \tilde{g}_l \widetilde{v}_{1,t-2l \bmod T},$$

$t = 1, \ldots, T$. Keeping all vectors of wavelet coefficients, and the final level of scaling coefficients, we have the following length $T$ decomposition: $\widetilde{\mathbf{w}} = [\widetilde{\mathbf{w}}_1\ \widetilde{\mathbf{w}}_2\ \widetilde{\mathbf{v}}_2]'$. After the third iteration of the pyramid algorithm, where we apply filtering operations to $\mathbf{v}_2$, the decomposition now looks like $\widetilde{\mathbf{w}} = [\widetilde{\mathbf{w}}_1\ \widetilde{\mathbf{w}}_2\ \widetilde{\mathbf{w}}_3\ \widetilde{\mathbf{v}}_3]'$. This procedure may be repeated up to $J$ times where $J = \log_2(T)$ and gives the vector of MODWT coefficients in equation (20).

Similar to DWT, MODWT wavelet and scaling coefficients are variance preserving

$$\|\widetilde{\mathbf{w}}\|^2 = \sum_{t=1}^{T} \widetilde{v}_{t,J}^2 + \sum_{j=1}^{J} \left(\sum_{t=1}^{T} \widetilde{w}_{t,j}^2\right) = \sum_{t=1}^{T} y_t^2 = \|\mathbf{y}\|^2.$$

and a partial decomposition $J_p < J$ may be performed when it deems necessary.

The following properties are important for distinguishing the MODWT from the DWT. The MODWT can accommodate any sample size $T$, while the $J_p$th order partial DWT restricts the sample size to a multiple of $2^{J_p}$. The detail and smooth coefficients of a MODWT are associated with zero phase filters. Thus, events that feature in the original time series can be properly aligned with features in the MODWT multiresolution analysis. The MODWT is invariant to circular shifts in the original time series. This property does not hold for the DWT. The MODWT wavelet variance estimator is asymptotically more efficient than the same estimator based on the DWT. For both MODWT and DWT, the scaling coefficients contain the lowest frequency information. But each level's wavelet coefficients contain progressively lower frequency information.

## Appendix A2. Proofs of Theorems

THEOREM 1 *Assuming zero drift and constant spot volatility in the underlying price process and assuming A1 and A2 are satisfied, if there is no jump in* $(t_{i-1}, t_i)$, *as* $\Delta t \to 0$

$$sup_i\ |J_W(i) - \widehat{J}_W(i)| = O_p(\Delta t^{\frac{3}{2}-\delta-\epsilon}),$$
$$where\ \delta\ satisfies\ 0 < \delta < 3/2\ and\ \widehat{J}_W(i) = \frac{U_i}{d}. \quad (22)$$

*where* $W_t$ *follows a Brownian motion process. And* $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, *a standard normal variable, and a constant* $d = \frac{\sqrt{E[BPV_G]}}{\sqrt{\sum_{l=0}^{L-2}(\sum_{j=0}^{l} h_j)^2}}$, *where*

$$\begin{aligned}
&E[BPV_G] \\
&= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} E[BPV_G|x_1, \ldots, x_{L-2}]\phi(x_1) \\
&\quad \ldots \phi(x_{L-2}) dx_1 \ldots dx_{L-2}, \\
&E[BPV_G|\widetilde{r}_1, \ldots, \widetilde{r}_{L-2}] \\
&= 2\sigma^2 f\left(\frac{-\sum_{l=1}^{L-2}\left(\sum_{j=0}^{l} h_j\right)(\widetilde{r}_{L-l-l})}{h_0}\right) \\
&\quad \times f\left(\frac{-\sum_{l=0}^{L-3}\left(\sum_{j=0}^{l} h_j\right)(\widetilde{r}_{L-l-l})}{\left(\sum_{j=0}^{L-2} h_j\right)}\right),
\end{aligned}$$

*where* $f(a) = 2\phi(a) + 2a\Phi(a) - a$, *and* $\widetilde{r}_t s$ *are independently and identically distributed (iid) standard normal variables.*

*Proof* Assuming zero drift, using MODWT with a general filter to transform the data, the wavelet coefficients at scale 1 are

$$P_{1,t_i} = \sum_{l=0}^{L-1} h_l (P_{t_{i-l}} - P_{t_{i-l-1}}). \quad (23)$$

Since $\sum_{l=0}^{L-1} h_l = 0$, we have $h_{L-1} = -\sum_{l=0}^{L-2} h_l$. Therefore, equation (23) can be rewritten as

$$P_{1,t_i} = \sum_{l=0}^{L-2} \left(\sum_{j=0}^{l} h_j\right)\left(P_{t_{i-l}} - P_{t_{i-l-1}}\right). \quad (24)$$

Hence we have

$$\begin{aligned}
P_{t_{i-l}} - P_{t_{i-l-1}} &= \log S_{t_i - l} - \log S_{t_{i-l-1}} \\
&= \int_{t_{i-l-1}}^{t_{i-l}} \mu_u du + \int_{t_{i-l-1}}^{t_{i-l}} \sigma_u dW(u). \quad (25)
\end{aligned}$$

Imposing assumption A1, we have

$$\int_{t_{i-1}}^{t_i} \mu_u du - \mu_{t_{i-1}} \Delta t = O_p\left(\Delta t^{\frac{3}{2}-\epsilon}\right) \quad (26)$$

This implies

$$\sup\left|\int_{t_{i-1}}^{t_i} [\mu_u - \mu_{u-1}] du\right| = O_p\left(\Delta t^{\frac{3}{2}-\epsilon}\right) \quad (27)$$

Similarly, applying Burkholder's inequality (Protter 2004),† we have

$$\sup|\int_{t_{i-1}}^{t_i}[\sigma_u-\sigma_{u-1}]dW(u)| = O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right) \quad (28)$$

where $\delta$ can be any number in $0<\delta<3/2$. Therefore over the window, for $t\in[t_{i-1},t_i]$, $d\log S_t$ can be approximated by $d\log S_t^i$, such that

$$d\log S_t^i = \mu_{t_{i-1}}dt+\sigma_{t_{i-1}}dW(t) \quad (29)$$

because

$$\begin{aligned}&|(\log S_{t_i}-\log S_{t_{i-1}})-(\log S_{t_i}^i-\log S_{t_{i-1}}^i)|\\ &=|\int_{t_{i-1}}^{t_i}\left(\mu_u-\mu_{t_{i-1}}\right)du+\int_{t_{i-1}}^{t_i}\left(\sigma_u-\sigma_{t_{i-1}}\right)dW(u)|\\ &=O_P\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right)\end{aligned} \quad (30)$$

Therefore, for all $i$, $j$ and $t_j\in[t_{i-1},t_i]$, the numerator is

$$\begin{aligned}\log S_{t_j}-\log S_{t_{j-1}} &= \log S_{t_j}^i-\log S_{t_{j-1}}^i+O_P\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right)\\ &=\sigma_{t_{i-1}}W_{\Delta t}+O_P\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right)\\ &=\sigma_{t_{i-1}}\sqrt{\Delta t}U_i+O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right)\end{aligned} \quad (31)$$

where $W_{\Delta t}=W_{t_i}-W_{t_{i-1}}$, $U_i=\frac{1}{\sqrt{\Delta t}}(W_{t_i}-W_{t_{i-1}})$, which is an independently identically distributed ($iid$) normal.

Hence, $P_{1,t_i}$ is a linear combination of $iid$ normal variables $U_i$s. Therefore, $P_{1,t_i}$ is normally distributed with mean 0 and variance

$$\sigma_{P,t_i}=\sqrt{\sum_{l=0}^{L-2}\left(\sum_{j=0}^{l}h_j\right)^2\sigma_{t_{i-l-1}}^2\Delta t} \quad (32)$$

And it is $\sigma\sqrt{\sum_{l=0}^{L-2}\left(\sum_{j=0}^{l}h_j\right)^2\Delta t}$ if $\sigma_t=\sigma$. Formally,

$$P_{1,t_i}=\left(\sqrt{\sum_{l=0}^{L-2}\left(\sum_{j=0}^{l}h_j\right)^2}\right)\sigma\sqrt{\Delta t}\widehat{U}_i+O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right)$$

where $\widehat{U}_i=\frac{1}{\sqrt{\sum_{l=0}^{L-2}\left(\sum_{j=0}^{l}h_j\right)^2}}\sum_{l=0}^{L-2}\left(\sum_{j=0}^{l}h_j\right)(U_{i-l})$, which is standard normal but not independent anymore.

For the denominator, notice that increment of the numerator is not independent anymore, i.e. a direct application of Barndorff-Nielsen and Shephard (2004) is not applicable. For expositional purpose, let $r_{L-l-1}=\log S_{t_{i-l}}-\log S_{t_{i-l-1}}$ and $\widetilde{r}_{L-l-1}=\frac{r_{L-l-1}}{\sigma\Delta t}\sim iid$ N(0,1). Consequently, we have:

$$\begin{aligned}&E[BPV_G]\\ &=\sigma^2E|\sum_{l=0}^{L-2}\left(\sum_{j=0}^{l}h_j\right)(\widetilde{r}_{L-l-1})\,||\sum_{l=0}^{L-2}\left(\sum_{j=0}^{l}h_j\right)(\widetilde{r}_{L-l-2})\,|\end{aligned}$$

Conditional on $\widetilde{r}_1,\ldots,\widetilde{r}_{L-2}$ and we can get‡

$$\begin{aligned}&E[BPV_G|\widetilde{r}_1,\ldots,\widetilde{r}_{L-2}]\\ &=2\sigma^2 f\left(\frac{-\sum_{l=1}^{L-2}(\sum_{j=0}^{l}h_j)(\widetilde{r}_{L-l-1})}{h_0}\right)\\ &\quad\times f\left(\frac{-\sum_{l=0}^{L-3}(\sum_{j=0}^{l}h_j)(\widetilde{r}_{L-l-1})}{(\sum_{j=0}^{L-2}h_j)}\right)\end{aligned}$$

with

$$f(a)=2\phi(a)+2a\Phi(a)-a$$

where $a$ is a constant, $x\sim iid$ N(0,1) and $\phi(.)$ and $\Phi(.)$ denote the PDF and CDF of a standard normal. Notice that $\widetilde{r}_1,\ldots,\widetilde{r}_{L-2}$ are $iid$ standard normals, Thus,

$$\begin{aligned}&E[BPV_G]\\ &=\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}E[BPV_G|r_1,\ldots,r_{L-2}]\phi(r_1)\\ &\ldots\phi(r_{L-2})dr_1\ldots dr_{L-2}\end{aligned}$$

That is,

$$J_W(i)=\frac{\sigma\sqrt{\sum_{l=0}^{L-2}(\sum_{j=0}^{l}h_j)^2}}{\sigma\sqrt{E[BPV_G]}}U_i+O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right)$$

Let $d=\frac{\sqrt{E[BPV_G]}}{\sqrt{\sum_{l=0}^{L-2}(\sum_{j=0}^{l}h_j)^2}}$, we have

$$=\frac{U_i}{d}+O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right).$$

□

Proposition 2 (*Haar filter case*) *Assuming zero drift and A1 and A2 are satisfied. If there is no jump in* $(t_{i-1},t_i)$, *as* $\Delta t\to 0$,

$$\begin{aligned}&sup_i\ |J_W(i)-\widehat{J}_W(i)|=O_p(\Delta t^{\frac{3}{2}-\delta-\epsilon}),\\ &\text{where } \delta \text{ satisfies } 0<\delta<3/2 \text{ and } \widehat{J}_W(i)=\frac{U_i}{d}.\end{aligned} \quad (33)$$

*where* $W_t$ *follows a Brownian motion process. Here* $U_i=\frac{1}{\sqrt{\Delta t}}(W_{t_i}-W_{t_{i-1}})$, *a standard normal variable, and* $d=E[|U_i|]=\frac{\sqrt{2}}{\sqrt{\pi}}$ *is a constant.*

*Proof* This proof follows a similar line of reasoning to Theorem 1 and 1.1 in Lee and Mykland (2008). Using MODWT with the $Haar$ filter to transform the data, the wavelet coefficients at scale level 1 are

$$P_{1,t_i}=\frac{1}{2}\left(P_{t_i}-P_{t_{i-1}}\right)=\frac{1}{2}\left(\log S_{t_i}-\log S_{t_{i-1}}\right). \quad (34)$$

That is, our test statistic is of the form

$$J_W(i)=\frac{\frac{1}{2}(\log S_{t_i}-\log S_{t_{i-1}})}{\frac{1}{2}\widehat{\sigma}_{t_{i-1}}} \quad (35)$$

†This inequality gives bounds for the maximum of a martingale in terms of the quadratic variation. For a local martingale $M$ starting at zero, with maximum denoted by $M_t^*=sup_{s\le t}|M_s|$, and for any real number $p\ge 1$, the inequality is $c_pE([M]_t^{\frac{p}{2}})\le E((M_t^*)^p)\le C_pE([M]_t^{\frac{p}{2}})$, where $c_p\le C_p$ are constant depending on the choice of $p$, but not depending on the martingale $M$ and $t$ used. If $M$ is continuous martingale, then this inequality holds for any $p$.

‡Given a normally distributed random variable $X$ with mean $\mu$ and variance $\sigma^2$, the random variable $Y=|X|$ has a folded normal distribution. For further reference, please see Leone *et al.* (1961).

which is equivalent to the Lee and Mykland (2008) test with the window size of one data point prior to the location to be detected. The numerator, under the no-drift assumption, is

$$\log S_{t_i} - \log S_{t_{i-1}} = \int_{t_{i-1}}^{t_i} \sigma_u dW(u). \tag{36}$$

Applying Burkholder's inequality (Protter 2004), we have

$$\sup| \int_{t_{i-1}}^{t_i} [\sigma_u - \sigma_{u-1}] dW(u)| = O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right) \tag{37}$$

where $\delta$ can be any number in $0 < \delta < 3/2$. Therefore, for $t \in [t_{i-1}, t_i]$, $d \log S_t$ can be approximated by $d \log S_t^i$, such that

$$d \log S_t^i = \sigma_{t_{i-1}} dW(t) \tag{38}$$

because

$$\begin{aligned} &|(\log S_{t_i} - \log S_{t_{i-1}}) - (\log S_{t_i}^i - \log S_{t_{i-1}}^i)| \\ &= \left| \int_{t_{i-1}}^{t_i} (\sigma_u - \sigma_{t_{i-1}}) dW(u) \right| \\ &= O_P\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right). \end{aligned} \tag{39}$$

Therefore, for all $i$, $j$ and $t_j \in [t_{i-1}, t_i]$, the numerator of the test statistic is

$$\begin{aligned} \log S_{t_j} - \log S_{t_{j-1}} &= \log S_{t_j}^i - \log S_{t_{j-1}}^i + O_P\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right) \\ &= \sigma_{t_{i-1}} W_{\Delta t} + O_P\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right) \\ &= \sigma_{t_{i-1}} \sqrt{\Delta t} U_i + O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right) \end{aligned} \tag{40}$$

where $W_{\Delta t} = W_{t_i} - W_{t_{i-1}}, U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, which is an independently identically distributed normal random variable. Notice the second equality follows directly from equation (38) and third equality follows directly from our definition. For the denominator,

$$\begin{aligned} &\frac{1}{i-2} \sum_{k=2}^{i-1} |P_{1,t_k}||P_{1,t_{k-1}}| \\ &= \frac{1}{4} \frac{1}{(i-2)} \sum_{k=2}^{i-1} |\log S_{t_k} - \log S_{t_{k-1}}||\log S_{t_{k-1}} - \log S_{t_{k-2}}| \end{aligned} \tag{41}$$

Following Barndorff-Nielsen and Shephard (2004), we have

$$\begin{aligned} &plim_{\Delta t \to 0} c^2 \widehat{\sigma}_{t_{i-1}}^2 \\ &= plim_{\Delta t \to 0} \frac{1}{(i-2)\Delta t} \sum_{k=2}^{i-1} \\ &\quad |\log S_{t_k} - \log S_{t_{k-1}}||\log S_{t_{k-1}} - \log S_{t_{k-2}}| \\ &= d^2 \sigma_{t_{i-1}}^2 \end{aligned} \tag{42}$$

where $d = E[|U_i|] = \frac{\sqrt{2}}{\sqrt{\pi}}$. Hence

$$plim_{\Delta t \to 0} \frac{1}{(i-2)\Delta t} \sum_{k=2}^{i-1} |P_{1,t_k}||P_{1,t_{k-1}}| = \frac{1}{4} d^2 \sigma_{t_{i-1}}^2 \tag{43}$$

Therefore, the test statistics is

$$J_W(i) = \frac{U_i}{d} + O_p(\Delta t^{\frac{3}{2}-\delta-\epsilon}). \tag{44}$$

□

Proposition 3 ($D4$ filter) *Assuming zero drift and constant spot volatility in the underlying price process and assuming A1 and A2 are satisfied, if there is no jump in $(t_{i-1}, t_i)$, as $\Delta t \to 0$*

$$\begin{aligned} &sup_i \ |J_W(i) - \widehat{J}_W(i)| = O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right), \\ &\text{where } \delta \text{ satisfies } 0 < \delta < 3/2 \text{ and } \widehat{J}_W(i) = \frac{U_i}{d}. \end{aligned} \tag{45}$$

*where $W_t$ follows a Brownian motion process. Here $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, a standard normal variable, and a constant $d \approx \frac{\sqrt{2}}{\sqrt{3}}$.*

*Proof* Assuming zero drift and using MODWT with the $D(4)$ filter to transform the data, the wavelet coefficients at scale 1 are

$$\begin{aligned} P_{1,t_i} &= \frac{1}{8}((1-\sqrt{3})P_{t_i} + (\sqrt{3}-3)P_{t_{i-1}} + (3+\sqrt{3})P_{t_{i-2}} \\ &\quad + (-1-\sqrt{3})P_{t_{i-3}}) \\ &= \frac{1}{8}((1-\sqrt{3})(P_{t_i} - P_{t_{i-1}}) + (1+\sqrt{3})(P_{t_{i-2}} - P_{t_{i-3}}) \\ &\quad - 2(P_{t_{i-1}} - P_{t_{i-2}})) \\ &= \frac{1}{8}((1-\sqrt{3})(\log S_{t_i} - \log S_{t_{i-1}}) + (1+\sqrt{3}) \\ &\quad (\log S_{t_{i-2}} - \log S_{t_{i-3}}) - 2(\log S_{t_{i-1}} - \log S_{t_{i-2}})). \end{aligned} \tag{46}$$

Hence, we have

$$\begin{aligned} \log S_{t_i} - \log S_{t_{i-1}} &= \int_{t_{i-1}}^{t_i} \sigma_u dW(u) \\ \log S_{t_{i-1}} - \log S_{t_{i-2}} &= \int_{t_{i-2}}^{t_{i-1}} \sigma_u dW(u) \\ \log S_{t_{i-2}} - \log S_{t_{i-3}} &= \int_{t_{i-3}}^{t_{i-2}} \sigma_u dW(u). \end{aligned} \tag{47}$$

From Burkholder's inequality, we have

$$\sup| \int_{t_{i-1}}^{t_i} [\sigma_u - \sigma_{u-1}] dW(u)| = O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right) \tag{48}$$

Using the results in the proof of Theorem 2, we have

$$\begin{aligned} \log S_{t_i} - \log S_{t_{i-1}} &= \sigma_{t_{i-1}} \sqrt{\Delta t} U_i + O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right) \\ \log S_{t_{i-1}} - \log S_{t_{i-2}} &= \sigma_{t_{i-2}} \sqrt{\Delta t} U_{i-1} + O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right) \\ \log S_{t_{i-2}} - \log S_{t_{i-3}} &= \sigma_{t_{i-3}} \sqrt{\Delta t} U_{i-2} + O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right) \end{aligned} \tag{49}$$

where $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, which is an independently identically distributed normal random variable.

Hence, $P_{1,t_i}$ is approximated by a linear combination of $iid$ normal variables $U_i$s. Therefore, $P_{1,t_i}$ is normally distributed with mean 0 and standard deviation $\sigma_{P,t_i} = \frac{\sqrt{3}}{4}\sigma\sqrt{\Delta t}$ when

$\sigma_{t_i} = \sigma$.† Formally,

$$P_{1,t_i} = \frac{\sqrt{3}}{4}\sigma\sqrt{\Delta t}\widehat{U}_i + O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right)$$

where $\widehat{U}_i = \frac{1}{\sqrt{12}}((1-\sqrt{3})U_i + (1+\sqrt{3})U_{i-2} - 2U_{i-1})$, which is standard normal variable but is no longer independent.

As noted by Jiang and Oomen (2008), with such dependence, BPV is still robust to the presence of jumps but is biased. If we know the explicit form of the bias, we can conduct bias correction as in Jiang and Oomen (2008). For expositional purpose, let us assume constant variance.

$$\begin{aligned}
&E[BPV_{D4}] \\
&= \frac{1}{64}E|(1-\sqrt{3})(\log S_{t_i} - \log S_{t_{i-1}}) \\
&\quad + (1+\sqrt{3})(\log S_{t_{i-2}} - \log S_{t_{i-3}}) \\
&\quad - 2(\log S_{t_{i-1}} - \log S_{t_{i-2}})| \\
&\quad |(1-\sqrt{3})(\log S_{t_{i-1}} - \log S_{t_{i-2}}) \\
&\quad + (1+\sqrt{3})(\log S_{t_{i-3}} - \log S_{t_{i-4}}) \\
&\quad - 2(\log S_{t_{i-2}} - \log S_{t_{i-3}})|
\end{aligned}$$

Following Jiang and Oomen (2008), let $r_j = \log S_{t_{i+j-3}} - \log S_{t_{i+j-4}}$ and $\widetilde{r}_j = \frac{r_j}{\sigma\sqrt{\Delta t}} \sim iid\,\mathrm{N}(0,1)$. Consequently, we have:

$$\begin{aligned}
&E[BPV_{D4}] \\
&= \frac{\sigma^2}{64}E|(1-\sqrt{3})\widetilde{r}_3 + (1+\sqrt{3})\widetilde{r}_1 - 2\widetilde{r}_2||(1-\sqrt{3})\widetilde{r}_2 \\
&\quad + (1+\sqrt{3})\widetilde{r}_0 - 2\widetilde{r}_1|
\end{aligned}$$

Conditional on $\widetilde{r}_2$ and $\widetilde{r}_1$, $E[BPV_{D4}]$ is the expectation of a product of two independent variables so it is equal to the product of the expectations. Therefore, we can get

$$\begin{aligned}
&E[BPV_{D4}|\widetilde{r}_1, \widetilde{r}_2] \\
&= \frac{\sigma^2}{64}E|(1-\sqrt{3})\widetilde{r}_3 + (1+\sqrt{3})\widetilde{r}_1 \\
&\quad - 2\widetilde{r}_2|\widetilde{r}_1, \widetilde{r}_2|E|(1-\sqrt{3})\widetilde{r}_2 + (1+\sqrt{3})\widetilde{r}_0 - 2\widetilde{r}_1|\widetilde{r}_1, \widetilde{r}_2| \\
&= \frac{\sigma^2}{64}(\sqrt{3}+1)(\sqrt{3}-1)|\widetilde{r}_3 - \frac{2\widetilde{r}_2 - (1+\sqrt{3})\widetilde{r}_1}{1-\sqrt{3}}|\widetilde{r}_1, \widetilde{r}_2||\widetilde{r}_0 \\
&\quad - \frac{2\widetilde{r}_1 - (1-\sqrt{3})\widetilde{r}_2)}{1+\sqrt{3}}|\widetilde{r}_1, \widetilde{r}_2| \\
&= \frac{\sigma^2}{32}f\left(\frac{2\widetilde{r}_2 - (1+\sqrt{3})\widetilde{r}_1}{1-\sqrt{3}}\right)f\left(\frac{2\widetilde{r}_1 - (1-\sqrt{3})\widetilde{r}_2}{\sqrt{3}+1}\right)
\end{aligned}$$

with

$$f(a) = E[|x-a|] = 2\phi(a) + 2a\Phi(a) - a$$

where $a$ is a constant, $x \sim iid\,\mathrm{N}(0,1)$ and $\phi(.)$ and $\Phi(.)$ denote the PDF and CDF of a standard normal. Notice $f(.)$ is a direct application of the expectation of a folded normal formula. Notice that $\widetilde{r}_2$ and $\widetilde{r}_1$ are $iid$ standard normal variables. Thus, we calculate the unconditional expectation by integrating the conditional expectation over the bi-variate normal distribution of $\widetilde{r}_2$ and $\widetilde{r}_1$. That is,

$$E[BPV_{D4}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[BPV_{D4}|x, y]\phi(x)\phi(y)dxdy$$

Therefore,

$$E[BPV_{D4}] \approx \frac{\sigma^2}{8}$$

Since

$$\begin{aligned}
plim_{\Delta t\to 0}BPV_{D4} &= plim_{\Delta t\to 0}\frac{1}{(i-2)\Delta t}\sum_{k=2}^{i-1}|P_{1,t_k}||P_{1,t_{k-1}}| \\
&= E[BPV_{D4}] \approx \frac{\sigma^2}{8}
\end{aligned}$$

That is,

$$J_W(i) = \frac{P_{1,t}}{\sqrt{BPV_{D4}}} = \frac{\sqrt{3}/4\sigma}{1/(2\sqrt{2})\sigma}U_i + O_p\left(\Delta^{\frac{3}{2}-\delta-\epsilon}\right)$$

Let $d = \frac{1/(2\sqrt{2})\sigma}{\sqrt{3}/4\sigma} = \frac{\sqrt{2}}{\sqrt{3}}$, we have

$$= \frac{U_i}{d} + O_p\left(\Delta t^{\frac{3}{2}-\delta-\epsilon}\right).$$

□

†To see this, notice that

$$\begin{aligned}
&P_{1,t_i} \\
&= \frac{1}{8}((1-\sqrt{3})(\log S_{t_i} - \log S_{t_{i-1}}) \\
&\quad + (1+\sqrt{3})(\log S_{t_{i-2}} - \log S_{t_{i-3}}) \\
&\quad - 2(\log S_{t_{i-1}} - \log S_{t_{i-2}})) \\
&= \frac{1}{8}((1-\sqrt{3})(\sigma\sqrt{\Delta t}U_i) \\
&\quad + (1+\sqrt{3})(\sigma\sqrt{\Delta t}U_{i-2}) \\
&\quad - 2(\sigma\sqrt{\Delta t}U_{i-1})).
\end{aligned}$$

Because $U_i$'s are $iid$,

$$\begin{aligned}
&\mathrm{Var}(P_{1,t_i}) \\
&= \frac{1}{64}((4-2\sqrt{3})\sigma^2\Delta t + (4+2\sqrt{3})\sigma^2\Delta t + 4\sigma^2\Delta t)) \\
&= \frac{3}{16}\sigma^2\Delta t.
\end{aligned}$$