Volatility estimation and jump detection for drift–diffusion processes

Sébastien Laurent, Shuping Shi

Volatility Estimation and Jump Detection for Drift-diffusion Processes*

Sébastien Laurent[1] and Shuping Shi[2]

[1]Aix-Marseille University
[2]Macquarie University

September 4, 2019

Abstract

The logarithmic prices of financial assets are conventionally assumed to follow a drift-diffusion process. While the drift term is typically ignored in the infill asymptotic theory and applications, the presence of temporary nonzero drifts is an undeniable fact. The finite sample theory for integrated variance estimators and extensive simulations provided in this paper reveal that the drift component has a nonnegligible impact on the estimation accuracy of volatility, which leads to a dramatic power loss for a class of jump identification procedures. We propose an alternative construction of volatility estimators and observe significant improvement in the estimation accuracy in the presence of nonnegligible drift. The analytical formulas of the finite sample bias of the realized variance, bipower variation, and their modified versions take simple and intuitive forms. The new jump tests, which are constructed from the modified volatility estimators, show satisfactory performance. As an illustration, we apply the new volatility estimators and jump tests, along with their original versions, to 21 years of 5-minute log returns of the NASDAQ stock price index.

*The authors gratefully acknowledge the editors of this special issue, two anonymous referees, Yacine Ait-Sahalia, Jihyun Kim, Nour Meddahi, Peter B.C. Phillips, Roberto Renò, Orimar Sauri, Jun Yu, Yichong Zhang, Colin Bowers, and participants at the econometric study group at Singapore Management University, the Frontiers in Econometrics workshop at Macquarie University, and the 2019 Toulouse Financial Econometrics conference for helpful discussions. Shi acknowledges research support from the Australian Research Council under project No. DE190100840. Sébastien acknowledges research support by the French National Research Agency Grant ANR-17-EURE-0020. Sébastien Laurent, Aix-Marseille School of Economics, Aix-Marseille University; E-mail: sebastien.laurent@univ-amu.fr. Shuping Shi, Department of Economics, Macquarie University; E-mail: shuping.shi@mq.edu.au.

Keywords: Diffusion process, nonzero drift, finite sample theory, volatility estimation, jumps

JEL classification: C12, C14.

1 Introduction

Consider the conventional setup of an Ito semimartingale process of log prices y_t such that

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

where W_t is an $\mathcal{F}_t$-adapted standard Brownian motion, with $\{\mathcal{F}_t : t \in [0, T]\}$ being a right-continuous information filtration. With locally bounded coefficients μ_t and σ_t (as in the volatility estimation and jump detection literature), the drift term is dominated by the diffusion process. For this reason, most infill asymptotics are unaffected by the presence of a drift. The drift component is therefore typically ignored in the high-frequency literature.

There is, however, substantial empirical evidence documenting that asset prices might have a nonzero drift component. In the low-frequency (such as monthly, weekly and daily) framework, prolonged periods of mildly explosive trends have been identified in many financial assets, e.g., the stock market during the dot-com bubble period in the late 1990s (Phillips et al., 2011, 2015; Shi and Song, 2016) and the commodity markets over the preceding decade (Etienne et al., 2014; Gutierrez, 2012; Phillips and Yu, 2011). Figure 1 displays the daily median[1] of the 5-minute log returns of the NASDAQ stock market index for the period 1996–2016. The daily median of the 5-minute log returns is observed to deviate from zero for a substantial period of time in the early 2000s when the dot-com bubble burst and around the subprime mortgage crisis period.

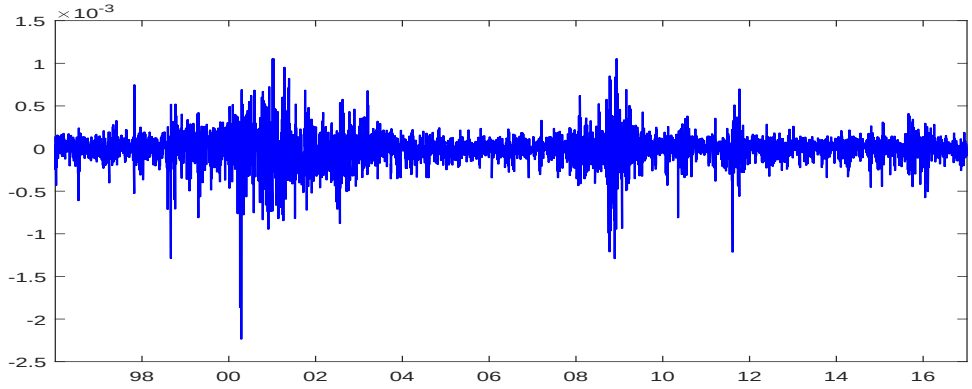
Evidence of a nonnegligible drift has also been observed by Phillips and Shi (2017) in the log prices of the S&P 500 index during the 2008 subprime mortgage crisis period and in the bond yields and CDS spreads of most European countries during the 2010 debt crisis. Additionally, there is extensive literature documenting the temporary deviations in log prices from the random walk.[2] In the high-frequency framework, motivated by the large number of flash crashes,[3] Christensen

[1] We calculate the median and not the sample mean to reduce the impact of jumps in asset prices.

[2] See, for example, Bekaert and Hodrick (1992); Bessembinder and Chan (1992); Campbell and Ammer (1993); Campbell and Hamao (1992); Lo and MacKinlay (1988); Fama and French (1988); Balvers et al. (2000); Chaudhuri and Wu (2003).

[3] See, for example, Nanex Research: `http://www.nanex.net/NxResearch/`.

Figure 1: Daily median of the 5-minute log returns of the NASDAQ stock index during 1996 - 2016.



et al. (2016) propose a drift burst hypothesis, which postulates the existence of short-lived locally explosive trends in log prices. This hypothesis was tested using a nonparametric approach and applied to tick-by-tick data. The authors observe that drift bursts (especially negative drifts) form an integral part of price dynamics in equities, fixed income, currencies and commodities.

Our paper investigates the finite sample impact of nonzero drifts on the estimation accuracy of volatility in the high-frequency setting. The literature on volatility estimation focuses on the asymptotic properties and, apart from some notable exceptions (Meddahi, 2002; Bandi and Russell, 2005), does not study its finite sample properties. We derive the analytical formulas of the bias of the realized variance (Andersen and Bollerslev, 1998) and bipower variation (Barndorff-Nielsen and Shephard, 2004) under a constant drift-diffusion process and a linear drift-diffusion process. The two processes capture the important forms of nonzero drift and have been studied extensively in the literature.[4] We also investigate via Monte Carlo simulations the impact of a nonzero drift on a noise-robust volatility estimator (Podolskij and Vetter, 2009). The finite sample theory, together with extensive simulations, reveals that a nonzero drift causes a substantial bias in volatility estimation. It is important to highlight that the integrated variance estimators are typically computed over a short period of time (e.g., one day), and therefore, we do not mean by nonzero drift that the drift needs to be nonzero over a long span (e.g., one month or one year) to cause a bias in these estimators, but rather it deviates from zero during the period over which it is computed.

We propose computing volatility estimators on centered log returns instead of raw log returns.

[4]See, for example, Lo and Wang (1995); Barndorff-Nielsen and Shephard (2001); Nicolato and Venardos (2003); Aalen and Gjessing (2004); Zhou and Yu (2015); Wang and Yu (2016).

For centering, we consider either the sample mean or the median. The former has well-known finite sample distributions but is sensitive to outliers (or additive jumps in our context), while the latter has the advantage of being robust to outliers, but little is known about its finite sample distribution. We derive the analytical bias of the modified realized variance estimators under both drift-diffusion processes and provide an analytical expression for the bias of the modified bipower volatility estimators under the constant drift-diffusion process.[5] The theoretical and simulation results suggest that both modifications lead to a dramatic improvement in the estimation accuracy of volatility, especially when the drift deviates far from zero. In particular, estimators relying on centered log returns (relative to the median) have the overall best performance in the presence of jumps.

Volatility plays a central role in finance. It is the most important type of market risk and is fundamental to asset pricing, portfolio choice, and financial market regulation. Volatility estimators have been used extensively in this field. The finite sample bias of volatility estimation arising from the presence of a nonzero drift is therefore expected to have many secondary impacts. As an example, we demonstrate in this paper that the finite sample bias of the volatility estimators could lead to the unsatisfactory performance of jump detection procedures. We show that in the presence of a nonzero drift, the Lee and Mykland (2008, LM08 hereafter) and Lee and Mykland (2012, LM12 hereafter) tests are severely undersized, which translates into a dramatic loss of power. To address this finite sample problem, we propose an alternative construction of the test statistic, which relies on the proposed modified volatility estimators. Despite its ease of implementation, our test improves the finite sample performance significantly.

As an illustration, we apply the new bipower variation estimator and the new LM08 jump test, along with their original versions, to 5-minute log returns of the NASDAQ stock index from 1996 to 2016. The main conclusion is that the bipower variation tends to overestimate the volatility in the presence of nonzero drift by, on average, 2.5% but, sometimes, by as much as 40% or even more. The proposed new jump test allows for the identification of more jumps. These additional jumps occur during periods with upward or downward trends in log prices which are likely due to

[5]The derivation for the modified bipower variations under the linear drift-diffusion process is very complicated, as it involves deriving the first moment of the absolute value of the product of two correlated variables and the distributions of order statistics (i.e., median) for nonidentically distributed and dependent variables. Such derivation is therefore left for future research.

the presence of nonzero drift.

The paper is organized as follows. The theoretical results on the impact of a constant drift on the finite sample bias of realized variance and bipower variation computed on raw and centered log returns are derived in Section 2, while the case of a linear drift-diffusion process is considered in Section 3. Section 4 illustrates the impact of the drift on a noise-robust volatility estimator. Section 5 studies the finite sample performance of two popular jump tests and the modified jump tests in the presence of a nonzero drift. An application is proposed in Section 6. Section 7 concludes. All the proofs are reported in Appendix A. Appendix B compares the estimation accuracy of the sample mean and the median in the presence of jumps.

2 Volatility Estimation under a Constant Drift-Diffusion Process

Let $\{0 < t_1 < \cdots < t_T < N\}$ be a set of T equally spaced observation times spanning N days. The distance between two consecutive observation times is denoted by $\Delta = t_i - t_{i-1} = N/T$. For expositional purpose, we first consider a simple data generating process, where both the drift and diffusion coefficients are constant, i.e.,

$$dy_t = \mu dt + \sigma dW_t. \tag{2}$$

The log returns $r_{t_i} = y_{t_i} - y_{t_{i-1}}$ (computed using log prices at equally spaced observation times) can be written as

$$r_{t_i} = \mu\Delta + \sigma\sqrt{\Delta}\varepsilon_{t_i} \text{ with } \varepsilon_{t_i} \sim N(0,1). \tag{3}$$

It is obvious that the log return process is asymptotically dominated by the volatility component $\sigma\sqrt{\Delta}\varepsilon_{t_i}$. In other words, the drift term is asymptotically negligible, and as shown by Barndorff-Nielsen and Shephard (2002), $r_{t_i}^2 \to \int_{t_{i-1}}^{t_i} \sigma^2 ds$. However, in practice, one very often has to rely on low-frequency data for estimations and hypothesis testing because ultrahigh-frequency data for asset prices are not always available or because lower-frequency data (such as 5- or 10-minute data) are preferred to mitigate the impact of microstructure noise (Park and Linton, 2011).

2.1 Realized Variance Estimator

The realized variance estimator is defined as

$$RV_{t_i}(K) = \sum_{j=i-K+1}^{i} r_{t_j}^2,$$

where K is the number of observations included in the estimation. See Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002), etc.

Proposition 2.1 *Under the drift-diffusion process (2), the bias of the realized variance is*

$$\mathbb{E}\left[RV_{t_i}(K) - K\sigma^2\Delta\right] = K\mu^2\Delta^2.$$

The bias of the realized variance estimator is zero if the drift coefficient is zero (i.e., $\mu = 0$). In practice, parameter K is often set to the total number of intraday observations available at the end of each trading day (i.e., $1/\Delta$). See Andersen et al. (2001, 2003); Boudt et al. (2011); Bauwens et al. (2012), etc. Thus, the bias of the daily realized volatility computed at the end of the day is $\mu^2\Delta$, which increases with the magnitude of μ and decreases with the sampling frequency Δ.

One natural way to mitigate this bias is to remove the drift component before calculating volatilities or, equivalently, to compute the volatility estimators on centered log returns. We consider two different ways of centering, using either the sample mean or the median of K log returns involved in the computation of realized variance, i.e.,

$$\hat{m}_{t_i}(K) = \frac{1}{K}\sum_{j=i-K+1}^{i} r_{t_j}, \qquad \text{and} \qquad \hat{m}^*_{t_i}(K) = median\left(r_{t_{i-K+1}}, \cdots, r_{t_i}\right).$$

Lemma 2.1 *Under the drift-diffusion process (2), both the sample mean and median are unbiased estimators of* $\mathbb{E}(r_{t_i})$*, i.e.,*

$$\mathbb{E}\left(\hat{m}_{t_i}(K)\right) = \mathbb{E}\left(\hat{m}^*_{t_i}(K)\right) = \mu\Delta. \tag{4}$$

While the proof of the unbiasedness of the sample mean is trivial, it is less straightforward for the median. The calculation of the median is equivalent to running a quantile regression (for the 50^{th} quantile) with only a constant as a regressor. From the proof of Lemma 2.1 in Appendix A,

under the DGP (2), the finite sample distribution of the median is a beta-normal (BN) distribution (Eugene et al., 2002), i.e., $\hat{m}^*_{t_i}(K) \sim BN(p, K-p+1, \mu\Delta, \sigma\sqrt{\Delta})$, where $p = \lceil 0.5K \rceil$ is the smallest integer greater than $0.5K$.

The modified realized variance estimators are based on centered log returns and defined as

$$RV^{\dagger}_{t_i}(K) = \sum_{j=i-K+1}^{i} \left[r_{t_j} - \hat{m}_{t_i}(K)\right]^2 \qquad \text{and} \qquad RV^{*}_{t_i}(K) = \sum_{j=i-K+1}^{i} [r_{t_j} - \hat{m}^*_{t_i}(K)]^2.$$

Proposition 2.2 *Under the drift-diffusion process (2),*

(i) the bias of the modified realized variance $RV^{\dagger}$ is

$$\mathbb{E}\left[RV^{\dagger}_{t_i}(K) - K\sigma^2\Delta\right] = -K\,\mathbb{V}\left(\hat{m}\left(K\right)\right) = -\sigma^2\Delta, \tag{5}$$

where $\mathbb{V}(.)$ represents the variance of the argument;

(ii) the bias of the modified realized variance RV^{*} is

$$\mathbb{E}\left[RV^{*}_{t_i}(K) - K\sigma^2\Delta\right] = K\left[\mathbb{V}\left(\hat{m}^*\left(K\right)\right) - 2\,\mathrm{cov}\left(\hat{m}\left(K\right), \hat{m}^*\left(K\right)\right)\right], \tag{6}$$

where $\mathrm{cov}(.,.)$ is the covariance between the two arguments.

The finite sample biases of both RV and $RV^{\dagger}$ take very simple forms, i.e., $K\mu^2\Delta^2$ for RV, and $-\sigma^2\Delta$ for $RV^{\dagger}$. Note that if $K = 1/\Delta$, the bias of RV is $\mu^2\Delta$, which is of the same order of magnitude as that of $RV^{\dagger}$. For a given sampling frequency, the biases of $RV^{\dagger}$ and RV are positively related to, respectively, σ^2 and μ^2. It is obvious that the bias of RV will be larger than that of $RV^{\dagger}$ if $\mu > \sigma$ and will be smaller otherwise. We discuss empirically realistic settings of μ and σ in Section 2.3.

The bias of RV^{*} takes a slightly more complicated form. It depends on the variance of the median and the covariance between the sample mean and the median. Since $\hat{m}^*_{t_i}(K) \sim BN(p, K-p+1, \mu\Delta, \sigma\sqrt{\Delta})$, from Gupta and Nadarajah (2005), we have

$$\mathbb{V}\left(\hat{m}^*_{t_i}(K)\right) = \mu^2\Delta^2 K \binom{K-1}{p-1} \sum_{j=0}^{K-p} (-1)^j \binom{K-p}{j} \sum_{i=1}^{2} 2\left(\frac{\sigma}{\mu\sqrt{\Delta}}\right)^i$$

$$\times \left[\sum_{k=0}^{p+j-1} (-1)^k \binom{p+j-1}{k} I_{i,k} + (-1)^i I_{i,p+j-1} \right],$$

with $I_{i,k} = \int_0^\infty \nu^i \phi(\nu) [1 - \Phi(\nu)]^k d\nu$. The finite sample covariance between the sample mean and the median (i.e., $\hat{m}(K)$ and $\hat{m}^*_{t_i}(K)$) is unknown but can be approximated by its asymptotic counterpart. See, for example, Ferguson (1999) for the joint asymptotic distribution of $\hat{m}(K)$ and $\hat{m}^*_{t_i}(K)$.

2.2 Bipower Variation

The bipower variation is probably the most popular jump-robust estimator of the integrated variance. It is defined as

$$BV_{t_i}(K) = \frac{\pi}{2} \frac{K}{K-1} \sum_{j=i-K+2}^{i} |r_{t_j}||r_{t_{j-1}}|. \tag{7}$$

Despite its well-behaved asymptotic properties, we demonstrate below that, similar to the case of the realized variance, the finite sample performance of the bipower variation is unsatisfactory in the presence of a large nonzero drift.

We propose an alternative construction of the bipower variation based on centered log returns (using either the sample mean or the median). The new estimators of the integrated variance, denoted by $BV^\dagger_{t_i}(K)$ and $BV^*_{t_i}(K)$, are defined as follows:

$$\begin{aligned} BV^\dagger_{t_i}(K) &= \frac{\pi}{2} \frac{K}{K-1} \sum_{j=i-K+2}^{i} |r_{t_j} - \hat{m}_{t_i}(K)||r_{t_j} - \hat{m}_{t_i}(K)|, \\ BV^*_{t_i}(K) &= \frac{\pi}{2} \frac{K}{K-1} \sum_{j=i-K+2}^{i} |r_{t_j} - \hat{m}^*_{t_i}(K)||r_{t_j} - \hat{m}^*_{t_i}(K)|. \end{aligned}$$

Since the drift component of log returns is asymptotically dominated by the variance, using the centered log returns for the calculation of the bipower variation will not alter its limiting property but should improve its finite sample properties, as is the case for the modified realized variance estimators $RV^\dagger$ and RV^*.

Proposition 2.3 *Under the drift-diffusion process (2),*

(i) the bias of the bipower variation BV is

$$\begin{aligned}\mathbb{E}\left[BV_{t_i}(K) - K\sigma^2\Delta\right] &= K\sigma^2\Delta\left(e^{-\frac{\mu^2\Delta}{\sigma^2}} - 1\right) + \frac{\pi}{2}K\mu^2\Delta^2\left[1 - 2\Phi\left(-\frac{\mu\sqrt{\Delta}}{\sigma}\right)\right]^2 \\ &\quad + 2K\mu\Delta^{3/2}\sigma\sqrt{\frac{\pi}{2}}e^{-\frac{\mu^2\Delta}{2\sigma^2}}\left[1 - 2\Phi\left(-\frac{\mu\sqrt{\Delta}}{\sigma}\right)\right];\end{aligned}$$

(ii) the bias of the modified bipower variation $BV^\dagger$ is

$$\mathbb{E}\left[BV^\dagger_{t_i}(K) - K\sigma^2\Delta\right] \leq \left(\frac{\pi}{2} - 1\right)K\sigma^2\Delta - \frac{\pi}{2}\sigma^2\Delta;$$

(iii) the bias of the modified bipower variation BV^* is

$$\mathbb{E}\left[BV^*_{t_i}(K) - K\sigma^2\Delta\right] \leq \left(\frac{\pi}{2} - 1\right)K\sigma^2\Delta + \frac{\pi}{2}K\left[\mathbb{V}\left(\hat{m}^*\left(K\right)\right) - 2\,\mathrm{cov}\left(\hat{m}\left(K\right), \hat{m}^*\left(K\right)\right)\right].$$

The magnitude of the bias of BV is related to values of K, σ, Δ and μ. If $\mu = 0$, the bias of BV is zero. In contrast, the biases of $BV^\dagger$ and BV^* are bounded above, where the upper bounds do not depend on μ.

2.3 Visualization of the Bias

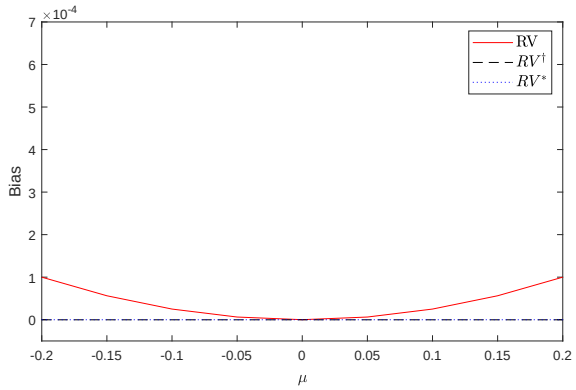
For comparison, we compute the biases of the volatility estimators in this subsection. We set $K = 1/\Delta$, so that volatility is computed over one day. We allow μ to vary from -0.2 to 0.2 with an increment of 0.05. The daily unconditional mean of log returns is thus between $\pm 2.5 \times 10^{-3}$ at the 5-minute frequency, which aligns with the order of magnitude of the daily median of 5-minute NASDAQ stock returns (see Figure 1). As mentioned above, we do not assume μ to be nonzero over a long time span but do so only during the period over which the volatility estimators are computed. The diffusion coefficient σ is set to 0.01, which implies an annualized variance of $10^{-4} \times 252 = 0.152^2$ (a reasonable value for equity returns).

The biases of the realized variance estimators (RV, $RV^\dagger$, and RV^*) are computed from Proposition 2.1 and 2.2, with the quantities $\mathbb{V}(\hat{m}(K))$, $\mathbb{V}(\hat{m}^*(K))$, and $\mathrm{cov}(\hat{m}(K), \hat{m}^*(K))$ obtained from

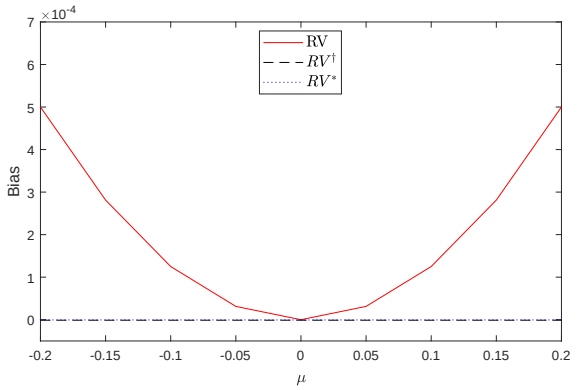
Monte Carlo simulations.[6] Since we do not have an exact formula for the bias of the modified bipower variations but only an upper bound, we obtain them by Monte Carlo simulations. For all simulations conducted in this paper, we assume that the asset is traded 6.5 hours per day as is the case on the NASDAQ stock exchange (from 9:30am to 16:00pm). That is, there are $24,000$ observations over one day at the one-second frequency. We pick one observation every 60 (300) data points to obtain the 1-minute log prices (5-minute log prices). The sampling interval is set to $\Delta = 1/400$ and $\Delta = 1/80$ for the 1- and 5-minute data, respectively. The simulations are repeated for $10,000$ times. In this simulations, log returns are assumed to follow Equation (3).

Figure 2: Bias of the realized variance estimators under the constant drift-diffusion process.

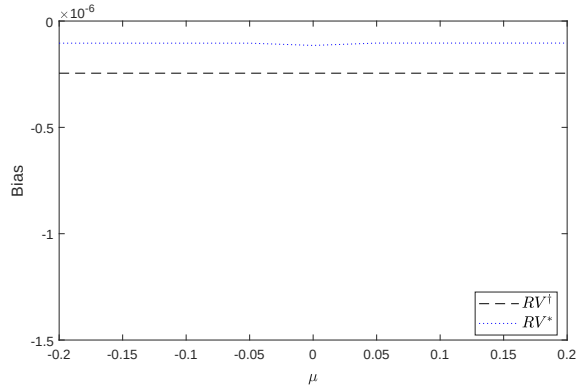
(a) Realized variances: 1-minute $\Delta = 1/400$



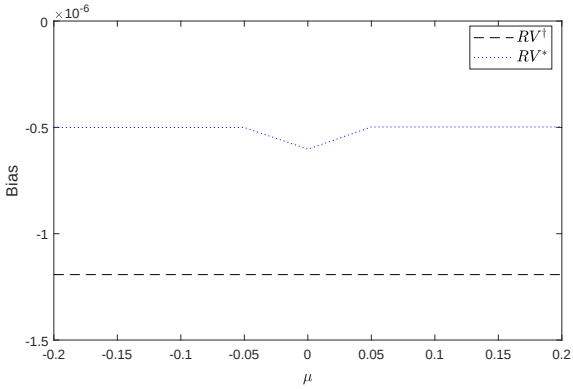
(b) Realized variances: 5-minute $\Delta = 1/80$



(c) Modified realized variances: 1-minute $\Delta = 1/400$



(d) Modified realized variances: 5-minute $\Delta = 1/80$

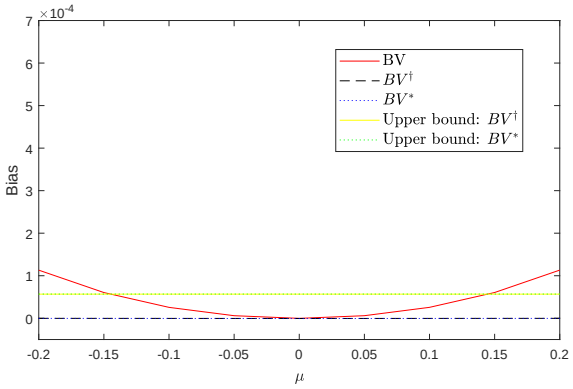


Figures 2 and 3 display the biases of RV, $RV^{\dagger}$ and RV^{*} and the biases of the two modified realized volatility estimators in the bottom panel for further comparison. Figure 3 presents the biases of the bipower variations (BV, $BV^{\dagger}$ and BV^{*}) obtained from Monte Carol simulations,

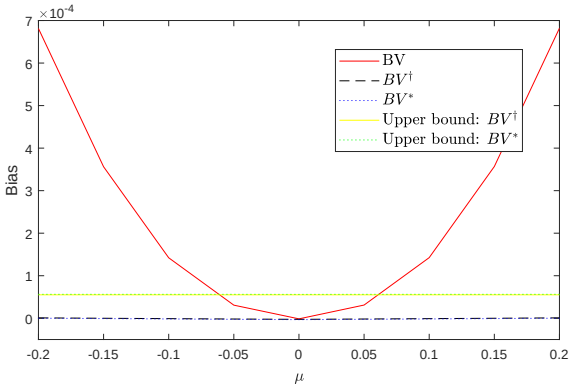
[6] Although the analytical forms of $\mathbb{V}(\hat{m}(K))$ and $\mathbb{V}(\hat{m}^{*}(K))$ are known, for fairness of comparison between $RV^{\dagger}$ and RV^{*}, we compute them by simulations.

Figure 3: Bias of the bipower variations under the constant drift-diffusion process.

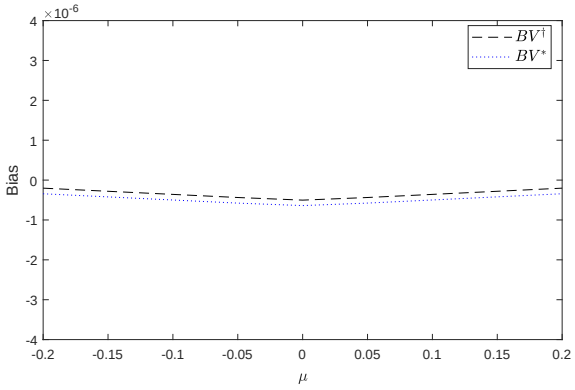
(a) Bipower variations: 1-minute $\Delta = 1/400$



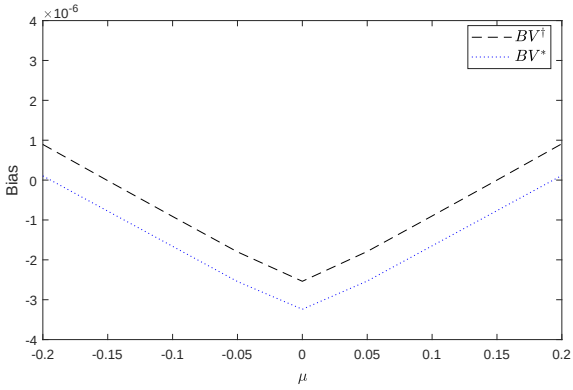
(b) Bipower variations: 5-minute $\Delta = 1/80$



(c) Modified bipower variations: 1-minute $\Delta = 1/400$



(d) Modified bipower variations: 5-minute $\Delta = 1/80$



along with the upper bounds of the modified bipower variations computed from Proposition 2.3, in the top panel. It is clear that the biases of the modified volatility estimators are much smaller than those of their respective original estimators in finite samples. The discrepancy becomes increasingly visible as the drift value deviates further away from zero. Additionally, the biases of the modified volatility estimators are extremely close to zero and indistinguishable in the top panels. One can see from the bottom panels that if $\Delta = 1/80$, the order of magnitude of the biases is 10^{-6} (compared to 10^{-4} for the original estimators). This suggests that although Propositions 2.2 and 2.3 show that the biases of the modified estimators are nonzero if there is a nonzero drift, they are negligible compared to those of the original volatility estimators. The modified realized volatility estimators are downward biased with the bias of $RV^{\dagger}$ being slightly larger than that of RV^{*}. The biases of the modified bipower variations increase from negative to positive values as μ derivates further from zero. The upper bounds are always above the actual values and far below the bias curve of BV

when μ is large.

3 Volatility Estimation under a Linear Drift-Diffusion Process

We consider an alternative drift-diffusion process, where the drift coefficient is a linear function of the log price, and the diffusion coefficient is time-varying:

$$dy_t = \theta(y_t - \rho)dt + \sigma_t dW_t, \tag{8}$$

where θ and ρ are constants and σ_t is an adapted and cádlág volatility process with $E\left(\sigma_t^2\right) = d_0$. The assumption on σ_t is very general. It includes a wide range of volatility models, e.g., the eigenfunction stochastic volatility process of Meddahi (2002), the log-normal model, the square root model, and the GARCH(1,1) model of Nelson (1991). It also allows for jumps and intraday periodical patterns in volatility. For simplicity, we assume the independence of σ_t and W_t and hence do not allow for the presence of leverage effect.

The exact discrete solution of (8) is (Arnold, 1974, Corollary 8.2.4)

$$y_{t_{i+1}} = g(\theta) + \alpha(\theta)\, y_{t_i} + \eta_{t_{i+1}}, \tag{9}$$

where $g(\theta) = \mu[1 - \exp(\theta\Delta)]$, $\alpha(\theta) = \exp(\theta\Delta)$, and $\eta_{t_{i+1}} = \int_{t_i}^{t_{i+1}} e^{\theta(t_i+\Delta-s)}\sigma_s dW_s$. The intercept $g(\theta)$ converges to zero at the rate of Δ. The autoregressive coefficient is $\alpha(\theta) = 1$ if $\theta = 0$; hence, the log price process has a random walk dynamic. If $\theta \neq 0$, we obtain that

$$\alpha(\theta) = \exp(\theta\Delta) = 1 + \theta\Delta + O\left(\Delta^2\right)$$

converges to unity as $\Delta \to 0$ at the rate of Δ. The order of magnitude of the autoregressive coefficient $O(\Delta)$ can be written as $O(1/T)$, given that $\Delta = N/T$ and N is a constant. Therefore, the dynamic in (9) is *local-to-unity* (Phillips, 1987) in the explosive direction if $\theta > 0$ and in the stationary direction if $\theta < 0$.

The continuous-time solution of the stochastic differential equation (8) is

$$y_t = e^{\theta t}y_0 + \rho(1 - e^{\theta t}) + \int_0^t e^{\theta(t-s)}\sigma_s dW_s, \tag{10}$$

where y_0 is the initial value.

Lemma 3.1 *Under the linear drift-diffusion process* (8)*, log return* r_{t_i} *can be written as*

$$\begin{aligned} r_{t_i} &= \Delta_i A + \int_{t_{i-1}}^{t_i} \sigma_s dW_s + \int_{t_{i-1}}^{t_i} B(t_i, s)\sigma_s dW_s + \int_0^{t_{i-1}} [B(t_i, s) - B(t_{i-1}, s)]\sigma_s dW_s \quad (11) \\ &= \int_{t_{i-1}}^{t_i} \sigma_s dW_s\{1 + o_p(1)\}, \end{aligned}$$

where $\Delta_i A = A(t_i) - A(t_{i-1})$*,* $A(t) = \int_0^t a(r)dr$ *and* $B(t,s) = \int_s^t b(r,s)dr$ *with* $t \geq s$*,* $a(t) = \theta\left(y_0 - \rho\right)e^{\theta t}$ *and* $b(t,s) = \theta e^{\theta(t-s)}$*;*

The log return process is asymptotically dominated by the volatility component $\int_{t_{i-1}}^{t_i} \sigma_s dW_s$ as $\Delta \to 0$. Consequently, the squared return $r_{t_i}^2$ converges to the integrated variance $\int_{t_{i-1}}^{t_i} \sigma_s^2 ds$ (Barndorff-Nielsen and Shephard, 2002).

The unconditional expectation of r_{t_i}, denoted by m_{t_i}, is

$$m_{t_i} = \Delta_i A = \int_{t_{i-1}}^{t_i} a(r)dr = (y_0 - \rho)e^{\theta t_{i-1}}(e^{\theta\Delta} - 1).$$

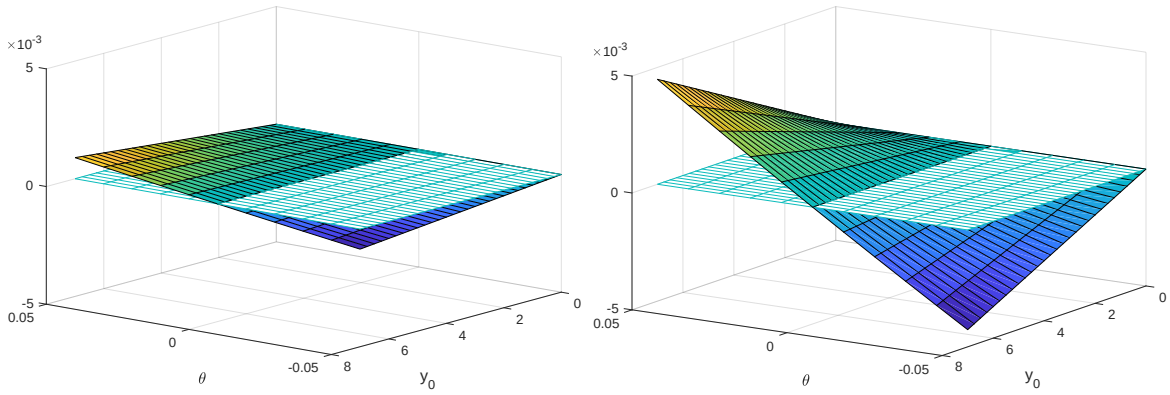
One can see that it depends on the values of ρ, θ, and Δ, as well as the initial value y_0. It is noted that $m_{t_i} = 0$ if either $y_0 = 0$ or $\theta = 0$. Furthermore, the unconditional expectation of the averaged log returns over the past K observations, denoted by $\bar{m}_{t_i}(K)$, is

$$\bar{m}_{t_i}(K) = \frac{1}{K}\sum_{j=i-K+2}^{i} m_{t_j} = \frac{1}{K}\int_{t_i-K}^{t_i} a(r)dr = \frac{1}{K}\left(y_0 - \rho\right)\left[e^{\theta t_i} - e^{\theta t_{i-K}}\right].$$

Suppose the sample runs over one day (i.e., $T = 1/\Delta$) and $K = T$. The unconditional expectation of the daily averaged return is $\bar{m}_1(K) = \frac{1}{K}\left(y_0 - \rho\right)(e^{\theta} - 1)$. Figure 4 shows the value $\bar{m}_1(K)$ for several combinations of values of y_0 and θ at 1-minute and 5-minute frequencies. We consider a wide range of values of θ and allow for the initial value to vary from 0 to 7. We set ρ to zero for simplicity. The magnitude of the drift increases under three circumstances: 1) if the sampling frequency is lower, 2) if the initial value becomes larger, and 3) if θ moves away from zero. Additionally, the unconditional mean $\bar{m}_1(K)$ has the order of magnitude of 10^{-3}, which is the same as that presented in Figure 1. This result suggests that the parameter settings being considered are empirically realistic and hence will be used for the simulations performed subsequently.

Figure 4: Average drift of log returns over one day for various combinations of θ, y_0 and Δ.

(a) 1-minute $\Delta = 1/400$ (b) 5-minute $\Delta = 1/80$



3.1 Unbiasedness of the Sample Mean and Median

Lemma 3.2 *Under the linear drift-diffusion process* (8)*, both the sample mean and the median are unbiased estimators of the unconditional mean* $\bar{m}_{t_i}(K)$*, i.e.,*

$$\mathbb{E}\left(\hat{m}_{t_i}(K)\right) = \mathbb{E}\left(\hat{m}^*_{t_i}(K)\right) = \bar{m}_{t_i}(K). \tag{12}$$

The unbiasedness of the sample mean follows directly from (11). The derivation of the unbiasedness of the median is, however, rather difficult. The literature on the finite sample distribution of order statistics with correlated and nonidentically distributed underlying variables is rather thin (Rychlik, 1994; Gupta et al., 1973; Chen, 2014). This question is further complicated by the time-varying correlation structure among returns. Nevertheless, the asymptotic distribution of quantile estimators in non i.i.d settings is well known (see, e.g., Koenker et al. (2005); Wu et al. (2005); Dominicy et al. (2013)). The median converges to the true mean $\bar{m}_{t_i}(K)$ at a rate of $O_p(\sqrt{K})$ and follows a normal distribution in the limit.

We resort to simulations to show the unbiasedness of the median. The data generating process is (9). The settings for ρ, θ and y_0 are the same as in Figure 4. The error term $\eta_{t_{i+1}}$ is specified by

$$\eta_{t_{i+1}} = \sigma_{t_{i+1}}\sqrt{\Delta}\varepsilon_{t_{i+1}} \tag{13}$$

$$\sigma^2_{t_{i+1}} = \alpha_0 + \sigma^2_{t_i}(\beta_1 + \alpha_1\sqrt{\Delta}v_{t_{i+1}}), \tag{14}$$

where $\varepsilon_{t_{i+1}}$ and $v_{t_{i+1}}$ are two independent standard normal random variables. The volatility dynamic

(14) is a Euler discretization of the GARCH(1,1) diffusion process of Nelson (1991), which is

$$d\sigma_t^2 = \kappa\left(\omega - \sigma_t^2\right)dt + \sqrt{2\lambda\kappa}\sigma_t^2 dW_t, \tag{15}$$

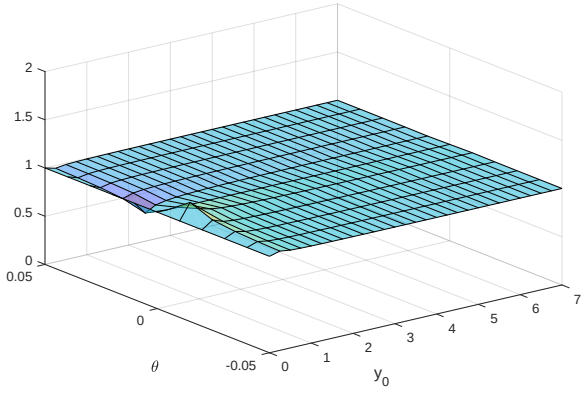
where $\kappa > 0$, $\omega > 0$, and $0 < \lambda < 1$. The parameters are related as follows: $\alpha_0 = \kappa\omega\Delta$, $\beta_1 = 1-\kappa\Delta$, and $\alpha_1 = \sqrt{2\lambda\kappa}$. We follow Andersen and Bollerslev (1998) and choose the parameters $\kappa = 0.035$ and $\lambda = 0.296$ to simulate a realistic log price process with very persistent GARCH effects and set $\omega = 10^{-4}$ such that $E(\sigma_{t_{i+1}}^2) = 10^{-4}$. The initial value of the volatility dynamic is set to be the unconditional volatility.

For each parameter constellation, we simulate 10^4 data series at the one second frequency and obtain data at the 1-minute ($\Delta = 1/400$) and 5-minute ($\Delta = 1/80$) frequencies by aggregation. The sample size is $T = 1/\Delta$. The median of the log returns of each data series is denoted by $\hat{m}_{t_T}^* (K)^j$ with $j = 1, \cdots, 10^4$ and $K = T$. The expected value of the median estimator is approximated by its sample counterpart, i.e.,

$$\mathbb{E}\left(\hat{m}_{t_i}^*(K)\right) \simeq \frac{1}{10^4}\sum_{j=1}^{10^4} \hat{m}_{t_i}^* (K)^j .$$

Figure 5 plots the ratio $\mathbb{E}\left(\hat{m}_{t_i}^*(K)\right) / \bar{m}_{t_i}(K)$. It is clear from the graph that the ratio is close to one for all parameter settings, which suggests the unbiasedness of the median as an estimator of the unconditional mean.

Figure 5: Ratio of $\mathbb{E}\left(\hat{m}_{t_i}^*(K)\right)$ and the unconditional mean $\bar{m}_{t_i}(K)$.



3.2 Realized Volatility Estimators

Proposition 3.1 *Under the linear diffusion process* (8)*, the finite sample bias of the realized variance is zero if* $\theta = 0$*. If* $\theta \neq 0$*,*

$$\begin{aligned}\mathbb{E}\left(RV_{t_i}(K) - \int_{t_i-K}^{t_i} \sigma_u^2 du\right) &= (y_0-\rho)^2\left(e^{\theta\Delta}-1\right)^2 e^{2\theta t_{i-K}}\frac{1-e^{2\theta\Delta K}}{1-e^{2\theta\Delta}} + \frac{d_0 K}{\theta}\left(e^{\theta\Delta}-1\right) \\ &\quad + \left(e^{\theta\Delta}-1\right)^2 \frac{d_0}{2\theta} e^{2\theta t_{i-K}}\frac{1-e^{2\theta\Delta K}}{1-e^{2\theta\Delta}} - d_0 K\Delta \\ &\equiv \mathcal{E},\end{aligned}$$

where $d_0 = E(\sigma_t^2)$*.*

Similar to the constant drift-diffusion process (2), the drift coefficient $\mu_t = \theta(y_t - \rho)$ in (8) has a finite sample impact on the realized variance estimator if both y_0 and $\theta \neq 0$.[7] The bias $\mathcal{E}$ arises from the nonzero mean of r_t and the interaction between the drift and diffusion terms. The finite sample biases of the modified realized volatility estimators under the linear drift-diffusion process are illustrated in Proposition 3.2.

Proposition 3.2 *Under the linear diffusion process* (8)*, the finite sample biases of the modified realized variance estimators* $RV^{\dagger}$ *and* RV^{*} *are zero if* $\theta = 0$*. If* $\theta \neq 0$*,*

(1) the bias of $RV^{\dagger}$ *is*

$$\mathbb{E}\left(RV_{t_i}^{\dagger}(K) - \int_{t_i-K}^{t_i} \sigma_u^2 du\right) = \mathcal{E} - K\,\mathbb{E}\left(\hat{m}_{t_i}^2\right).$$

with $K\,\mathbb{E}\left(\hat{m}_{t_i}^2\right)$ *equal*

$$\frac{1}{K}\left\{(y_0-\rho)^2 e^{2\theta t_{i-K}}\left(e^{\theta K\Delta}-1\right)^2 - \frac{d_0}{2\theta}\left(e^{\theta\Delta K}-1\right)^2\left(1-e^{2\theta t_{i-K}}\right) - \frac{d_0}{2\theta}\left(1-e^{2\theta\Delta}\right)\frac{e^{2\theta\Delta K}-1}{e^{2\theta\Delta}-1}\right\};$$

(2) the bias of RV^{*} *is*

$$\mathbb{E}\left(RV_{t_i}^{*}(K) - \int_{t_i-K}^{t_i} \sigma_u^2 du\right) = \mathcal{E} - 2K\,\mathbb{E}\left(\hat{m}_{t_i}^{*}\hat{m}_{t_i}\right) + K\,\mathbb{E}(\hat{m}_{t_i}^{*2}), \tag{16}$$

[7] The finite sample bias of the realized variance estimator was first documented in Meddahi (2002), where the drift coefficient was assumed to be a square-integrable function of the state variable S_t such that $\mu_t = \sum_{i=0}^{p} g_i \Pi_{i,S_t}$, where $\sum_{i=0}^{p} |g_i| < \infty$.

where the quantities $\mathbb{E}\left(\hat{m}^*_{t_i}\hat{m}_{t_i}\right)$ *and* $\mathbb{E}(\hat{m}^{*2}_{t_i})$ *can be obtained via simulations.*

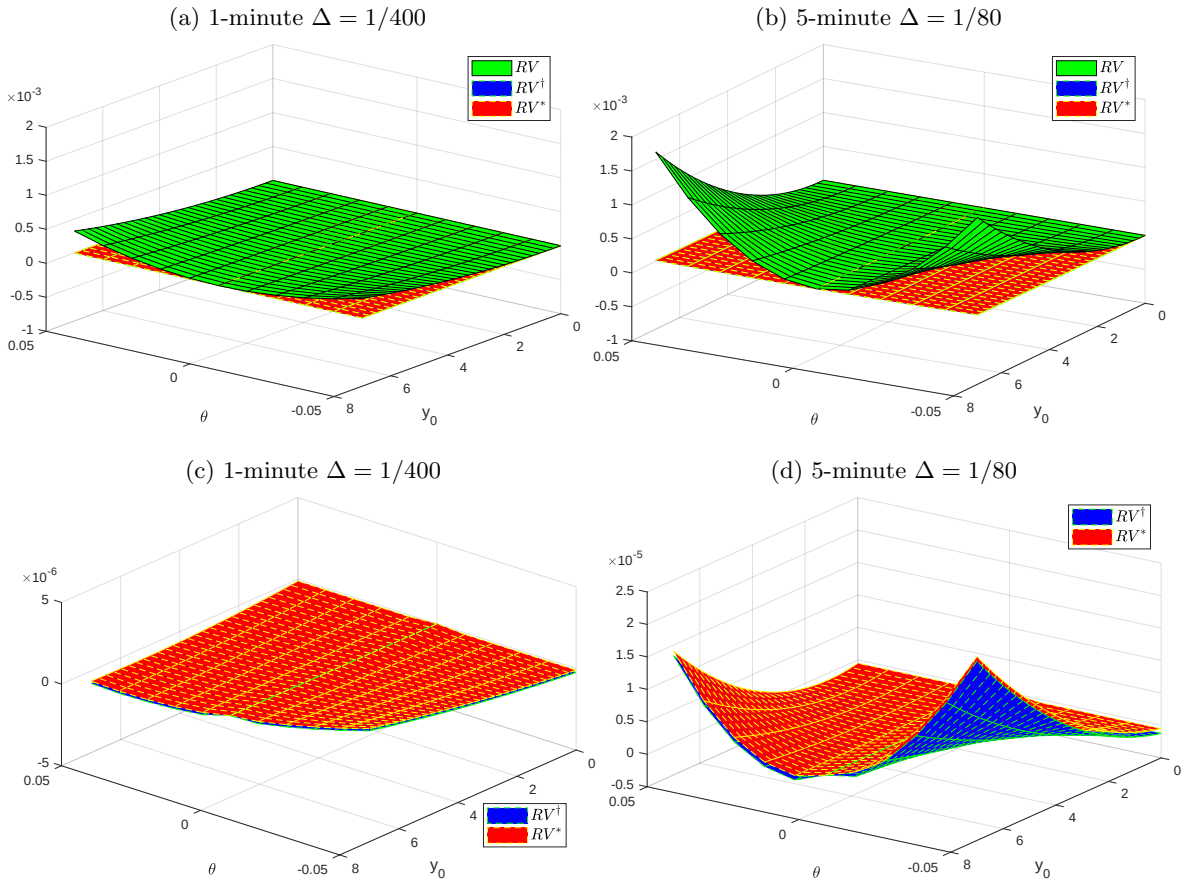
The overall bias is reduced substantially due to the use of centered log returns in the calculation. The magnitude of reduction is visualized in the next subsection.

3.3 Visualization of the Bias

3.3.1 Realized Volatilities

Figure 6 plots the bias of the realized volatilities RV, $RV^\dagger$ and RV^*, calculated from Propositions 3.1 and 3.2. The top panel plots all three estimators, while the bottom panel shows only the modified ones for further comparison. The parameter settings are the same as those in Figure 5. The unconditional variance d_0 is set to 10^{-4}. The quantities $\mathbb{E}\left(\hat{m}^2_{t_i}\right)$, $\mathbb{E}\left(\hat{m}^*_{t_i}\hat{m}_{t_i}\right)$, and $\mathbb{E}(\hat{m}^{*2}_{t_i})$ are obtained via Monte Carlo simulations with DGP (9) and 10^4 replications.

Figure 6: Bias of the realized variance and modified realized variances under the linear drift-diffusion process.



Recall that the unconditional mean (drift) of log returns increases with $|\theta|$ and the initial value y_0 (see Figure 4). It is clear from Figure 6 that the bias of the realized variance increases nonlinearly with the magnitude of the drift, especially at the 5-minute frequency. Furthermore, the biases of the modified volatility estimators $RV^{\dagger}$ and RV^{*} are of a much smaller magnitude. Similar to the constant drift case, the bias of $RV^{\dagger}$ is slightly smaller than that of RV^{*}.

3.4 Bipower Variations

As explained above, the derivation of the finite sample bias of the bipower variation and its modified versions under the linear drift-diffusion process is rather complicated and left for future research. Here, we compare the performance of the bipower variation and its modified versions via simulations when the DGP is a linear drift-diffusion process with additive jumps.

We consider a DGP that generalizes (9) by allowing for k additive jumps:

$$y_{t_{i+1}} = g(\theta) + \alpha(\theta) y_{t_i} + \sum_{j=1}^{k} \phi^{j}_{t_{i+1}} I^{j}_{t_{i+1}} + \eta_{t_{i+1}}, \tag{17}$$

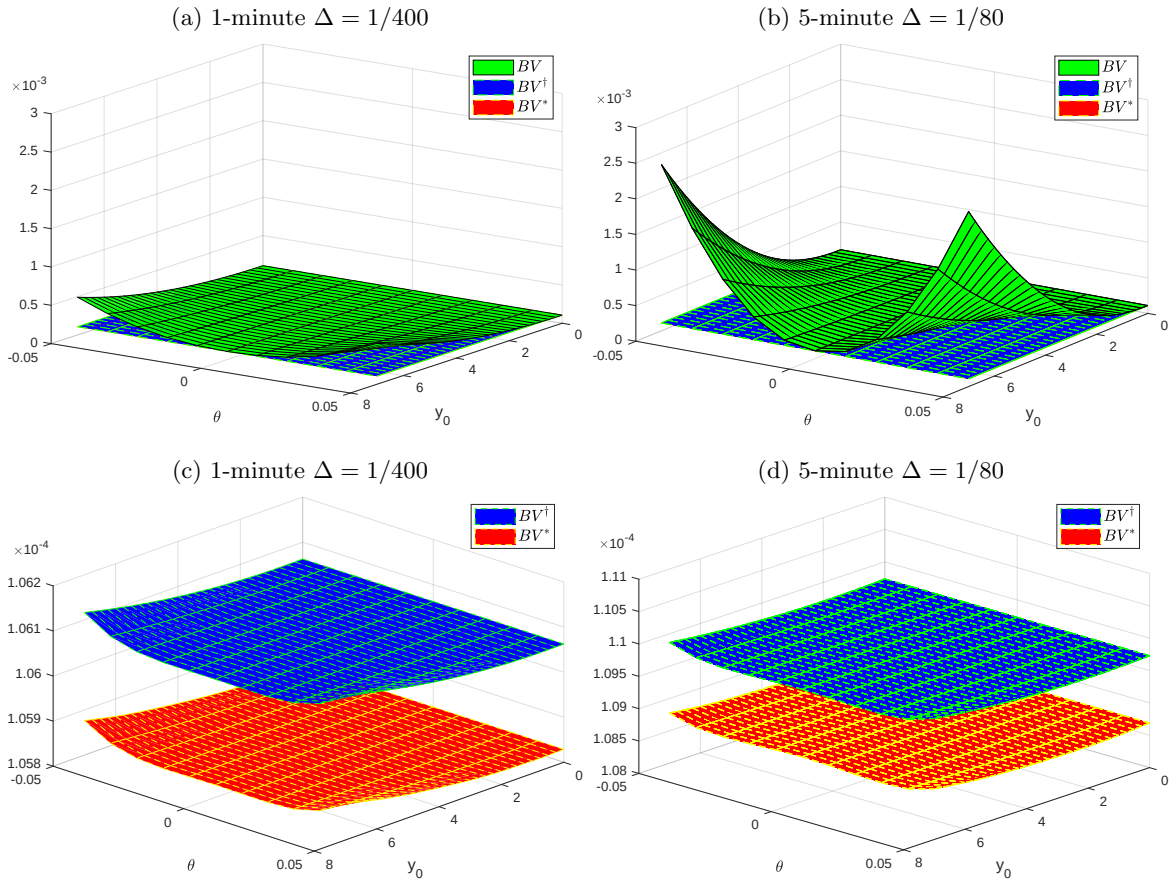
where $I^{j}_{t_{i+1}}$ is a dummy variable indicating the location of the j^{th} jump (the occurrence of which is random), and $\phi^{j}_{t_{i+1}}$ is the corresponding jump size. For simplicity, we allow for one negative jump per day with the size equal to 60% of the spot volatility (i.e., $\phi^{1}_{t_{i+1}} = -0.6\sigma_{t_{i+1}}$). The volatility dynamic and the settings of the remaining parameters are as in Section 3.1.

We report in Figure 7 the bias of the bipower variation and the modified bipower volatility estimators for the last observation of the day (i.e., t_T), denoted by $BV_{t_T}(K)$, $BV^{\dagger}_{t_T}(K)$ and $BV^{*}_{t_T}(K)$, respectively. They are calculated at the end of each day using the $K = T$ observations of that day. The top panel displays the biases of all three estimators, and the bottom panel presents only the two modified bipower volatility estimators.

As in Figure 2, there is almost no difference in the estimation accuracy among the three estimators if either θ or y_0 equals zero. We observe an upward bias of bipower variation when both θ or y_0 are not zero. The bias becomes larger as the sampling frequency decreases and as $|\theta|$ and y_0 increase. As expected, the modified bipower volatility estimators $BV^{\dagger}$ and BV^{*} provide much more accurate estimates of the integrated variance. Indeed, the biases of $BV^{\dagger}$ and BV^{*} are very small

for all combinations of parameters considered in this simulation. Furthermore, the bias of BV^* is always smaller than that of $BV^\dagger$. This is because the median provides more accurate estimate of the unconditional expectation of log returns than does the sample mean in the presence of jumps (see Appendix B for an illustration).[8]

Figure 7: Empirical bias of the bipower and the modified bipower volatility estimators under the linear drift-diffusion process.



A similar correction can be applied to other jump-robust estimators of the integrated variance. We have also performed Monte Carlo simulations, considering the MedRV of Andersen et al. (2012) and the threshold realized variance of Mancini (2009). The results are qualitatively the same as for BV, $BV^\dagger$ and BV^* in Figure 7 and are therefore not reported to save space.

[8]It is well known in the robust statistics literature that the median is less sensitive to outliers (i.e., jumps in our framework) than is the sample mean. Indeed, the asymptotic breakdown point is 0 for the sample mean and 1/2 for the median (see Maronna et al., 2006, etc.).

4 Volatility Estimation with Ultrahigh-frequency Data

One might expect that with ultrahigh-frequency data, the drift component will be extremely close to zero; hence, the discrepancy between the original volatility estimators (based on log returns) and the modified ones (based on centered log returns) will diminish. We show in this section that this is unfortunately untrue for the noise-robust volatility estimator of Podolskij and Vetter (2009).

Assume that the noise-contaminated log price $y^{\circ}_{t_i}$ is

$$y^{\circ}_{t_i} = y_{t_i} + \eta_{t_i}, \tag{18}$$

where $(\eta_{t_i})_{1\leq i\leq T}$ is a noise process with mean zero and variance q^2, independent of y_{t_i}. Assume also that the noise process is serially correlated of order $s-1$.

The estimator of Podolskij and Vetter (2009) is constructed as follows. Define Γ-return $r^{(\Gamma)}_{t_i}$ as

$$r^{(\Gamma)}_{t_i} = y^{\circ}_{t_i} - y^{\circ}_{t_{i-\Gamma}},$$

where $\Gamma = \gamma_1 K^{1/2}$ with $\gamma_1 > 0$. We divide the past K Γ-returns into B nonoverlapping blocks. The number of blocks is $B = \gamma_2\Gamma$ with $\gamma_2 > 1$, and the size of the blocks is $S = K/B$. The average log return of block b is

$$\bar{r}_{t_i,b} = \frac{1}{S}\sum_{j=i-(B-b+1)S+1}^{i-(B-b)S} r^{(\Gamma)}_{t_j}. \tag{19}$$

Podolskij and Vetter (2009) propose computing the bipower variation for noise-contaminated data using pre-averaged returns, i.e.,

$$BV^N_{t_i}(K) = \frac{\pi}{2}\frac{B}{B-1}\sum_{b=2}^{B}|\bar{r}_{t_i,b}||\bar{r}_{t_i,b-1}|. \tag{20}$$

One can obtain a consistent estimator of the integrated variance $\widehat{IV}_{t_i}(K)$ by removing the variation induced by the market microstructure noise from the bipower variation:

$$\widehat{IV}_{t_i}(K) := \frac{\gamma_1\gamma_2 BV^N_{t_i}(K) - v_2\hat{q}^2_{t_i}(K)}{v_1} \to \int_{t_{i-K}}^{t_i}\sigma^2_u du, \tag{21}$$

where $\hat{q}^2_{t_i}(K)$ is a consistent estimator of the noise variance q^2 and is defined as

$$\hat{q}^2_{t_i}(K) = \frac{1}{2(K-s)} \sum_{j=i-K+s+1}^{i} (y^\circ_{t_j} - y^\circ_{t_{j-s}})^2,$$

$$v_1 = \frac{\gamma_1[3\gamma_2 - 4 + \max\{(2-\gamma_2^3), 0\}]}{3(\gamma_2-1)^2} \text{ and } v_2 = \frac{2\min\{(\gamma_2-1), 1\}}{\gamma_1(\gamma_2-1)^2}.$$

While the drift component of log returns is extremely close to zero in the ultrahigh-frequency setting, the drift of pre-averaged returns $\bar{r}_{t_i,b}$ (of which the bipower variation is computed) might be of a nonnegligible magnitude. For the sake of illustration, we consider a data generating process following (17), (13)-(14) and (18). The parameter settings of the non-noise components remain unchanged. We assume i.i.d. noise (i.e., $s = 1$) for simplicity. The variance of the noise is set to be proportional to the variance of the underlying process as in Bandi and Russell (2006); Boudt et al. (2017); Lee and Mykland (2012). Specifically, $q^2 = 0.01\sqrt{\int_0^1 \sigma^4(s)ds}$ and hence $q \approx 0.1\%$. We simulate one day of 1-second data with one small jump per day (for a total of 10^4 replications).

Figure 8: Median of $\bar{r}_{t_i,b}$ for various combinations of θ and y_0.

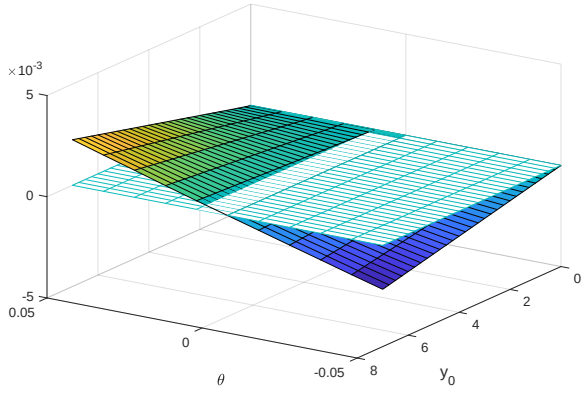


Figure 8 displays the average (over 10^4 replications) of the median of pre-averaged returns, denoted by $\tilde{m}^*_{t_i}(K) = median(\bar{r}_{t_i,1}, \cdots, \bar{r}_{t_i,B})$, for various combinations of θ and y_0. This figure shows that the median increases as the process deviates from the random walk, and the magnitude is comparable to the median of 5-minute log returns plotted in Figure 4. As before, this deviation is expected to affect the performance of the integrated variance estimator.

Analogously, we propose modifying the estimator of Podolskij and Vetter (2009) by computing

the bipower variation using centered pre-averaged returns, i.e.,

$$BV_{t_i}^{N*}(K) = \frac{\pi}{2}\frac{B}{B-1}\sum_{b=2}^{B}|\bar{r}_{t_i,b} - \tilde{m}_{t_i}^{*}(K)||\bar{r}_{t_i,b-1} - \tilde{m}_{t_i}^{*}(K)|, \tag{22}$$

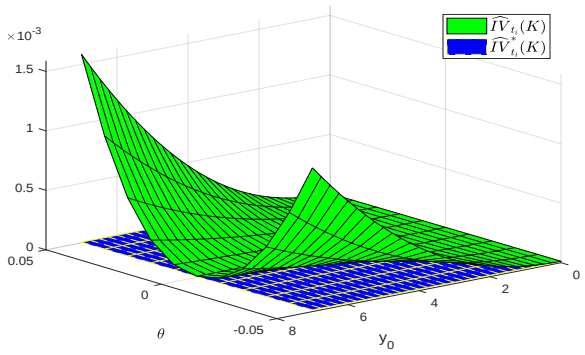
and

$$\widehat{IV}_{t_i}^{*}(K) := \frac{\gamma_1\gamma_2 BV_{t_i}^{N*}(K) - v_2\hat{q}_{t_i}^2(K)}{v_1} \to \int_{t_{i-K}}^{t_i}\sigma_u^2 du.$$

The consistency of $\widehat{IV}_{t_i}^{*}(K)$ follows directly from Podolskij and Vetter (2009).

We compute the bias of the two noise- and jump-robust estimators $\widehat{IV}_{t_i}(K)$ and $\widehat{IV}_{t_i}^{*}(K)$ in the same simulation setting. The results are presented in Figure 9. We observe patterns similar to those in Figure 7. While the estimation accuracy of the original volatility estimator deteriorates substantially as $|\theta|$ and y_0 deviate from zero, the new estimator is much more accurate.[9]

Figure 9: Estimation bias of $\widehat{IV}_{t_i}(K)$ and $\widehat{IV}_{t_i}^{*}(K)$.



5 Jump Tests

Testing for jumps and precisely identifying their occurrences is of overwhelming importance in finance since jumps have implications in risk management, portfolio allocation and derivative pricing (Aït-Sahalia, 2004). Several tests have been proposed in the literature (see Mancini and Calvori, 2012 for a survey). The most popular test is probably the test for finite-activity jumps[10] proposed

[9]The results for $\widehat{IV}_{t_i}^{\dagger}(K)$ that uses the sample mean of the pre-average return (i.e., $\frac{1}{K}\sum_{b=1}^{B}\bar{r}_{t_i,b}$) for centering are similar to those of $\widehat{IV}_{t_i}^{*}(K)$ and hence are not reported for brevity.

[10]Lee and Hannig (2010) propose a test for the presence of infinite-activity jumps (i.e., a Levy jump-diffusion process).

independently by Andersen et al. (2007) and Lee and Mykland (2008) and extended by Lee and Mykland (2012) to account for the presence of microstructure noise. These tests (especially the LM08 test) have been shown to have the overall best performance by Dumitru and Urga (2012) in a comprehensive Monte Carlo simulation comparing nine jump detection procedures available in the literature.

We show in this section that the finite sample bias of volatility estimators (due to nonzero drift components) leads to a significant downward size distortion and power loss for the LM08 test when applied to relatively low-frequency (e.g., 5-minute) data. Despite the fact that the LM12 test is designed for ultrahigh-frequency (e.g., 1-second) data and can be applied to very short time spans (e.g., 1 hour), it is observed to be also undersized if the log price process has a nonzero drift. We propose a modification of both tests that relies on modified bipower variations and show the importance of this correction for the finite sample performance of the tests through Monte Carlo simulations.

5.1 Lee and Mykland (2008) Tests

Andersen et al. (2007) and Lee and Mykland (2008) independently proposed a test statistic for jumps, denoted by J_{t_i} below, for which they derived the asymptotic distribution in the zero drift case, while Lee and Mykland (2008) also proposed another test statistic, denoted by $\tilde{J}_{t_i}$ below, for the nonzero drift case. The two statistics are defined as follows:

$$J_{t_i} = \frac{r_{t_i}}{\hat{\sigma}_{t_i}(K)} \quad \text{and} \quad \tilde{J}_{t_i} = \frac{r_{t_i} - \hat{m}_{t_i}(K)}{\hat{\sigma}_{t_i}(K)}. \tag{23}$$

Lee and Mykland (2008) proposed estimating the instantaneous volatility $\hat{\sigma}_{t_i}(K)$ using a rolling window of K log returns as follows:

$$\hat{\sigma}_{t_i}(K) = \sqrt{\frac{1}{K} BV_{t_i}(K)}. \tag{24}$$

The construction of both test statistics, J_{t_i} and $\tilde{J}_{t_i}$, involves the bipower variation. As discussed in the previous section, the finite sample performance of this volatility estimator is unsatisfactory if the process has a nonzero drift. Additionally, the demeaned test statistic $\tilde{J}$ is based on the sample

mean, the performance of which is inferior to that of the median in the presence of jumps, as shown in Appendix B. To improve the finite sample performance of the tests, we propose a correction to these jump test statistics. The new test statistic is denoted by $J^*_{t_i}$ and defined as[11]

$$J^*_{t_i} = \frac{r_{t_i} - \hat{m}^*_{t_i}(K)}{\hat{\sigma}^*_{t_i}(K)}. \tag{25}$$

We replace $\hat{m}_{t_i}(K)$ by the median $\hat{m}^*_{t_i}(K)$ and $\hat{\sigma}_{t_i}(K)$ by $\hat{\sigma}^*_{t_i}(K)$, an estimator of the instantaneous volatility based on the bipower variation computed from centered log returns, i.e.,

$$\hat{\sigma}^*_{t_i}(K) = \sqrt{\frac{1}{K} BV^*_{t_i}(K)}. \tag{26}$$

The asymptotic properties of both test statistics have been studied by Lee and Mykland (2008). More specifically, they show that in the absence of jumps, J_{t_i} and $\tilde{J}_{t_i}$ converge to a standard normal distribution as the sampling interval Δ tends to zero, provided that K is sufficiently large and the drift and diffusion coefficients in (1) do not change dramatically over a short time interval (i.e., $O_p(\Delta^{1/2})$). Since the drift component is asymptotically negligible, the proposed correction will not alter the limiting distribution of the test statistic. Therefore, if $K = O_p(\Delta^{\alpha})$ with $-1 < \alpha < -0.5$, we have

$$\sup_{i \in \{1, \cdots, T\}} |S_{t_i} - U| = O_p(\Delta^{\eta}),$$

where $S_{t_i} = \left\{ J_{t_i}, \tilde{J}_{t_i}, J^*_{t_i} \right\}$ and $-\epsilon < \eta < \frac{3}{2} + \alpha - \epsilon$ for any $\epsilon \geq 0$.

The jump test is implemented for each individual observation within the day. To control for the size of multiple tests, while Andersen et al. (2007) use a Bonferroni correction, Lee and Mykland (2008) suggest using critical values based on the extreme value theory. The maximum of a set of L i.i.d. realizations of the absolute value of the standard normal random distribution U_i (for $i = 1, \ldots, L$) asymptotically follows a Gumbel distribution (see, e.g., Aldous, 1989; Mutangi and Matarise, 2011), i.e.,

$$\frac{\max_i |U_i| - C_L}{S_L} \to \xi, \tag{27}$$

[11] A similar extension of the Lee and Hannig (2010) test, which is robust to infinite-activity jumps, is possible but beyond the scope of this paper.

where $C_L = (2\log L)^{1/2} - \frac{1}{2}(2\log L)^{-1/2}[\log\pi + \log(\log L)]$, $S_L = (2\log L)^{-1/2}$ and ξ is the standard Gumbel distribution with the cumulative distribution function $P\{\xi \leq x\} = \exp[-\exp(-x)]$.

Since, under the null hypothesis of no jump, S_{t_i} follows a standard normal distribution, the probability of $\max_i |S_{t_i}|$ (over a set of L values) exceeding the critical value $cv_{L,\beta}$ is $100\beta\%$ such that

$$P\left\{\max_i |S_{t_i}| > cv_{L,\beta}\right\} = 1 - \exp\left[-\exp\left(-\frac{cv_{L,\beta} - C_L}{S_L}\right)\right] = \beta,$$

and hence,

$$cv_{L,\beta} = C_L - S_L \log\left[-\log\left(1-\beta\right)\right]. \tag{28}$$

Therefore, we declare that there is a jump at time t_i according to J_{t_i} (respectively, $\tilde{J}_{t_i}$ and $J^*_{t_i}$) statistic if $|J_{t_i}| > cv_{L,\beta}$ (respectively, $\left|\tilde{J}_{t_i}\right| > cv_{L,\beta}$ and $\left|J^*_{t_i}\right| > cv_{L,\beta}$). If we set L to be the number of observations per day, the probability of finding at least one spurious jump (either positive or negative) within each day is $100\beta\%$.[12]

5.2 Finite Sample Performance of Jump Tests

In this section, we investigate the finite sample performance of the LM08 tests J and $\tilde{J}$, and our modified test J^*. The data generating process used to study the size of the tests is (9), while we rely on (17) to study their power. The volatility dynamic is specified by (13)-(14). The parameter settings are the same as in Section 3.1. Under the alternative, there is one jump per day with the magnitude of $\phi^1_{t_{i+1}} = -0.6\sigma_{t_{i+1}}$.

We generate $48,000$ observations corresponding to two days of 1-second data of an asset and aggregate them at the 1-minute and 5-minute frequencies as above. The first day is used as a burn-in period; we focus on the detection results of the second day. Therefore, the time span of interest is one day ($N = 1$), the sample size T equals the number of observations per day, and the time interval is $\Delta = 1/T$. The sample mean and median and the instantaneous volatility are estimated

[12]Given the rolling window calculation of the test statistics, the i.i.d assumption required by the extreme value theorem is likely to be violated for the jump tests. As a consequence, the multiplicity issue might not have been perfectly controlled for with the proposed critical values. One can observe from Tables 1 and 2 that there is a small upward size distortion remaining for the modified LM tests. This is, however, not the focus of this paper. A solid investigation of this problem is left to future research.

from a rolling window of K log returns, which is required to be between $\sqrt{T}$ and T. We choose K to be closer to the upper bound (i.e., $K = T - 1$) to ensure that there are sufficient observations to estimate the integrated variance before rescaling it to obtain an estimate of the instantaneous volatility of r_{t_i}.

For the critical values, we set $L = T$ and $\beta = 0.1\%$ so that the probability of finding at least one spurious jump within each day is 0.1%, or equivalently, we expect that one out of 1000 days contains at least one spurious jump. Therefore, the critical values used in this simulation are $cv_{24000,0.1\%} = 5.64$, $cv_{400,0.1\%} = 5.03$ and $cv_{80,0.1\%} = 4.85$ when the tests are applied to 1-second, 1-minute and 5-minute data, respectively. Note that the asymptotic critical values $cv_{L,\beta}$ in (28) depend on the values chosen for both L and β. Here, we control for the overall size of the test over a day ($L = T$) and set $\beta = 0.1\%$. Alternatively, we could control the size of the test over a longer period (e.g., a month or a year).

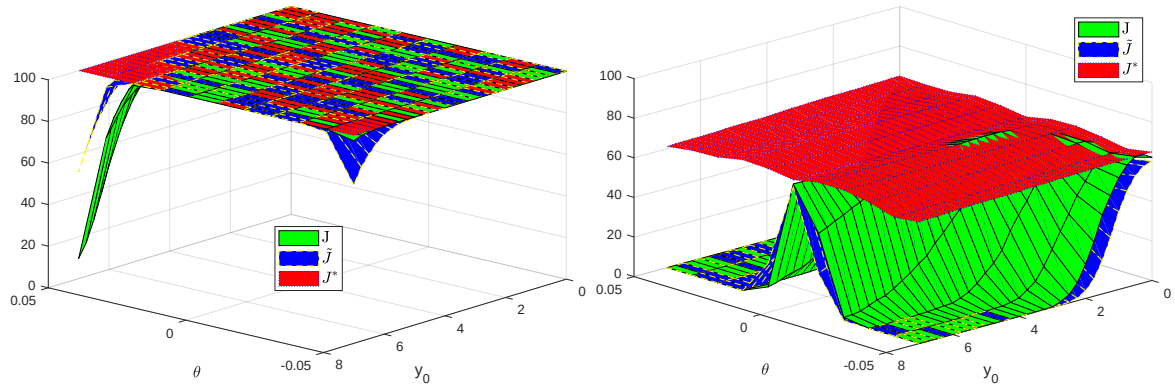
Table 1: Empirical sizes (%) of jump tests for a nominal size of 0.1%.

θ	-0.05	-0.04	-0.03	-0.02	-0.01	**0.0**	0.01	0.02	0.03	0.04	0.05
1-minute: $\Delta = 1/400$											
$y_0 = 1$											
J	0.037	0.038	0.046	0.043	0.044	**0.042**	0.042	0.054	0.030	0.039	0.042
$\tilde{J}$	0.027	0.026	0.040	0.042	0.051	**0.045**	0.044	0.052	0.027	0.034	0.028
J^*	0.049	0.045	0.051	0.053	0.055	**0.048**	0.049	0.065	0.042	0.058	0.062
$y_0 = 3$											
J	0.001	0.009	0.023	0.035	0.046	**0.042**	0.039	0.035	0.014	0.004	0.001
$\tilde{J}$	0.000	0.000	0.004	0.019	0.039	**0.045**	0.032	0.020	0.002	0.000	0.000
J^*	0.048	0.046	0.052	0.054	0.055	**0.048**	0.049	0.065	0.041	0.058	0.060
$y_0 = 6$											
J	0.000	0.000	0.000	0.010	0.035	**0.042**	0.031	0.010	0.000	0.000	0.000
$\tilde{J}$	0.000	0.000	0.000	0.001	0.016	**0.045**	0.016	0.002	0.000	0.000	0.000
J^*	0.051	0.047	0.053	0.054	0.056	**0.048**	0.049	0.064	0.042	0.056	0.060
5-minute: $\Delta = 1/80$											
$y_0 = 1$											
J	0.034	0.047	0.066	0.083	0.079	**0.082**	0.082	0.070	0.052	0.050	0.028
$\tilde{J}$	0.010	0.014	0.031	0.062	0.071	**0.091**	0.077	0.051	0.021	0.018	0.006
J^*	0.134	0.151	0.150	0.152	0.130	**0.150**	0.127	0.140	0.137	0.144	0.127
$y_0 = 3$											
J	0.000	0.000	0.000	0.022	0.064	**0.082**	0.071	0.017	0.000	0.000	0.000
$\tilde{J}$	0.000	0.000	0.000	0.004	0.037	**0.091**	0.041	0.001	0.000	0.000	0.000
J^*	0.140	0.152	0.151	0.154	0.129	**0.150**	0.126	0.141	0.140	0.154	0.135
$y_0 = 6$											
J	0.000	0.000	0.000	0.000	0.019	**0.082**	0.019	0.000	0.000	0.000	0.000
$\tilde{J}$	0.000	0.000	0.000	0.000	0.004	**0.091**	0.003	0.000	0.000	0.000	0.000
J^*	0.142	0.150	0.149	0.154	0.129	**0.150**	0.125	0.143	0.135	0.158	0.138

The empirical sizes (with 10^5 replications) of the LM08 statistics J and $\tilde{J}$ and the modified

Figure 10: Empirical performance of (modified) LM08 tests.

(a) Power: 1-minute $\Delta = 1/400$ (b) Power: 5-minute $\Delta = 1/80$



LM08 statistic J^* are reported in Table 1, while the powers are plotted in Figure 10 for the two sampling frequencies. We observe that if log prices follow a random walk (i.e., $\theta = 0$), the empirical sizes of the LM08 tests J and $\tilde{J}$ are close to the nominal size of 0.1%.[13] On the contrary, we observe a significant downward size distortion for both tests as the dynamic of log price deviates from the random walk (i.e., $\theta \neq 0$) and the initial value y_0 increases.[14] The undersize problem becomes even more severe if the sampling frequency is lower. Indeed, if the tests are applied to 5-minute data, the null hypothesis of no jump is almost never rejected for both tests[15] if $|\theta| \geq 0.03$ and $y_0 \geq 3$.

Figure 10a shows that the powers of both tests are close to 100% for all specifications if the tests are applied to data at the 1-minute frequency (except the bottom left and right corners when both θ and y_0 are large). For 5-minute data, power values are approximately 65% if asset prices follow a random walk with additive jumps. As expected, in the absence of microstructure noise, jumps are easier to detect if the sampling frequency increases. However, the assumption of no microstructure noise for data sampled at a frequency higher than 1-minute is unrealistic. Therefore, the LM08 tests are usually applied to 5-minute data to reduce the impact of microstructure noise at the cost of a slight power loss. If 5-minute data are used (see Figure 10b), the downward size distortion problem of both tests translates into a dramatic loss of power if the process has a large nonzero drift. In

[13] We obtained qualitatively the same results for the case of constant volatility and for other quantiles (5% and 0.1%). The results are not reported to save space.

[14] When the LM08 tests are applied to 1-second data (without microstructure noise), both tests have an empirical size close to the nominal size of 0.1% for all combinations of parameters and a power of 100%. The results are not reported here to save space.

[15] Importantly, a close-to-zero rate of rejection is also observed when using a higher critical value of $\beta = 5\%$.

particular, both tests have a power close to 0 if $|\theta| \geq 0.03$ and $y_0 \geq 3$. This result is consistent with our expectation that the nonnegligible mean of log returns affects the performance of the LM08 tests in finite samples. Importantly, the demeaned version of the test (i.e., $\tilde{J}$), which relies on $\hat{m}$ and BV (rather than $\hat{m}^*$ and BV^*), does not improve the performance of the test.

The new jump test has an outstanding performance. The empirical size of the test is reasonably close to the nominal size of 1% for both sampling frequencies. The empirical power of the test is 100% if the sampling frequency is high (i.e., 1-minute) and approximately 65% if applied to 5-minute data. This result is in sharp contrast to those of J and $\tilde{J}$, where both tests suffer from serious size distortion toward 0 and a lack of power for $1/\Delta = 80$ and $|\theta| \neq 0$.

5.3 Intraday Periodicity

For ease of exposition, we have so far ignored intraday periodicity effects in the spot volatility. However, since the studies of Taylor and Xu (1997) and Andersen and Bollerslev (998b), it has been well known that the opening, lunch period and closing of financial markets induce a strong periodic pattern in the volatility of high-frequency returns. More recently, Boudt et al. (2011) proposed several nonparametric robust-to-jumps estimators of the intraday (or intraweek) periodicity and a correction to the LM08 jump statistics by allowing for the spot volatility to depend on the estimated periodicity. They show that this modification helps increase the power to detect the relatively small jumps occurring at times of the volatility being periodically low and to reduce the number of spurious jump detections at times when the volatility is periodically high.

The most efficient nonparametric periodicity estimator of Boudt et al. (2011), denoted by $\hat{f}_{t_i}$, is the weighted standard deviation (WSD). Assuming for simplicity that the length of the periodicity cycle is one day and that the intraday periodicity is estimated on 5-minute data, the WSD estimator corresponds to the standard deviation of weighted standardized log returns J_{t_i} computed on all the observations belonging to the same 5-minute interval (across T days) and multiplied by a correction factor to ensure its consistency in the absence of jumps. Observation J_{t_i} receives a weight of either zero if $J_{t_i}^2$ is higher than a high quantile (e.g., 99%) of χ_1^2 distribution (i.e., the distribution of $J_{t_i}^2$ in the absence of jumps) or one otherwise. We refer the reader to Boudt et al. (2011) for details

on the WSD estimator.[16] It is important to note that this estimator of intraday periodicity relies on the assumption that, in the absence of jumps, J_{t_i} follows a standard normal distribution, while we have observed in Section 5.2 that this assumption is likely to be violated in finite samples if the process has a nonzero drift.

Consequently, we propose a modified WSD estimator, denoted by $\hat{f}^*_{t_i}$, where the periodicity is estimated on $J^*_{t_i}$ rather than J_{t_i}. Finally, as in Boudt et al. (2011), we also modify our newly proposed J^* statistic by multiplying $\hat{\sigma}^*_{t_i}$ in (25) by the estimated intraday periodicity $\hat{f}^*_{t_i}$. The periodicity-adjusted jump test statistic is denoted by $J^{*P}_{t_i}$ and defined as

$$J^{*P}_{t_i} = \frac{r_{t_i} - \hat{m}^*_{t_i}(K)}{\hat{f}^*_{t_i}\hat{\sigma}^*_{t_i}(K)}. \tag{29}$$

Periodicity-adjusted jump test statistics J_{t_i} and $\tilde{J}_{t_i}$ can be obtained similarly, i.e., $J^P_{t_i} = J_{t_i}/\hat{f}_{t_i}$ and $\tilde{J}^P_{t_i} = \tilde{J}_{t_i}/\hat{f}_{t_i}$.

To study the finite sample properties of the modified WSD estimator and the periodicity-adjusted jump test statistic, we extend the previous simulation by introducing intraday periodicity in addition to GARCH(1,1) dynamics in the conditional variance. To this end, we simulate the data according to (17), where

$$\eta_{t_{i+1}} = \sigma_{t_i} f_{t_i} \sqrt{\Delta} \varepsilon_{t_{i+1}}, \tag{30}$$

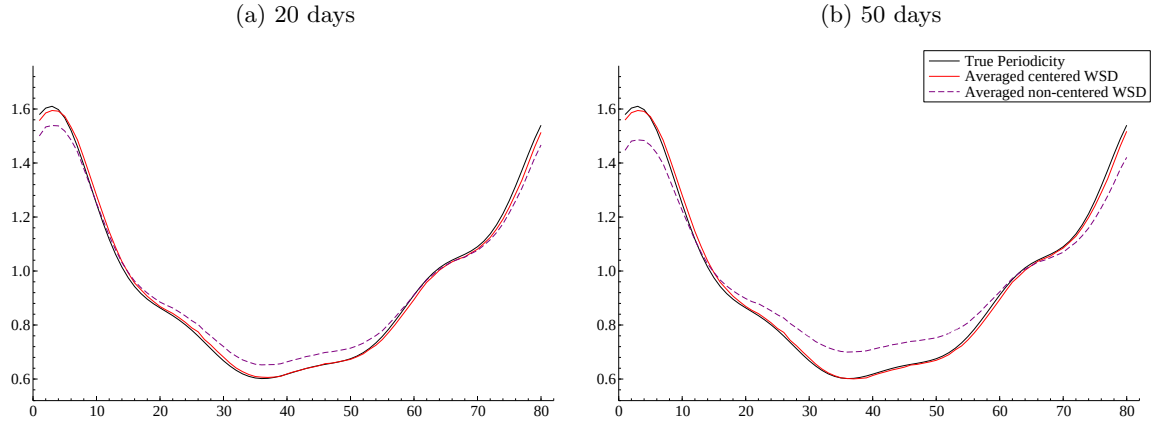
with σ_{t_i} following (14) and f_{t_i} capturing the volatility periodicity. The periodicity component depicts the usual U-shaped pattern during the day and is restricted to be the same on all days. We simulate 252 days of one-second data. The first 251 days serve as a burn-in period to estimate the periodicity using 5-minute data, while the last day is used to study the size and power of $J^{*P}_{t_i}$ test statistic (also computed on 5-minute data). The burn-in period contains on average one jump per day (with size $\phi^1_{t_{i+1}} = -0.6\sigma_{t_{i+1}} f_{t_i}$). The last day contains no jump under the null hypothesis and one jump under the alternative.

The true periodicity of the simulated 5-minute data (i.e., f_{t_i}) is plotted in Figure 11 together with the averages (over 10^5 replications) of $\hat{f}_{t_i}$ and $\hat{f}^*_{t_i}$. To study the impact of a nonzero drift on the

[16]Note that, as in Boudt et al. (2011), we normalize $\hat{f}_{t_i}$ so that $\hat{f}^2_{t_i}$ averages to the inverse of the length of the periodicity cycle (e.g., one day or one week).

estimation of f_{t_i}, θ is set to a nonzero value for either 20 (left panel) or 50 (right panel) randomly chosen days (out of the first 251 days) and to 0 on the remaining days. Figure 11 corresponds to the most extreme case, i.e., $y_0 = 7$ and $\theta = 0.05$ for 20 or 50 days. The results suggest that unlike the bias of $\hat{f}_{t_i}$, the bias of the modified WSD estimator $\hat{f}^*_{t_i}$ is negligible in both cases, even in this very extreme scenario.

Figure 11: Simulated intraday periodicity and periodicity estimators.



Finally, we explore the performance of the periodicity-adjusted jump test. Table 2 contains the rejection frequency of J^{*P} test under the null hypothesis of no jump, while its rejection frequency under the alternative of one jump is plotted in Figure 12. The same DGP is used as above. We consider the same range of values of θ and y_0 as in the previous simulations for the last day of the simulated sample (i.e., day 252). For the first 251 days, we assume that 20 random days deviate from the random walk (i.e., $\theta \neq 0$).[17] The general conclusion from Table 2 and Figure 12 is that our periodicity-adjusted jump test J^{*P} behaves similarly to the unadjusted one using J^* in the absence of periodicity. The test is slightly oversized for 5-minute data but is reasonably close to the nominal size for 1-minute data, and the power is satisfactory.

5.4 Ultrahigh-Frequency Jump Tests

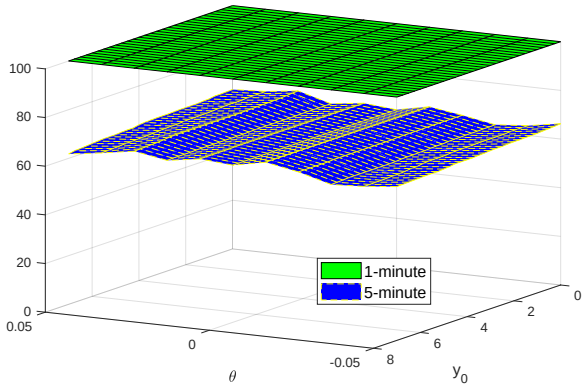
Lee and Mykland (2012) extend the LM08 tests to the ultrahigh-frequency setting, allowing for the presence of market microstructure noise. Interestingly, examining noise-contaminated one-second data, we observe similar patterns of size distortion and power loss of the LM12 test if θ and y_0

[17] The results are qualitatively the same for 50 days and are not reported here to save space.

Table 2: Empirical size (%) of the periodicity-adjusted jump test J^{*P}. The nominal size is 0.1%.

θ	-0.05	-0.04	-0.03	-0.02	-0.01	**0.0**	0.01	0.02	0.03	0.04	0.05
1-minute: $\Delta = 1/400$											
$y_0 = 1$	0.118	0.102	0.115	0.107	0.106	**0.116**	0.118	0.115	0.124	0.133	0.115
$y_0 = 3$	0.113	0.105	0.116	0.107	0.106	**0.116**	0.118	0.115	0.123	0.133	0.114
$y_0 = 6$	0.114	0.108	0.117	0.107	0.106	**0.116**	0.118	0.116	0.122	0.128	0.116
5-minute: $\Delta = 1/80$											
$y_0 = 1$	0.277	0.286	0.311	0.240	0.280	**0.246**	0.267	0.274	0.303	0.304	0.257
$y_0 = 3$	0.288	0.289	0.311	0.240	0.281	**0.246**	0.266	0.275	0.299	0.306	0.256
$y_0 = 6$	0.298	0.288	0.312	0.242	0.280	**0.246**	0.265	0.279	0.300	0.301	0.274

Figure 12: Power of the periodicity-adjusted jump test J^{*P}.



deviate from zero and show that a similar correction to the test statistic dramatically improves its finite sample performance.

The noise-contaminated log price is $y^{\circ}_{t_i}$, as defined by (18). Recall that the noise is assumed to be serially correlated of order s. Let $\tilde{\Gamma} = sM$ with $\tilde{M} = CT_1^{1/2}$ (where C is a constant and $T_1 = T/s$). For every s^{th} observation, we calculate the $\tilde{\Gamma}-$differenced log return as

$$r^{(\tilde{\Gamma})}_{t_{js}} = y^{\circ}_{t_{js}} - y^{\circ}_{t_{js-\tilde{\Gamma}}}$$

with $j = \tilde{M}+1, \cdots, T_1$. Let us divide the sequence $\{r^{(\tilde{\Gamma})}_{t_{js}}\}_{j=M+1}^{T_1}$ into $\tilde{B}$ blocks of size $\tilde{M}$. For each block b, we calculate the average log return over the block as follows:

$$\bar{r}_b = \frac{1}{\tilde{M}} \sum_{j=1}^{\tilde{M}} r^{(\tilde{\Gamma})}_{t_{bMs+js}} \quad \text{with} \quad b = 1, 2, \cdots, \tilde{B}.$$

The Lee and Mykland (2012) test statistics for the presence of jumps in log prices between t_{bsM}

and $t_{(b+1)sM}$ are

$$L_b = \sqrt{\tilde{M}}\frac{\bar{r}_b}{\sqrt{\hat{V}_b(K)}} \quad \text{and} \quad \tilde{L}_b = \sqrt{\tilde{M}}\frac{\bar{r}_b - \hat{m}_b(K)}{\sqrt{\hat{V}_b(K)}}, \tag{31}$$

where $\hat{V}_b(K)$ is an estimate of the variance of $\sqrt{M}\bar{r}_b$, and $\hat{m}_b(K)$ is the empirical mean of $\bar{r}_b$. Both $\hat{V}_b(K)$ and $\hat{m}_b(K)$ are calculated using a rolling window of K returns (over D days). Specifically, let $B^0 = K/\tilde{M}$. For $b \geq B^0$, we have $\hat{m}_b(K) = \frac{1}{B^0}\sum_{j=0}^{B^0-1}\bar{r}_{b-j}$ and

$$\hat{V}_b(K) = \frac{2}{3}\widehat{IV}_b(K)C^2D + 2\hat{q}_b^2(K), \tag{32}$$

where $\widehat{IV}_b(K)$ and $\hat{q}_b^2(K)$ are as in Section 4.

To improve the finite sample accuracy of the two test statistics in (31), we propose a modification similar to that in (25). The new test statistic is

$$L_b^* = \sqrt{M}\frac{\bar{r}_b - \hat{m}_b^*(K)}{\sqrt{\hat{V}_b^*(K)}}, \tag{33}$$

where $\hat{m}_b^*(K) = median(\bar{r}_{b-B^0+1}, \cdots, \bar{r}_b)$, and $\hat{V}_b^*(K)$ is defined similarly to $\hat{V}_b(K)$ with $\widehat{IV}_b(K)$ replaced by $\widehat{IV}_b^*(K)$. All three test statistics (L_b, $\tilde{L}_b$, and $L_{t_b}^*$) converge to the standard normal distribution under the null hypothesis of no jump.

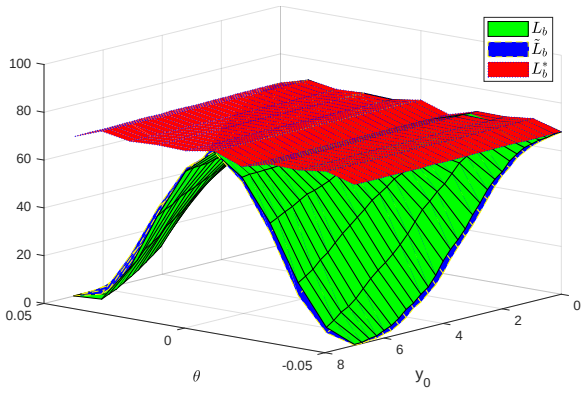
To study the finite sample properties of the three tests, we consider the same data generating process and parameter settings as in Section 4. However, for consistency with the simulations and application in Lee and Mykland (2012), we apply the tests to 1-second data over a very short period of 1 hour (rather than 1 day as in Sections 5.2 and 5.3). To this end, we simulate observations over a 2-hour period and use the first hour as a burn-in period. The rolling window size is set to $T - 1$. Parameter C is set according to Table 5 of Lee and Mykland (2012). The optimal value of C is 1/18 for the value of $q \approx 0.1\%$. We set $\gamma_1 = 1$ and $\gamma_2 = 1.6$ as in Podolskij and Vetter (2009). The nominal size of the test is again 0.1%.

The rejection frequencies of the L_b, $\tilde{L}_b$, and $L_{t_b}^*$ tests under the null hypothesis of no jump are reported in Table 3 for three values of y_0 (i.e., $1, 3$ and 6), while the rejection frequencies under the alternative of one jump are plotted in Figure 13 for a larger and finer range of values of y_0.

Table 3: Empirical size (%) of the L_b, $\tilde{L}_b$ and L_b^* tests. The nominal size is 0.1%.

θ	-0.05	-0.04	-0.03	-0.02	-0.01	**0.0**	0.01	0.02	0.03	0.04	0.05
$y_0 = 1$											
L_b	0.070	0.100	0.110	0.040	0.170	**0.100**	0.100	0.100	0.070	0.070	0.050
$\tilde{L}_b$	0.060	0.100	0.140	0.040	0.180	**0.100**	0.090	0.110	0.060	0.050	0.040
L_b^*	0.090	0.140	0.140	0.060	0.200	**0.100**	0.120	0.110	0.080	0.110	0.120
$y_0 = 3$											
L_b	0.000	0.020	0.040	0.000	0.130	**0.100**	0.070	0.050	0.010	0.000	0.000
$\tilde{L}_b$	0.000	0.000	0.030	0.000	0.120	**0.100**	0.070	0.040	0.010	0.000	0.000
L_b^*	0.090	0.140	0.140	0.060	0.200	**0.100**	0.100	0.110	0.080	0.080	0.120
$y_0 = 6$											
L_b	0.000	0.000	0.000	0.000	0.050	**0.100**	0.040	0.000	0.000	0.000	0.000
$\tilde{L}_b$	0.000	0.000	0.000	0.000	0.050	**0.100**	0.040	0.000	0.000	0.000	0.000
L_b^*	0.090	0.150	0.140	0.050	0.200	**0.100**	0.100	0.110	0.060	0.070	0.090

Figure 13: Power of the L_b, $\tilde{L}_b$ and L_b^* tests.



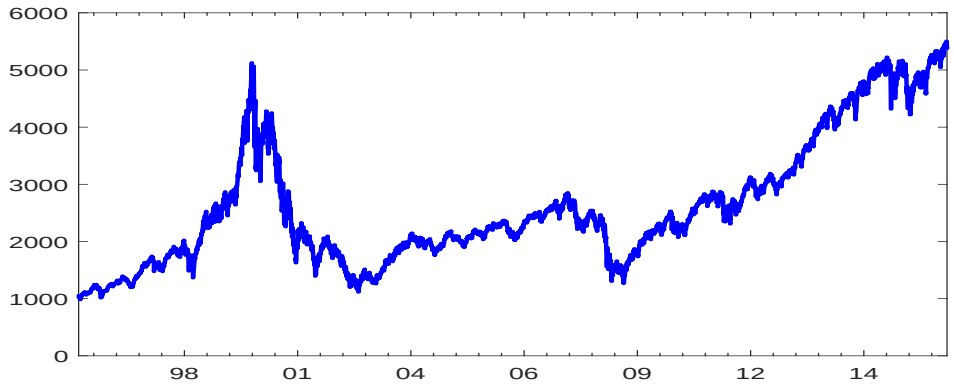
First, there is no visible difference in the empirical performance (in both size and power) of the L_b and $\tilde{L}_b$ tests. Second, we observe a downward size distortion and a power deterioration in both tests as the process deviates from the random walk and the initial value y_0 increases. Importantly, the size of the new test L_b^* fluctuates around the nominal size 0.1%, while its power is approximately 65% for all configurations of θ and y_0. These results are unsurprising, given our discussion and findings in Section 4 regarding the ultrahigh-frequency volatility estimators based on pre-averaging.

6 Empirical Application

The NASDAQ stock price index is sampled from 1996 to 2016 at the 5-minute frequency. All trades before 9:30 am or after 4:00 pm and the first trade after 9:30 am are discarded, which is the usual method of avoiding the overnight effect. The choice of this series is dictated by the fact that several studies (see Phillips et al., 2011, Homm and Breitung, 2012, and Shi and Song, 2016, etc.) have

shown evidence of deviations from the unit root in weekly and monthly data of the NASDAQ in the late 1990s. The log prices are plotted in Figure 14. We observe a very rapid expansion in the log prices in the late 1990s. The NASDAQ stock price has been rising steadily after the global financial crisis in 2008.

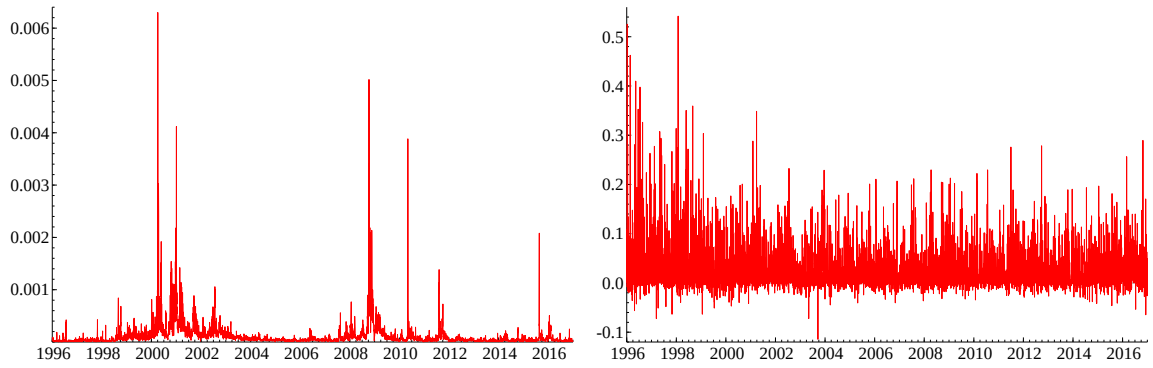
Figure 14: 5-minute log prices of the NASDAQ stock index over the period 1996-2016.



The daily bipower variation BV is plotted in the left panel of Figure 15, while the discrepancy between BV and the modified version BV^*, measured as $(BV - BV^*)/BV^*$, is plotted in the right panel. As expected, the daily bipower variation is of a much higher magnitude during crisis periods (e.g., after the dot-com bubble in the early 2000s and during the 2008 subprime mortgage crisis). Interestingly, Figure 15b suggests that the conventional bipower variation very often overestimates the integrated variance (on average by 2.5% but, in some cases, by more than 40%).

Figure 15: Estimated bipower variation (BV) of the NASDAQ stock market index and the discrepancy between BV and the modified bipower variation (BV^*).

(a) Bipower Variation BV (b) $(BV - BV^*)/BV^*$



To compute the periodicity-adjusted J test statistics, we first estimate the intraday periodicity

$\hat{f}_{t_i}$ and $\hat{f}^*_{t_i}$ year-by-year with a cycle length of one week (to allow for different day-of-week effects as in Boudt et al., 2011) with the procedures described in Section 5.3. The results of the three tests (J^P, $\tilde{J}^P$ and J^{*P}) are reported in Table 4. For the critical value $cv_{L,\beta}$, we set L to the total number of 5-minute log returns per day (i.e., $L = 78$ and $\beta = 0.1\%$). The values are reported in the three columns below: 'Significant jumps' correspond to the number of jump statistics greater than the critical value, while those below 'Significant days' correspond to the number of days in which at least one significant jump is detected.

Table 4: Descriptive statistics of significant jumps.

	Significant jumps			Significant days		
	J^P	$\tilde{J}^P$	J^{*P}	J^P	$\tilde{J}^P$	J^{*P}
1996	42	39	50	32	30	35
1997	41	44	51	34	34	40
1998	41	47	50	30	35	37
1999	22	21	28	19	18	24
2000	25	30	35	21	24	28
2001	27	28	28	24	23	23
2002	20	25	29	18	23	27
2003	36	39	44	31	33	38
2004	43	48	54	35	39	44
2005	73	75	81	53	54	57
2006	60	58	61	44	43	45
2007	60	59	71	42	41	48
2008	34	32	33	29	27	28
2009	45	39	49	35	31	38
2010	56	52	57	40	37	41
2011	34	44	43	30	37	37
2012	42	41	45	34	34	37
2013	47	48	49	35	36	36
2014	47	46	51	38	38	40
2015	31	32	37	25	26	30
2016	38	36	38	26	26	26
Total	864	883	984	675	689	759

The first conclusion we can draw from Table 4 is that for each year, the numbers of jumps detected by the three tests are not dramatically different. However, our proposed J^{*P} statistic allows us to almost systematically detect more jumps. Over 21 years, 864, 883 and 984 jumps are detected using the statistics J^P, $\tilde{J}^P$ and J^{*P}, respectively, which correspond to 675, 689 and 759

days with at least one significant jump. This result is consistent with our simulation findings that if the log price process has a nonzero drift, the newly proposed test J^{*P} has a higher power than that of the other two tests.

We can identify 87 days during which a jump is only detected by the J^{*P} statistic (i.e., $J^{P*} > cv_{78,0.1\%}$ while J^P and $\tilde{J}^P \leq cv_{78,0.1\%}$).[18] To better understand the difference between the tests, the log prices on 4 randomly selected days (out of these 87 days) are plotted in Figure 16. On each graph we plot the log prices of the day for which a jump is detected using our modified LM test as well as the log prices of the proceeding day (recall that the LM test statistics are backward looking). Log prices for which a jump is detected by J^P, $\tilde{J}^P$ and J^{*P} are highlighted by a cross, a plus and a circle, respectively. There is one jump in each panel that is identified by J^{*P} but not by J^P and $\tilde{J}^P$. The shaded area corresponds to the data period entering the calculation of the test statistics for this jump. Note that multiple jumps might be detected within a day. For instance, two jumps are identified on September 19, 2016 (bottom right panel). The first one is detected only by J^{*P} while the second one is identified by all three tests.

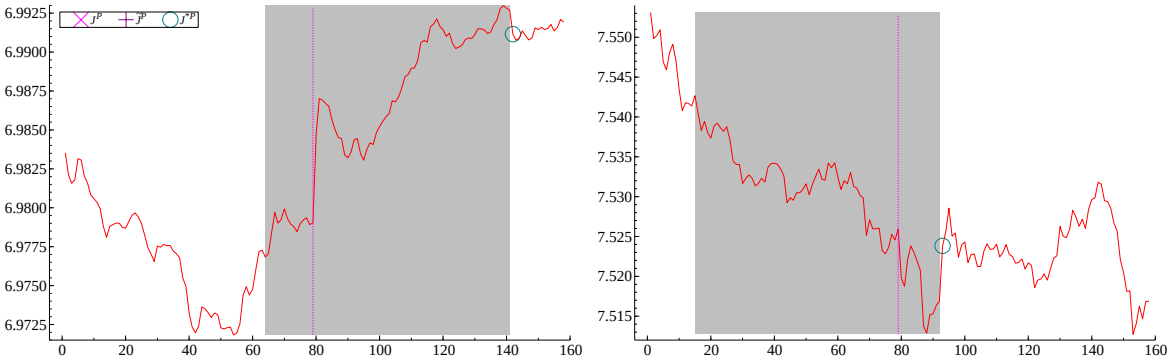
The vertical (purple) dotted line corresponds to the first price of the second day. Recall that log returns corresponding to the first prices of the day (i.e., overnight returns) are removed from all calculations. Interestingly, all four shaded areas are characterized by strong upward or downward trends which we attribute to nonzero drifts. This observation is consistent with our simulation results in Section 5 that if the process has a nonzero drift, the original LM (2008) tests applied to 5-minute data are undersized and have significantly lower power than J^*.

[18] There is on average one jump per year where the LM tests detect the jump but J^{*P} does not. We attribute such results to type I errors and do not report them here for brevity.

Figure 16: Log prices on four randomly selected days on which there is one jump identified by J^{*P} but not by J^P and $\tilde{J}^P$. The shaded areas correspond to the data periods entering the calculation of the test statistics for the jump identified only by J^{*P}.

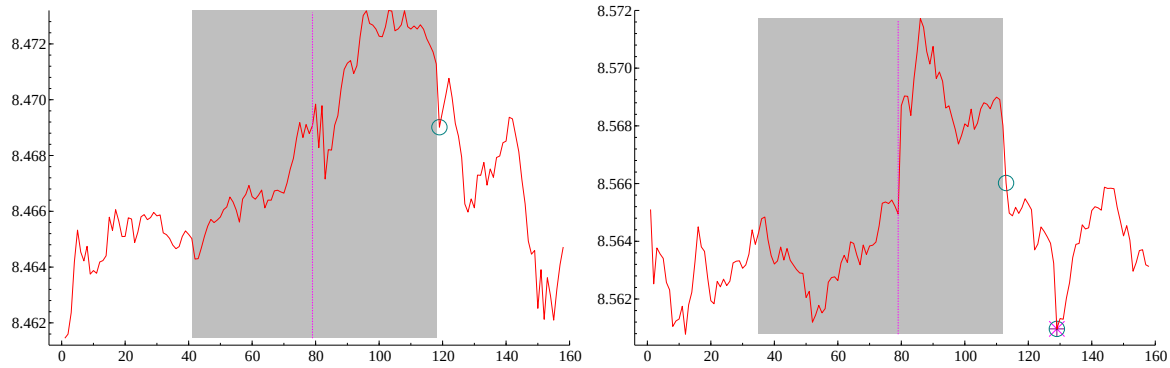
(a) Mar 12-13, 1996

(b) Feb 04-05, 2002



(c) Feb 05-06, 2015

(d) Sep 16-19, 2016



7 Conclusions

The finite sample theory, together with extensive simulations, reveal that with a rather low but realistic sampling frequency (e.g., 5 minutes), the realized variance and the bipower variation tend to overestimate the integrated variance in the presence of a nonzero drift, and the bias increases with the magnitude of the drift. Moreover, despite the drift becoming extremely close to zero as the sampling frequency increases, our simulations show that the volatility estimator of Podolskij and Vetter (2009), which is robust to microstructure noise and designed for ultrahigh-frequency data, suffers from the same problem due to the use of pre-averaged returns. Consequently, the procedures derived from these integrated variance estimators such as the intraday periodicity estimator of Boudt et al. (2011) and the jump tests of Lee and Mykland (2008, 2012) have unsatisfactory performance in finite samples if log prices have a nonzero drift. In particular, we demonstrate that in the presence

of a nonzero drift, the jump tests have strong size distortions and power losses.

We propose an alternative construction of these realized estimators and jump test statistics. The finite sample theory for the modified volatility estimators and simulations show significant improvement in the estimation accuracy of volatility, while simulation studies of the jump tests reveal a dramatic increase in powers. The newly proposed volatility estimators and the jump test, along with their original versions, are applied to 5-minute log returns of the NASDAQ for the period from 1996 to 2016. For most observations, the new estimator provides a lower estimate of the daily integrated variance than does the bipower variation, which is consistent with our theory and simulations. Furthermore, more jumps are detected using the new jump test. Interestingly, on days when jumps are detected only by the new test, log prices exhibit clear upward or downward trend movements. We attribute these trends to the presence of a relatively large drift, which explains why the original tests of Lee and Mykland (2008) fail to detect jumps on these days.

References

Aalen, O. O. and H. K. Gjessing (2004). Survival models based on the Ornstein-Uhlenbeck process. *Lifetime data analysis 10*(4), 407–423.

Aït-Sahalia, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics 74*, 487–528.

Aldous, D. (1989). *Probability approximations via the Poisson clumping heuristic*, Volume 77. New York: Springer-Verlag.

Andersen, T. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 885–905.

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

Andersen, T. G. and T. Bollerslev (1998b). Deutsch mark-dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance 53*, 219–265.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association 96*(453), 42–55.

Andersen, T. G., T. Bollerslev, and D. Dobrev (2007). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications. *Journal of Econometrics 138*(1), 125–180.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics 169*(1), 75–93.

Arnold, L. (1974). *Stochastic differential equations*. A Wiley-Interscience publication. Wiley.

Balvers, R., Y. Wu, and E. Gilliland (2000). Mean reversion across national stock markets and parametric contrarian investment strategies. *The Journal of Finance 55*(2), 745–772.

Bandi, F. M. and J. R. Russell (2005). On the finite sample properties of kernel-based integrated variance estimators. *Unpublished paper, Graduate School of Business, University of Chicago*.

Bandi, F. M. and J. R. Russell (2006). Separating microstructure noise from volatility. *Journal of Financial Economics 79*(3), 655–692.

Barndorff-Nielsen, O. E. and Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B 64*(2), 253–280.

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(2), 167–241.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics 2*(1), 1–37.

Bauwens, L., C. M. Hafner, and S. Laurent (2012). *Handbook of volatility models and their applications*, Volume 3. John Wiley & Sons.

Bekaert, G. and R. J. Hodrick (1992). Characterizing predictable components in excess returns on equity and foreign exchange markets. *The Journal of Finance 47*(2), 467–509.

Bessembinder, H. and K. Chan (1992). Time-varying risk premia and forecastable returns in futures markets. *Journal of Financial Economics 32*(2), 169–193.

Boudt, K., C. Croux, and S. Laurent (2011). Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance 18*(2), 353–367.

Boudt, K., S. Laurent, A. Lunde, R. Quaedvlieg, and O. Sauri (2017). Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. *Journal of Econometrics 196*(2), 347–367.

Campbell, J. Y. and J. Ammer (1993). What moves the stock and bond markets? A variance decomposition for long-term asset returns. *The Journal of Finance 48*(1), 3–37.

Campbell, J. Y. and Y. Hamao (1992). Predictable stock returns in the United States and Japan: A study of long-term capital market integration. *The Journal of Finance 47*(1), 43–69.

Chaudhuri, K. and Y. Wu (2003). Random walk versus breaking trend in stock prices: Evidence from emerging markets. *Journal of Banking and Finance 27*(4), 575–592.

Chen, E. J. (2014). Selection and order statistics from correlated normal random variables. *Discrete Event Dynamic Systems 24*(4), 659–668.

Christensen, K., R. Oomen, and R. Renò (2016). The drift burst hypothesis. Technical report, Department of Economics and Business Economics, Aarhus University.

Dominicy, Y., S. Hörmann, H. Ogata, and D. Veredas (2013). On sample marginal quantiles for stationary processes. *Statistics & probability letters 83*(1), 28–36.

Dumitru, A. and G. Urga (2012). Identifying jumps in financial assests: a comparison between nonparametric jump tests. *Journal of Business and Economic Statistics 30*, 242–255.

Etienne, X. L., S. H. Irwin, and P. Garcia (2014). Bubbles in food commodity markets: Four decades of evidence. *Journal of International Money and Finance 42*, 129–155.

Eugene, N., C. Lee, and F. Famoye (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and methods 31*(4), 497–512.

Fama, E. F. and K. R. French (1988). Permanent and temporary components of stock prices. *Journal of Political Economy 96*(2), 246–273.

Ferguson, T. S. (1999). Asymptotic joint distribution of sample mean and a sample quantile. *unpublished: http://www. math. ucla. edu/ tom/papers/unpublished/meanmed. pdf 5*.

Gupta, A. K. and S. Nadarajah (2005). On the moments of the beta normal distribution. *Communications in Statistics-Theory and Methods 33*(1), 1–13.

Gupta, S. S., K. Nagel, and S. Panchapakesan (1973). On the order statistics from equally correlated normal random variables. *Biometrika 60*(2), 403–413.

Gutierrez, L. (2012). Speculative bubbles in agricultural commodity markets. *European Review of Agricultural Economics 40*(2), 217–238.

Homm, U. and J. Breitung (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics 10*(1), 198–231.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33–50.

Koenker, R., A. Chesher, and M. Jackson (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

Lee, S. S. and J. Hannig (2010). Detecting jumps from levy jump diffusion processes. *Journal of Financial Economics 96*, 271–290.

Lee, S. S. and P. A. Mykland (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies 21*(6), 2535–2563.

Lee, S. S. and P. A. Mykland (2012). Jumps in equilibrium prices and market microstructure noise. *Journal of Econometrics 168*(2), 396–406.

Lo, A. W. and A. C. MacKinlay (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies 1*(1), 41–66.

Lo, A. W. and J. Wang (1995). Implementing option pricing models when asset returns are predictable. *The Journal of Finance 50*(1), 87–129.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics 36*, 270–296.

Mancini, C. and F. Calvori (2012). Jumps. In L. Bauwens, C. Hafner, and S. Laurent (Eds.), *Handbook of Volatility Models and Their Applications*. Wiley.

Maronna, R. A., D. R. Martin, and V. J. Yohai (2006). *Robust Statistics: Theory and Methods*. Wiley.

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics 17*(5), 479–508.

Mutangi, K. and D. F. Matarise (2011). On how to find the norming constants for the maxima of a folded normally distributed variable. *Journal of Statistical Research*.

Nelson, D. B. (1991). ARCH models as diffusion approximations. *Journal of Econometrics 45*, 7–38.

Nicolato, E. and E. Venardos (2003). Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type. *Mathematical Finance 13*(4), 445–466.

Park, S. and O. Linton (2011). Realized volatility: Theory and application. In L. Bauwens, C. Hafner, and S. Laurent (Eds.), *Handbook of Volatility Models and Their Applications*. Wiley.

Phillips, P. C. (1987). Time series regression with a unit root. *Econometrica*, 277–301.

Phillips, P. C. and S. Shi (2017). Detecting financial collapse and ballooning sovereign risk. *Cowles Foudation Discussion Paper No. 3010*.

Phillips, P. C., S. Shi, and J. Yu (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review 56*(4), 1043–1078.

Phillips, P. C., Y. Wu, and J. Yu (2011). Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? *International Economic Review 52*(1), 201–226.

Phillips, P. C. and J. Yu (2011). Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics 2*(3), 455–491.

Podolskij, M. and M. Vetter (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli 15*(3), 634–658.

Rychlik, T. (1994). Distributions and expectations of order statistics for possibly dependent random variables. *Journal of multivariate analysis 48*(1), 31–42.

Shi, S. and Y. Song (2016). Identifying speculative bubbles using an infinite hidden Markov model. *Journal of Financial Econometrics 14*(1), 159–184.

Taylor, S. and X. Xu (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance 4*, 317–340.

Wang, X. and J. Yu (2016). Double asymptotics for explosive continuous time models. *Journal of Econometrics 193*, 35–53.

Wu, W. B. et al. (2005). On the bahadur representation of sample quantiles for dependent sequences. *The Annals of Statistics 33*(4), 1934–1963.

Zhou, Q. and J. Yu (2015). Asymptotic theory for linear diffusions under alternative sampling schemes. *Economics Letters 128*, 1–5.

Appendix A: Proofs

Proof Proposition 2.1

Proof. The bias of realized volatility is

$$\mathbb{E}\left(\sum_{j=i-K+1}^{i} r_{t_j}^2 - K\sigma^2\Delta\right) = \sum_{j=i-K+1}^{i} \mathbb{E}(r_{t_j}^2) - K\sigma^2\Delta = K\mu^2\Delta^2$$

because

$$\begin{aligned}\mathbb{E}(r_{t_j}^2) &= \mathbb{E}\left(\mu\Delta + \sigma\sqrt{\Delta}\varepsilon_{t_j}\right)^2 = \mathbb{E}\left(\mu^2\Delta^2 + \sigma^2\Delta\varepsilon_{t_j}^2 + 2\mu\Delta\sigma\sqrt{\Delta}\varepsilon_{t_j}\right) \\ &= \mu^2\Delta^2 + \sigma^2\Delta\,\mathbb{E}\,\varepsilon_{t_j}^2 + 2\mu\Delta\sigma\sqrt{\Delta}\,\mathbb{E}\,\varepsilon_{t_j} \\ &= \mu^2\Delta^2 + \sigma^2\Delta.\end{aligned}$$

■

The Proof of Lemma 2.1

Proof. The proof of unbiasedness of the sample mean is straightforward. We focus on the proof of the unbiased property of the median under the DGP (2).

The estimation of the median is equivalent to running a quantile regression with the 50% quantile and only a constant regressor. Let $Q_r(\tau) = \xi_\tau$ be the τ^{th} quantile of r_{t_i} and $\rho_\tau(u)$ denote the check function:

$$\rho_\tau(u) = u\left(\tau - 1_{(u<0)}\right)$$

as in Koenker and Bassett (1978). The quantile ξ_τ can be estimated by solving the problem below:

$$\xi_\tau \in \underset{\xi}{\operatorname{argmin}}\,\mathbb{E}\left[\rho_\tau(r-\xi)\right] = \mathbb{E}\left[(r-\xi)\left(\tau - 1_{(r<\xi)}\right)\right],$$

or equivalently by solving the first order condition $\mathbb{E}[1_{(r<\xi_\tau)}] = \tau$. For a sample containing K observations (i.e., $r_{t_1}, \ldots, r_{t_K}$), the first order condition implies

$$g(\xi_\tau) = \frac{1}{K}\sum_{i=1}^{K} 1_{(r_{t_i}<\xi_\tau)} - \tau = 0.$$

The solution $\hat{\xi}_\tau$ satisfies the condition $g\left(\hat{\xi}_\tau\right) = 0$.

It is obvious that the gradient function $g(.)$ is monotonically increasing with ξ_τ. Therefore,

$$\Pr\left(\hat{\xi}_\tau > \xi_\tau\right) = \Pr\left[g(\xi_\tau) < 0\right] = \Pr\left[\sum_{i=1}^{K} 1_{(r_{t_i}<\xi_\tau)} < K\tau\right].$$

Let $X = \sum_{i=1}^{K} 1_{\left(r_{t_i} < \xi_\tau\right)}$ be the number of successes (i.e. $r_{t_i} < \xi_\tau$) among K trials. Since $r_{t_i} \overset{i.i.d.}{\sim} N(\mu\Delta, \sigma^2\Delta)$, we have

$$F_r\left(\xi_\tau\right) = \Pr\left(r_{t_i} < \xi_\tau\right) = \Phi\left(\frac{\xi_\tau - \mu\Delta}{\sigma\sqrt{\Delta}}\right) \text{ for all } i, \tag{34}$$

where $\Phi\left(.\right)$ is the CDF of the standard normal distribution. Therefore, X follows a binomial distribution with probability $F_r\left(\xi_\tau\right)$, which is denoted by $X \sim Bin(K, F_r\left(\xi_\tau\right))$. Let $p = \lceil K\tau \rceil$ be the smallest integer that is greater than $K\tau$. The cumulative distribution of $\hat{\xi}_\tau$ takes the form of a regularized incomplete beta function $I_{F_r(\xi_\tau)}\left(p, K-p+1\right)$. That is,

$$\begin{aligned}
\Pr\left(\hat{\xi}_\tau \leq \xi_\tau\right) &= 1 - \Pr\left[\sum_{i=1}^{K} 1_{\left(r_{t_i} < \xi_\tau\right)} < K\tau\right] \\
&= 1 - \Pr(X \leq p-1) \\
&= I_{F_r(\xi_\tau)}\left(p, K-p+1\right).
\end{aligned}$$

Since $F_r\left(\xi_\tau\right) = \Phi\left(\frac{\xi_\tau - \mu\Delta}{\sigma\sqrt{\Delta}}\right)$, we have $\hat{\xi}_\tau$ follows a beta-normal distribution (Eugene et al., 2002) such that $\hat{\xi}_\tau \sim BN(p, K-p+1, \mu\Delta, \sigma\sqrt{\Delta})$.

We denote the median by $\hat{m}^*\left(K\right) := \hat{\xi}_{0.5}$. If $\tau = 0.5$, $p \simeq K - p + 1$. Therefore, $\hat{m}^*\left(K\right)$ is a symmetric beta-normal distribution with mean $\mu\Delta$. That is,

$$E\left[\hat{m}^*\left(K\right)\right] = \mu\Delta.$$

■

Proof of Proposition 2.2

Proof. (i) The sample mean of K log returns is

$$\hat{m}\left(K\right) = \frac{1}{K}\sum_{j=i-K+1}^{i} r_{t_j} = \frac{1}{K}\sum_{j=i-K+1}^{i}\left(\mu\Delta + \sigma\sqrt{\Delta}\varepsilon_{t_j}\right) = \mu\Delta + \sigma\sqrt{\Delta}\frac{1}{K}\sum_{j=i-K+1}^{i}\varepsilon_{t_j}.$$

Since $\varepsilon_{t_j} \overset{i.i.d.}{\sim} N\left(0,1\right)$, the linear combination $\sum_{j=i-K+1}^{i}\varepsilon_{t_j} \sim N\left(0, K\right)$ and hence

$$\hat{m}\left(K\right) \sim N\left(\mu\Delta, \frac{1}{K}\sigma^2\Delta\right).$$

The bias of $RV^{\dagger}$ is therefore

$$\begin{aligned}
& \mathbb{E}\left\{\sum_{j=i-K+1}^{i}\left[r_{t_j}-\hat{m}\left(K\right)\right]^2-K\sigma^2\Delta\right\} \\
= & \sum_{j=i-K+1}^{i}\mathbb{E}\left[r_{t_j}^2+\hat{m}\left(K\right)^2-2r_{t_j}\hat{m}\left(K\right)\right]-K\sigma^2\Delta \\
= & \left(\sum_{j=i-K+1}^{i}\mathbb{E}\,r_{t_j}^2-K\sigma^2\Delta\right)+K\,\mathbb{E}\,\hat{m}\left(K\right)^2-2\,\mathbb{E}\left[\left(\sum_{j=i-K+1}^{i}r_{t_j}\right)\hat{m}\left(K\right)\right] \\
= & K\mu^2\Delta^2-K\,\mathbb{E}\left(\hat{m}\left(K\right)^2\right) \\
= & -K\,\mathbb{V}\left(\hat{m}\left(K\right)\right)=-\sigma^2\Delta.
\end{aligned}$$

(ii) The bias of RV^{*} is

$$\begin{aligned}
& \mathbb{E}\left\{\sum_{j=i-K+1}^{i}\left(r_{t_j}-\hat{m}^{*}\left(K\right)\right)^2-K\sigma^2\Delta\right\} \\
= & \sum_{j=i-K+1}^{i}\mathbb{E}\left[r_{t_j}^2+\hat{m}^{*}\left(K\right)^2-2r_{t_j}\hat{m}^{*}\left(K\right)\right]-K\sigma^2\Delta \\
= & \left(\sum_{j=i-K+1}^{i}\mathbb{E}\,r_{t_j}^2-K\sigma^2\Delta\right)+K\,\mathbb{E}\,\hat{m}^{*}\left(K\right)^2-2\,\mathbb{E}\left[\left(\sum_{j=i-K+1}^{i}r_{t_j}\right)\hat{m}^{*}\left(K\right)\right] \\
= & K\mu^2\Delta^2+K\,\mathbb{E}\left(\hat{m}^{*}\left(K\right)^2\right)-2K\,\mathbb{E}\left[\hat{m}\left(K\right)\hat{m}^{*}\left(K\right)\right] \\
= & K\left[\mathbb{V}\left(\hat{m}^{*}\left(K\right)\right)-2\,\mathrm{cov}\left(\hat{m}\left(K\right),\hat{m}^{*}\left(K\right)\right)\right].
\end{aligned}$$

■

Proof of Proposition 2.3

Proof. (i) Since $r_{t_i}\overset{i.i.d.}{\sim}N(\mu\Delta,\sigma^2\Delta)$, $|r_{t_i}|$ follows a folded normal distribution and

$$\mathbb{E}\left|r_{t_i}\right|=\sigma\sqrt{\frac{2}{\pi}\Delta}e^{-\frac{\mu^2\Delta}{2\sigma^2}}+\mu\Delta\left[1-2\Phi\left(-\frac{\mu\sqrt{\Delta}}{\sigma}\right)\right].$$

The bias of bipower variation is

$$\begin{aligned}
&\mathbb{E}\left[\frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\left|r_{t_j}\right|\left|r_{t_{j-1}}\right| - K\sigma^2\Delta\right] \\
= \;& \frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\mathbb{E}\left|r_{t_j}\right|\mathbb{E}\left|r_{t_{j-1}}\right| - K\sigma^2\Delta \\
= \;& \frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\left\{\sigma\sqrt{\frac{2}{\pi}\Delta}e^{-\frac{\mu^2\Delta}{2\sigma^2}} + \mu\Delta\left[1-2\Phi\left(-\frac{\mu\sqrt{\Delta}}{\sigma}\right)\right]\right\}^2 - K\sigma^2\Delta \\
= \;& K\sigma^2\Delta\left(e^{-\frac{\mu^2\Delta}{\sigma^2}}-1\right) + \frac{\pi}{2}K\mu^2\Delta^2\left[1-2\Phi\left(-\frac{\mu\sqrt{\Delta}}{\sigma}\right)\right]^2 \\
& +2K\mu\Delta^{3/2}\sigma\sqrt{\frac{\pi}{2}}e^{-\frac{\mu^2\Delta}{2\sigma^2}}\left[1-2\Phi\left(-\frac{\mu\sqrt{\Delta}}{\sigma}\right)\right].
\end{aligned}$$

(ii) The bias of the modified bipower variation $BV^{\dagger}$ is

$$\begin{aligned}
&\mathbb{E}\left(\frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\left|r_{t_j}-\hat{m}_{t_i}(K)\right|\left|r_{t_{j-1}}-\hat{m}_{t_i}(K)\right|\right) - K\sigma^2\Delta \\
= \;& \frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\mathbb{E}\left|r_{t_j}-\hat{m}_{t_i}(K)\right|\left|r_{t_{j-1}}-\hat{m}_{t_i}(K)\right| - K\sigma^2\Delta \\
\leq \;& \frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\sqrt{\mathbb{E}\left|r_{t_j}-\hat{m}_{t_i}(K)\right|^2}\sqrt{\mathbb{E}\left|r_{t_{j-1}}-\hat{m}_{t_i}(K)\right|^2} - K\sigma^2\Delta \\
= \;& \frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\mathbb{E}\left[r_{t_j}-\hat{m}_{t_i}(K)\right]^2 - K\sigma^2\Delta
\end{aligned}$$

by Hölder's inequality. Since, by definition,

$$r_{t_j}-\hat{m}_{t_i}(K) = \sigma\sqrt{\Delta}\left[\varepsilon_{t_j} - \frac{1}{K}\sum_{s=i-K+1}^{i}\varepsilon_{t_s}\right] = \sigma\sqrt{\Delta}\left[\frac{K-1}{K}\varepsilon_{t_j} - \frac{1}{K}\sum_{s=i-K+1,s\neq j}^{i}\varepsilon_{t_s}\right],$$

we have

$$\mathbb{E}\left[r_{t_j}-\hat{m}_{t_i}(K)\right]^2 = \frac{K-1}{K}\sigma^2\Delta.$$

The bias of $BV^{\dagger}$ is, therefore,

$$\mathbb{E}\left(\frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\left|r_{t_j}-\hat{m}_{t_i}(K)\right|\left|r_{t_{j-1}}-\hat{m}_{t_i}(K)\right|\right)-K\sigma^2\Delta$$

$$\leq \frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\frac{K-1}{K}\sigma^2\Delta=\left(\frac{\pi}{2}-1\right)K\sigma^2\Delta-\frac{\pi}{2}\sigma^2\Delta.$$

(iii) Similarly, by Hölder's inequality, the bias of the modified bipower variation computed on centered (using the median) log returns is

$$\mathbb{E}\left(\frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\left|r_{t_j}-\hat{m}^*(K)\right|\left|r_{t_{j-1}}-\hat{m}^*(K)\right|\right)-K\sigma^2\Delta$$

$$\leq \frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\mathbb{E}\left[r_{t_j}-\hat{m}^*(K)\right]^2-K\sigma^2\Delta.$$

Since $\mathbb{E}\,\hat{m}^*(K)=\mu\Delta$, the expected value of $r_{t_j}-\hat{m}^*(K)$ is zero and

$$\begin{aligned}\mathbb{E}\left[r_{t_j}-\hat{m}^*(K)\right]^2 &= \mathbb{V}\left[r_{t_j}-\hat{m}^*(K)\right]\\ &= \mathbb{V}\left(r_{t_j}\right)+\mathbb{V}\left(\hat{m}^*(K)\right)-2\left[\mathbb{E}\left(r_{t_j}\hat{m}^*(K)\right)-\mu^2\Delta^2\right]\\ &= \sigma^2\Delta+\mathbb{V}\left(\hat{m}^*(K)\right)-2\left[\mathbb{E}\left(r_{t_j}\hat{m}^*(K)\right)-\mu^2\Delta^2\right].\end{aligned}$$

Consequently,

$$\sum_{j=i-K+2}^{i}\mathbb{E}\left[r_{t_j}-\hat{m}^*(K)\right]^2$$

$$= \sum_{j=i-K+2}^{i}\left\{\sigma^2\Delta+\mathbb{V}\left(\hat{m}^*(K)\right)-2\left[\mathbb{E}\left(r_{t_j}\hat{m}^*(K)\right)-\mu^2\Delta^2\right]\right\}$$

$$= (K-1)\sigma^2\Delta+(K-1)\mathbb{V}\left(\hat{m}^*(K)\right)-2\sum_{j=i-K+2}^{i}\mathbb{E}\left(r_{t_j}\hat{m}^*(K)\right)+2(K-1)\mu^2\Delta^2$$

$$= (K-1)\sigma^2\Delta+(K-1)\mathbb{V}\left(\hat{m}^*(K)\right)-2(K-1)\mathbb{E}\left[\hat{m}(K)\hat{m}^*(K)\right]+2(K-1)\mu^2\Delta^2$$

$$= \quad (K-1)\,\sigma^2\Delta + (K-1)\,\mathbb{V}\left(\hat{m}^*(K)\right) - 2\,(K-1)\,\mathrm{cov}\left(\hat{m}(K),\hat{m}^*(K)\right).$$

Therefore,

$$\mathbb{E}\left(\frac{\pi}{2}\frac{K}{K-1}\sum_{j=i-K+2}^{i}\left|r_{t_j}-\hat{m}^*(K)\right|\left|r_{t_{j-1}}-\hat{m}^*(K)\right|\right) - K\sigma^2\Delta$$

$$\leq \quad \frac{\pi}{2}\frac{K}{K-1}\left[(K-1)\,\sigma^2\Delta + (K-1)\,\mathbb{V}\left(\hat{m}^*(K)\right) - 2\,(K-1)\,\mathrm{cov}\left(\hat{m}(K),\hat{m}^*(K)\right)\right] - K\sigma^2\Delta$$

$$= \quad \left(\frac{\pi}{2}-1\right)K\sigma^2\Delta + \frac{\pi}{2}K\left[\mathbb{V}\left(\hat{m}^*(K)\right) - 2\,\mathrm{cov}\left(\hat{m}(K),\hat{m}^*(K)\right)\right].$$

■

The proof of Lemma 3.1

Proof. (i) The error term

Replacing y_t in the drift coefficient μ_t with (10), we have

$$\mu_t = a(t) + \int_0^t b(t,s)\sigma_s dW_s,$$

where $a(t) = \theta\,(y_0-\rho)\,e^{\theta t}$ and $b(t,s) = \theta e^{\theta(t-s)}$. The process y_t can be rewritten as

$$\begin{aligned} y_t &= y_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s \\ &= y_0 + A(t) + \int_0^t B(t,s)\sigma_s dW_s + \int_0^t \sigma_s dW_s, \end{aligned}$$

where $A(t) = \int_0^t a(r)dr$ and $B(t,s) = \int_s^t b(r,s)dr$ with $t \geq s$. The log return r_{t_i} follows a dynamic of

$$\begin{aligned} r_{t_i} &= \Delta_i A + \int_{t_{i-1}}^{t_i}\sigma_s dW_s + \int_{t_{i-1}}^{t_i} B(t_i,s)\sigma_s dW_s + \int_0^{t_{i-1}}[B(t_i,s)-B(t_{i-1},s)]\sigma_s dW_s \qquad (35) \\ &= \int_{t_{i-1}}^{t_i}\sigma_s dW_s\{1+o_p(1)\}, \end{aligned}$$

where $\Delta_i A = \int_{t_{i-1}}^{t_i} a(r)dr$. The second equality arises from the fact that $\Delta_i A = O_p(\Delta)$, $\int_{t_{i-1}}^{t_i}\sigma_s dW_s = O_p(\Delta^{1/2})$, and the third and fourth terms on the right-hand side of the equation are $O_p(\Delta^{3/2})$.

■

Proof of Proposition 3.1

Proof. The bias of realized volatility is

$$\mathbb{E}\left[\sum_{j=i-K+1}^{i} r_{t_j}^2 - \int_{t_i-K}^{t_i} \sigma_u^2 du\right] = \sum_{j=i-K+1}^{i} \mathbb{E}\, r_{t_j}^2 - \mathbb{E}\left(\int_{t_i-K}^{t_i} \sigma_u^2 du\right).$$

From Lemma 3.1,

$$\begin{aligned} r_{t_i} &= \Delta_i A + \int_{t_{i-1}}^{t_i} \sigma_s dW_s + \int_{t_{i-1}}^{t_i} B(t_i, s)\sigma_s dW_s + \int_0^{t_{i-1}} [B(t_i, s) - B(t_{i-1}, s)]\sigma_s dW_s \quad (36) \\ &= \int_{t_{i-1}}^{t_i} a(r)dr + e^{\theta t_i} \int_{t_{i-1}}^{t_i} e^{-\theta s}\sigma_s dW_s + e^{\theta t_{i-1}} \left(e^{\theta\Delta} - 1\right)\left(\int_0^{t_{i-1}} e^{-\theta s}\sigma_s dW_s\right). \quad (37) \end{aligned}$$

since, by construction,

$$\begin{aligned} B(t_j, s) &= \int_s^{t_j} \theta e^{\theta(r-s)} dr = e^{\theta(t_j - s)} - 1, \\ B(t_j, s) - B(t_{j-1}, s) &= e^{\theta(t_{j-1} - s)} \left(e^{\theta\Delta} - 1\right). \end{aligned}$$

It is straightforward that when $\theta = 0$, the bias of the realized volatility is zero. We focus on the case of $\theta \neq 0$ in the subsequent derivations. The expected value of the squared return is

$$\mathbb{E}\, r_{t_i}^2 = \left(\int_{t_{i-1}}^{t_i} a(r)dr\right)^2 + e^{2\theta t_i}\, \mathbb{E}\left(\int_{t_{i-1}}^{t_i} e^{-\theta s}\sigma_s dW_s\right)^2 + e^{2\theta t_{i-1}} \left(e^{\theta\Delta} - 1\right)^2 \mathbb{E}\left(\int_0^{t_{i-1}} e^{-\theta s}\sigma_s dW_s\right)^2.$$

The first term can be rewritten as

$$\left(\int_{t_{i-1}}^{t_i} a(r)dr\right)^2 = (y_0 - \rho)^2 \left(e^{\theta t_i} - e^{\theta t_{i-1}}\right)^2.$$

Since by Ito's lemma,

$$\left(\int_{t_{i-1}}^{t_i} e^{-\theta s}\sigma_s dW_s\right)^2 = 2\int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{u} e^{-\theta s}\sigma_s dW_s\right) e^{-\theta u}\sigma_u dW_u + \int_{t_{i-1}}^{t_i} e^{-2\theta u}\sigma_u^2 du, \quad (38)$$

the second term

$$e^{2\theta t_i}\,\mathbb{E}\left(\int_{t_{i-1}}^{t_i} e^{-\theta s}\sigma_s dW_s\right)^2 = e^{2\theta t_i}\int_{t_{i-1}}^{t_i} e^{-2\theta u}\,\mathbb{E}\left(\sigma_u^2\right)du = d_0\frac{1}{2\theta}\left(e^{2\theta\Delta}-1\right).$$

The third term

$$\begin{aligned} e^{2\theta t_{i-1}}\left(e^{\theta\Delta}-1\right)^2\mathbb{E}\left(\int_0^{t_{i-1}} e^{-\theta s}\sigma_s dW_s\right)^2 &= e^{2\theta t_{i-1}}\left(e^{\theta\Delta}-1\right)^2\int_0^{t_{i-1}} e^{-2\theta u}\,\mathbb{E}\left(\sigma_u^2\right)du \\ &= \left(e^{\theta\Delta}-1\right)^2\frac{d_0}{2\theta}\left(e^{2\theta t_{i-1}}-1\right). \end{aligned}$$

Therefore,

$$\mathbb{E}\,r_{t_i}^2 = (y_0-\rho)^2\left(e^{\theta t_j}-e^{\theta t_{j-1}}\right)^2 + d_0\frac{1}{2\theta}\left(e^{2\theta\Delta}-1\right) + \left(e^{\theta\Delta}-1\right)^2\frac{d_0}{2\theta}\left(e^{2\theta t_{i-1}}-1\right)$$

and

$$\begin{aligned} &\sum_{j=i-K+1}^{i}\mathbb{E}\,r_{t_j}^2 \\ =\;& (y_0-\rho)^2\left(e^{\theta\Delta}-1\right)^2\sum_{j=i-K+1}^{i} e^{2\theta t_{j-1}} + d_0\frac{K}{2\theta}\left(e^{2\theta\Delta}-1\right) + \left(e^{\theta\Delta}-1\right)^2\frac{d_0}{2\theta}\left(\sum_{j=i-K+1}^{i} e^{2\theta t_{j-1}} - K\right) \\ =\;& (y_0-\rho)^2\left(e^{\theta\Delta}-1\right)^2 e^{2\theta t_{i-K}}\frac{1-e^{2\theta\Delta K}}{1-e^{2\theta\Delta}} + d_0\frac{K}{\theta}\left(e^{\theta\Delta}-1\right) + \left(e^{\theta\Delta}-1\right)^2\frac{d_0}{2\theta}e^{2\theta t_{i-K}}\frac{1-e^{2\theta\Delta K}}{1-e^{2\theta\Delta}}. \end{aligned}$$

Furthermore, we have

$$\mathbb{E}\left(\int_{t_i-K}^{t_i}\sigma_u^2 du\right) = \left(\int_{t_i-K}^{t_i}\mathbb{E}\left(\sigma_u^2\right)du\right) = d_0K\Delta.$$

Therefore, the bias of the realized volatility is zero when $\theta = 0$ and when $\theta \neq 0$

$$\begin{aligned} \mathbb{E}\left[\sum_{j=i-K+1}^{i} r_{t_j}^2 - \int_{t_i-K}^{t_i}\sigma_u^2 du\right] &= (y_0-\rho)^2\left(e^{\theta\Delta}-1\right)^2 e^{2\theta t_{i-K}}\frac{1-e^{2\theta\Delta K}}{1-e^{2\theta\Delta}} + \frac{d_0K}{\theta}\left(e^{\theta\Delta}-1\right) \\ &\quad + \left(e^{\theta\Delta}-1\right)^2\frac{d_0}{2\theta}e^{2\theta t_{i-K}}\frac{1-e^{2\theta\Delta K}}{1-e^{2\theta\Delta}} - d_0K\Delta \\ &\equiv \mathcal{E}. \end{aligned}$$

■

Proof of Proposition 3.2

Proof. (1) For notational simplicity, we write $\hat{m}_{t_i}(K)$ as $\hat{m}_{t_i}$. The bias of the modified realized variance estimator $RV^{\dagger}$ is

$$\begin{aligned}
& \mathbb{E}\left\{\sum_{j=i-K+1}^{i}\left(r_{t_j}-\hat{m}_{t_i}\right)^2-\int_{t_i-K}^{t_i}\sigma_u^2du\right\} \\
= \; & \mathbb{E}\left\{\sum_{j=i-K+1}^{i}r_{t_j}^2-2\hat{m}_{t_i}\sum_{j=i-K+1}^{i}r_{t_j}+K\hat{m}_{t_i}^2-\int_{t_i-K}^{t_i}\sigma_u^2du\right\} \\
= \; & \mathbb{E}\left(\sum_{j=i-K+1}^{i}r_{t_j}^2-\int_{t_i-K}^{t_i}\sigma_u^2du\right)-2K\,\mathbb{E}\left(\hat{m}_{t_i}^2\right)+K\,\mathbb{E}(\hat{m}_{t_i}^2) \\
= \; & \mathbb{E}\left(\sum_{j=i-K+1}^{i}r_{t_j}^2-\int_{t_i-K}^{t_i}\sigma_u^2du\right)-K\,\mathbb{E}\left(\hat{m}_{t_i}^2\right).
\end{aligned}$$

From Lemma 3.1,

$$\sum_{j=i-K+1}^{i}r_{t_j}=\int_{t_i-K}^{t_i}a(r)dr+\sum_{j=i-K+1}^{i}e^{\theta t_j}\int_{t_{j-1}}^{t_j}e^{-\theta s}\sigma_s dW_s+\left(e^{\theta\Delta}-1\right)\sum_{j=i-K+1}^{i}e^{\theta t_{j-1}}\left(\int_0^{t_{j-1}}e^{-\theta s}\sigma_s dW_s\right).$$

The second term of $\sum_{j=i-K+1}^{i}r_{t_j}$

$$\begin{aligned}
& \sum_{j=i-K+1}^{i}e^{\theta t_j}\int_{t_{j-1}}^{t_j}e^{-\theta s}\sigma_s dW_s \\
= \; & e^{\theta t_{i-K+1}}\int_{t_{i-K}}^{t_{i-K+1}}e^{-\theta s}\sigma_s dW_s+e^{\theta t_{i-K+2}}\int_{t_{i-K+1}}^{t_{i-K+2}}e^{-\theta s}\sigma_s dW_s+\cdots+e^{\theta t_i}\int_{t_{i-1}}^{t_i}e^{-\theta s}\sigma_s dW_s.
\end{aligned}$$

The third term of $\sum_{j=i-K+1}^{i}r_{t_j}$

$$\begin{aligned}
& \left(e^{\theta\Delta}-1\right)\sum_{j=i-K+1}^{i}e^{\theta t_{j-1}}\left(\int_0^{t_{j-1}}e^{-\theta s}\sigma_s dW_s\right) \\
= \; & \left(e^{\theta\Delta}-1\right)\left[e^{\theta t_{i-K}}\int_0^{t_{i-K}}e^{-\theta s}\sigma_s dW_s+e^{\theta t_{i-K+1}}\int_0^{t_{i-K+1}}e^{-\theta s}\sigma_s dW_s+\ldots+e^{\theta t_{i-1}}\int_0^{t_{i-1}}e^{-\theta s}\sigma_s dW_s\right]
\end{aligned}$$

$$
\begin{aligned}
&= \left(e^{\theta\Delta}-1\right)e^{\theta t_{i-K}}\int_0^{t_{i-K}} e^{-\theta s}\sigma_s dW_s + \left(e^{\theta\Delta}-1\right)e^{\theta t_{i-K+1}}\left[\int_0^{t_{i-K}} e^{-\theta s}\sigma_s dW_s + \int_{t_{i-K}}^{t_{i-K+1}} e^{-\theta s}\sigma_s dW_s\right] + \ldots \\
&\quad + \left(e^{\theta\Delta}-1\right)e^{\theta t_{i-1}}\left[\int_0^{t_{i-K}} e^{-\theta s}\sigma_s dW_s + \int_{t_{i-K}}^{t_{i-K+1}} e^{-\theta s}\sigma_s dW_s + \ldots + \int_{t_{i-2}}^{t_{i-1}} e^{-\theta s}\sigma_s dW_s\right] \\
&= \left(e^{\theta\Delta}-1\right)\left(e^{\theta t_{i-K}} + e^{\theta t_{i-K+1}} + \ldots + e^{\theta t_{i-1}}\right)\int_0^{t_{i-K}} e^{-\theta s}\sigma_s dW_s + \\
&\quad \left(e^{\theta\Delta}-1\right)\left(e^{\theta t_{i-K+1}} + e^{\theta t_{i-K+2}}\ldots + e^{\theta t_{i-1}}\right)\int_{t_{i-K}}^{t_{i-K+1}} e^{-\theta s}\sigma_s dW_s + \ldots + \left(e^{\theta\Delta}-1\right)e^{\theta t_{i-1}}\int_{t_{i-2}}^{t_{i-1}} e^{-\theta s}\sigma_s dW_s \\
&= e^{\theta t_{i-K}}\left(e^{\theta\Delta K}-1\right)\int_0^{t_{i-K}} e^{-\theta s}\sigma_s dW_s + e^{\theta t_{i-K+1}}\left(e^{\theta\Delta(K-1)}-1\right)\int_{t_{i-K}}^{t_{i-K+1}} e^{-\theta s}\sigma_s dW_s \\
&\quad + \ldots + e^{\theta t_{i-1}}\left(e^{\theta\Delta}-1\right)\int_{t_{i-2}}^{t_{i-1}} e^{-\theta s}\sigma_s dW_s.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{j=i-K+1}^{i} r_{t_j}\right)^2 \\
&= \left(\int_{t_{i-K}}^{t_i} a(r)dr\right)^2 + \mathbb{E}\left[\sum_{j=i-K+1}^{i} e^{\theta t_j}\int_{t_{j-1}}^{t_j} e^{-\theta s}\sigma_s dW_s\right]^2 \\
&\quad + \left(e^{\theta\Delta}-1\right)^2\mathbb{E}\left[\sum_{j=i-K+1}^{i} e^{\theta t_{j-1}}\left(\int_0^{t_{j-1}} e^{-\theta s}\sigma_s dW_s\right)\right]^2 \\
&\quad + 2\left[e^{2\theta t_{i-K+1}}\left(e^{\theta\Delta(K-1)}-1\right)\mathbb{E}\left(\int_{t_{i-K}}^{t_{i-K+1}} e^{-\theta s}\sigma_s dW_s\right)^2 + \cdots e^{2\theta t_{i-1}}\left(e^{\theta\Delta}-1\right)\mathbb{E}\left(\int_{t_{i-2}}^{t_{i-1}} e^{-\theta s}\sigma_s dW_s\right)^2\right].
\end{aligned}
$$

Next, we derive, one-by-one, the theoretical properties of terms on the right-hand side of the above equation. The first term

$$
\left(\int_{t_{i-K}}^{t_i} a(r)dr\right)^2 = (y_0-\rho)^2 e^{2\theta t_{i-K}}\left(e^{\theta K\Delta}-1\right)^2.
$$

The second term

$$\begin{aligned}\mathbb{E}\left[\sum_{j=i-K+1}^{i} e^{2\theta t_j}\left(\int_{t_{j-1}}^{t_j} e^{-\theta s}\sigma_s dW_s\right)\right]^2 &= \sum_{j=i-K+1}^{i} e^{2\theta t_j}\,\mathbb{E}\left(\int_{t_{j-1}}^{t_j} e^{-\theta s}\sigma_s dW_s\right)^2 \\ &= d_0 \sum_{j=i-K+1}^{i} e^{2\theta t_j}\int_{t_{j-1}}^{t_j} e^{-2\theta u}du \\ &= \frac{d_0}{2\theta}K\left(e^{2\theta\Delta}-1\right).\end{aligned}$$

The third term

$$\begin{aligned}&\left(e^{\theta\Delta}-1\right)^2\mathbb{E}\left[\sum_{j=i-K+1}^{i} e^{\theta t_{j-1}}\left(\int_0^{t_{j-1}} e^{-\theta s}\sigma_s dW_s\right)\right]^2 \\ =\;& e^{2\theta t_{i-K}}\left(e^{\theta\Delta K}-1\right)^2\mathbb{E}\left(\int_0^{t_{i-K}} e^{-\theta s}\sigma_s dW_s\right)^2 + e^{2\theta t_{i-K+1}}\left(e^{\theta\Delta(K-1)}-1\right)^2\mathbb{E}\left(\int_{t_{i-K}}^{t_{i-K+1}} e^{-\theta s}\sigma_s dW_s\right)^2 \\ &+\ldots+e^{2\theta t_{i-1}}\left(e^{\theta\Delta}-1\right)^2\mathbb{E}\left(\int_{t_{i-2}}^{t_{i-1}} e^{-\theta s}\sigma_s dW_s\right)^2 \\ =\;& -\frac{d_0}{2\theta}\left[\left(e^{\theta\Delta K}-1\right)^2\left(1-e^{2\theta t_{i-K}}\right)+\left(e^{\theta\Delta(K-1)}-1\right)^2\left(1-e^{2\theta\Delta}\right)+\ldots+\left(e^{\theta\Delta}-1\right)^2\left(1-e^{2\theta\Delta}\right)\right] \\ =\;& -\frac{d_0}{2\theta}\left(e^{\theta\Delta K}-1\right)^2\left(1-e^{2\theta t_{i-K}}\right)-\frac{d_0}{2\theta}\left(1-e^{2\theta\Delta}\right)\left[\left(e^{\theta\Delta(K-1)}-1\right)^2+\ldots+\left(e^{\theta\Delta}-1\right)^2\right] \\ =\;& -\frac{d_0}{2\theta}\left(e^{\theta\Delta K}-1\right)^2\left(1-e^{2\theta t_{i-K}}\right)-\frac{d_0}{2\theta}\left(1-e^{2\theta\Delta}\right)\left[(K-1)-2e^{\theta\Delta}\frac{e^{\theta\Delta(K-1)}-1}{e^{\theta\Delta}-1}+e^{2\theta\Delta}\frac{e^{2\theta\Delta(K-1)}-1}{e^{2\theta\Delta}-1}\right].\end{aligned}$$

The fourth term

$$\begin{aligned}&2\left[e^{2\theta t_{i-K+1}}\left(e^{\theta\Delta(K-1)}-1\right)\mathbb{E}\left(\int_{t_{i-K}}^{t_{i-K+1}} e^{-\theta s}\sigma_s dW_s\right)^2+\cdots e^{2\theta t_{i-1}}\left(e^{\theta\Delta}-1\right)\mathbb{E}\left(\int_{t_{i-2}}^{t_{i-1}} e^{-\theta s}\sigma_s dW_s\right)^2\right] \\ =\;& -\frac{d_0}{\theta}\left(1-e^{2\theta\Delta}\right)\left[\left(e^{\theta\Delta(K-1)}-1\right)+\cdots+\left(e^{\theta\Delta}-1\right)\right] \\ =\;& -\frac{d_0}{\theta}\left(1-e^{2\theta\Delta}\right)\left[e^{\theta\Delta}\frac{e^{\theta\Delta(K-1)}-1}{e^{\theta\Delta}-1}-(K-1)\right].\end{aligned}$$

Therefore,

$$
\begin{aligned}
& \mathbb{E}\left(\sum_{j=i-K+1}^{i} r_{t_j}\right)^2 \\
= \; & (y_0-\rho)^2 e^{2\theta t_{i-K}}\left(e^{\theta K\Delta}-1\right)^2 + \frac{d_0}{2\theta}K\left(e^{2\theta\Delta}-1\right) - \frac{d_0}{2\theta}\left(e^{\theta\Delta K}-1\right)^2\left(1-e^{2\theta t_{i-K}}\right) \\
& -\frac{d_0}{2\theta}\left(1-e^{2\theta\Delta}\right)\left[(K-1)-2e^{\theta\Delta}\frac{e^{\theta\Delta(K-1)}-1}{e^{\theta\Delta}-1}+e^{2\theta\Delta}\frac{e^{2\theta\Delta(K-1)}-1}{e^{2\theta\Delta}-1}\right] \\
& -\frac{d_0}{\theta}\left(1-e^{2\theta\Delta}\right)\left[e^{\theta\Delta}\frac{e^{\theta\Delta(K-1)}-1}{e^{\theta\Delta}-1}-(K-1)\right] \\
= \; & (y_0-\rho)^2 e^{2\theta t_{i-K}}\left(e^{\theta K\Delta}-1\right)^2 - \frac{d_0}{2\theta}\left(e^{\theta\Delta K}-1\right)^2\left(1-e^{2\theta t_{i-K}}\right) - \frac{d_0}{2\theta}\left(1-e^{2\theta\Delta}\right)\frac{e^{2\theta\Delta K}-1}{e^{2\theta\Delta}-1}.
\end{aligned}
$$

Since

$$
K\,\mathbb{E}\left(\hat{m}_{t_i}^2\right) = \frac{1}{K}\,\mathbb{E}\left(\sum_{j=i-K+1}^{i} r_{t_j}\right)^2,
$$

the bias of the modified realized variance estimator $RV^{\dagger}$ is

$$
\begin{aligned}
& \mathbb{E}\left\{\sum_{j=i-K+1}^{i}\left(r_{t_j}-\hat{m}_{t_i}\right)^2 - \int_{t_i-K}^{t_i}\sigma_u^2 du\right\} = \mathcal{E} - K\,\mathbb{E}\left(\hat{m}_{t_i}^2\right) \\
= \; & \mathcal{E} - \frac{1}{K}(y_0-\rho)^2 e^{2\theta t_{i-K}}\left(e^{\theta K\Delta}-1\right)^2 + \frac{1}{K}\frac{d_0}{2\theta}\left(e^{\theta\Delta K}-1\right)^2\left(1-e^{2\theta t_{i-K}}\right) \\
& +\frac{d_0}{2\theta K}\left(1-e^{2\theta\Delta}\right)\frac{e^{2\theta\Delta K}-1}{e^{2\theta\Delta}-1}
\end{aligned}
$$

if $\theta \neq 0$.

(2) We write $\hat{m}_{t_i}^*(K)$ as $\hat{m}_{t_i}^*$ for simplicity. The bias of the modified realized variance estimator RV^* is

$$
\mathbb{E}\left\{\sum_{j=i-K+1}^{i}\left(r_{t_j}-\hat{m}_{t_i}^*\right)^2 - \int_{t_i-K}^{t_i}\sigma_u^2 du\right\}
$$

$$
\begin{aligned}
&= \mathbb{E}\left\{ \sum_{j=i-K+1}^{i} r_{t_j}^2 - 2\hat{m}_{t_i}^* \sum_{j=i-K+1}^{i} r_{t_j} + K\hat{m}_{t_i}^{*2} - \int_{t_i-K}^{t_i} \sigma_u^2 du \right\} \\
&= \mathbb{E}\left(\sum_{j=i-K+1}^{i} r_{t_j}^2 - \int_{t_i-K}^{t_i} \sigma_u^2 du \right) - 2\,\mathbb{E}\left(\hat{m}_{t_i}^* \sum_{j=i-K+1}^{i} r_{t_j} \right) + K\,\mathbb{E}(\hat{m}_{t_i}^{*2}) \\
&= \mathcal{E} - 2K\,\mathbb{E}\left(\hat{m}_{t_i}^* \hat{m}_{t_i}\right) + K\,\mathbb{E}(\hat{m}_{t_i}^{*2}).
\end{aligned}
$$

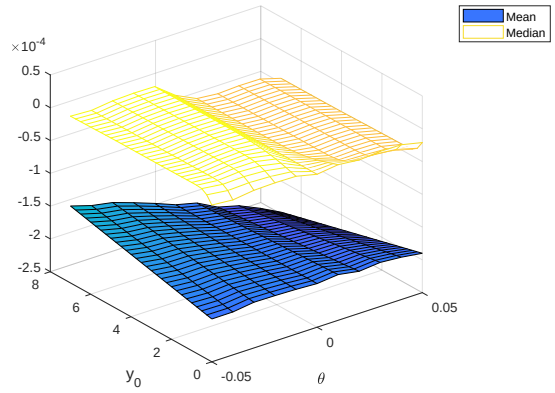
■

Appendix B: Sample Mean and Median

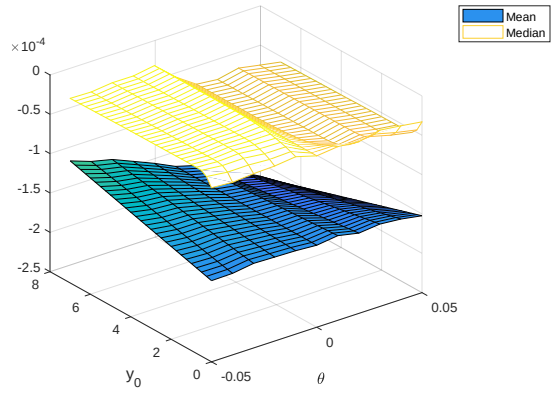
To compare the accuracy of $\hat{m}_{t_i}^*(K)$ and $\hat{m}_{t_i}(K)$ as estimators of the drift component in the presence of jumps, we consider the data generating process as in Section 3.1. Two different settings for jumps are considered. There is either one single large negative jump within a day with $\phi_{t_{i+1}}^1 = -1.5\sigma_{t_{i+1}}$ or two small jumps with $\phi_{t_{i+1}}^j = -0.6\sigma_{t_{i+1}}$ for $j = 1, 2$.[19] The remaining parameters are the same as in Section 3.1.

Figure 17: Bias of the sample mean and median in the presence of one single large jump (i.e., $\phi_{t_{i+1}}^1 = -1.5\sigma_{t_{i+1}}$) and two small jumps ($\phi_{t_{i+1}}^j = -0.6\sigma_{t_{i+1}}$ for $j = 1, 2$) at the 5-minute frequency.

(a) One large jump



(b) Two small jumps



The empirical mean and median of the last observation of each day, denoted by $\hat{m}_{t_T}(T)$ and $\hat{m}_{t_T}^*(T)$, respectively, are compared to the true value m_{t_T}. The superiority of the median is clear from Figure 17, which shows that the bias is systematically smaller for the median than for the sample

[19] The simulation results are qualitatively the same for positive jumps and jumps with a random sign.

mean for all combinations of parameters considered in the simulation. Therefore, we recommend the use of the median instead of the sample mean in empirical applications if the presence of jumps is suspected.