

Volatility Estimation and Jump Detection for drift-diffusion Processes

Sébastien Laurent
Shuping Shi

WP 2018 - Nr 43

Volatility Estimation and Jump Detection for drift-diffusion Processes*

Sébastien Laurent¹ and Shuping Shi²

¹Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE and Graduate
School of Management

²Macquarie University and CAMA

December 4, 2018

Abstract

Logarithms of prices of financial assets are conventionally assumed to follow drift-diffusion processes. While the drift term is typically ignored in the infill asymptotic theory and applications, the presence of nonzero drifts is an undeniable fact. The finite sample theory and extensive simulations provided in this paper reveal that the drift component has a nonnegligible impact on the estimation accuracy of volatility and leads to a dramatic power loss of a class of jump identification procedures. We propose an alternative construction of volatility estimators and jump tests and observe significant improvement of both in the presence of nonnegligible drift. 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.

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

JEL classification: C12, C14.

*The authors gratefully acknowledge the editors of this special issue, two anonymous referees as well as Yacine Ait-Sahalia, Peter B.C. Phillips, Orimar Sauri, and Jun Yu and participants at the econometric study group at the Singapore Management University and the Frontiers in Econometrics workshop at Macquarie University for helpful discussions. Sébastien Laurent, Aix-Marseille School of Economics, Aix-Marseille University; Email: sebastien.laurent@univ-amu.fr. Shuping Shi, Department of Economics, Macquarie University and the Centre for Applied Macroeconomic Analysis; Email: shuping.shi@mq.edu.au.

1 Introduction

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

$$dy_t = \mu_t dt + \sigma_t dW_t, \quad (1)$$

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 , 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 framework (such as monthly, weekly and daily), prolonged periods of mildly explosive trends were identified in many financial assets, for example, 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 last decade (Etienne et al., 2014; Gutierrez, 2012; Phillips and Yu, 2011). Evidence of a non-negligible drift was also observed by Phillips and Shi (2017) in log-prices of the S&P 500 index during the 2008 subprime mortgage crisis period and in bond yields and CDS spreads of most European countries during the 2010 debt crisis. Additionally, there is extensive literature documenting temporary deviations of log-prices from the random walk.¹ In the high-frequency framework, motivated by the large number of flash crashes,² Christensen et al. (2016) propose a drift bust hypothesis, which postulates the existence of short-lived locally explosive trends in log-prices. This hypothesis was tested using a nonparametric approach and with tick data. The authors find 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 ac-

¹See, 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).

²See, for example, Nanex Research <http://www.nanex.net/NxResearch/>.

curacy of volatility in the high-frequency setting, namely the realized volatility (Barndorff-Nielsen and Shephard, 2002), the bipower variation (Barndorff-Nielsen and Shephard, 2004), and the noise-robust volatility estimator (Podolskij and Vetter, 2009). We consider a linear drift-diffusion process, which captures an important form of nonzero drifts and has been studied extensively in the literature. 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). The discrete-time model of this process allows for both the random walk and local-to-unity dynamics³ (Phillips, 1987).

The theoretical literature on realized volatility merely focuses on the asymptotic properties and apart from some notable exceptions (Meddahi, 2002; Bandi and Russell, 2005) do not study their finite sample properties. One important contribution made by this paper is deriving the finite sample bias of the realized volatility under the linear drift diffusion process. The finite sample theory, together with extensive simulations, reveal that the nonzero drift causes a substantial bias in the volatility estimation. We propose a correction that leads to a significant improvement in the estimation accuracy.

Furthermore, we study the impact of nonzero drifts on the performance of 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 derivatives 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 jumps⁴ proposed independently by Andersen et al. (2007) and Lee and Mykland (2008, LM08 hereafter) and extended by Lee and Mykland (2012, LM12 hereafter) 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 demonstrate that the finite sample performance of the LM08 and LM12 tests is

³The local-to-unity process deviates mildly from the random walk in both stationary and explosive directions. It bridges the gap between a random walk and stationary/explosive processes.

⁴Lee and Hannig (2010) propose a test for the presence of infinite-activity jumps (i.e., a Levy jump diffusion process).

unsatisfactory if the drift is nonzero. Indeed, under this circumstance, the 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. Despite its ease of implementation, our test improves the finite sample performance significantly.

As an illustration, we apply the new bipower volatility 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 up to more than 40%. The proposed new jump test allows the identification of more jumps. These additional jumps occur during periods with upward or downward trends in log-prices.

2 The Drift-Diffusion Process

Let $\{0 < t_1 < \dots < t_T < N\}$ be a set of T equally spaced observation times spanning N days. The distance between two consecutive observation times is denoted by $\Delta = t_i - t_{i-1} = N/T$.

Assumption 2.1 *Assume that the drift and diffusion coefficients in (1) do not change dramatically over a short time interval such that*

$$(1) \sup_i \sup_{t_i \leq s \leq t_{i+1}} |\mu_s - \mu_{t_i}| = O_p(\Delta^{1/2-\epsilon})$$

$$(2) \sup_i \sup_{t_i \leq s \leq t_{i+1}} |\sigma_s - \sigma_{t_i}| = O_p(\Delta^{1/2-\epsilon})$$

for any $\epsilon \geq 0$.

These conditions are satisfied by most Ito processes and allow both the drift and stochastic volatility to depend on the log-price process itself. For the convenience of exposition, we focus on a linear drift-diffusion process with μ_t and σ_t specified as follows.

Assumption 2.2 *The drift coefficient in (1) takes the form of*

$$\mu_t = \theta(y_t - m) \text{ with constants } \theta \text{ and } m.$$

Note that the drift coefficient μ_t is a linear function of the log-price y_t and satisfies Assumption 2.1-(1) because for any $t_i \leq s \leq t_{i+1}$,

$$\begin{aligned} \mu_s - \mu_{t_i} &= \theta^2 \int_{t_i}^s y_t dt - \theta^2 m(s - t_i) + \theta \int_{t_i}^s \sigma_t dW_t \\ &= \theta \int_{t_i}^s \sigma_t dW_t \{1 + o_p(1)\} = O_p(\Delta^{1/2}). \end{aligned}$$

Assumption 2.3 *The diffusion coefficient is given by*

$$\sigma_t^2 = \sum_{i=0}^p d_i \Pi_{i,S_t},$$

where p is an integer number, the d_i 's are real numbers and the Π_{i,S_t} 's are the eigenfunctions of the infinitesimal generator⁵ associated to the state variable S_t whose dynamic is

$$dS_t = \mu(S_t) dt + \sigma(S_t) dW_t.$$

Assumption 2.3 includes the most popular volatility models such as the log-normal model, the squared root model and the GARCH(1,1) model of Nelson (1991) as special cases. See Meddahi (2002) for more details. Note that Assumption 2.3 implies that $E(\sigma_t^2) = d_0$.

The volatility process satisfies Assumption 2.1-(2) that $\sigma_s - \sigma_{t_i} = O_p(\Delta^{1/2})$ for any $t_i \leq s \leq t_{i+1}$. For example, the GARCH(1,1) volatility model is defined as

$$\sigma_t^2 = S_t \text{ and } dS_t = \kappa(\omega - S_t) dt + \sqrt{2\lambda\kappa} S_t dW_t, \quad (2)$$

where $\kappa > 0$, $\omega > 0$, and $0 < \lambda < 1$. The changes of the diffusion coefficient $\sigma_s - \sigma_{t_i} =$

⁵See Hansen and Scheinkman (1995) and Ait-Sahalia et al. (2003) for a review and Appendix A for a summary of their theoretical properties.

$O_p(\Delta^{1/2})$ because $\sigma_s + \sigma_{t_i} = O_p(1)$; hence,

$$\begin{aligned}\sigma_s^2 - \sigma_{t_i}^2 &= \kappa\omega(s - t_i) - \kappa \int_{t_i}^s \sigma_t^2 dt + \sqrt{2\lambda\kappa} \int_{t_i}^s \sigma_t^2 dW_t \\ &= \sqrt{2\lambda\kappa} \int_{t_i}^s \sigma_t^2 dW_t \{1 + o_p(1)\} = O_p(\Delta^{1/2}),\end{aligned}$$

and $\sigma_s - \sigma_{t_i} = (\sigma_s^2 - \sigma_{t_i}^2)/(\sigma_s + \sigma_{t_i})$.

The linear drift-diffusion process is

$$dy_t = \theta (y_t - m) dt + \sigma_t dW_t, \quad (3)$$

with an exact discrete solution (Arnold, 1974, Corollary 8.2.4)

$$y_{t_{i+1}} = g(\theta) + \alpha(\theta) y_{t_i} + V_{t_{i+1}}, \quad (4)$$

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

$$\alpha(\theta) = \exp(\theta\Delta) = 1 + \theta\Delta + O_p(\Delta^2)$$

converges to unity as $\Delta \rightarrow 0$ at a rate of Δ . The order of magnitude of the autoregressive coefficient $O_p(\Delta)$ can be written as $O_p(1/T)$, given that $\Delta = N/T$ and N is a constant. Therefore, the dynamic in (4) is *local-to-unity* (Phillips, 1987) in the explosive direction if $\theta > 0$ and in the stationary direction if $\theta < 0$. One can show that $V_{t_{i+1}}$ follows a linear drift process such that

$$dV_{t_{i+1}} = \theta V_{t_{i+1}} ds + \sigma_{t_{i+1}} dW_{t_{i+1}}. \quad (5)$$

Under (3), we have $\int_{t_i}^{t_i + \Delta} \sigma_s dW_s = O_p(\Delta^{1/2})$; hence, $V_{t_{i+1}} = O_p(\Delta^{1/2})$. See Phillips and Yu (2011) for the derivation.

2.1 The Return Process

The solution of the stochastic differential equation (3) is

$$y_t = e^{\theta t} y_0 + m(1 - e^{\theta t}) + \int_0^t e^{\theta(t-s)} \sigma_s dW_s, \quad (6)$$

where y_0 is the initial value. The return process is denoted by r_{t_i} and defined as $y_{t_i} - y_{t_{i-1}}$.

Lemma 2.1 *Under the linear drift-diffusion process (3), r_{t_i} has the following properties:*

(i) r_{t_i} can be written as

$$\begin{aligned} r_{t_i} &= \Delta_i A + \int_{t_{i-1}}^{t_i} \sigma_s dW_s + \int_{t_{i-1}}^{t_i} B(t_i, s) \sigma_s dW_s + \int_0^{t_{i-1}} [B(t_i, s) - B(t_{i-1}, s)] \sigma_s dW_s \quad (7) \\ &= \int_{t_{i-1}}^{t_i} \sigma_s dW_s \{1 + o_p(1)\}, \end{aligned}$$

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 $t \geq s$,
 $a(t) = \theta(y_0 - m)e^{\theta t}$ and $b(t, s) = \theta e^{\theta(t-s)}$;

(ii) The expectation of r_{t_i} , denoted by m_{t_i} , is

$$m_{t_i} = \Delta_i A = (y_0 - m)e^{\theta t_{i-1}}(e^{\theta \Delta} - 1).$$

See Appendix A for the proof.

From Lemma 2.1(i), the return process is asymptotically dominated by the volatility component $\int_{t_{i-1}}^{t_i} \sigma_s dW_s$. In other words, the drift term is asymptotically negligible and as shown by Barndorff-Nielsen and Shephard (2002),

$$r_{t_i}^2 \rightarrow \int_{t_{i-1}}^{t_i} \sigma_s^2 ds.$$

Although the drift component is negligible when the sampling interval Δ approaches zero,

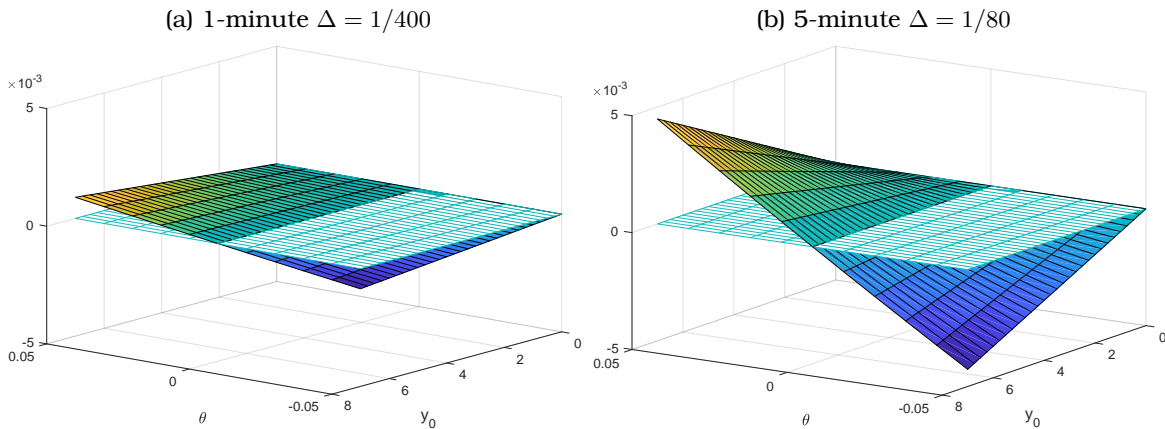
very often, one has to rely on low-frequency data for estimations and hypothesis testing. This is because ultrahigh-frequency data for asset prices are not always available, or because lower-frequency data (such as 5- or 10-minute data) are preferred to mitigate the impact of microstructure noise (Park and Linton, 2011).

Lemma 2.1(ii) suggests that the expected value of the log-return process depends on the values of m , θ , Δ , as well as the initial value y_0 . The mean of the average log-returns over $N = 1$ day is

$$\Delta \sum_{i=1}^{1/\Delta} m_{t_i} = \Delta(y_0 - m)(e^\theta - 1). \quad (8)$$

Figure 1 shows the value of (8) for several combinations of values for y_0 and θ at the 1-minute ($\Delta = 1/400$) and 5-minute ($\Delta = 1/80$) frequencies. We consider a wide range of values of θ and allow the initial value to vary from 0 to 7. We set m to zero for simplicity. As we observe, when the process deviates from the random walk (i.e., with nonzero θ), the drift component has an order of magnitude of 10^{-3} . Additionally, 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.⁶

Figure 1: The average drift of log-returns over one day for various combinations of θ , y_0 and Δ .



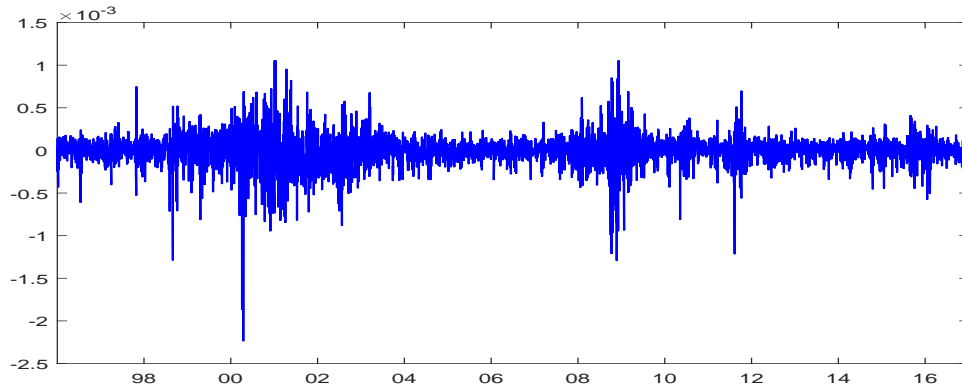
The parameter settings are empirically realistic. As a comparison, the daily median⁷ of

⁶If the sampling frequency is one second ($\Delta = 1/24000$), the drift component is negligible for all combinations of parameters and is therefore not reported to save space.

⁷We calculate the median instead of the sample mean to reduce the impact of jumps in asset prices (see

5-minute log-returns of the NASDAQ stock market index is plotted in Figure 2 for the period 1996–2016. This figure shows that the daily median of 5-minute log-returns ranges from -2.2×10^{-3} to 1.1×10^{-3} , which is of the same order of magnitude as the drift component presented in Figure 1. In particular, the daily median of the 5-minute log-returns deviates from zero for a substantial period of time in early 2000 when the dot-com bubble burst and around the subprime mortgage crisis period.

Figure 2: Daily median of 5-minute log-returns of the NASDAQ stock index from 1996 to 2016.



2.2 The Realized Volatility Estimator

The realized volatility estimator (see Andersen and Bollerslev, 1998 and Barndorff-Nielsen and Shephard, 2002 among others) 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. The bias of the realized volatility is therefore

$$E \left(RV_{t_i}(K) - \int_{t_i-K}^{t_i} \sigma_u^2 du \right).$$

Proposition 2.1 *Under the linear diffusion process (3), the finite sample bias of the realized*
 Section 3.1 for a detailed discussion).

volatility is zero if $\theta = 0$. If $\theta \neq 0$,

$$E \left(RV_{t_i}(K) - \int_{t_i-K}^{t_i} \sigma_u^2 du \right) = \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3, \quad (9)$$

where

$$\mathcal{E}_1 = (y_0 - m)^2 \left(e^{\theta\Delta} - 1 \right)^2 e^{\theta t_i - K} \frac{1 - e^{2\theta\Delta K}}{1 - e^{2\theta\Delta}}, \quad (10)$$

$$\mathcal{E}_2 = d_0 K \left[\frac{1}{2\theta} \left(e^{2\theta\Delta} - 1 \right) + \frac{2}{\theta} \left(1 - e^{\theta\Delta} \right) + 1 \right], \quad (11)$$

$$\mathcal{E}_3 = \frac{d_0}{2\theta} \left(e^{\theta\Delta} - 1 \right)^2 \left(e^{2\theta t_i - K} \frac{1 - e^{2\theta\Delta K}}{1 - e^{2\theta\Delta}} - K \right). \quad (12)$$

The proof of the proposition is in Appendix A. One can see that although the drift term μ_t is asymptotically negligible, it has a finite sample impact on the realized volatility estimator. To minimize the impact of the drift term, we propose calculating the realized volatility from centered log-returns. We study two modified realized volatility estimators. The first one is infeasible because it is based on the unknown quantity m_{t_j} , while the second one is feasible because it relies on the sample mean of the K log-returns entering in the computation of the realized volatility, i.e.,

$$\hat{m}_{t_i}(K) = \frac{1}{K} \sum_{j=i-K+1}^i r_{t_j}.$$

The two modified realized volatility estimators are

$$\text{Infeasible: } RV_{t_i}^\dagger(K) = \sum_{j=i-K+1}^i (r_{t_j} - m_{t_j})^2;$$

$$\text{Feasible: } RV_{t_i}^*(K) = \sum_{j=i-K+1}^i [r_{t_j} - \hat{m}_{t_i}(K)]^2.$$

Proposition 2.2 *Under the linear diffusion process (3), the finite sample bias of the infea-*

sible modified realized volatility estimator RV^\dagger is zero when $\theta = 0$; while when $\theta \neq 0$,

$$E \left(RV_{t_i}^\dagger(K) - \int_{t_i-K}^{t_i} \sigma_u^2 du \right) = \mathcal{E}_2 + \mathcal{E}_3. \quad (13)$$

Proposition 2.3 *Under the linear diffusion process (3), the finite sample bias of the feasible modified realized volatility estimator RV^* is zero when $\theta = 0$; while when $\theta \neq 0$,*

$$E \left(RV_{t_i}^*(K) - \int_{t_i-K}^{t_i} \sigma_u^2 du \right) = (\mathcal{E}_1 - \mathcal{E}_4) + \frac{K+1}{K}(\mathcal{E}_2 + \mathcal{E}_3) + d_0 \Delta, \quad (14)$$

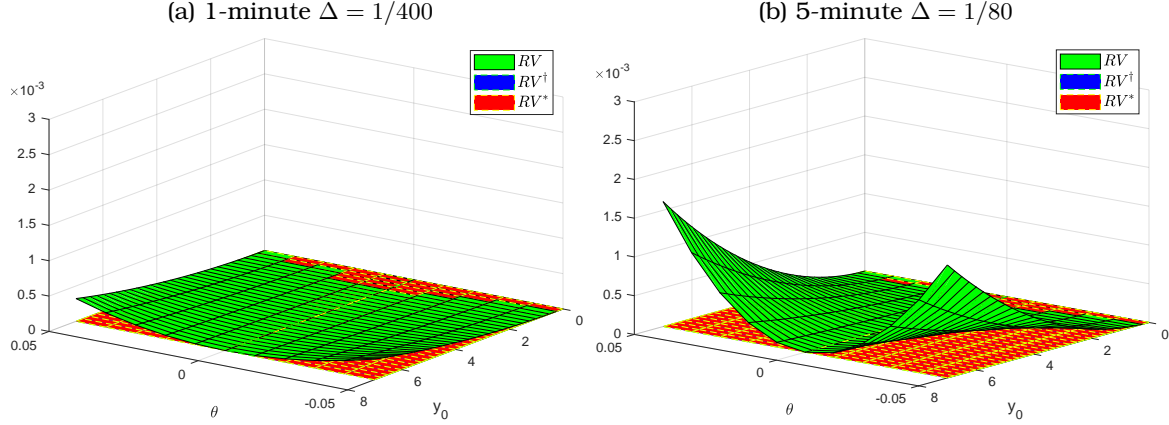
where $\mathcal{E}_4 = \frac{1}{K} (y_0 - m)^2 e^{\theta t_{i-K}} (e^{\theta \Delta K} - 1)^2$.

The proofs of the above two propositions are in Appendix A. The first term in (9) disappears when the realized volatility is calculated from centered log-returns (i.e., $r_{t_j} - m_{t_j}$). When using the RV^* estimator, while we observe additional noise terms arising from the estimation of the sample mean, the first bias term in (9) is reduced by \mathcal{E}_4 . To visualize the relative magnitude of the bias, we simulate the quantities in Proposition 2.1-2.3 with the same parameter setting as for Figure 1 but also set the unconditional variance d_0 to 10^{-4} (as in the subsequent simulations), $K = 1/\Delta$ and $t_{i-K} = 0$.

The bias of the realized volatilities RV , RV^\dagger and RV^* are plotted in Figure 3. It is obvious that the biases of the modified volatility estimators are much smaller than that of the original estimator in finite sample. The discrepancy becomes increasingly obvious as the process deviates further away from the random walk and when the initial value increases. Additionally, the bias of the modified volatility estimator is extremely close to zero with an order of magnitude of 10^{-9} for RV^\dagger and 10^{-6} for RV^* when $\Delta = 1/80$. This suggests that although Propositions 2.2 and 2.3 show that the biases of the modified estimators are nonzero when the process deviates from the random walk, they are negligible compared to that of the realized volatility estimator.

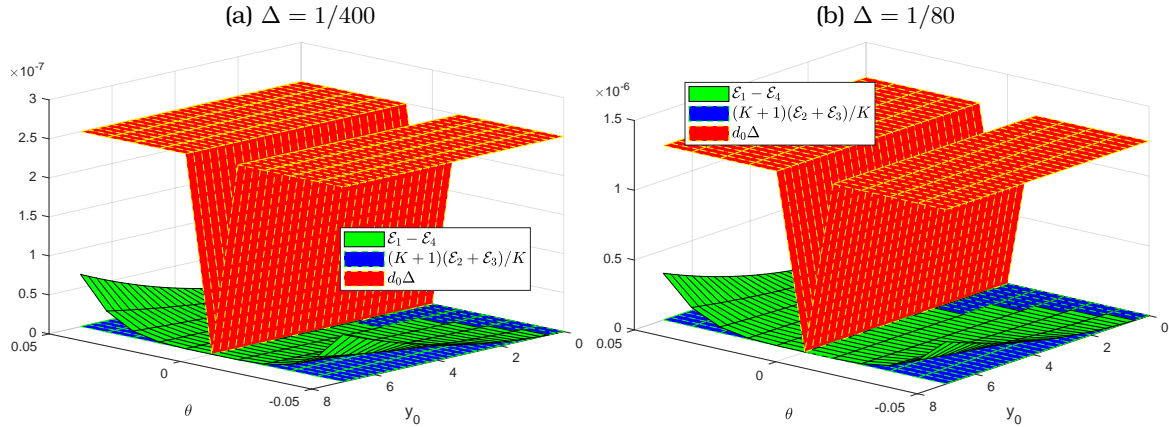
Furthermore, we plot each of the three terms characterizing the bias of RV^* in Proposition 2.3 in Figure 4. First, the term $(K+1)/(\mathcal{E}_2 + \mathcal{E}_3)K$ is the smallest among the three

Figure 3: The bias of the (modified) realized volatility estimators.



terms. The order of $(K+1)/(\mathcal{E}_2 + \mathcal{E}_3)K$, therefore $\mathcal{E}_2 + \mathcal{E}_3$, is 10^{-9} for the 5-minute data and 10^{-10} for the 1-minute data, which is negligible compared with the total bias of RV (i.e., 10^{-3} as shown in Figure 3). This allows us to infer that the bias of RV is mostly due to \mathcal{E}_1 . Second, the $\mathcal{E}_1 - \mathcal{E}_4$ term has the same order as $(K+1)/(\mathcal{E}_2 + \mathcal{E}_3)K$ and hence also negligible (although it increases with both $|\theta|$ and y_0). Lastly, the term that dominates the bias of RV^* is $d_0\Delta$. It takes value 2.5×10^{-7} and 1.25×10^{-6} , respectively, when $\Delta = 1/400$ and $1/80$, which are much smaller than the unconditional variance (recall that $E(\sigma_t^2) = 10^{-4}$).

Figure 4: Components of the bias of RV^* .



The finite sample bias of the realized volatility estimator was first documented in Meddahi (2002), where the drift coefficient is 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 $\sum_{i=0}^p |g_i| < \infty$. This specification

includes, for example, the case where the drift coefficient is an affine function of the variance σ_t^2 and also the case of constant drift.

Remark 2.1 Assume that the drift coefficient $\mu_t = m$. The log-price y_t can be rewritten as

$$y_t = y_0 + mt + \int_0^t \sigma_s dW_s$$

and therefore $r_{t_i} = m\Delta + \int_{t_{i-1}}^{t_i} \sigma_s dW_s$. The biases of RV , RV_{\dagger} and RV^* are respectively

$$\begin{aligned} E \left(RV_{t_i}(K) - \int_{t_{i-1}}^{t_i} \sigma_u^2 du \right) &= Km^2 \Delta^2, \\ E \left(RV_{t_i}^{\dagger}(K) - \int_{t_{i-1}}^{t_i} \sigma_u^2 du \right) &= 0, \\ E \left(RV_{t_i}^*(K) - \int_{t_{i-1}}^{t_i} \sigma_u^2 du \right) &= d_0 \Delta. \end{aligned}$$

The proof is given in Appendix A. See Meddahi (2002, Proposition 4.1) for the exact expression for the bias of the realized volatility estimator under the general assumption that $\mu_t = \sum_{i=0}^p g_i \Pi_{i,S_t}$. It is important to note that while the realized volatility estimator is biased, the modified estimator RV^{\dagger} is unbiased. The feasible RV^* estimator is biased but the magnitude of its bias is marginal.⁸

3 Jump Robust Drift and Volatility Estimators

There is ample evidence on the presence of jumps in asset prices. This questions the relevance of the empirical mean for the estimation of $E(r_{t_j})$ and also calls for the use of robust to jumps estimators of the integrated variance. In this section, we study the finite sample properties of the median and the jump robust volatility estimators computed on centered log-returns via a comprehensive Monte-Carlo simulation.

⁸Deriving a bias correction might be possible but is beyond the scope of this paper.

3.1 Jump Robust Drift Estimator

It is well known in the robust statistics literature that the median is less sensitive to outliers or jumps 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, among others). In presence of jumps, it is therefore natural to center the log-returns using the median of the past K log-returns, denoted by

$$\hat{m}_{t_i}^*(K) = \text{median}(r_{t_{i-K+1}}, \dots, r_{t_i}),$$

rather than their empirical mean $\hat{m}_{t_i}(K)$.

To compare the accuracy of $\hat{m}_{t_i}^*(K)$ and $\hat{m}_{t_i}(K)$ as estimators for the drift component in the presence of jumps, we consider a data generating process that generalizes (4) by allowing for p additive jumps:

$$y_{t_{i+1}} = g(\theta) + \alpha(\theta)y_{t_i} + \sum_{j=1}^p \phi_{t_{i+1}}^j I_{t_{i+1}}^j + V_{t_{i+1}}, \quad (15)$$

where $I_{t_{i+1}}^j$ is a dummy variable indicating the location of the j^{th} jump (the occurrence of which is random), and $\phi_{t_{i+1}}^j$ is the corresponding jump size. The error term $V_{t_{i+1}}$ is specified by

$$V_{t_{i+1}} = \sigma_{t_{i+1}} \sqrt{\Delta} \varepsilon_{t_{i+1}} \quad (16)$$

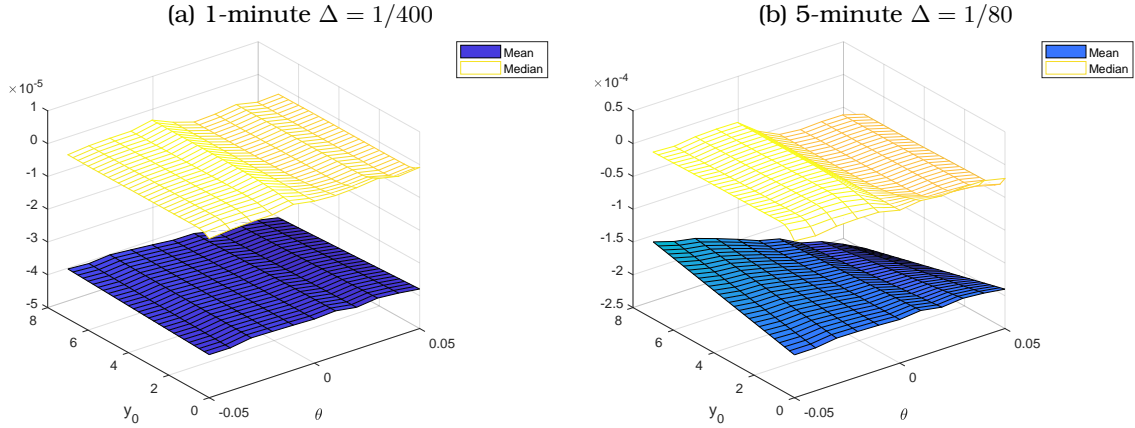
$$\sigma_{t_{i+1}}^2 = \alpha_0 + \sigma_{t_i}^2 (\beta_1 + \alpha_1 \sqrt{\Delta} v_{t_{i+1}}), \quad (17)$$

where $\varepsilon_{t_{i+1}}$ and $v_{t_{i+1}}$ are two independent $\overset{i.i.d.}{\sim} N(0, 1)$ random variables. Equation (16) is consistent with the specification of $V_{t_{i+1}}$ in (5) if the value of the error term within the time interval $[t_i, t_{i+1}]$ is assumed to be fixed. The volatility dynamic (17) is a standard Euler discretization of the GARCH(1,1) diffusion process (2). The parameters are linked as follows: $\alpha_0 = \kappa\omega\Delta$, $\beta_1 = 1 - \kappa\Delta$, and $\alpha_1 = \sqrt{2\lambda\kappa}$.

To illustrate the impact of jumps on the empirical mean and the robustness of the median, we generate 24,000 observations corresponding to one day of 1-second data of an asset traded during 6.5 hours per day.⁹ We pick one observation every 60 and 300 data points to obtain the 1-minute and 5-minute data, respectively. That is, 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 settings of y_0 , μ and θ are as in Figure 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 $j = 1, 2$.¹⁰ For the volatility dynamic, 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 $E(\sigma_{t_{i+1}}^2) = 10^{-4}$.

Figure 5: The bias of the sample mean and median in the presence of one single large jump (i.e. $\phi_{t_{i+1}}^1 = -1.5\sigma_{t_{i+1}}$).



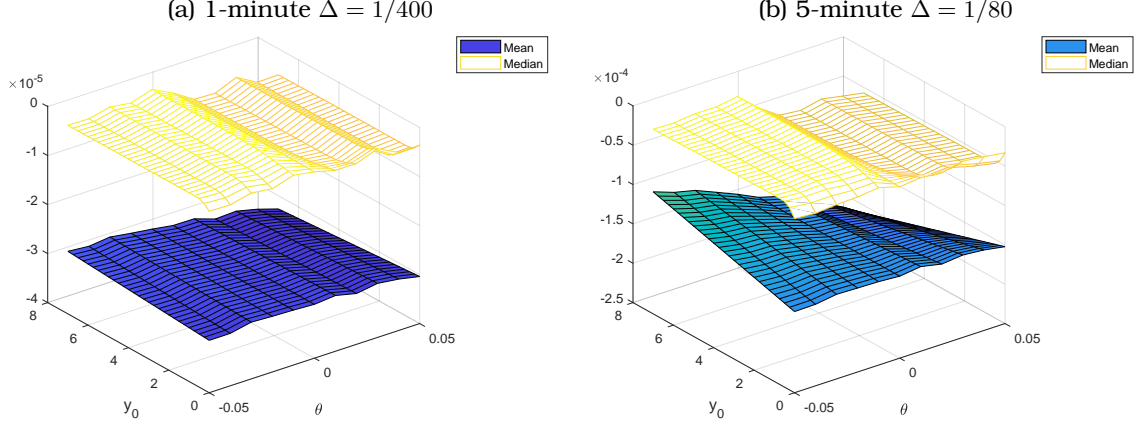
The empirical mean and median of the last observation of each day, denoted respectively by $\hat{m}_{t_T}(T)$ and $\hat{m}_{t_T}^*(T)$, are compared to the true value m_{t_T} . The superiority of the median is clear from Figures 5 and 6, which shows that the bias is systematically smaller for the median than for the sample mean for all combinations of parameters considered in the simulation.¹¹ Therefore, we recommend the use of the median instead of the sample

⁹This is the case for the NASDAQ stock market, which trades from 9:30 to 16:00.

¹⁰The simulation results are qualitatively the same for positive jumps and jumps with a random sign.

¹¹Similar figures are available for the root mean squared errors (RMSE) but not reported here to save space and because they follow more or less closely the shape of the figures on the bias.

Figure 6: The bias of the sample mean and median in the presence of two small jumps ($\phi_{t_{i+1}}^j = -0.6\sigma_{t_{i+1}}$ for $j = 1, 2$).



mean in empirical applications if the presence of jumps is suspected.

3.2 Jump Robust Volatility Estimators

The bipower variation proposed by Barndorff-Nielsen and Shephard (2004) is probably the most popular jump-robust estimator of the integrated variance. It has been extended by Podolskij and Vetter (2009) to accommodate the presence of market microstructure noise in ultrahigh-frequency data. Despite the well-behaved asymptotic properties of these estimators, we demonstrate below that like the realized volatility estimator their finite sample performance is unsatisfactory in the presence of a nonzero drift. We propose an alternative construction of these estimators that leads to a significant improvement in the estimation accuracy.

The bipower variation of Barndorff-Nielsen and Shephard (2004) is defined as

$$BV_{t_i}(K) = \frac{1}{c_1^2} \sum_{j=i-K+2}^i |r_{t_j}| |r_{t_{j-1}}|, \quad (18)$$

where $c_1 = E(|U|) = 0.7979$ is a normalizing constant (U denoting the standard normal distribution). As $\Delta \rightarrow 0$,

$$BV_{t_i}(K) \rightarrow_p \int_{t_{i-K}}^{t_i} \sigma_u^2 du.$$

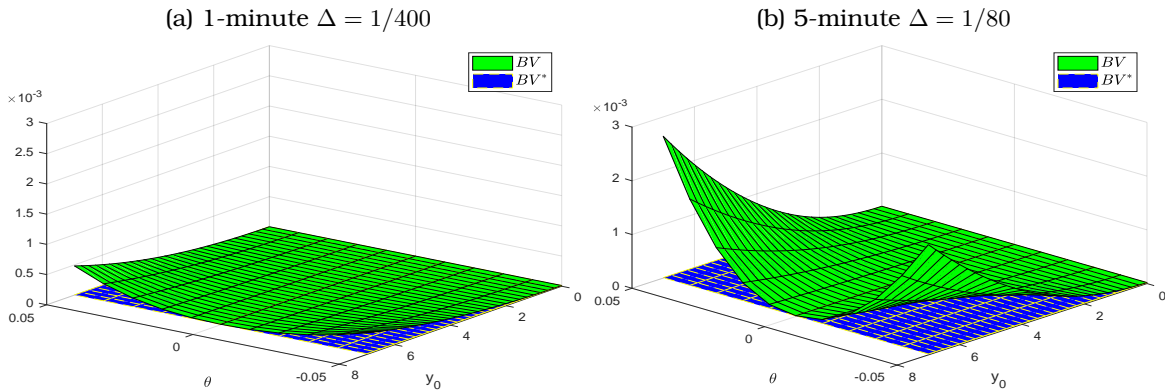
The new estimator of the integrated variance we propose, denoted by $BV_{t_i}^*(K)$, corresponds to the bipower variation computed from centered log-returns, i.e.,

$$BV_{t_i}^*(K) = \frac{1}{c_1^2} \sum_{j=i-K+2}^i |r_{t_j}^*| |r_{t_{j-1}}^*|,$$

where $r_{t_i}^* = r_{t_i} - \hat{m}_{t_i}^*(K)$. Since the drift component of asset returns is asymptotically dominated by the variance, using the centered log-returns $r_{t_i}^*$ for the calculation of the bipower variation will not alter its limiting property but should improve its finite sample properties, as for the modified realized volatility RV^* .

To illustrate the robustness of our estimator, we consider the data generating process (15)-(17) with the same parameter settings as in Section 3.1. For simplicity, we allow for one negative jump per day with the magnitude equal to 60% of the spot volatility (i.e., $\phi_{t_{i+1}}^1 = -0.6\sigma_{t_{i+1}}$). We report in Figure 7 the bias of both the bipower variation and modified bipower variation for the last observation of the day (i.e., t_T) from observations over the past day (i.e., respectively $BV_{t_T}(K)$ and $BV_{t_T}^*(K)$, with $K = T$).

Figure 7: The estimation bias of the bipower and modified bipower variation.



First, if $\theta = 0$, there is almost no difference in the estimation accuracy between the two estimators. Second, the bias of the bipower variation (BV) increases dramatically as the sampling frequency decreases and as $|\theta|$ and y_0 become larger. On the other hand, as expected, BV^* provides a much more accurate estimation of the integrated variance.

Indeed, the bias of BV^* is very small for all combinations of parameters considered in this simulation.

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

4 The Jump and Noise Robust Volatility Estimators

One might expect that with ultrahigh-frequency data, the drift component will be extremely close to zero; hence, the discrepancy between the bipower variation and the modified bipower variation will diminish. We show in this section that this is unfortunately not true for the noise-robust volatility estimator of Podolskij and Vetter (2009).

Assume that the noise-contaminated log-price $y_{t_i}^\dagger$ is

$$y_{t_i}^\dagger = y_{t_i} + \eta_{t_i}, \quad (19)$$

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 with an order of $s - 1$.

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

$$r_{t_i}^{(\Gamma)} = y_{t_i}^\dagger - y_{t_{i-\Gamma}}^\dagger,$$

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

$$\bar{r}_{t_i, b} = \frac{1}{S} \sum_{j=i-(B-b+1)S+1}^{i-(B-b)S} r_{t_j}^{(\Gamma)}. \quad (20)$$

Podolskij and Vetter (2009) propose to compute the bipower variation for noise-contaminated data using preaveraged returns, i.e.,

$$BV_{t_i}^N(K) = \frac{1}{c_1^2} \sum_{b=2}^B |\bar{r}_{t_i,b}| |\bar{r}_{t_i,b-1}|. \quad (21)$$

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

$$\widehat{IV}_{t_i}(K) := \frac{\gamma_1 \gamma_2 BV_{t_i}^N(K) - v_2 \hat{q}_{t_i}^2(K)}{v_1} \rightarrow \int_{t_i-K}^{t_i} \sigma_u^2 du, \quad (22)$$

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

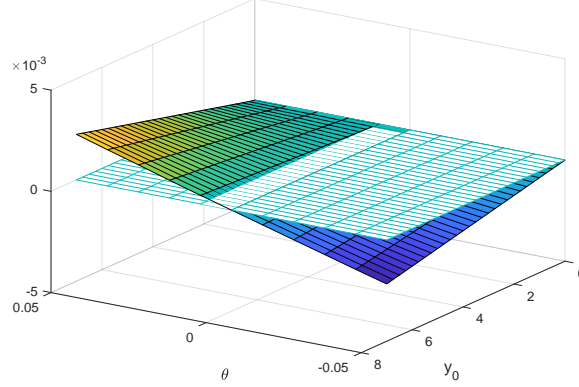
$$\hat{q}_{t_i}^2(K) = \frac{1}{2(K-s)} \sum_{j=i-K+s+1}^i (y_{t_j}^\dagger - y_{t_{j-s}}^\dagger)^2,$$

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

While the drift component of log-returns is extremely close to zero in the ultrahigh-frequency setting, the drift of the preaveraged returns $\bar{r}_{t_i,b}$ (using which the bipower variation is computed) might be of a nonnegligible magnitude. For the sake of illustration, we consider a data generating process following (15), (16)-(17) and (19). Parameter settings for 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 displays the average of the median of $\bar{r}_{t_i,b}(K)$, denoted by $\tilde{m}_{t_i}^*(K) = \text{median}(\bar{r}_{t_i,1}, \dots, \bar{r}_{t_i,B})$, for various combinations of θ and y_0 . This figure shows that the median increases as the

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



process deviates from the random walk, and the magnitude is comparable to the median of 5-minute returns plotted in Figure 1. As before, this deviation is expected to affect the performance of the integrated variance estimator.

As an alternative, we propose modifying the estimator of Podolskij and Vetter (2009) by computing the bipower variation using centered preaveraged returns, i.e.,

$$BV_{t_i}^{N*}(K) = \frac{1}{c_1^2} \sum_{b=2}^B |\bar{r}_{t_i,b} - \tilde{m}_{t_i}^*(K)| |\bar{r}_{t_i,b-1} - \tilde{m}_{t_i}^*(K)|, \quad (23)$$

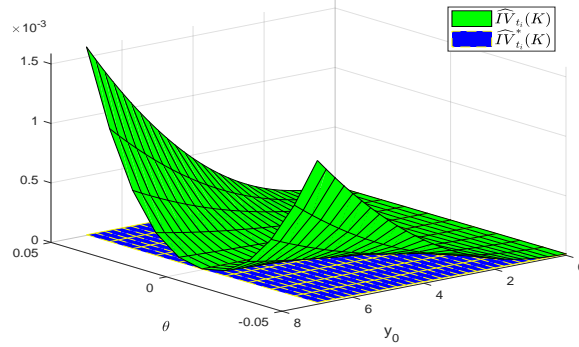
and

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

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

With the same simulation setting, we compute the bias of the two noise and jumps robust estimators $\widehat{IV}_{t_i}(K)$ and $\widehat{IV}_{t_i}^*(K)$ (see 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.

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



5 Jump Tests

The popularity of the LM08 and LM12 tests is due to their ease of implementation, their good size and power under fairly general conditions (e.g., allowing both the drift and the volatility to be stochastic) and the fact that they allow for testing the presence of jumps within a short period of time (e.g., 5 minutes or 1 hour).

We show in this section that the nonzero drift component 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 also undersized if the log-price process has a nonzero drift. We propose a modification of both tests and show the importance of this correction for the finite sample performance of the tests through Monte-Carlo simulations.

5.1 The Lee and Mykland (2008) Tests

Andersen et al. (2007) and Lee and Mykland (2008) independently proposed a test statistic for jumps, denoted 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:

$$J_{t_i} = \frac{r_{t_i}}{\hat{\sigma}_{t_i}(K)} \quad \text{and} \quad \tilde{J}_{t_i} = \frac{r_{t_i} - \hat{m}_{t_i}(K)}{\hat{\sigma}_{t_i}(K)}. \quad (24)$$

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

$$\hat{\sigma}_{t_i}(K) = \sqrt{\frac{1}{K-1}BV_{t_i}(K)}. \quad (25)$$

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 Section 3.1. 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

$$J_{t_i}^* = \frac{r_{t_i} - \hat{m}_{t_i}^*(K)}{\hat{\sigma}_{t_i}^*(K)}. \quad (26)$$

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.,

$$\hat{\sigma}_{t_i}^*(K) = \sqrt{\frac{1}{K-1}BV_{t_i}^*(K)}. \quad (27)$$

The properties of both test statistics have been studied by Lee and Mykland (2008). More specifically, they show that in the absence of jumps and if log-prices follow a diffusion process as in (1), and satisfying the two conditions stated in Assumption 2.1, 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. Given that 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

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

where $S_{t_i} = \{J_{t_i}, \tilde{J}_{t_i}, J_{t_i}^*\}$ 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. Recall that the maximum of a set of L i.i.d. realizations of the absolute value of the standard normal random distribution U_i (for $i = 1, \dots, L$) follows asymptotically a Gumbel distribution (see, for example, Aldous, 1989; Mutangi and Matarise, 2011), i.e.,

$$\frac{\max_i |U_i| - C_L}{S_L} \rightarrow \xi \quad (28)$$

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 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,

$$cv_{L,\beta} = C_L - S_L \log[-\log(1 - \beta)]. \quad (29)$$

Therefore, we declare that there is a jump at time t_i according to the J_{t_i} (resp. \tilde{J}_{t_i} and $J_{t_i}^*$) statistic if $|J_{t_i}| > cv_{L,\beta}$ (resp. $|\tilde{J}_{t_i}| > cv_{L,\beta}$ and $|J_{t_i}^*| > cv_{L,\beta}$). 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 The Finite Sample Performance of the 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 (4) while we rely on (15) to study their power. The volatility dynamic is specified by (16)-(17). The parameter settings are the same as in Section 3.1. Under the alternative, there is one jump per day with a magnitude of $\phi_{t_{i+1}}^1 = -0.6\sigma_{t_{i+1}}$.

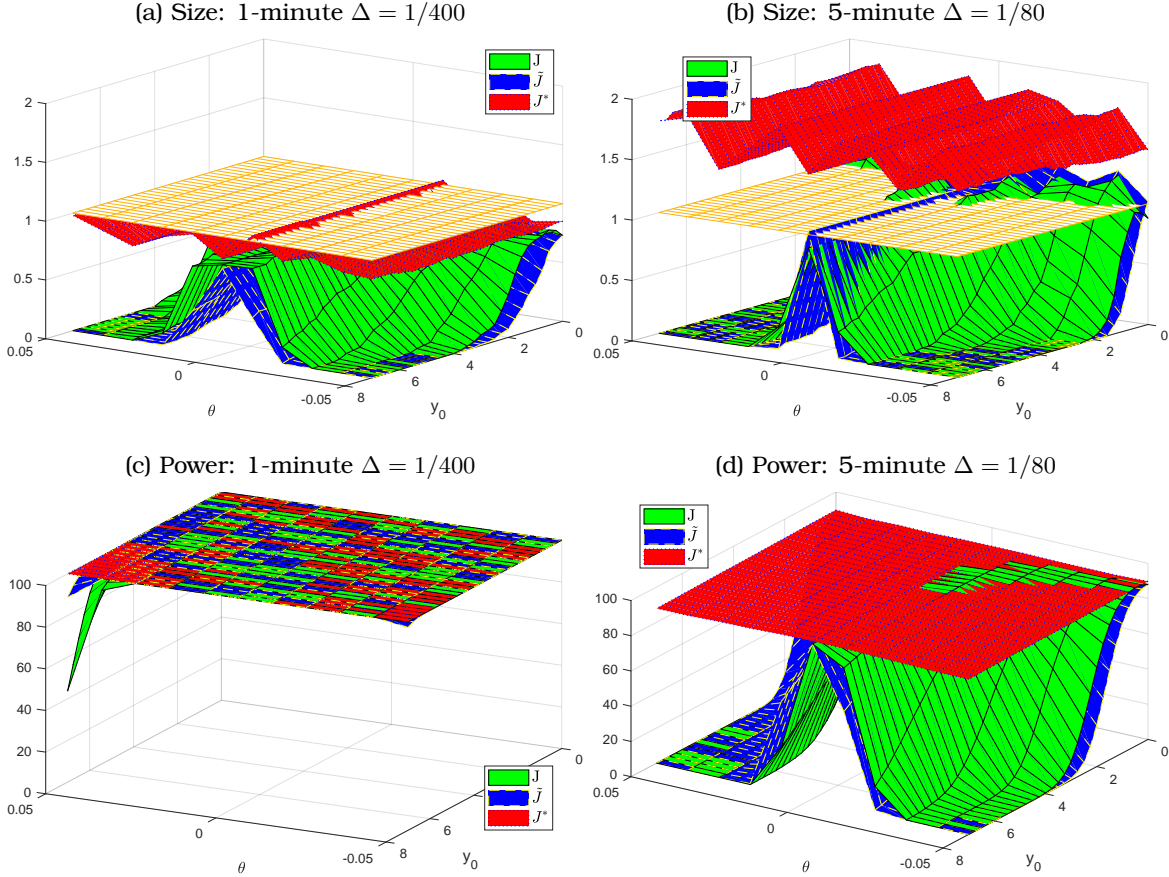
We generate 48,000 observations corresponding to two days of 1-second data of an asset and aggregate them at the 1-minute and 5-minute frequencies as above. The first day is used as a burn-in period; we focus on the detection results of the second day. Therefore, the time span of interest is one ($N = 1$), 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 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 $\beta = 1\%$ so that the probability of finding at least one spurious jump within each day is 1% or, equivalently, we expect that one out of 100 days contains at least one spurious jump. Therefore, the critical values used in this simulation are $cv_{24000,1\%} = 5.13$, $cv_{400,1\%} = 4.37$ and $cv_{80,1\%} = 4.07$ when the tests are applied to 1-second, 1-minute and 5-minute data, respectively. Note that the asymptotic critical values (29) 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 $\beta = 1\%$. Alternatively, we could control the size of the test over a longer period (e.g., a month or a year). Interestingly, when setting $L = 20T$ and $\beta = 10\%$, the critical value is 4.02 for 5-minute data and is slightly smaller than the

¹²Given 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 see from Figure 10(b) and 12(a) 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 for future work.

one of the above setting (i.e., 4.07).

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



The empirical sizes and powers (with 10^4 replications) of the LM08 statistics J and \tilde{J} , and the modified LM08 statistic J^* are plotted in Figure 10 for two sampling frequencies (1- and 5-minute data).¹³ 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 1%.¹⁴ On the contrary, we observe a significant downward size distortion for both tests as the dynamic of log-prices deviates from the random walk (i.e., $\theta \neq 0$) and the initial value y_0 increases.¹⁵ 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

¹³A table display of results in Figure 10, 12, and 13 are provided in an online supplement.

¹⁴We obtained qualitatively the same results for the case of constant volatility and for other quantiles (5% and 0.1%). Results are not reported to save space.

¹⁵When the LM08 tests are applied to 1-second data (without microstructure noise), both tests have an empirical size close to the nominal size of 1% for all combinations of parameters and a power of 100%. The results are not reported here to save space.

never rejected for both tests¹⁶ if $|\theta| \geq 0.02$ and $y_0 \geq 3$ (Figure 10b).

Figure 10c shows that the power of both tests is close to 100% for all specifications (except the bottom left corner) if the tests are applied to data at the 1-minute frequency. For 5-minute data, power values are approximately 88% 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 10d), the downward size distortion problem of both tests translates into a dramatic loss of power if the process has a nonzero drift. In particular, both tests have a power close to 0 if $|\theta| \geq 0.02$ and $y_0 \geq 4$. 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 88% if applied to 5-minute data. This result is in sharp contrast to Figure 10, where both J and \tilde{J} tests suffer from serious size distortion towards 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, it has been well known since the studies of Taylor and Xu (1997) and Andersen and Bollerslev (1998b) 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

¹⁶Importantly, a close-to-zero rate of rejection is also observed if using a higher critical value of $\beta = 5\%$.

of the intraday (or intraweek) periodicity and a correction to the LM08 jump statistics by allowing 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 volatility being periodically low and to reduce the number of spurious jump detections at times when volatility is periodically high.

The most efficient nonparametric periodicity estimator of Boudt et al. (2011), denoted \hat{f}_{t_i} , is the weighted standard deviations (WSD). Assuming for simplicity that the length of the periodicity cycle is one day and that we dispose of 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. The observation J_{t_i} receives either a weight of either zero when $J_{t_i}^2$ is higher than a high quantile (e.g., 99%) of the 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.¹⁷ It is important to note that this estimator of the 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 Boudt et al. (2011), we also modify our newly proposed J^* statistic by multiplying $\hat{\sigma}_{t_i}^*$ in (26) by the estimated intraday periodicity $\hat{f}_{t_i}^*$. The periodicity-adjusted jump test statistic is denoted by $J_{t_i}^{*P}$ and defined as

$$J_{t_i}^{*P} = \frac{r_{t_i} - \hat{m}_{t_i}^*(K)}{\hat{f}_{t_i}^* \hat{\sigma}_{t_i}^*(K)}. \quad (30)$$

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

¹⁷Note that, as Boudt et al. (2011), we normalize \hat{f}_{t_i} such that $\hat{f}_{t_i}^2$ averages to one over the length of the periodicity cycle (e.g., one day or one week).

J_{t_i}/\hat{f}_{t_i} and $\tilde{J}_{t_i}^P = \tilde{J}_{t_i}/\hat{f}_{t_i}$.

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 intra-periodicity in addition to GARCH(1,1) dynamics in the conditional variance. To this end, we simulate data according to (15) and (17), where

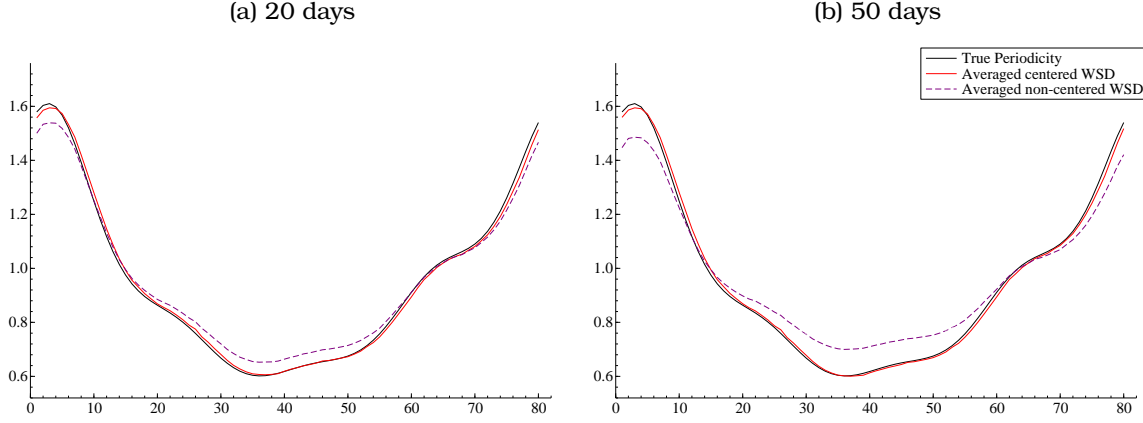
$$V_{t_{i+1}} = \sigma_{t_i} f_{t_i} \sqrt{\Delta} \varepsilon_{t_{i+1}}, \quad (31)$$

and f_{t_i} , i.e., the true periodicity, depicts the usual U-shaped pattern during the day and is restricted to be 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 the $J_{t_i}^{*P}$ test statistic (also computed on 5-minute data). The burn-in period contains on average one jump per day (with size $\phi_{t_{i+1}}^1 = -0.6\sigma_{t_{i+1}}f_{t_i}$). 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 10^4 replications) of \hat{f}_{t_i} and $\hat{f}_{t_i}^*$. To study the impact of a nonzero drift on the 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 0 on the remaining days. Figure 11 corresponds to the most extreme case, i.e. $y_0 = 7$ and $\theta = 0.05$ for the 20 or 50 days. The results suggest that unlike that for \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.

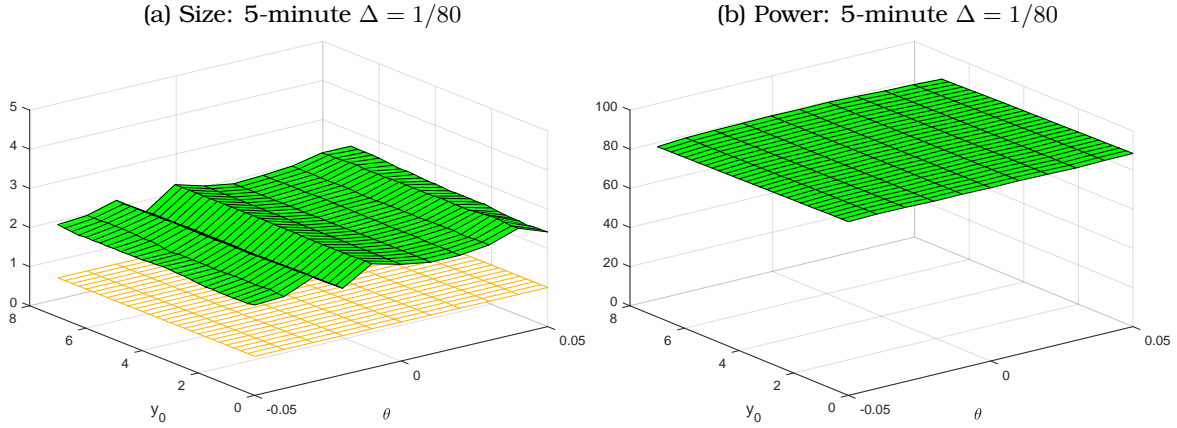
Finally, we explore the performance of the periodicity-adjusted jump test in Figure 12. The same DGP is used as above but now we consider the same range of values for θ and y_0 as in the previous simulations for the last day of the simulated sample (i.e., day 252). We assume that 20 random days (out of the first 251 days) deviate from the random walk (i.e., $\theta \neq 0$) in the burn-in period. Results are qualitatively the same for 50 days and are not reported here to save space. The general conclusion from Figure 12 is that our

Figure 11: The simulated intraday periodicity and the periodicity estimators.



periodicity-adjusted jump test J^{*P} behaves similarly to the unadjusted one using J^* in the absence of periodicity. The test is slightly oversized but reasonably close to the nominal size, and the power is satisfactory.

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



5.4 The (Modified) Ultrahigh-Frequency Jump Test

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 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 $y_{t_i}^\dagger$, as defined by (19). Recall that the noise is assumed to be serially correlated of order s . Let $\tilde{\Gamma} = sM$ with $\tilde{M} = CT_1^{1/2}$ (C is a constant and $T_1 = T/s$). For every s^{th} observation, we calculate the $\tilde{\Gamma}$ -differenced log-return as

$$r_{t_{js}}^{(\tilde{\Gamma})} = y_{t_{js}}^\dagger - y_{t_{js-\tilde{\Gamma}}}^\dagger$$

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

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

The Lee and Mykland (2012) test statistics for the presence of jumps in the log-prices between t_{bsM} and $t_{(b+1)sM}$ are

$$L_b = \sqrt{\tilde{M}} \frac{\bar{r}_b}{\sqrt{\hat{V}_b(K)}} \quad \text{and} \quad \tilde{L}_b = \sqrt{\tilde{M}} \frac{\bar{r}_b - \hat{m}_b(K)}{\sqrt{\hat{V}_b(K)}}, \quad (32)$$

where $\hat{V}_b(K)$ is an estimate of the variance of $\sqrt{\tilde{M}}\bar{r}_b$ 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

$$\hat{V}_b(K) = \frac{2}{3} \widehat{IV}_b(K) C^2 D + 2\hat{q}_b^2(K), \quad (33)$$

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 (32), we propose a

modification similar to the one in (26). The new test statistic is

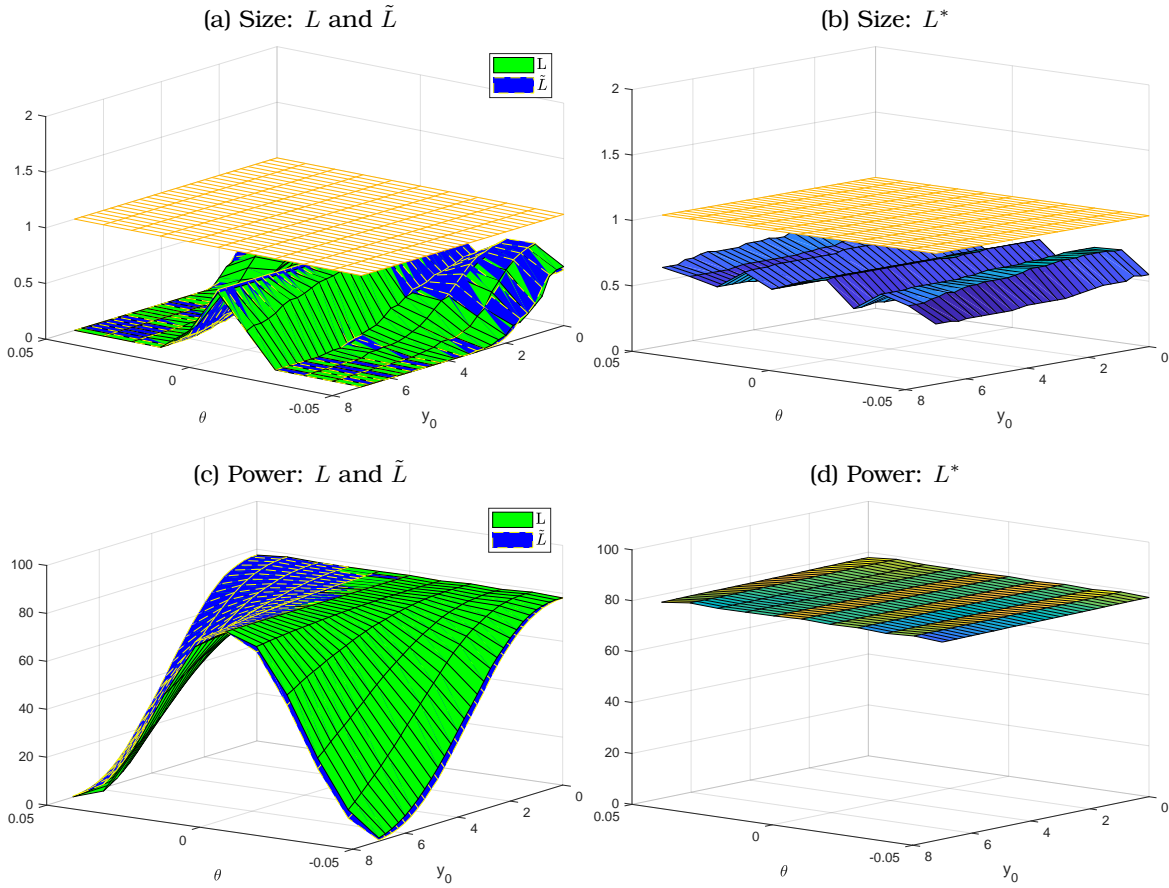
$$L_b^* = \sqrt{M} \frac{\bar{r}_b - \hat{m}_b^*(K)}{\sqrt{\hat{V}_b^*(K)}}, \quad (34)$$

where $\hat{m}_b^*(K) = \text{median}(\bar{r}_{b-B^0+1}, \dots, \bar{r}_b)$ 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_{tb}^*) converge to the standard normal distribution under the null hypothesis of no jumps.

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 a 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 1%.

The empirical performance of L_b , \tilde{L}_b and L_b^* statistics is illustrated in Figure 13. 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, although the magnitudes of the size distortion and the power loss are smaller than those observed in the right panel of Figure 10, that correspond to the original J and \tilde{J} tests of Lee and Mykland (2008) applied to 5-minute data and the span of one day. Importantly, the size of the new test L_b^* fluctuates slightly below the nominal size of 1%, while its power is approximately 80% for all configurations of θ and y_0 . These results are not surprising given our discussion and findings in Section 4 regarding the ultrahigh-frequency volatility estimators.

Figure 13: Empirical performance of L , \tilde{L} and L^* tests.

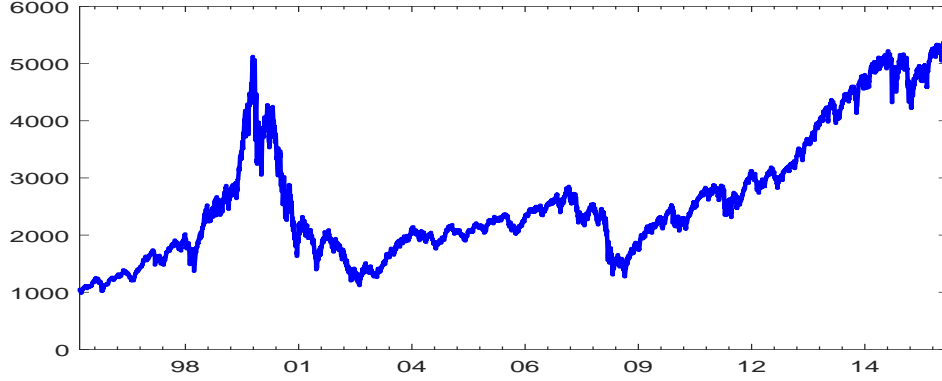


6 Empirical Application

In this section, we compute the three J test statistics using 5-minute data of the NASDAQ stock price index during the period from 1996 to 2016. 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, among others) 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.

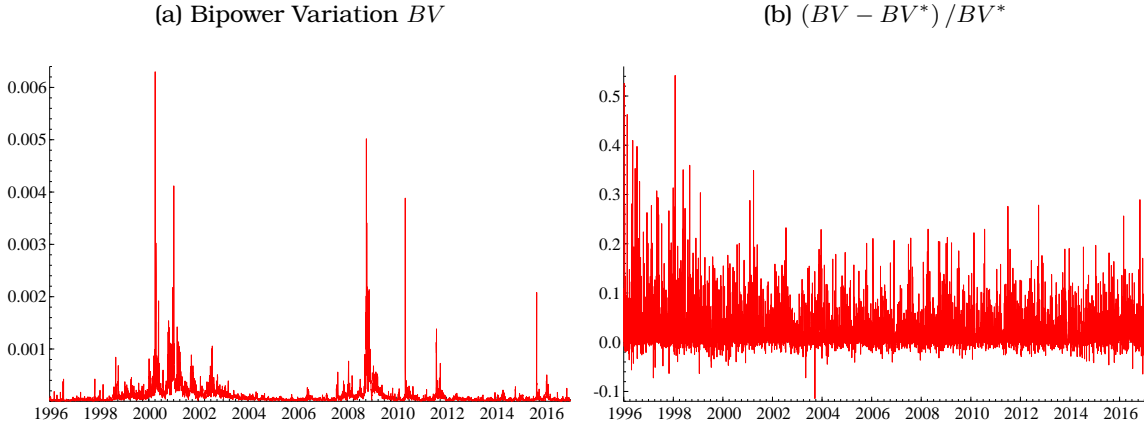
The daily bipower variation BV is plotted in the left panel of Figure 15 while the dis-

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



crepancy 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 and the discrepancy between BV and the modified bipower variation (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. Results of the three tests (J^P , \tilde{J}^P and J^{*P}) are reported in Table 1. 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 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.

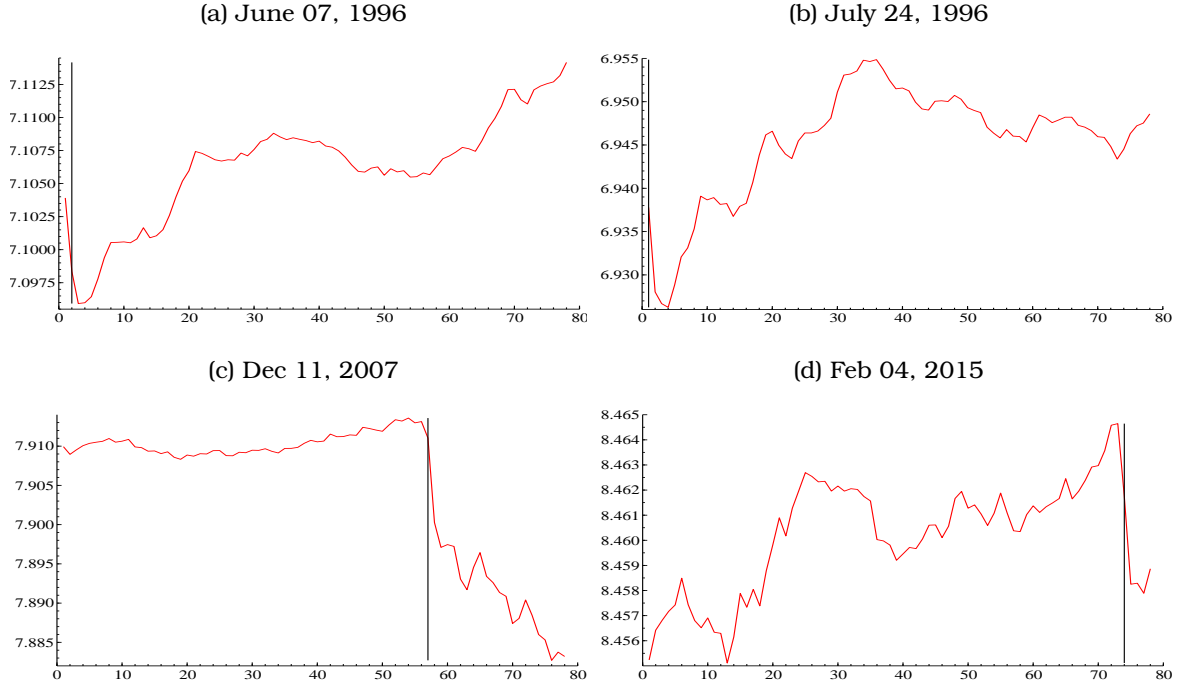
Table 1: Descriptive statistics of significant jumps.

	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 1 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 on which a jump is only detected by the J^{*P} statistic (i.e., $J^{P*} > cv_{78,0.1\%}$ while J^P and $\tilde{J}^P \leq cv_{78,0.1\%}$). To better understand the difference between the tests, the log-prices of 4 out of these 87 days (selected randomly) are plotted in Figure 16. Interestingly, these 4 days are characterized by a strong upward or downward trends that we attribute to a nonzero drift. Recall from Figure 10 that if the process has a nonzero drift, the original LM (2008) tests (both J and \tilde{J}) applied to 5-minute data are observed to be undersized and to have less power than J^* , while J^* has good size and very good power (the same comment applies to periodicity adjusted tests).

Figure 16: The log-prices on 4 days on which jumps are identified by J^{*P} but not by J^P and \tilde{J}^P . The vertical lines indicate the arrival times of the detected jumps.



7 Conclusion

This paper derives the finite sample theory for the realized volatility under the drift-diffusion process. The finite sample theory and extensive simulations reveal that with a fixed sampling interval, the realized volatility but also the jump robust volatility estimators (including the bipower variation, the MedRV, and the threshold realized variance) tend to overestimate the integrated variance in the presence of a nonzero drift, and the bias rises with the magnitude of the drift term. Moreover, despite the drift term 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 preaveraged returns.

Consequently, procedures derived from these integrated variance estimators (such as the intraday periodicity estimator of Boudt et al., 2011 and the Lee and Mykland (2008, 2012) jump tests have unsatisfactory performance in finite samples when 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 the integrated variance estimator, intraday periodicity estimator and jump test statistics, which leads to significant improvements of their performance.

The newly proposed 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 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., Gjessing, H. K., 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.
- Aït-Sahalia, Y., Hansen, L., Scheinkman, J., 2003. Operator methods for continuous-time markov models. *Handbook of Financial Econometrics*. Amsterdam: North-Holland. Forthcoming.
- Aldous, D., 1989. Probability approximations via the Poisson clumping heuristic. Vol. 77. Springer-Verlag, New York.
- Andersen, T., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 885–905.
- Andersen, T. G., Bollerslev, T., 1998b. Deutsch mark-dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance* 53, 219–265.
- Andersen, T. G., Bollerslev, T., Dobrev, D., 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., Dobrev, D., Schaumburg, E., 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., Wu, Y., Gilliland, E., 2000. Mean reversion across national stock markets and parametric contrarian investment strategies. *The Journal of Finance* 55 (2), 745–772.

- Bandi, F. M., Russell, J. R., 2005. On the finite sample properties of kernel-based integrated variance estimators. Unpublished paper, Graduate School of Business, University of Chicago.
- Bandi, F. M., Russell, J. R., 2006. Separating microstructure noise from volatility. *Journal of Financial Economics* 79 (3), 655–692.
- Barndorff-Nielsen, O. E., 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., Shephard, N., 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., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics* 2 (1), 1–37.
- Bekaert, G., Hodrick, R. J., 1992. Characterizing predictable components in excess returns on equity and foreign exchange markets. *The Journal of Finance* 47 (2), 467–509.
- Bessembinder, H., Chan, K., 1992. Time-varying risk premia and forecastable returns in futures markets. *Journal of Financial Economics* 32 (2), 169–193.
- Boudt, K., Croux, C., Laurent, S., 2011. Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance* 18 (2), 353–367.
- Boudt, K., Laurent, S., Lunde, A., Quaadvlieg, R., Sauri, O., 2017. Positive semidefinite integrated covariance estimation, factorizations and asynchronicity. *Journal of Econometrics* 196 (2), 347–367.
- Campbell, J. Y., Ammer, J., 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., Hamao, Y., 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., Wu, Y., 2003. Random walk versus breaking trend in stock prices: Evidence from emerging markets. *Journal of Banking and Finance* 27 (4), 575–592.
- Christensen, K., Oomen, R., Renò, R., 2016. The drift burst hypothesis. Tech. rep., Department of Economics and Business Economics, Aarhus University.
- Dumitru, A., Urga, G., 2012. Identifying jumps in financial assests: a comparison between nonparametric jump tests. *Journal of Business and Economic Statistics* 30, 242–255.
- Etienne, X. L., Irwin, S. H., Garcia, P., 2014. Bubbles in food commodity markets: Four decades of evidence. *Journal of International Money and Finance* 42, 129–155.
- Fama, E. F., French, K. R., 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 96 (2), 246–273.
- Gutierrez, L., 2012. Speculative bubbles in agricultural commodity markets. *European Review of Agricultural Economics* 40 (2), 217–238.
- Hansen, L. P., Scheinkman, J. A., 1995. Back to the future: Generating moment implications for continuous-time markov processes. *Econometrica: Journal of the Econometric Society*, 767–804.
- Homm, U., Breitung, J., 2012. Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics* 10 (1), 198–231.
- Lee, S. S., Hannig, J., 2010. Detecting jumps from levy jump diffusion processes. *Journal of Financial Economics* 96, 271–290.
- Lee, S. S., Mykland, P. A., 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies* 21 (6), 2535–2563.
- Lee, S. S., Mykland, P. A., 2012. Jumps in equilibrium prices and market microstructure noise. *Journal of Econometrics* 168 (2), 396–406.

- Lo, A. W., MacKinlay, A. C., 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., Wang, J., 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., Calvori, F., 2012. Jumps. In: Bauwens, L., Hafner, C., Laurent, S. (Eds.), *Handbook of Volatility Models and Their Applications*. Wiley.
- Maronna, R. A., Martin, D. R., Yohai, V. J., 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., Matarise, D. F., 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., Venardos, E., 2003. Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type. *Mathematical Finance* 13 (4), 445–466.
- Park, S., Linton, O., 2011. Realized volatility: Theory and application. In: Bauwens, L., Hafner, C., Laurent, S. (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., Magdalinos, T., 2007. Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136 (1), 115–130.

- Phillips, P. C., Shi, S., 2017. Detecting financial collapse and ballooning sovereign risk. Cowles Foudation Discussion Paper No. 3010.
- Phillips, P. C., Shi, S., Yu, J., 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., Wu, Y., Yu, J., 2011. Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? *International Economic Review* 52 (1), 201–226.
- Phillips, P. C., Yu, J., 2011. Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics* 2 (3), 455–491.
- Podolskij, M., Vetter, M., 2009. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* 15 (3), 634–658.
- Shi, S., Song, Y., 2016. Identifying speculative bubbles using an infinite hidden Markov model. *Journal of Financial Econometrics* 14 (1), 159–184.
- Taylor, S., Xu, X., 1997. The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance* 4, 317–340.
- Wang, X., Yu, J., 2016. Double asymptotics for explosive continuous time models. *Journal of Econometrics* 193, 35–53.
- Zhou, Q., Yu, J., 2015. Asymptotic theory for linear diffusions under alternative sampling schemes. *Economics Letters* 128, 1–5.

Appendix A

The proof of Lemma 2.1

Proof. Replacing y_t in the drift coefficient μ_t with (6), we have

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

where $a(t) = \theta (y_0 - m) e^{\theta t}$ 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 $t \geq s$.

(i) The return of the asset follows a dynamic of the type

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

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 is $O_p(\Delta^{3/2})$.

(ii) Take expectation of both sides of the return process, we have

$$E(r_{t_i}) = \Delta_i A = (y_0 - m) e^{\theta t_{i-1}} (e^{\theta \Delta} - 1).$$

■

Theoretical Properties of the Eigenfunctions

The eigenfunctions has the following properties: (1) The eigenfunctions are normalized such that $\Pi_{0, S_t} = 1$ and $Var(\Pi_{i, S_t}) = 1$ for all $i \neq 0$.

(2) Any non-constant eigenfunction is centered such that

$$E[\Pi_{i, S_t}] = 0.$$

(3) Two eigenfunctions that are associated with two different eigenvalues are orthogonal

$$E [\Pi_{i,S_t} \Pi_{j,S_t}] = 0 \text{ for } i \neq j.$$

(4) Any eigenfunction is an autoregressive process of order one,

$$E [\Pi_{i,S_{t+h}} | S_t, \tau \leq t] = \exp(-\delta_i h) \Pi_{i,S_t}, \forall h > 0.$$

(5) Any square-integrable function g , i.e. $E [g(S_t)^2] < \infty$, may be written as a linear combination of the eigenfunctions, i.e.

$$g(S_t) = \sum_{i=0}^{\infty} d_i \Pi_{i,S_t},$$

where $d_i = E [g(S_t) \Pi_{i,S_t}]$ and $\sum_{i=0}^{\infty} d_i^2 = E [g(S_t)^2] < \infty$.

Proof of Proposition 2.1

Proof. The bias of the realized volatility estimator is

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

The first step is to calculate the squared return from (35) such that

$$\begin{aligned} r_{t_j}^2 &= \underbrace{(\Delta_j A)^2}_{(1)} + \underbrace{\left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right)^2}_{(2)} + \underbrace{\left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right)^2}_{(3)} \\ &\quad + \underbrace{\left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right)^2}_{(4)} \end{aligned}$$

$$\begin{aligned}
& \underbrace{+ 2\Delta_j A \left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right)}_{(5)} + \underbrace{2 \left(\Delta_j A + \int_{t_{j-1}}^{t_j} \sigma_s dW_s \right) \int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s}_{(6)} \\
& + \underbrace{2 \left(\Delta_j A + \int_{t_{j-1}}^{t_j} \sigma_s dW_s \right) \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right)}_{(7)} \\
& + \underbrace{2 \left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right) \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right)}_{(8)},
\end{aligned}$$

which consists of eight terms. The first term

$$(\Delta_j A)^2 = \left(\int_{t_{j-1}}^{t_j} a(r) dr \right)^2.$$

We apply the Ito's lemma to the squared terms (2), (3) and (4).¹⁸ The second term

$$\left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right)^2 = s \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u \sigma_s dW_s \right) \sigma_u dW_u + \int_{t_{j-1}}^{t_j} \sigma_u^2 du;$$

the third term

$$\begin{aligned}
\left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right)^2 &= 2 \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u B(t_j, s) \sigma_s dW_s \right) B(t_j, u) \sigma_u dW_u \\
&+ \int_{t_{j-1}}^{t_j} B(t_j, u)^2 \sigma_u^2 du;
\end{aligned}$$

and the fourth term

$$\left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right)^2$$

¹⁸If $\eta(x) = x^2$, we have $d\eta = \frac{\partial \eta}{\partial x} dx + \frac{1}{2} \frac{\partial^2 \eta}{\partial x^2} (dx)^2$.

$$\begin{aligned}
&= 2 \int_0^{t_{j-1}} \left(\int_0^u [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) [B(t_j, u) - B(t_{j-1}, u)] \sigma_u dW_u \\
&\quad + \int_0^{t_{j-1}} [B(t_j, u) - B(t_{j-1}, u)]^2 \sigma_u^2 du.
\end{aligned}$$

Similarly, we apply the Ito's lemma to all the cross product terms (5), (6), (7) and (8).¹⁹

The fifth term

$$\begin{aligned}
&2\Delta_j A \left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right) \\
&= 2 \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u a(r) dr \right) \sigma_u dW_u + 2 \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u \sigma_s dW_s \right) a(u) du;
\end{aligned}$$

the sixth term

$$\begin{aligned}
&2 \left(\Delta_j A + \int_{t_{j-1}}^{t_j} \sigma_s dW_s \right) \left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right) \\
&= 2\Delta_j A \int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s + 2 \left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right) \left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right) \\
&= 2 \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u a(r) dr \right) B(t_j, u) \sigma_u dW_u + 2 \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u B(t_j, s) \sigma_s dW_s \right) a(u) du \\
&\quad + 2 \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u \sigma_s dW_s \right) B(t_j, u) \sigma_u dW_u + 2 \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u B(t_j, s) \sigma_s dW_s \right) \sigma_u dW_u;
\end{aligned}$$

the seventh term

$$\begin{aligned}
&2 \left(\Delta_j A + \int_{t_{j-1}}^{t_j} \sigma_s dW_s \right) \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) \\
&= 2\Delta_j A \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right)
\end{aligned}$$

¹⁹If $\eta(x, y) = 2xy$, we have $d\eta = \frac{\partial \eta}{\partial x} dx + \frac{\partial \eta}{\partial y} dy$.

$$\begin{aligned}
& +2 \left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right) \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) \\
= & 2 \int_0^{t_{j-1}} \left(\int_{t_{j-1}}^u a(r) dr \right) [B(t_j, u) - B(t_{j-1}, u)] \sigma_u dW_u \\
& +2 \int_{t_{j-1}}^{t_j} \left(\int_0^u [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) a(u) du \\
& +2 \int_{t_{j-1}}^{t_j} \left(\int_0^u [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) \sigma_u dW_u \\
& +2 \int_0^{t_{j-1}} \left(\int_{t_{j-1}}^u \sigma_s dW_s \right) [B(t_j, u) - B(t_{j-1}, u)] \sigma_u dW_u;
\end{aligned}$$

and the eighth term

$$\begin{aligned}
& 2 \left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right) \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) \\
= & 2 \int_{t_{j-1}}^{t_j} \left(\int_0^u [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) B(t_j, u) \sigma_u dW_u \\
& +2 \int_0^{t_{j-1}} \left(\int_{t_{j-1}}^u B(t_j, s) \sigma_s dW_s \right) [B(t_j, u) - B(t_{j-1}, u)] \sigma_u dW_u.
\end{aligned}$$

The expected value of $\left(r_{t_j}^2 - \int_{t_{j-1}}^{t_j} \sigma_u^2 du \right)$ is therefore

$$\begin{aligned}
E \left(r_{t_j}^2 - \int_{t_{j-1}}^{t_j} \sigma_u^2 du \right) &= \left(\int_{t_{j-1}}^{t_j} a(r) dr \right)^2 + \int_{t_{j-1}}^{t_j} B(t_j, u)^2 E(\sigma_u^2) du \\
&+ \int_0^{t_{j-1}} [B(t_j, u) - B(t_{j-1}, u)]^2 E(\sigma_u^2) du \\
&+ 2 \int_{t_{j-1}}^{t_j} E \left(\int_{t_{j-1}}^u \sigma_s dW_s \right) a(u) du \\
&+ 2 \int_{t_{j-1}}^{t_j} E \left(\int_{t_{j-1}}^u B(t_j, s) \sigma_s dW_s \right) a(u) du
\end{aligned}$$

$$+2 \int_{t_{j-1}}^{t_j} E \left(\int_0^u [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) a(u) du.$$

From Assumption 2.3, we have

$$E(\sigma_t^2) = E \left(\sum_{i=0}^p d_i \Pi_{i, S_t} \right) = \sum_{i=0}^p d_i E(\Pi_{i, S_t}) = d_0.$$

Therefore,

$$\begin{aligned} E \left(r_{t_j}^2 - \int_{t_{j-1}}^{t_j} \sigma_u^2 du \right) &= \left(\int_{t_{j-1}}^{t_j} a(r) dr \right)^2 + d_0 \int_{t_{j-1}}^{t_j} B(t_j, u)^2 du \\ &\quad + d_0 \int_0^{t_{j-1}} [B(t_j, u) - B(t_{j-1}, u)]^2 du. \end{aligned}$$

By construction,

$$\left(\int_{t_{j-1}}^{t_j} a(r) dr \right)^2 = (y_0 - m)^2 \left(\int_{t_{j-1}}^{t_j} \theta e^{\theta r} dr \right)^2 = (y_0 - m)^2 \left[e^{\theta t_j} - e^{\theta t_{j-1}} \right]^2,$$

which equal zero when $\theta = 0$. By construction,

$$B(t_j, u) = \int_u^{t_j} b(r, u) dr = \int_u^{t_j} \theta e^{\theta(r-u)} dr = e^{\theta(t_j-u)} - 1,$$

and hence

$$\begin{aligned} d_0 \int_{t_{j-1}}^{t_j} B(t_j, u)^2 du &= d_0 \int_{t_{j-1}}^{t_j} \left(e^{\theta(t_j-u)} - 1 \right)^2 du \\ &= d_0 \int_{t_{j-1}}^{t_j} \left(e^{2\theta(t_j-u)} - 2e^{\theta(t_j-u)} + 1 \right) du \\ &= d_0 \left[\int_{t_{j-1}}^{t_j} e^{2\theta(t_j-u)} du - 2 \int_{t_{j-1}}^{t_j} e^{\theta(t_j-u)} du + \Delta \right] \end{aligned}$$

$$= \begin{cases} d_0 \left[-\frac{1}{2\theta} (1 - e^{2\theta\Delta}) + \frac{2}{\theta} (1 - e^{\theta\Delta}) + \Delta \right] & \text{if } \theta \neq 0 \\ 0 & \text{if } \theta = 0. \end{cases}$$

The last term

$$\begin{aligned} & d_0 \int_0^{t_{j-1}} [B(t_j, u) - B(t_{j-1}, u)]^2 du \\ &= d_0 \int_0^{t_{j-1}} [e^{\theta(t_j-u)} - e^{\theta(t_{j-1}-u)}]^2 du \\ &= d_0 [e^{\theta\Delta} - 1]^2 e^{2\theta t_{j-1}} \int_0^{t_{j-1}} e^{-2\theta u} du \\ &= \begin{cases} \frac{d_0}{2\theta} (e^{\theta\Delta} - 1)^2 [e^{2\theta t_{j-1}} - 1] & \text{if } \theta \neq 0 \\ 0 & \text{if } \theta = 0. \end{cases} \end{aligned}$$

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

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

■

Proof of Proposition 2.2

Proof. The bias of the infeasible realized volatility estimator is

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

where $m_{t_j} = E(r_{t_j})$. From (35), the squared returns

$$\begin{aligned} r_{t_j}^2 &= \underbrace{\left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right)^2}_{(1)} + \underbrace{\left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right)^2}_{(2)} \\ &\quad + \underbrace{\left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right)^2}_{(3)} \\ &\quad + 2 \underbrace{\left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right) \int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s}_{(4)} \\ &\quad + 2 \underbrace{\left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right) \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right)}_{(5)} \\ &\quad + 2 \underbrace{\left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right) \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right)}_{(6)}, \end{aligned}$$

which consists of six terms. From the proof of Proposition 2.1, the (1), (2) and (3) terms are

$$\left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right)^2 = \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u \sigma_s dW_s \right) \sigma_u dW_u + \int_{t_{j-1}}^{t_j} \sigma_u^2 du;$$

$$\begin{aligned}
\left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right)^2 &= 2 \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u B(t_j, s) \sigma_s dW_s \right) B(t_j, u) \sigma_u dW_u \\
&\quad + \int_{t_{j-1}}^{t_j} B(t_j, u)^2 \sigma_u^2 du;
\end{aligned}$$

and

$$\begin{aligned}
&\left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right)^2 \\
&= 2 \int_0^{t_{j-1}} \left(\int_0^u [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) [B(t_j, u) - B(t_{j-1}, u)] \sigma_u dW_u \\
&\quad + \int_0^{t_{j-1}} [B(t_j, u) - B(t_{j-1}, u)]^2 \sigma_u^2 du.
\end{aligned}$$

Similarly, the cross product terms (4), (5) and (6) are

$$\begin{aligned}
&2 \left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right) \left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right) \\
&= 2 \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u \sigma_s dW_s \right) B(t_j, u) \sigma_u dW_u + 2 \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u B(t_j, s) \sigma_s dW_s \right) \sigma_u dW_u; \\
&2 \left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right) \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) \\
&= 2 \int_{t_{j-1}}^{t_j} \left(\int_0^u [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) \sigma_u dW_u \\
&\quad + 2 \int_0^{t_{j-1}} \left(\int_{t_{j-1}}^u \sigma_s dW_s \right) [B(t_j, u) - B(t_{j-1}, u)] \sigma_u dW_u;
\end{aligned}$$

and

$$\begin{aligned}
& 2 \left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right) \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) \\
&= 2 \int_{t_{j-1}}^{t_j} \left(\int_0^u [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right) B(t_j, u) \sigma_u dW_u \\
&\quad + 2 \int_0^{t_{j-1}} \left(\int_{t_{j-1}}^u B(t_j, s) \sigma_s dW_s \right) [B(t_j, u) - B(t_{j-1}, u)] \sigma_u dW_u.
\end{aligned}$$

The expected value of the bias is therefore

$$\begin{aligned}
E \left[(r_{t_j} - m_{t_j})^2 - \int_{t_{j-1}}^{t_j} \sigma_u^2 du \right] &= \int_{t_{j-1}}^{t_j} B(t_j, u)^2 E(\sigma_u^2) du \\
&\quad + \int_0^{t_{j-1}} [B(t_j, u) - B(t_{j-1}, u)]^2 E(\sigma_u^2) du.
\end{aligned}$$

Therefore,

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

when $\theta \neq 0$ and equals zero when $\theta = 0$. ■

Proof of Proposition 2.3

Proof. The bias of the feasible realized volatility estimator is

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

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

The first term is the bias of the realized volatility estimator derived in Proposition 2.1. We only need to derive the second and third terms. From Lemma 2.1,

$$\begin{aligned}
\hat{m}_{t_i} &= \frac{1}{K} \int_{t_i-K}^{t_i} a(r) dr + \frac{1}{K} \int_{t_i-K}^{t_i} \sigma_s dW_s + \frac{1}{K} \sum_{j=i-K+1}^i \int_{t_{i-j}}^{t_j} B(t_j, s) \sigma_s dW_s \\
&\quad + \frac{1}{K} \sum_{j=i-K+1}^i \int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s.
\end{aligned}$$

Following the proof of Proposition 2.1, we can show that

$$-2 \sum_{j=i-K+1}^i E(\hat{m}_{t_i} r_{t_j}) = -2 \frac{1}{K} \left(\int_{t_i-K}^{t_i} a(r) dr \right)^2.$$

Moreover,

$$\begin{aligned}
KE(\hat{m}_{t_i}^2) &= \frac{1}{K} \left(\int_{t_i-K}^{t_i} a(r) dr \right)^2 + \frac{1}{K} E \left(\int_{t_i-K}^{t_i} \sigma_s dW_s \right)^2 \\
&\quad + \frac{1}{K} \sum_{j=i-K+1}^i E \left(\int_{t_{i-j}}^{t_j} B(t_j, s) \sigma_s dW_s \right)^2 \\
&\quad + \frac{1}{K} \sum_{j=i-K+1}^i \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right)^2 \\
&= \frac{1}{K} \left(\int_{t_i-K}^{t_i} a(r) dr \right)^2 + d_0 + \frac{d_0}{K} \sum_{j=i-K+1}^i \int_{t_{j-1}}^{t_j} B(t_j, u)^2 du
\end{aligned}$$

$$+\frac{d_0}{K} \sum_{j=i-K+1}^i \int_0^{t_{j-1}} [B(t_j, u) - B(t_{j-1}, u)]^2 du$$

since

$$\begin{aligned} \frac{1}{K} E \left(\int_{t_{i-K}}^{t_i} \sigma_s dW_s \right)^2 &= \frac{1}{K} \int_{t_{i-K}}^{t_i} E(\sigma_u^2) du; \\ \frac{1}{K} \sum_{j=i-K+1}^i E \left(\int_{t_{j-1}}^{t_j} B(t_j, s) \sigma_s dW_s \right)^2 &= \frac{1}{K} \sum_{j=i-K+1}^i \int_{t_{j-1}}^{t_j} B(t_j, u)^2 E(\sigma_u^2) du; \\ \frac{1}{K} \sum_{j=i-K+1}^i \left(\int_0^{t_{j-1}} [B(t_j, s) - B(t_{j-1}, s)] \sigma_s dW_s \right)^2 &= \frac{1}{K} \sum_{j=i-K+1}^i \int_0^{t_{j-1}} [B(t_j, u) - B(t_{j-1}, u)]^2 E(\sigma_u^2) du. \end{aligned}$$

Therefore,

$$\begin{aligned} &E \left\{ \sum_{j=i-K+1}^i (r_{t_j} - \hat{m}_{t_i})^2 - \int_{t_{i-K}}^{t_i} \sigma_u^2 du \right\} \\ &= E \left(\sum_{j=i-K+1}^i r_{t_j}^2 - \int_{t_{i-K}}^{t_i} \sigma_u^2 du \right) - \frac{1}{K} \left(\int_{t_{i-K}}^{t_i} a(r) dr \right)^2 + d_0 \Delta \\ &\quad + \frac{d_0}{K} \sum_{j=i-K+1}^i \int_{t_{j-1}}^{t_j} B(t_j, u)^2 du + \frac{d_0}{K} \sum_{j=i-K+1}^i \int_0^{t_{j-1}} [B(t_j, u) - B(t_{j-1}, u)]^2 du. \end{aligned}$$

Furthermore,

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

$$\frac{d_0}{K} \sum_{j=i-K+1}^i \int_{t_{j-1}}^{t_j} B(t_j, u)^2 du = \frac{1}{K} \sum_{j=i-K+1}^i d_0 \int_{t_{j-1}}^{t_j} \left(e^{\theta(t_j-u)} - 1 \right)^2 du$$

$$\begin{aligned}
&= \frac{1}{K} \sum_{j=i-K+1}^i d_0 \int_{t_{j-1}}^{t_j} \left(e^{2\theta(t_j-u)} - 2e^{\theta(t_j-u)} + 1 \right) du \\
&= \frac{d_0}{K} \left[\sum_{j=i-K+1}^i \int_{t_{j-1}}^{t_j} e^{2\theta(t_j-u)} du - 2 \sum_{j=i-K+1}^i \int_{t_{j-1}}^{t_j} e^{\theta(t_j-u)} du + \Delta K \right] \\
&= \begin{cases} d_0 \left[\frac{1}{2\theta} (e^{2\theta\Delta} - 1) + \frac{2}{\theta} (1 - e^{\theta\Delta}) + \Delta \right] & \text{if } \theta \neq 0 \\ 0 & \text{if } \theta = 0 \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
\frac{d_0}{K} \sum_{j=i-K+1}^i \int_0^{t_{j-1}} [B(t_j, u) - B(t_{j-1}, u)]^2 du &= \frac{d_0}{K} \sum_{j=i-K+1}^i \int_0^{t_{j-1}} [e^{\theta(t_j-u)} - e^{\theta(t_{j-1}-u)}]^2 du \\
&= d_0 [e^{\theta\Delta} - 1]^2 \frac{1}{K} \sum_{j=i-K+1}^i e^{2\theta t_{j-1}} \int_0^{t_{j-1}} e^{-2\theta u} du \\
&= \begin{cases} \frac{d_0}{2\theta K} (e^{\theta\Delta} - 1)^2 \left[e^{2\theta t_{i-K}} \frac{1 - e^{2\theta\Delta K}}{1 - e^{2\theta\Delta}} - K \right] & \text{if } \theta \neq 0 \\ 0 & \text{if } \theta = 0. \end{cases}
\end{aligned}$$

Therefore,

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

when $\theta \neq 0$ and equals zero when $\theta = 0$. ■

Proof of Remark 2.1

Proof. (1) The squared return of the constant drift process is

$$r_{t_j}^2 = \underbrace{\left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right)^2}_{(1)} + \underbrace{(m\Delta)^2}_{(2)} + \underbrace{2m\Delta \int_{t_{j-1}}^{t_j} \sigma_s dW_s}_{(3)}.$$

The first term

$$\left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right)^2 = \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u \sigma_s dW_s \right) \sigma_u dW_u + \int_{t_{j-1}}^{t_j} \sigma_u^2 du.$$

Therefore,

$$E \left\{ r_{t_j}^2 - \int_{t_{j-1}}^{t_j} \sigma_u^2 du \right\} = m^2 \Delta^2.$$

and the bias of RV is

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

(2) Similarly, the square term of the demeaned returns

$$[r_{t_j} - E(r_{t_j})]^2 = \left(\int_{t_{j-1}}^{t_j} \sigma_s dW_s \right)^2 = \int_{t_{j-1}}^{t_j} \left(\int_{t_{j-1}}^u \sigma_s dW_s \right) \sigma_u dW_u + \int_{t_{j-1}}^{t_j} \sigma_u^2 du.$$

Therefore, the bias of $RV_{t_i}^\dagger(K)$ is zero.

(3) The bias of the feasible modified realized volatility estimator RV^* is

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

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

The sample average

$$\hat{m}_{t_i} = m\Delta + \frac{1}{K} \int_{t_i-K}^{t_i} \sigma_s dW_s.$$

We can show that

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

$$KE(\hat{m}_{t_i}^2) = Km^2\Delta^2 + \frac{1}{K} \int_{t_i-K}^{t_i} E(\sigma_u^2) du = Km^2\Delta^2 + d_0\Delta.$$

Therefore,

$$E \left\{ \sum_{j=i-K+1}^i (r_{t_j} - \hat{m}_{t_i})^2 - \int_{t_i-K}^{t_i} \sigma_u^2 du \right\} = d_0\Delta.$$

■