CARDIFF BUSINESS SCHOOL
WORKING PAPER SERIES

Exploring drift bias in asymmetric jump estimation and its implications for volatility forecasting

Kevin P. Evans, Dudley Gilder, and Kefu Liao [1]

December 2023

Cardiff Business School
Cardiff University
Colum Drive Cardiff CF10 3EU
United Kingdom
t: +44 (0)29 2087 4000
f: +44 (0)29 2087 4419
www.cardiff.ac.uk/carbs

[1] Contact author. Email: LiaoK2@cardiff.ac.uk. Kefu acknowledges the research support of Cardiff University and the China Scholarship Council (CSC).

Exploring drift bias in asymmetric jump estimation and its implications for volatility forecasting

Abstract

The intraday log returns of financial assets are conventionally assumed to follow a drift-diffusion process. While the drift term is ignored by the infill asymptotic theory, under the assumption…, it may, in fact, be large for realistic samples of observations. Consequently, the volatility estimators relying on assembling intraday logarithm returns may be biased. In an extensive simulation study, we show that for the realistic samples the drift component has a non-negligible impact on the estimation accuracy of upside and downside volatility estimators, which leads to a dramatic bias for the jump asymmetry estimators made from the difference between upside and downside volatility estimators. We propose an alternative construction of the volatility estimators and observe significant improvements in the estimation accuracy in the presence of non-negligible drift. The new jump asymmetry estimators, which are constructed from the modified volatility estimators, also have significantly reduced drift bias.

In an empirical application, we compare the new jump asymmetry estimators with their original versions that include drift bias, for forecasting volatility using a sample of 25 years of high frequency log returns of the S&P 500 stock price index. The empirical results show that the predictability of the original jump asymmetry estimators found in the literature is almost exclusively due to drift bias, whereas the new jump asymmetry estimators, which measure the jump asymmetry much more precisely, show only limited forecasting power. We show that extracting the drift bias from the jump asymmetry estimators results in significant out-of-sample volatility forecast improvements.

Contents

1 Introduction

It is widely believed that asset prices obey an Ito semi-martingale process. In the high-frequency financial econometrics literature, asset prices are typically modelled as an Ito semi-martingale process with two main components: drift and diffusion. With locally bounded coefficients of drift and volatility (as in the volatility estimation literature), the drift term is dominated by the diffusion process. In other words, the price move (or intraday log return) is always dominated by the diffusion process if the time interval of the move is small. Because the drift component is omittable in the intraday log returns, these volatility estimators, which attempt to estimate volatility via an aggregate function of intraday log returns over an estimation window (based on the asymptotic theories), are unaffected by the presence of a non-zero drift. Typical volatility estimators include the Realized Variance (RV) by Andersen and Bollerslev (1998), which attempts to measure volatility by aggregating intraday squared log returns over an estimation window, the Bipower Variation (BV) by Barndorff-Nielsen and Shephard (2004), which attempts to measure volatility without the interference of occasional price jumps by aggregating the multiplication of neighbouring intraday absolute log returns over an estimation window, and Realized Semivariance (RS) by Barndorff-Nielsen et al. (2008)), whose positive version (RS^+) attempts to measure the upside volatility by aggregating positive intraday squared log returns and negative version (RS^-) attempts to measure the downside volatility by aggregating negative intraday squared log returns over an estimation window.

For the real financial market, there is often only a finite sample of price observations as the time interval of the price move is not possible to be infinitesimal in practise. A realistic and popular time interval in the literature for intraday log returns is 5 minutes. Recent research by Laurent and Shi (2020) shows that the 5-minute interval is not sufficiently small to ensure that

the drift component in log returns is ignorable. As a consequence, the volatility estimators, built upon aggregated log returns, will capture both volatility and drift components and thus be biased in estimating the volatility. For example, RV and BV based on the 5-minute log returns will capture both volatility and drift variation thus overstating the volatility level, and this estimation issue has been evidenced by Laurent and Shi (2020). Additionally, although the bias of RS due to finite sample drift is also expected to be large, existing studies have not yet studied the bias due to a nonzero drift in RS. My first contribution to the literature in this paper is to investigate the bias of RS in the presence of a nonzero drift via a Monte Carlo study. As expected, the results show that the drift is non-zero and the bias of RS is not ignorable. RS^+ tends to strongly overstate (weakly understate) the upside volatility in the presence of a positive (negative) drift. RS^- tends to strongly overstate (weakly understate) the downside volatility in the presence of a negative (positive) drift.

To reduce this finite sample bias of the volatility estimators, Laurent and Shi (2020) suggest removing the drift component in the log returns before computing these estimators. In other words, they propose computing the volatility estimators on centred log returns as opposed to raw log returns. Their theoretical and simulation results suggest that this modification leads to a dramatic improvement in the estimation accuracy of volatility, especially when the drift deviates far away from zero. However, Laurent and Shi (2020) focus on the RV and BV estimators and do not consider semi-variances. To reduce this finite sample bias of RS^+ and RS^-, We compute the RS^+ and RS^- based on the centred log returns. The Monte Carlo results show that this modification leads to a dramatic improvement in the estimation accuracy of upside volatility and downside volatility.

Volatility plays a central role in finance. It is the most important type of market risk and is fundamental to asset pricing, portfolio choice, risk management, and financial market regulation. Volatility estimators have been used extensively in this field. The finite sample bias

of volatility estimation arising from the presence of a non-zero drift is therefore expected to have many consequences. For example, Laurent and Shi (2020) demonstrate that the finite sample bias of the volatility estimators could lead to the unsatisfactory performance of jump detection procedures by Lee and Mykland (2008). There may be other consequences of the finite sample bias of the volatility estimators. One consequence, which has not yet been explored in the literature, is the unsatisfactory accuracy of estimation of price jumps since combinations of the volatility estimators are commonly used for jump estimation (Barndorff-Nielsen and Shephard, 2004, Barndorff-Nielsen et al., 2008, Christensen et al., 2014).

We inspect the drift bias of all possible combinations of RV, BV, RS^+, and RS^- estimators for estimating jumps, based on the simulation results in this chapter, along with the simulation results in Laurent and Shi (2020). The inspection results show that only the estimation bias of $RS^+ - RS^-$ by Barndorff-Nielsen et al. (2008) for estimating the jump asymmetry (the difference between positive jumps and negative jumps) tend to be large, while the biases of remaining combinations are generally small. Motivated by this result, we investigate the bias of $RS^+ - RS^-$ in detail. Specifically, we conduct a Monte Carlo study for the bias of $RS^+ - RS^-$ in the presence of a nonzero drift under various cases of jump asymmetry. The main conclusion is that $RS^+ - RS^-$ will be significantly upward biased due to a positive drift and significantly downward biased due to a negative drift.

To address the finite sample issue of the jump asymmetry estimator, we propose an alternative construction of this estimator by the difference between modified RS^+ and RS^-. The Monte Carlo study shows that this modified estimator leads to a significant improvement in the estimation accuracy in the presence of a non-zero drift. Given this result, we propose measuring the jump asymmetry by the modified jump asymmetry estimator and defining the drift bias of the original jump asymmetry estimator by the gap between this estimator and its modified version.

The log prices are contaminated by noise for ultra high-frequency data, which results in an additional noise component in the ultra high-frequency log returns. Therefore, the volatility estimators including RV, BV, RS^+, and RS^- will largely overstate the volatility. To improve the estimation of volatility with noisy ultra high-frequency data, one popular method is calculating the volatility estimators using pre-averaged returns (Podolskij and Vetter, 2009, Jacod et al., 2009, Christensen et al., 2014), which average noisy ultrahigh-frequency log prices within an appropriate window.

Despite the drift of log returns becoming extremely close to zero as the sampling frequency increases, the drift of the pre-averaged returns may be of a non-negligible magnitude since the pre-averaged returns are fundamentally an average of some low-frequency returns. As a consequence, the bias of the noise-modified volatility estimators due to a non-zero drift will also be large. To reduce the bias, it is natural to compute these volatility estimators on centred pre-averaged returns. As expected, the result of the Monte Carlo study in this paper shows that the bias of the noise-modified RV, BV, RS^+, and RS^- volatility estimators tend to be large and have similar patterns as their original versions, and the bias of the (drift-) modified versions are substantially reduced. In addition, the simulation results also show that the drift bias of the noise-modified jump asymmetry estimator is large and has similar patterns as its original version, and the bias of the noise-modified jump asymmetry estimator which is computed on centred pre-averaged return is much smaller.

The jump asymmetry is one of the most important types of tail risk and is fundamental to volatility forecasting, asset pricing, and financial market regulation. The finite sample bias of the jump asymmetry estimation arising from the presence of a non-zero drift is therefore expected to have many secondary impacts. As an example, we demonstrate in this paper that the finite sample bias of the jump asymmetry estimator $RS^+ - RS^-$ could lead to the misleading predictability of the jump asymmetry on future volatility. We show that the

predictability of the jump asymmetry estimator $RS^+ - RS^-$ found in Patton and Sheppard (2015) is almost exclusively due to its bias due to a non-zero drift, and the modified estimator, which measures the jump asymmetry much more accurately, indicates only limited predictive power. Additionally, we show that extracting the drift bias from the $RS^+ - RS^-$ leads to significantly better out-of-sample forecasting accuracy. Finally, these forecasting results are consistent for ultra high-frequency data, using noise-modified realized semi-variances.

The findings of the empirical application in this paper imply that the jump information is not important to volatility forecasting, which provides more evidence to a debate in existing research on the predictability of jumps (Andersen et al., 2007, Corsi et al., 2010, Corsi and Renò, 2012, Santos and Ziegelmann, 2014, Sévi, 2014, Patton and Sheppard, 2015, Prokopczuk et al., 2016, Bollerslev et al., 2021, Caporin, 2023). My findings of the prominent predictability of drift bias are novel to the literature.

The rest of this paper is structured as follows. Section 2 revisits volatility estimation for the drift-diffusion processes. Section 3 introduces the jump asymmetry estimation for the drift-diffusion processes. Sections 4 and 5 present the estimation of volatility and jump asymmetry for ultra high-frequency data, respectively. Section 6 reports the descriptive statistics for the volatility and jump asymmetry estimators applied to the US stock markets. Section 7 shows the empirical application of these estimators for volatility forecasting. Section 8 concludes.

2 Volatility estimation for drift-diffusion processes

A conventional assumption is that the logarithmic prices process follows a semi-martingale with no arbitrage. Specifically, log prices p_t are assumed to follow a drift-diffusion process,

$$dp_t = \mu_t d_t + \sigma_t dW_t + \kappa_t d\varsigma_t, 1 \leq t \leq, \tag{2.1}$$

where μ_t is the drift process, σ_t is the càdlàg stochastic volatility process, W_t is a standard Brownian motion, κ_t represents the random jump size at time t, and the counting process ς_t denotes a finite activity counting process with intensity λ_t, which may be constant or stochastic (e.g., ς_t is an independent Poisson counting process with constant intensity, λ), and κ_t represents the random jump sizes. The price variation or volatility (risk) is defined in this paper by the quadratic variation of Equation (2.1), [2]

$$V_t = \int_0^T \sigma_{t_s}^2 + \sum_{1<s\leq T} \kappa_s^2, \tag{2.2}$$

where V_t indicates the quadratic variation, $\int_0^T \sigma_{t_s}^2$ denotes the component of quadratic variation due to the continuous part of the price process (or continuous variation, or integrated variation)T, $\sum_{1<s\leq T} \kappa_s^2$, which is the summation of squared jumps, denotes the component of quadratic variation due to the discontinuous part of the price process (or jump variation), with κ_s capturing the size of a jump if present. If the signs of jumps are considered, the jump variation of Equation (2.2) can be further decomposed based on the positive and negative jumps,

$$V_t = \int_0^T \sigma_{t_s}^2 + \sum_{1<s\leq T} \kappa_s^2 I(\kappa_s > 0) + \sum_{1<s\leq T} \kappa_s^2 I(\kappa_s < 0), \tag{2.3}$$

where $I[.]$ is an indicator function, $\sum_{1<s\leq T} \kappa_s^2 I(\kappa_s > 0)$ denotes the component of quadratic variation due to the discontinuous part of the upside price process (or positive jump variation),

[2] This chapter will use the terms volatility and quadratic variation interchangeably.

and $\sum_{1<s\leq T} \kappa_s^2 I(\kappa_s < 0)$ denotes the component of quadratic variation due to the discontinuous part of the downside price process (or negative jump variation).

Assume $M+1$ log prices p_{t_i} are observed at equally spaced intervals $0 = t_0 < t_1 < t_2 \ldots < t_M = T = 1$ (within $[0, T]$, there are overall M intervals) for the intraday session of a trading day t (e.g., 09:30-16:00 for the New York Stock Exchange). The ith intraday returns are calculated by $r_{t_i} = p_{t_i} - p_{t_{i-1}}$ thus there are overall M intraday returns within $[0, \mathrm{T}]$. This paper focuses on volatility for the intraday session and ignores the overnight return, following the vast realized volatility literature (Andersen et al., 2007, Corsi, 2009, Bollerslev et al., 2016, Duong and Swanson, 2015, Patton and Sheppard, 2015).

To estimate the quadratic variation V_t within $[0, T]$, Andersen and Bollerslev (1998) proposed a Realized Variation or Realized Volatility (RV), which is defined by the aggregation of M squared intraday returns within $[0, T]$,

$$RV_t = \sum_{i=1}^{M} r_{t_i}^2 \tag{2.4}$$

They show that the probability limit of RV converges to the quadratic variation as the number of observations (M) becomes larger and hence each interval gets smaller,

$$RV_t = \sum_{i=1}^{M} r_{t_i}^2 \xrightarrow{p} \int_0^T \sigma_{t_s}^2 + \sum_{1<s\leq T} \kappa_s^2 , \text{ as } M \to \infty \tag{2.5}$$

To estimate the continuous variation only in the presence of jumps, Barndorff-Nielsen and Shephard (2006) proposed Bipower Variation (BV), defined by the summation of appropriately scaled cross-products of adjacent high-frequency absolute returns,

$$BV_t = \frac{\pi}{2} \sum_{i=2}^{M} |r_{t_{i-1}}||r_{t_i}| \tag{2.6}$$

As noted by Barndorff-Nielsen and Shephard (2006), BV converges in probability to continuous variation if M becomes large,

$$BV_t = \frac{\pi}{2}\sum_{i=2}^{M}\left|r_{t_{i-1}}\right|\left|r_{t_i}\right| \xrightarrow{p} \int_0^T \sigma_{t_s}^2, \text{ as } M \to \infty. \tag{2.7}$$

Volatility can also be separated into upside and downside components as proposed by Barndorff-Nielsen et al. (2008). The upside part of the volatility enables one to know the possibility of prices rising. And the downside part of the volatility enables one to know the risk of prices falling. To measure the upside and downside parts of the volatility, Barndorff-Nielsen et al. (2008) suggest using the positive realized semi-variance and negative realized semi-variance, respectively. Specifically, the positive realized semi-variance (RS^+) is defined by aggregating all positive squared log returns,

$$RS_t^+ = \sum_{i=1}^{M} r_{t_i}^2 I\left(r_{t_i} > 0\right) \tag{2.8}$$

and the negative realized semi-variance (RS^-) is defined by aggregating all negative squared log returns,

$$RS_t^+ = \sum_{i=1}^{M} r_{t_i}^2 I\left(r_{t_i} > 0\right). \tag{2.9}$$

Additionally, Barndorff-Nielsen et al. (2008) investigate the asymptotic properties of both realized semi-variances, relating to the continuous variation and jumps. Specifically, they show that as the number of observations (M) increases, the probability limit of the positive realized semi-variance converges to one-half of the continuous variation plus the jump variation only due to positive jumps (or positive jump variation),

$$RS_t^+ \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_{t_s}^2 + \sum_{1<s\le t} \kappa_s^2 I(\kappa_s > 0), \text{ as } M \to \infty, \tag{2.10}$$

and the probability limit of the negative realized semi-variance converges to one-half of the continuous variation plus the jump variation only due to negative jumps (or negative jump variation),

$$RS_t^- \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_{t_s}^2 + \sum_{1<s\leq t} \kappa_s^2 I(\kappa_s < 0)\,, \text{ as } M \to \infty. \tag{2.11}$$

It is important to note that there is no variation due to drift in Equation (2.5) and Equation (2.11). This is because as $M \to \infty$ the drift component is ignored because … relative to the volatility in the intraday log returns. For example, for Equation (2.1), the drift and volatility coefficients of the Ito semi-martingale are bounded by stochastic processes. Let Δ_M denote the interval between observations, $\Delta_M = T/M$. For $\Delta_M \leq t \leq T$ and for $\Delta_M \to 0$ (or $\Delta_M \to \infty$), then

$$\int_{t-\Delta_M}^{t} \mu_s\, ds = O_p(\Delta_M), \tag{2.12}$$

and

$$\int_{t-\Delta_M}^{t} \sigma_s\, dW = O_p\left(\Delta_M^{1/2}\right), \tag{2.13}$$

where $O_p\big(g(x)\big)$ denotes an upper bound ($g(x)$) on the growth rate of the function. It can be seen from Equations (2.9) and (2.10) that the upper bound of drift Δ_M is omittable relative to that of the volatility $\Delta_M^{1/2}$ as $M \to \infty$ ($\Delta_M/\Delta_M^{1/2} = \Delta_M^{1/2} = 1/\sqrt{M} \to 0$ as $M \to \infty$). This indicates that drift is much smaller than the volatility as the number of observations becomes larger. In other words, the volatility always dominates the intraday returns.

The above negligibility of drift requires an infinite number of observations ($M \to \infty$). However, in practice, an infinite number of observations for asset prices is not possible. The low-frequency data is often used for volatility estimation. For example, the 5-minute frequency data, which defines only 79 observations for US stocks during the intraday trading sessions, is often used for volatility estimation, since 5 minutes are preferred to balance the trade-off

between the distortion from market microstructure noise and estimation efficiency from an increasing number of observations (Andersen et al., 2000, Andersen et al., 2001, Corsi, 2009, Park and Linton, 2012), or since ultra high-frequency data for asset prices are not always available. Under the low-frequency sampling scheme, the drift is not necessarily ignorable compared to that of the volatility in the log returns. For example, for 5-minute frequency sampling during the official trading session of the US stock market (78 intraday returns), the upper bound of the drift is not ignorable compared to that of the volatility in the log return as $\Delta_M/\Delta_M^{1/2} = \Delta_M^{1/2} = 1/\sqrt{78} = 11.3\%$. Moreover, as shown by Laurent and Shi (2020), the level of drift may be empirically much larger than the level of volatility (e.g., for the US stock market). This further supports that the proportion of drift in the intraday log returns is not ignorable.

Motivated by the suggestion that drift is a non-negligible component of the intraday log returns for the low-frequency sampling schemes, Laurent and Shi (2020) investigate whether this fact leads to bias in the estimation of volatility using the popular RV and BV measures, since they are both computed by aggregating intraday returns. To study the drift bias in RV and BV, Laurent and Shi (2020) assume that log prices follow a constant drift-diffusion process. Specifically, the constant drift-diffusion process assumes that both drift, μ_t, and volatility, σ_t, in Equation (2.1) are constant,

$$dp_t = \mu d_t + \sigma dW_t + \kappa_t dq_t, 1 \leq t \leq T, \tag{2.14}$$

where μ is the constant drift parameter and σ denotes the constant volatility. For the generality of the results, they extend the constant drift-diffusion process to a more comprehensive linear drift-diffusion process, where the drift coefficient is a linear function of the log prices and the diffusion coefficient σ_t is an adapted and cádlág volatility process,

$$dp_t = \theta p_t dt + \sigma_t \mathrm{d}W_t + \kappa_t dq_t, 1 \leq t \leq T, \tag{2.15}$$

where θ is a constant parameter. As drift is a linear function of the log price, its magnitude is time-varying. Laurent and Shi (2020) suggest measuring the magnitude of linear drift within an estimation window by its average (measured by median) within this window, and they show that given a fixed sampling frequency the drift is an increasing function of θ and the initial value of the log price of the day. And they also show that the sign of the drift is consistent with the sign of θ.

Based on the log-returns of the constant and linear drift-diffusion model, Laurent and Shi (2020) show that for low-frequency data, RV will inherit the dynamics of the drift sample path process in addition to the continuous sample path process and the jump process. Meanwhile, BV will inherit the dynamics of drift sample path process in addition to the continuous sample path process. Since RV attempts to estimate the continuous variation plus jump variation and BV attempts to estimate the continuous variation, the non-zero drifts will result in an upward bias of RV and BV for low-frequency data.

To mitigate the drift bias in volatility estimators under low-frequency data, Laurent and Shi (2020) suggest removing the drift component before calculating volatilities or, equivalently, computing the volatility estimators on centred log returns. For centring, Laurent and Shi (2020) suggest the median, as the median is much less sensitive to jumps, which is assumed to occur occasionally by the consensus in the literature. Then, the realized variance and bipower variation computed on the returns centred by the median within the estimation window (indicated by RV^* and BV^*, respectively) are defined by,

$$RV_t^* = \sum_{i=1}^{M} \left(r_{t_i} - \mathcal{M}_t\right)^2, \tag{2.16}$$

$$BV_t^* = \frac{\pi}{2} \sum_{i=2}^{M} \left|r_{t_{i-1}} - \mathcal{M}_t\right| \left|r_{t_i} - \mathcal{M}_t\right|, \tag{2.17}$$

where $\mathcal{M}_t = \text{median}(r_{t_1} \dots r_{t_M})$ denotes the median of M log returns for day t that are involved in the computation of the volatilities for that day. The derivation and simulation results by Laurent and Shi (2020) show that in the presence of non-zero drift, the bias of RV* and BV* is extremely close to zero and is ignorable relative to that of RV and BV, respectively.

Since RS^+ and RS^- are also calculated using log returns with a non-negligible drift component for low-frequency data, the drift bias in RS^+ and RS^- is expected to be large. This bias should be substantially reduced by computing RS^+ and RS^- based on centred log returns, when the drift component is largely attenuated. The modified realized semi-variances are defined as the sum of squared positive or negative returns centred by their median within the estimation window,

$$RS_t^{*,+} = \sum_{i=1}^{M} (r_{t_i} - \mathcal{M}_t)^2 \, I(r_{t_i} - \mathcal{M}_t > 0), \tag{2.18}$$

$$RS_t^{*,-} = \sum_{i=1}^{M} (r_{t_i} - \mathcal{M}_t)^2 \, I(r_{t_i} - \mathcal{M}_t < 0), \tag{2.19}$$

where $RS_t^{*,+}$ and $RS^{*,-}$ indicate the modified positive and negative realized semi-variances, respectively. Accordingly, the sum of these two modified realized semi-variances ($RS^{*,+}$ and $RS^{*,-}$) equals the modified realized variance,

$$RV_t^* = RS_t^{*,+} + RS_t^{*,-}. \tag{2.20}$$

The expected bias of RS^+, RS^-, $RS^{*,+}$ and $RS^{*,-}$ has not yet been considered in the literature. Therefore, my investigation provides a novel contribution. To study the bias in realised semi-variances in more detail, this paper conducts a Monte Carlo study when the Data-Generating Process (DGP) is a drift–diffusion process. For the drift–diffusion process, this paper only reports the linear version (Equation (2.12)) for two reasons. First, the linear drift-diffusion process contains a more general volatility dynamic and thus has a wider application

than the constant drift-diffusion process. Second, We avoid repetition and save space by reporting the results from the constant drift case in Appendix XX. Since the conclusions made from both simulation exercises are almost identical, We only report the most relevant framework in the chapter.

Additionally, the biases of the semi-variances are studied only via the simulation in this chapter. The exact derivation of the bias of the semivariances under the linear drift–diffusion process is rather complicated and left for future research. There is a precedent for this approach to research. The Monte Carlo approach without technical derivation has already been applied by Laurent and Shi (2020) to gauge the bias of Bipower Variation under a linear drift-diffusion process, since the derivation of such bias is also not available due to its complexity. Furthermore, the simulation results in the paper are not intended to be comprehensive, but rather to reflect a realistic application of the linear drift-diffusion models.

Following Laurent and Shi (2020), the DGP of the linear drift-diffusion process is defined by a discrete version of Equation (2.15),

$$p_{t_i} = \exp(\theta\Delta_M)p_{t_{i-1}} + \eta_{t_i} + \sum_{j=1}^{k} \phi_{t_i}^{j} I_{t_i}^{j}, \tag{2.21}$$

where p_{t_i} denotes the log price, $I_{t_i}^{j}$ is a dummy variable that randomly assigns the occurrence of the j^{th} jump with corresponding jump size $\phi_{t_i}^{j}$, k indicates the number of jumps, and η_{t_i} is the diffusive volatility process, which is obtained by the Euler discretization of the continuous $\text{GARCH}(1,1)$ process by Nelson (1990),

$$\eta_{t_i} = \sigma_{t_i}\sqrt{\Delta_M}\epsilon_{t_i},$$

$$\sigma_{t_i}^2 = \kappa\left(\omega - \sigma_{t_{i-1}}^2\right)\Delta_M + \sqrt{2\lambda\kappa}\sigma_{t_{i-1}}^2\sqrt{\Delta_M}\nu_{t_i},$$

with $\kappa > 0$, $\omega > 0$, and $0 < \lambda < 1$, and ϵ and v are two independent standard normal random variables. As shown by Andersen and Bollerslev (1998), the unconditional variance ($E\left(\sigma_{t_i}^2\right)$) of this discretized GARCH(1,1) model is ω.

The expected bias of RS^+ under the DGP is defined by this estimator minus half of the unconditional variance and positive jump variation,

$$\text{bias of } RS^+ = \mathbb{E}\left\{RS_t^+ - \left[\frac{1}{2}\omega + \sum_{i=1}^{M}\sum_{j=1}^{k}\left(\phi_{t_i}^j\right)^2 I_{t_i}^j\, I\left(\phi_{t_i}^j > 0\right)\right]\right\}, \tag{2.22}$$

and the expected bias of RS^- is defined by this estimator minus half of the unconditional variance and negative jump variation,

$$\text{bias of } RS^- = \mathbb{E}\left\{RS_t^- - \left[\frac{1}{2}\omega + \sum_{i=1}^{M}\sum_{j=1}^{k}\left(\phi_{t_i}^j\right)^2 I_{t_i}^j\, I\left(\phi_{t_i}^j < 0\right)\right]\right\}. \tag{2.23}$$

The expected bias of $RS^{*,+}$ under the DGP is defined by this estimator minus half of the unconditional variance and positive jump variation,

$$\text{bias of } RS^{*,+} = \mathbb{E}\left\{RS_t^{*,+} - \left[\frac{1}{2}\omega + \sum_{i=1}^{M}\sum_{j=1}^{k}\left(\phi_{t_i}^j\right)^2 I_{t_i}^j\, I\left(\phi_{t_i}^j > 0\right)\right]\right\}, \tag{2.24}$$

and the expected bias of $RS^{*,-}$ is defined by this estimator minus half of the unconditional variance and negative jump variation,

$$\text{bias of } RS^{*,-} = \mathbb{E}\left\{RS_t^{*,-} - \left[\frac{1}{2}\omega + \sum_{i=1}^{M}\sum_{j=1}^{k}\left(\phi_{t_i}^j\right)^2 I_{t_i}^j\, I\left(\phi_{t_i}^j < 0\right)\right]\right\}. \tag{2.25}$$

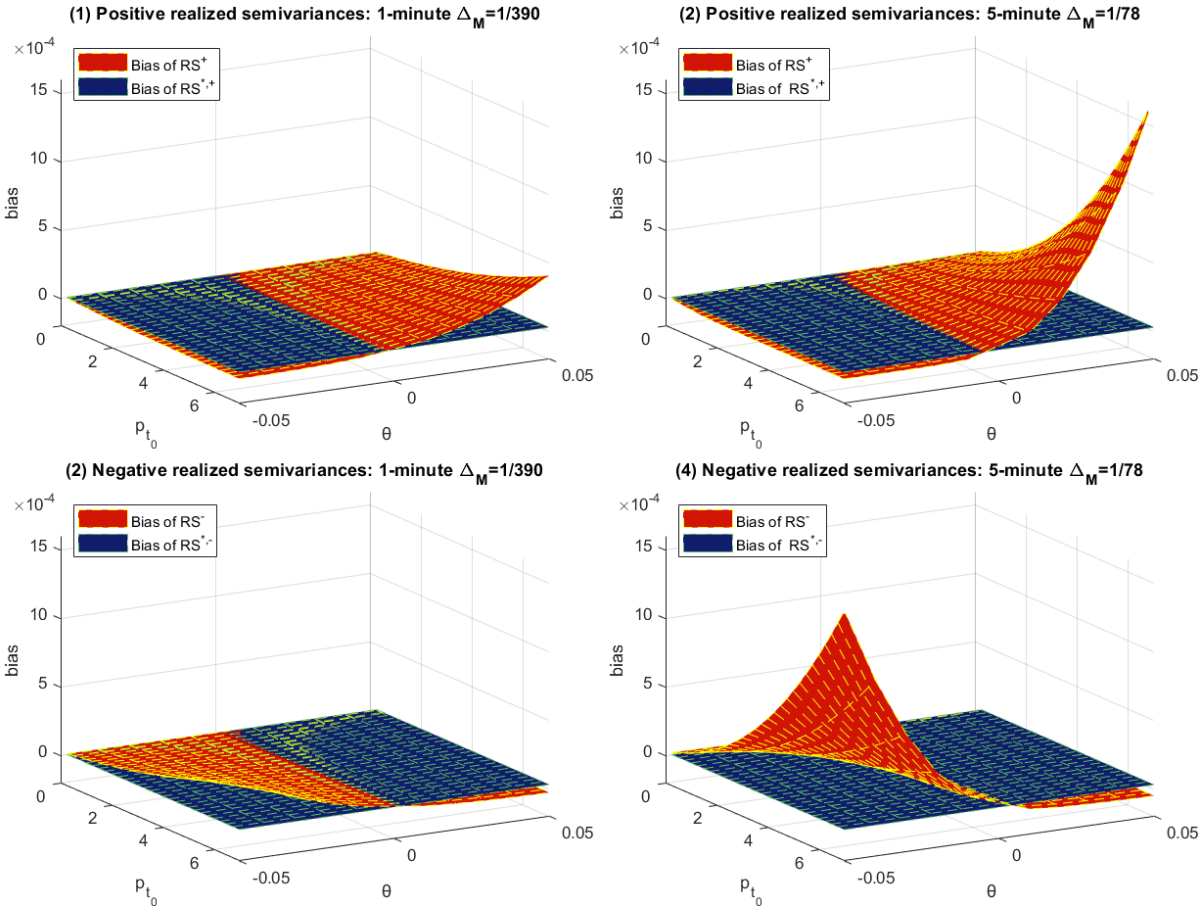
The parameter setting in Equation (2.16) is the same as in Laurent and Shi (2020) and Andersen and Bollerslev (1998): one negative jump is allowed per day with a size equal to 60% of the spot volatility ($\phi_{t_i}^1 = -0.6\sigma$), $\kappa = 0.035$, $\lambda = 0.296$, $\omega = 10^{-4}$, the initial volatility for the GARCH(1,1) process is set as the unconditional volatility of this process ($\sqrt{\omega}$), and W and B are independent standard Brownian motions. For the ranges of the values of θ and p_{t_0} for

determining the drift, this paper follows Laurent and Shi (2020): θ is set from -0.05 to 0.05, the initial log price p_{t_0} varies from 0 to 7. Besides, We assume that the asset is traded 6.5 hours per day as is the case on the NYSE and the Nasdaq stock exchanges (from 9:30am to 4:00pm). In other words, there are 23401 1-second simulated log prices over one day. The sampling interval is set to $\Delta_M = 1/390$ and $\Delta_M = 1/78$ for the 1- and 5-minute data, respectively. Following Laurent and Shi (2020), the simulations are repeated 10^4 times and the expectation ($\mathbb{E}$) is realized by averaging these repetitions.

Figure 2.1 plots the expected bias of realized semi-variance estimators (RS^+ and RS^-) as the red surfaces, along with the modified realized semi-variance estimators ($RS^{*,+}$ and $RS^{*,-}$) depicted in blue. Recall that the drift of log returns increases with $|\theta|$ and the initial value p_{t_0}, with the sign of the drift consistent with the sign of . There is an upward bias in RS^+ when the drift is positive (θ is positive and p_{t_0} is not zero), and this bias becomes larger as the sampling frequency decreases and as the magnitude of the positive drift increases. There is a small downward bias in RS^+ when the drift is negative (θ is negative and p_{t_0} is not zero), but the magnitude of this bias is not very sensitive to the changes of the negative drift. Additionally, Figure 2.1 shows that the bias in RS^- due to the non-zero drift is symmetrical to that of RS^+: there is an upward bias of RS^- for a negative drift, while there is a small downward bias of RS^+ for the positive drift. As expected, the biases of the modified realized semi-variances are much smaller than those of their respective original estimators in finite samples, with the biases extremely close to zero. The discrepancy becomes increasingly visible as the drift value deviates further away from zero.

Figure 2.1. Bias of RS^{+} and RS^{-} under the linear drift-diffusion process.

Notes: This figure reports the biases of RS^{+}, RS^{-}, $RS^{*,+}$, and $RS^{*,-}$ under the linear drift-diffusion process, with the biases calculated by Equations (2.22), (2.23), (2.24), and (2.25), respectively.



3 Jump asymmetry estimation for drift-diffusion processes

The imprecise estimation of RV, BV, RS^+ and RS^- due to a non-zero drift may lead to the unsatisfactory accuracy of estimation of price jumps, as the combination of these volatility estimators is commonly used for estimating jumps in the literature (Barndorff-Nielsen and Shephard, 2004, Barndorff-Nielsen et al., 2008, Christensen et al., 2014).

The first combination is the difference between RV and BV, which estimates the jump variation based on the asymptotic results by Barndorff-Nielsen and Shephard (2004),

$$RV_t - BV_t \xrightarrow{p} \sum_{1<s\leq T} \kappa_s^2, \text{ as } M \to \infty. \tag{3.1}$$

Since RV and BV estimators are biased due to a non-zero drift for finite samples, the estimation of the jump variation may also be biased. However, as shown by Laurent and Shi (2020), the bias in both RV and BV is positive and comparable in the presence of a non-zero drift. By taking the difference of RV and BV, a large proportion of the bias is removed. The drift bias thus has a small impact on the estimation of the jump variation.

The second combination of RV and BV relates to identifying upside and downside jump variation. Specifically, this is calculated as the difference between RS^+ and a half of BV ($RS^+ - 0.5BV$), which, according to the asymptotic results of Barndorff-Nielsen et al. (2008), estimates the positive jump variation,

$$RS_t^+ - \frac{1}{2}BV_t \xrightarrow{p} \sum_{1<s\leq t} \kappa_s^2 I(\kappa_s > 0), \text{ as } M \to \infty. \tag{3.2}$$

As both RS^+ and BV are biased due to a non-zero drift, the estimation of the jump variation may also be biased. However, comparing Figure 1.1 in this paper with Figure 7 by Laurent and Shi (2020), the bias of both RS^+ and a half of BV are upward and comparable when the drift is positive. This indicates that a large proportion of the bias due to a positive drift is eliminated when taking the difference between RS^+ and a half B.

Analogously, the third combination is the difference between RS^- and a half of BV (or $RS^- - 0.5BV$), which estimates the negative jump variation based on the asymptotic results by Barndorff-Nielsen et al. (2008),

$$RS_t^- - \frac{1}{2}BV_t \xrightarrow{p} \sum_{1<s\leq t} \kappa_s^2 I(\kappa_s < 0), \text{ as } M \to \infty. \tag{3.3}$$

As both RS^- and a half of BV are biased due to a non-zero drift for the finite sample, the estimation of the negative jump variation may also be biased. However, comparing Figure 1.1 in this paper with Figure 7 by Laurent and Shi (2020), the bias of both RS^- and a half of BV are upward and comparable when the drift is negative. This indicates that a large proportion of the bias due to a negative drift cancels out when making the difference between RS^- and a half B.

Finally, the fourth example of combining RV and BV for the estimation of jump variation is the calculation of the difference between RS^+ and RS^-. Based on the asymptotic results of Barndorff-Nielsen et al. (2008), $RS^+ - RS^-$ estimates the jump asymmetry or the signed jump variation, which is the difference between the positive jump variation and negative jump variation,

$$RS_t^+ - RS_t^- \xrightarrow{p} \sum_{1<s\leq t} \kappa_s^2 I(\kappa_s > 0) - \sum_{1<s\leq t} \kappa_s^2 I(\kappa_s < 0), \text{as } M \to \infty. \tag{3.4}$$

The imprecise RS^+ and RS^- estimators are likely to lead to a biased estimation of the signed jump variation. Recall that in the presence of a non-zero drift, the bias in RS^+ always has the opposite sign to the bias in RS^-. Thus, the bias magnitude of $RS^+ - RS^-$ equals the sum of the bias magnitude of RS^+ and RS^-, such that the bias in $RS^+ - RS^-$ cumulates the biases in the individual RS^+ and RS^- components. It is notable that the bias in $RS^+ - RS^-$ will be even greater than the bias in RV since the bias in RV adds the bias magnitudes of RS^+ and RS^-

(recall $RV = RS^+ + RS^-$) that are of opposite signs. This indicates that the drift bias of $RS^+ - RS^-$ is expected to be substantial for the finite sample.

Among the four examples above, only the drift bias of the signed jump estimator $RS^+ - RS^-$ seems to be influential for either positive or negative drift. This motivates me to study the bias of $RS^+ - RS^-$ in more detail. For presentation purposes, We use the notation J^{Δ} to denote signed jumps, $RS^+ - RS^-$ as

$$J_t^{\Delta} = RS_t^+ - RS_t^-. \tag{3.5}$$

Since the estimation biases of $RS^{*,+}$ and $RS^{*,-}$ are extremely close to zero, the estimation bias of the difference between $RS^{*,+}$ and $RS^{*,-}$ should be small, thus measuring the signed jump variation much more accurately. Therefore, We define a modified signed jump variation estimator as the difference between $RS^{*,+}$ and $RS^{*,-}$. We use the notation $J^{*,\Delta}$ to indicate $RS^{*,+} - RS^{*,-}$,

$$J_t^{*,\Delta} = RS_t^{*,+} - RS_t^{*,-}. \tag{3.6}$$

Again, the biases of $J^{*,\Delta}$ and J^{Δ} are calculated via a Monte Carlo study when the DGP is the linear drift–diffusion process defined in Equation (2.21), with the volatility dynamics and the settings of all remaining parameters the same as in section 2. The formula for calculating the expectation of the bias in J^{Δ} is defined by

$$\text{bias of } J^{\Delta} = \mathbb{E}\{J_t^{\Delta} - J_t^{\text{DGP},\Delta}\}, \tag{3.7}$$

where $J_t^{\text{DGP},\Delta}$ is the calculation of the signed jump variation and is expressed by the difference between the aggregation of squared positive jumps and the aggregation of squared negative jumps,

$$J_t^{\text{DGP},\Delta} = \sum_{i=1}^{M}\sum_{j=1}^{k}\left(\phi_{t_i}^j\right)^2 I_{t_i}^j I\left(\phi_{t_i}^j > 0\right) - \sum_{i=1}^{M}\sum_{j=1}^{k}\left(\phi_{t_i}^j\right)^2 I_{t_i}^j I\left(\phi_{t_i}^j < 0\right). \tag{3.8}$$

Analogously, the formula for calculating the bias of $J^{*,\Delta}$ is defined by,

$$\text{bias of } J^{*,\Delta} = \mathbb{E}\left\{J_t^{*,\Delta} - J_t^{\text{DGP},\Delta}\right\}, \tag{3.9}$$

Since this section focuses on jump variation, the simulation exercise considers more comprehensive settings for including alternative jump components in the linear drift-diffusion process. The detailed descriptions are as follows: (1) there is a single large positive jump within a day with the size of the jump equal to 150% of the spot volatility, $\phi_{t_i}^1 = 1.5\sigma_{t_i}$; (2) there is one large negative jump per day with the same size, $\phi_{t_i}^1 = -1.5\sigma_{t_i}$; (3) there are two small positive jumps within a day and the size of each jump is equal to 60% of the spot volatility, $\phi_{t_i}^j = 0.6\sigma_{t_i}$, for $j = 1,2$; and (4) there are two small negative jumps within a day of the same size, $\phi_{t_i}^j = -0.6\sigma_{t_i}$, for $j = 1,2$. The jump settings in (2) and (4) are identical to those applied by Laurent and Shi (2020), whereas the jump settings of (1) and (3) are identical except for the opposite jump signs.

Figure 3.1 depicts the bias of J^{Δ} and $J^{*,\Delta}$ (indicated by the red surface and blue surface, respectively) under a linear drift-diffusion process and for the four different jump settings. Panels (1) - (4) contain the simulation results for the jump settings (1) - (4), respectively. As shown in the figure, there is almost no difference between the two estimators if either θ or p_{t_0} equals zero, meaning there is no bias. The bias of J^{Δ} becomes larger as $|\theta|$ and p_{t_0} increase. When either θ or p_{t_0} are non-zero, the bias can be substantial relative to the signed jump variation. For example, for jump setting (1) in the top-left panel, the order of bias of J^{Δ} is 10^{-3}, which is considerable compared to the order of the signed jump variation (10^{-4}) when both θ or p_{t_0} are not zero. For the jump setting (4), the order of bias of J^{Δ} is 10^{-3}, which is again a big amount relative to the order of true signed jumps (10^{-5}) when both θ or p_{t_0} are not zero. From Figure 3.1 it is also clear that the bias of J^{Δ} increases as the magnitude of drift (the magnitude of θ and p_{t_0}) increases, and the sign of the bias of J^{Δ} is the same as the sign of the drift.

Figure 3.1. Bias of J_t^{Δ} and $J_t^{*,\Delta}$ under the linear drift-diffusion process.

Notes: Bias of J^{Δ} is defined by Equation (3.7) and the bias of $J^{*,\Delta}$ is defined by Equation (3.9). Panel (1) shows the bias of J^{Δ} and $J^{*,\Delta}$ when the log prices follow Equation (2.21) with one large positive jump during a day, $\phi_{t_i}^1 = 1.5\sigma_i$; Panel (2) reports the bias of J^{Δ} and $J^{*,\Delta}$ when the log prices follow Equation (2.21) with one large negative jump during a day $\phi_{t_i}^1 = -1.5\sigma_i$; Panel (3) plots the bias of J^{Δ} and $J^{*,\Delta}$ when the log prices follow Equation (2.21) with two small positive jumps during a day $\phi_{t_i}^j = 0.6\sigma_{t_i}$, for $j = 1,2$; and Panel (4) depicts the bias of J^{Δ} and $J^{*,\Delta}$ when the log prices follow Equation (2.21) with two small negative jumps during a day $\phi_{t_i}^j = -0.6\sigma_{t_i}$, for $j = 1,2$.

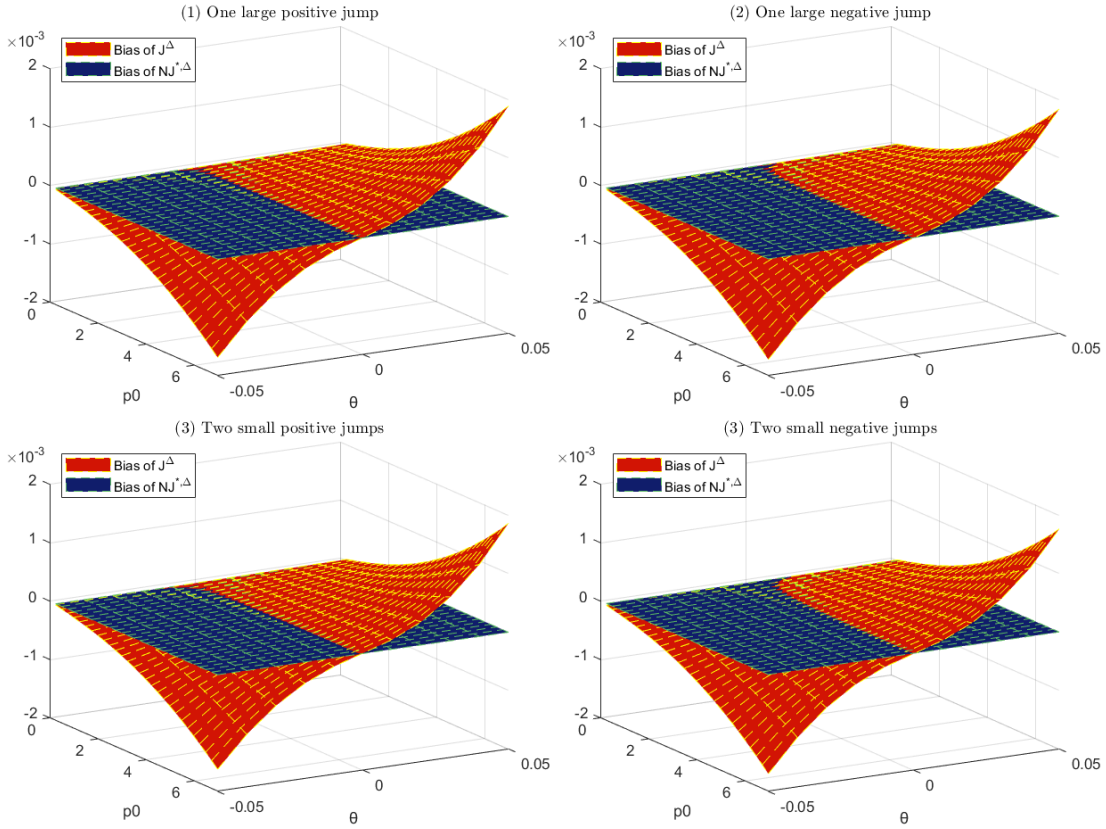


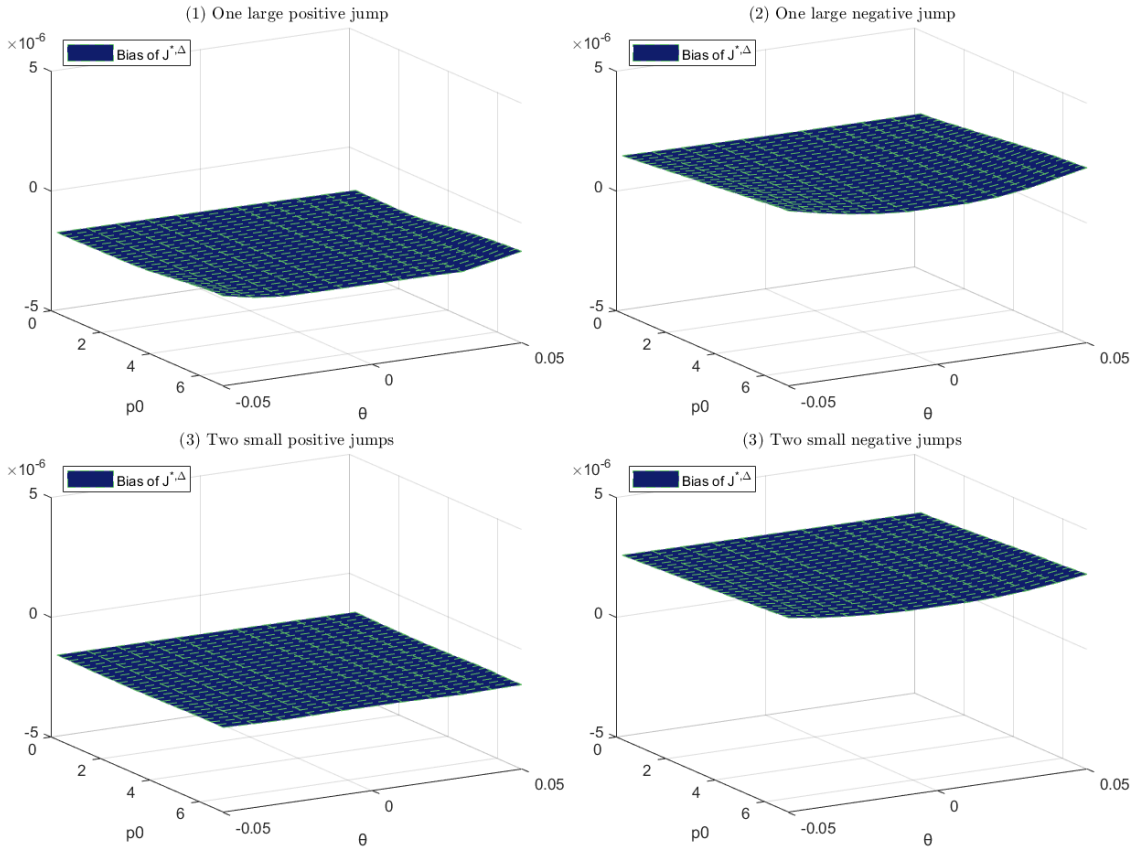
Figure 3.1 also shows that the bias in $J^{*,\Delta}$ is indistinguishable from zero, for non-zero θ and p_{t_0} and for all four jump settings. The gap between the bias of J^{Δ} and the bias of $J^{*,\Delta}$ becomes visible as θ and p_{t_0} move further away from zero. This indicates that the bias of $J^{*,\Delta}$ is much smaller than that of J^{Δ} in the presence of a non-zero drift.

Figure 3.2 reports the bias of $J^{*,\Delta}$ only, with the four jump settings coinciding with the four panel, identical to Figure 3.1 above. For all jump scenarios, the order of the bias of $J^{*,\Delta}$ is 10^6,

which is omittable compared to the order 10^3 of J^{Δ} as in Figure 3.2. This implies that the bias of the modified signed jump variation estimator is non-zero if there is a non-zero drift, it is negligible compared to that of the original estimator. Additionally, Figure 3.2 shows that the bias of $J^{*,\Delta}$ is negligible relative to the signed jump variation. For example, the bias of $J^{*,\Delta}$ is merely around $2 \times 10^6/(2.25 \times 10^{-4}) = 0.89\%$ of the signed jump variation for non-zero θ and p_{t_0} in setting (1) where there is one large positive jump. For setting (4), where there are two small negative jumps, the bias of $J^{*,\Delta}$ is only about $3 \times 10^{-6}/(7.2 \times 10^{-5}) = 4.17\%$ of the signed jump variation for nonzero θ and p_{t_0}. These suggest that $J^{*,\Delta}$ is always dominated by the signed jump variation, thus its drift bias is ignorable.

Figure 3.2. Bias of $J_t^{*,\Delta}$ for linear drift-diffusion process.

Notes: The bias of $J^{*,\Delta}$ is defined by Equation (3.9). The jump settings in the four panels are consistent with those in Figure 3.1.



The findings of the Monte Carlo study of the bias of J^{Δ} and $J^{*,\Delta}$ suggest the following four implications. First, the bias of J^{Δ} is not negligible in the presence of non-zero (average) drift. Second, the bias of J^{Δ} due to nonzero (average) drift is an increasing function of the size of the (average) drift. Third, in the presence of nonzero (average) drift, the bias of $J^{*,\Delta}$ is negligible compared to the bias of J^{Δ}. Fourth, $J^{*,\Delta}$ is always dominated by the signed jump variation.

The first implication suggests that J^{Δ} as an estimator of signed jump variation may be biased by drift. It is important to isolate the drift bias and obtain a more accurate estimator of the signed jump variation for those who are interested in the practical use of the signed jump variation. For example, a less biased or more accurate estimate for the signed jump variation will benefit those who are interested in applying the signed jump variation to predict volatility.

To separate the drift bias, first combine Equation (3.7) and Equation (3.9),

$$\begin{aligned}\mathbb{E}(\text{bias of } J^{\Delta}) &= \mathbb{E}(J^{\Delta}) - [\mathbb{E}(J^{*,\Delta}) - \mathbb{E}(\text{bias of } J^{*,\Delta})] \\ &= \mathbb{E}(J^{\Delta}) - \mathbb{E}(J^{*,\Delta}) + \mathbb{E}(\text{bias of } J^{*,\Delta}).\end{aligned} \tag{3.10}$$

Given that the bias of $J^{*,\Delta}$ is negligible compared to the bias of J^{Δ}, $\mathbb{E}(\text{bias of } J^{*,\Delta}) \approx 0$, Equation (3.10) becomes

$$\mathbb{E}(\text{bias of } J^{\Delta}) \approx \mathbb{E}\{J^{\Delta} - J^{*,\Delta}\}. \tag{3.12}$$

Equation (3.12) suggests that the bias of J^{Δ} can be proxied by the difference between J^{Δ} and $J^{*,\Delta}$. By using the notation D_t^{Δ} to indicate the bias of J^{Δ}, Equation (3.12) is expressed as

$$D_t^{\Delta} = J^{\Delta} - J^{*,\Delta}. \tag{3.13}$$

And Equation (3.13) can be easily transformed into

$$J^{\Delta} = J^{*,\Delta} + D_t^{\Delta}. \tag{3.14}$$

Equation (3.14) suggests that the signed jump variation estimator (J^{Δ}) can be decomposed into the drift-modified signed jump variation estimator ($J^{*,\Delta}$), which is a proxy for the signed jump variation, and a drift bias component (D^{Δ}).

4 Volatility estimation for ultra-high-frequency data

Recent studies argue that the 5-minute frequency sampling is too sparse and thus associated with substantial data loss, which results in the estimation inefficiency of volatility (Jacod et al., 2009, Podolskij and Vetter, 2009, Lee and Mykland, 2012, Aït-Sahalia et al., 2012, Hautsch and Podolskij, 2013, Christensen et al., 2014). As the estimation inefficiency tends to decrease in proportion to the reduced observed increments of the process as the sampling frequency decreases, these studies suggest using an ultra-high-frequency sampling scheme (e.g., tick-by-tick) as a better alternative for estimating volatility. However, for the ultra-high frequency sampling, prices are often contaminated by noise, induced by microstructure effects, that arise from market imperfections such as bid-ask spreads and price discreteness (e.g., Niederhoffer and Osborne, 1966, Roll, 1984, Black, 1986). These noise-contaminated prices invalidate the asymptotic properties of volatility estimators (see Bandi and Russell (2008) or Hansen and Lunde (2006) among others). To reduce the bias in volatility estimation for noisy ultra-high frequency prices, noise-modified estimators are commonly applied in these studies (Hautsch and Podolskij, 2013).

One might expect that with ultra-high frequency data, the drift component will be extremely close to zero, meaning that the drift bias in volatility estimators should be negligible. In other words, the drift bias measured as the discrepancy between the original volatility estimator and the modified estimator based on centred log returns will diminish. We show in this section that this is not true for the noise-modified volatility estimators built upon Christensen et al. (2014).[3]

Assume that the log price, p_{t_i}, is contaminated by noise, such that observed noisy price, $p^{\diamond}_{t_i}$, are expressed as

[3] Laurent and Shi (2020) study the drift bias of the noise-modified volatility estimators introduced by Podolskij and Vetter (2009), which are also based on the pre-averaged returns. This chapter focuses on the volatility estimators recommended by Christensen et al. (2014) since they are designed to estimate the jumps.

$$p_{t_i}^{\circ} = p_{t_i} + u_{t_i}, \text{with } 0 < t_i < , \tag{4.1}$$

where u_{t_i} is a noise process with mean zero and variance q^2, is independent of p_{t_i}, and is serially correlated with order $s-1$. Therefore, the noisy log prices can be obtained by combining the noise process with the constant drift–diffusion process (Equation (3.10)), and We term this the constant drift–diffusion plus noise process. A linear drift–diffusion plus noise process of the noisy log prices is obtained similarly by combining the noise process with the linear drift–diffusion process (Equation (3.11)).

To make inferences about volatility using noisy ultra-high frequency prices, Christensen et al. (2014) suggest making use of the pre-averaging approach, which was proposed by Jacod et al. (2009) and Podolskij and Vetter (2009). Intuitively, this approach locally smooths the observed price series, $p_{t_i}^{\circ}$, so that the microstructure component u_{t_i} (almost) disappears under averaging. Returns to this pre-averaged price series can then be used to construct noise consistent measures of the jump variation components. To implement pre-averaging, Christensen et al. (2014) suggest calculating returns on the log prices that are pre-averaged in a local neighbourhood of H observations,

$$r_{t_i}^{\circ} = \frac{1}{H}\left(\sum_{j=H/2}^{H-1} p_{t_{i+j}}^{\circ} - \sum_{j=0}^{H/2-1} p_{t_{i+j}}^{\circ}\right), \tag{4.2}$$

where $H = \left\lceil \theta\sqrt{M} \right\rceil$ with the parameter $\theta = 2$ following Christensen et al. (2014). From Equation (4.2), it can be seen that this pre-averaged return is essentially the average of returns at $H/2-1$ frequency, which can be much lower than the ultra-high frequency. For example, for the 6.5-hour trading session at a 1-second frequency, the pre-averaging window is $H = \left\lceil 2\sqrt{23400} \right\rceil = 305 \text{ seconds} \approx 5 \text{ minutes}$. From Equation (4.2), it can be seen that the pre-averaged return $r_{t_i}^{\circ}$ is the average of 2.5-minute frequency returns, where drift may not be small. Therefore, the drift can still be large in the pre-averaged returns.

Based on the pre-averaged return $r_{t_i}^{\circ}$, Christensen et al. (2014) suggest that the noise-modified RV and BV (indicated by NRV and NBV, respectively) are calculated as follows,

$$NRV_t = \frac{M}{M-H+2}\frac{1}{H\psi_H}\sum_{i=1}^{M-H+2}\left|r_{t_i}^{\circ}\right|^2 - \frac{\widehat{\omega}_t^2}{\theta^2\psi_H}, \tag{4.3}$$

$$NBV_t = \frac{M}{M-2H+2}\frac{1}{H\psi_H}\frac{\pi}{2}\sum_{i=1}^{M-2H+2}\left|r_{t_i}^{\circ}\right|\left|r_{t_{i+H}}^{\circ}\right| - \frac{\widehat{\omega}_t^2}{\theta^2\psi_H}, \tag{4.4}$$

where $\psi_H = (1+2H^{-2})/12$ and $\widehat{\omega}_t^2/\theta^2\psi_H$ is a bias correction, which compensates for any residual microstructure noise that may remain after pre-averaging. $\widehat{\omega}_t^2$ denotes the estimator for the noise variance given by

$$\widehat{\omega}_t^2 = \frac{1}{2(\mathrm{M}-1)}\sum_{i=2}^{M}\left|r_{t_i}^{\circ}\right|\left|r_{t_{i-1}}^{\circ}\right|. \tag{4.5}$$

Recall that RV can be decomposed into RS^+ and RS^-, respectively, based on decomposing the sign of returns. Analogously, based on splitting the sign of the pre-averaged returns, NRV may also be appropriately decomposed into the noise-modified RV^+ and RV^- (indicated by NRV^+ and NRV^-),

$$NRV_t^+ = \frac{M}{M-H+2}\frac{1}{H\psi_H}\sum_{i=1}^{M-H+2}\left|r_{t_i}^{\circ}\right|^2 I(r_{t_i}^{\circ}>0) - \frac{\widehat{\omega}_t^2}{2\theta^2\psi_H}, \tag{4.6}$$

$$NRV_t^- = \frac{M}{M-H+2}\frac{1}{H\psi_H}\sum_{i=1}^{M-H+2}\left|r_{t_i}^{\circ}\right|^2 I(r_{t_i}^{\circ}<0) - \frac{\widehat{\omega}_t^2}{2\theta^2\psi_H}. \tag{4.7}$$

For simplicity, We assume equally half bias-correction ($\widehat{\omega}_t^2/2\theta^2\psi_H$) for NRV^+ and NBV^- such that the bias correction drops out when taking the difference between NRV^+ and NRV^-.

Since the drift can be large in the pre-averaged returns, the biases of the noise-modified volatility estimators are expected to be large in the presence of a non-zero drift. Since the pre-averaged returns are simple averages of low-frequency returns, the biases of the noise-modified volatility estimators are expected to have similar patterns as those of the original volatility

estimators.[4] To reduce such drift bias, We modify the NRV and NBV estimators (indicated by NRV^* and NBV^*) by computing NRV and NBV on centred pre-averaged returns,

$$NRV_t^* = \frac{M}{M-H+2}\frac{1}{H\psi_H}\sum_{i=1}^{M-H+2}\left|r_{t_i}^{\diamond} - \mathcal{M}_t^{\diamond}\right|^2 - \frac{\widehat{\omega}_t^2}{\theta^2\psi_H}, \tag{4.8}$$

$$NBV_t^* = \frac{M}{M-2H+2}\frac{1}{H\psi_H}\frac{\pi}{2}\sum_{i=1}^{M-2H+2}\left|r_{t_i}^{\diamond} - \mathcal{M}_t^{\diamond}\right|\left|r_{t_{i+H}}^{\diamond} - \mathcal{M}_t^{\diamond}\right| - \frac{\widehat{\omega}_t^2}{\theta^2\psi_H}, \tag{4.9}$$

where $\mathcal{M}_t^{\diamond} = \text{median}\left(r_{t_1}^{\diamond} \dots r_{t_M}^{\diamond}\right)$ denotes the median of $M-H+2$ pre-averaged returns involved in the computation of the volatilities of day t. We also modify the NRV^+ and NRV^- estimators (indicated by $NRV^{*,+}$ and $NRV^{*,-}$) by computing NRV^+ and NRV^- (including the indicator function) on centred pre-averaged returns,

$$NRV_t^{*,+} = \frac{M}{M-H+2}\frac{1}{H\psi_H}\sum_{i=1}^{M-H+2}\left|r_{t_i}^{\diamond} - \mathcal{M}_t^{\diamond}\right|^2 I(r_{t_i}^{\diamond} - \mathcal{M}_t^{\diamond} > 0) - \frac{\widehat{\omega}_t^2}{2\theta^2\psi_H}, \tag{4.10}$$

$$NRV_t^{*,-} = \frac{M}{M-H+2}\frac{1}{H\psi_H}\sum_{i=1}^{M-H+2}\left|r_{t_i}^{\diamond} - \mathcal{M}_t^{\diamond}\right|^2 I(r_{t_i}^{\diamond} - \mathcal{M}_t^{\diamond} < 0) - \frac{\widehat{\omega}_t^2}{2\theta^2\psi_H}, \tag{4.11}$$

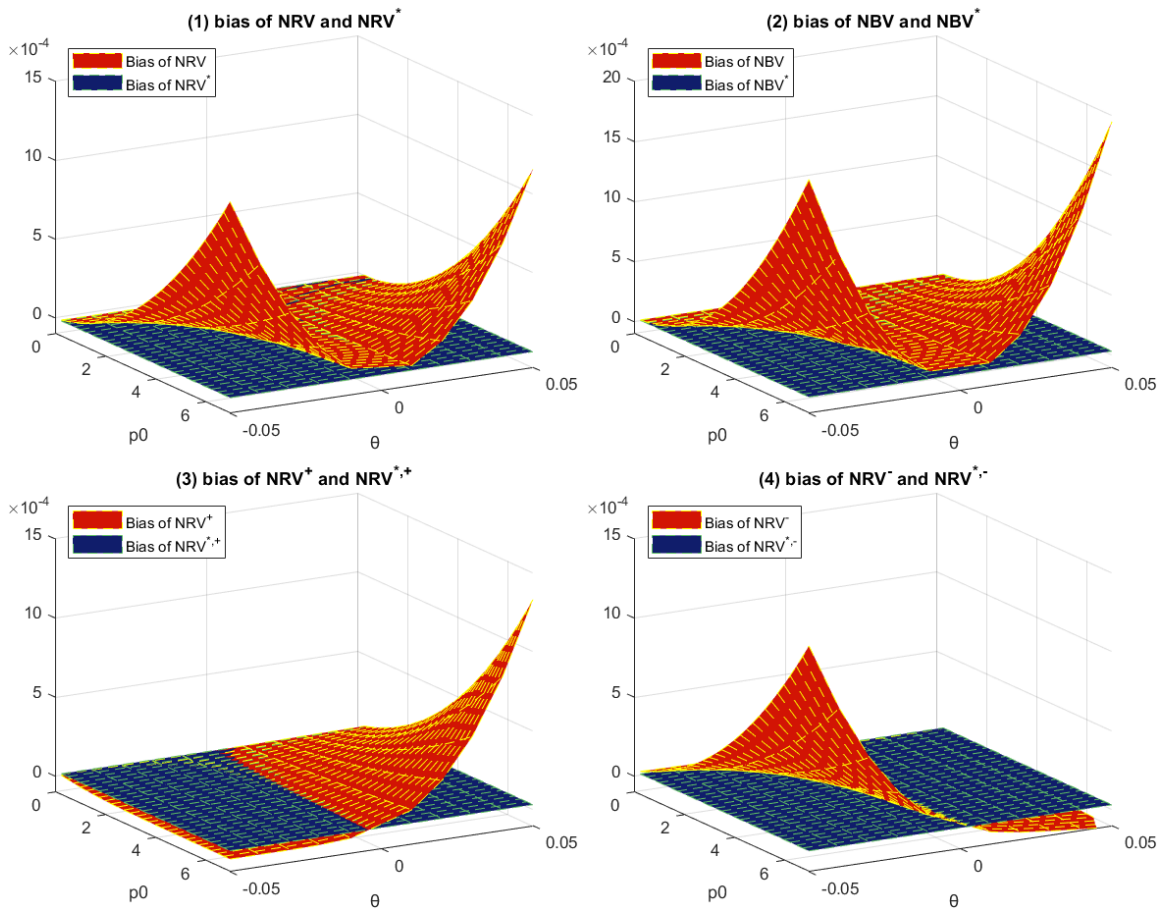
Due to the use of centred pre-average returns, the biases of these modified volatility estimators due to a non-zero drift are expected to be small.

As in section 3 above, the biases are defined by the difference between these volatility estimators and the volatility (or volatility and jumps) they attempt to estimate. We compute the bias of noise-modified volatility estimators and respective modified versions, via Monte Carlo simulations with 10^4 replications of the DGP (2.21&4.1) and with the parameter settings as in section 3. The sampling frequency is at 1 second as in Laurent and Shi (2020). Following Laurent and Shi (2020), the noise follows an independent and identically distributed process (i.e., $s = 1$), and the volatility of the noise is set $q = 0.1\%$. Figure 4.1 depicts the bias of

[4] Alternative ultra high frequency volatility estimators that may be subject to the drift bias are the sub-sampled volatility estimators proposed by Zhang et al. (2005). This is because the sub-sampled volatility estimators are computed by the average of the volatility estimators computed on some sub-samples of low frequency data.

noise-modified volatility estimators and respective modified versions. As expected, for the NRV and NBV estimators, We observe patterns similar to those in Figures 6 & 7 by Laurent and Shi (2020). While the estimation accuracy of the original volatility estimator deteriorates substantially as $|\theta|$ and p_{t_0} deviate from zero, the new estimator is much more accurate. For the NRV^+ and NRV^- estimators, We observe patterns similar to Figure 3.1 by this chapter. While the estimation accuracy of the original semi-variance deteriorates asymmetrically as θ and p_{t_0} deviate from zero, the new estimator is much more accurate.

Figure 4.1. The bias of the noise-modified estimators under the linear drift–diffusion process.



5 Jump asymmetry estimation for ultra-high-frequency data

Section 4 shows that the patterns of the drift biases of the ultra high frequency noise-modified volatility estimators are similar to those of lower frequency volatility estimators. Therefore, one can again conclude that the drift bias of the signed jump estimator $NRS^+ - NRS^-$ tend to be significant, while other jump estimators $NRV - NBV$, $NRS^+ - 0.5NBV$, and $NRS^- - 0.5NBV$ have a much smaller drift bias. This again motivates me to study the drift bias of $NRS^+ - NRS^-$ in more detail. Let NJ^{Δ} indicate $NRS^+ - NRS^-$,

$$NJ^{\Delta} = NRS^+ - NRS^-, \tag{5.1}$$

Again, We expect that NJ^{Δ} which are computed using the centred pre-averaged returns should have a small bias even though the drift is non-zero. The modified version of NJ^{Δ} that uses centred returns is indicated by $NJ^{*,\Delta}$,

$$NJ^{*,\Delta} = NRS^{*,+} - NRS^{*,-}. \tag{5.2}$$

For illustration, this section again visualizes the bias in NJ^{Δ} by using the same simulation setting as in section 4. The expected bias of NJ^{Δ} is calculated as

$$\text{bias of } NJ^{\Delta} = \mathbb{E}\{NJ^{*,\Delta} - J_t^{\text{DGP},\Delta}\}, \tag{5.3}$$

where $J_t^{\text{DGP},\Delta}$ is defined previously in Equation (3.8). And the expected bias of $NJ^{*,\Delta}$ is calculated by

$$\text{bias of } NJ^{*,\Delta} = \mathbb{E}\{NJ^{*,\Delta} - J_t^{\text{DGP},\Delta}\}. \tag{5.4}$$

Figure 4.3 depicts the bias of NJ^{Δ} and $NJ^{*,\Delta}$ (indicated by the red surface and blue surface, respectively). Figure 4.3 shows that the bias of NJ^{Δ} is large compared to the size of true signed jumps, when drift is non-zero. For example, for the jump setting (1) and (2) and for nonzero drift, the order of bias of NJ^{Δ} is 10^{-3}, which is considerable compared to the order of the signed jump variation (10^{-4}). And for the jump setting (3) and (4) and for nonzero drift, the

order of bias of NJ^{Δ} is 10^{-3}, which is again considerable relative to the order of true signed jumps (10^{-5}).

Figure 5.1. Bias of NJ_t^{Δ} and $NJ_t^{\Delta*,}$ for linear drift-diffusion process.

Notes: Bias of NJ^{Δ} is defined by Equation (5.3) and the bias of $NJ^{*,\Delta}$ is defined by Equation (5.4).

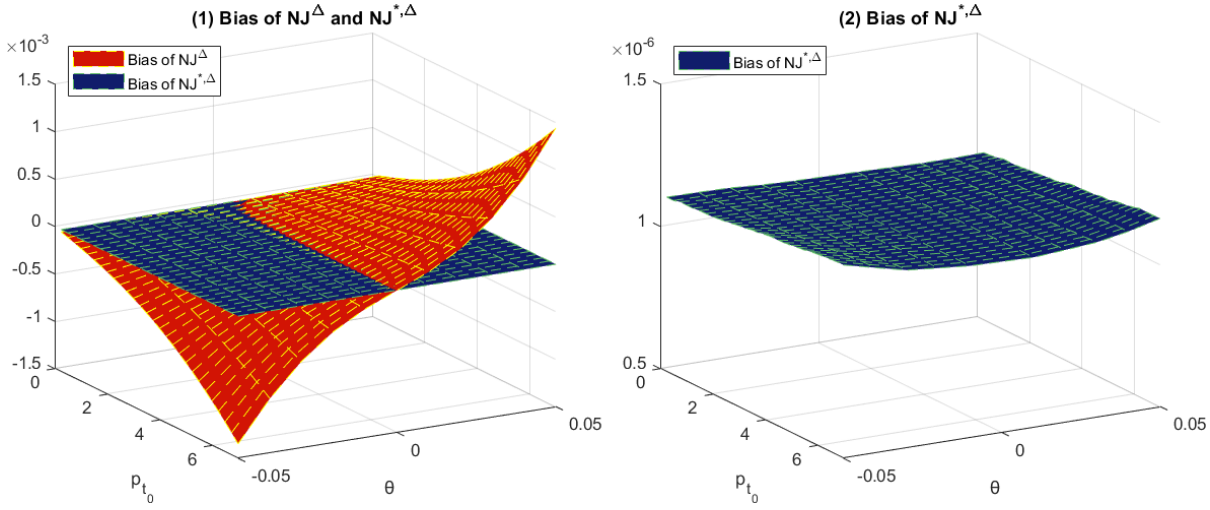


Figure 5.1 demonstrates the bias of NJ^{Δ} and $NJ^{*,\Delta}$ for ultrahigh frequency data. We observe patterns very similar to those in Figures 3.3 and 3.4. The estimation accuracy of the original signed jump variation estimator deteriorates substantially as drift deviates from zero, and the sign of the bias of NJ^{Δ} is consistent with the sign of the drift. In contrast, the new estimator is much more accurate, with the bias of the estimator $NJ^{*,\Delta}$ negligible relative to both the bias of NJ^{Δ} and the size of the signed jump variation.

Combining Equation (5.3) and Equation (5.4), one can obtain,

$$\mathbb{E}(\text{bias of } NJ^{\Delta}) = \mathbb{E}(NJ^{\Delta}) - \mathbb{E}(NJ^{*,\Delta}) + \mathbb{E}(\text{bias of } NJ^{*,\Delta}). \tag{5.5}$$

Given that the bias of $NJ^{*,\Delta}$ is negligible compared to the bias of NJ^{Δ}, $\mathbb{E}(\text{bias of } NJ^{*,\Delta}) \approx 0$, Then, Equation (5.5) becomes,

$$\mathbb{E}(\text{bias of } NJ^{\Delta}) \approx \mathbb{E}\{NJ^{\Delta} - NJ^{*,\Delta}\}. \tag{5.6}$$

Let ND^{Δ} indicate the bias of NJ^{Δ}, Equation (5.6) suggests that the bias of NJ^{Δ} approximately equals to the sum of $NJ^{*,\Delta}$ and ND^{Δ},

$$NJ^{\Delta} = NJ^{*,\Delta} + ND_t^{\Delta}. \tag{5.7}$$

6 Data Description

We present the results for US stock markets, to investigate empirical evidence for non-negligible drift bias in the signed jump variation estimator. Exchange-traded funds (ETFs) have been constructed to be a broad representation of the overall stock market and have been applied by previous literature as the market portfolio proxies (e.g., Barigozzi et al., 2014, Patton and Sheppard, 2015, Fan et al., 2016, Gao et al., 2018). Therefore, this paper uses SPDR S&P 500 Growth ETF (SPY) as the proxy for the US stock market. SPY is a broad proxy of the stock market based on a market capitalization of 500 large companies traded on the New York Stock Exchange (NYSE) and Nasdaq stock exchange. We begin this section with a brief discussion of the data sources, followed by descriptive statistics of the resulting measures for volatility, signed jumps and drift bias for these assets.

We obtain the tick-by-tick SPY from Tick Data Inc. The sample period is from January 2, 1997, to September 21, 2021, with a total of $n = 6222$ days. As in Barndorff-Nielsen et al. (2009) and Patton and Sheppard (2015), SPY tick data is cleaned according to the following rules:

1. Transactions outside 9:30:00 to 16:00:00 were removed.
2. Transactions with a 0 price or volume were removed.
3. Only retain the transaction prices from the most active exchange of each day (the transaction prices from other exchanges were dropped).
4. Only transaction prices from regular trades were retained (with prices related to irregular trades removed). The classification details of regular and irregular trades can be found on the official website of Tick Data Inc: https://www.tickdata.com/.
5. If multiple transactions have the same timestamp, use the median price.

6. Delete transaction prices related to corrected trades. The details of corrected trades can be found on the official website of Tick Data Inc.

This paper focuses on price volatility during the intraday session of a trading day. Thus, overnight returns are excluded. And this avoids the need to adjust prices for splits or dividends. After cleaning, the Tick-by-Tick transaction prices of SPY are then sampled at 5-minute frequency, and the log return based on the 5-minute prices is used for calculating the low-frequency estimators including RV, BV, J^{Δ}, $J^{*,\Delta}$, and D^{Δ}. The bipower variation may be biased due to the correlation in the adjacent returns induced by the microstructure noise, even for the 5-minute frequency (Andersen et al., 2007, Huang and Tauchen, 2005). To reduce such bias, the estimator for bipower variation BV defined in Equation (2.6) was modified by averaging its multiple "skip" versions. The "skip-q" BV estimator is defined as,

$$BV_t^{\mathrm{q}} = \frac{\pi}{2} \sum_{i=\mathrm{q}+2}^{M} \left| r_{t_{i-1-q}} \right| \left| r_{t_i} \right|. \tag{5.1}$$

Note that when $q = 0$, the "skip-q" BV is identical to the usual BV. This paper estimates the continuous variation by an averaged version of skip-q BV estimator, which is defined by averaging the skip-0 through skip-4 BV estimators,

$$BV_t = \frac{1}{4} \sum_{q=0}^{4} BV_t^{\mathrm{q}}. \tag{5.2}$$

This averaged version of skip-q BV estimator is able to balance the trade-off between locality ($q = 0$) and robustness to market microstructure noise ($q = 4$) (Patton and Sheppard, 2015).

Table 6.1 reports descriptive statistics of various estimators with the 5-minute frequency data. The statistics include the quantile at 0.25, median, mean, quantile at 0.75, and standard deviation. The first panel contains the descriptive statistics of volatility estimators. As this panel shows, the average value of daily RV for the SPY was 1.063, which is fairly close to the

levels of constant volatility and unconditional volatility assumed in the simulations study (The unconditional volatility of GARCH denotes the volatility of the population, and GARCH updates the conditional volatility in the sample to be trending toward the unconditional volatility). This indicates that the parameter settings of the simulations in this paper are consistent with the real market data. The second panel contains the descriptive statistics of the original signed jump variation estimators J^{Δ}, J^{+} and J^{-} from Patton and Sheppard (2015) (with $J^{+} = J^{\Delta}I(J^{\Delta} > 0)$ and $J^{-} = J^{\Delta}I(J^{\Delta} < 0)$) while the third panel contains the modified signed jump variation estimators $J^{*,\Delta}$, $J^{*,+}$ and $J^{*,-}$ (with $J^{+} = J^{*,\Delta}I(J^{*,\Delta} > 0)$ and $J^{-} = J^{*,\Delta}I(J^{*,\Delta} < 0)$). Comparing the original estimators J^{Δ}, J^{+}, and J^{-} with the corresponding modified estimators $J^{*,\Delta}$, $J^{*,+}$, and $J^{*,-}$, the mean of the original estimator is systematically more negative. This indicates that the original estimators are downward biased, on average.

Table 6.1. Data descriptive statistics

Notes: This table contains the descriptive statistics of various volatility estimators with 5-minute frequency SPDR S&P 500 ETF prices from 1997 January to 2021 September. The descriptive statistics include the quantile at 0.25, median, mean, quantile at 0.75, and standard deviation. The first panel contains the volatility estimators including Realized Variance (RV), and Bi-power Variation (BV). The second panel contains the original signed jump variation estimators following Patton and Sheppard (2015), including the signed jump variation estimator (J^{Δ}), the positive jump variation estimator $J^{+} = J^{\Delta}I(J^{\Delta} > 0)$, and the negative jump variation estimators $J^{-} = J^{\Delta}I(J^{\Delta} < 0)$. The third panel contains the modified signed jump variation estimators, including the modified signed jump variation estimator ($J^{*,\Delta}$), the modified positive jump variation estimator $J^{+} = J^{*,\Delta}I(J^{*,\Delta} > 0)$, and the modified negative jump variation estimators $J^{-} = J^{*,\Delta}I(J^{*,\Delta} < 0)$. The fourth panel contains the drift biases, including the drift bias (D^{Δ}), the positive drift bias $D^{+} = D^{*,\Delta}I(D^{*,\Delta} > 0)$, and the negative drift bias $D^{-} = D^{*,\Delta}I(D^{*,\Delta} < 0)$.

	$Q_{0.25}$	Median	Mean	$Q_{0.75}$	St. dev.
Volatility estimators					
RV	0.227	0.505	1.063	1.097	2.259
B	0.210	0.473	1.018	1.061	2.142
Signed jump variation estimators					
J^{Δ}	-0.078	0.001	0.011	0.073	0.661
J^{+}	0.025	0.072	0.219	0.178	0.769
J^{-}	-0.214	-0.080	-0.200	-0.029	0.436
Modified signed jump variation estimators					
$J^{*,\Delta}$	-0.067	-0.002	0.036	0.070	0.742
$J^{*,+}$	0.026	0.073	0.246	0.196	0.926
$J^{*,-}$	-0.169	-0.064	-0.166	-0.023	0.414

Drift biases					
D^{Δ}	-0.088	0.011	-0.033	0.068	0.515
D^{+}	0.029	0.061	0.154	0.144	0.356
D^{-}	-0.253	-0.100	-0.256	-0.043	0.581

The bottom panel of Table 6.1 reports the descriptive statistics of the drift bias D^{Δ}, with the positive drift bias defined by $D^{+} = D^{\Delta}I(D^{\Delta} > 0)$ and then negative drift bias defined by $D^{-} = D^{\Delta}I(D^{\Delta} < 0)$ also reported. First, the results show that the drift bias indeed has positive and negative signs, as found in the Monte Carlo study. Second, the results show that the drift bias in the signed jump variation estimator is empirically not small. For example, the mean of the negative drift bias D^{-} is -0.256. This magnitude is greater than the level of the mean of both modified signed jump variation estimators $J^{*,+}$ and $J^{*,-}$. The empirical evidence of large drift bias in the signed jump variation estimator corroborates the simulation results in section 3.

The cleaned Tick-by-Tick transaction prices of SPY are used to compute the pre-averaged returns, which are then used for calculating the noise-modified estimators, including NRV, NBV, NJ^{Δ}, $NJ^{*,\Delta}$ and ND^{Δ}. The sample shrinks to the recent decade (from 2010 January to 2021 September) to ensure that the SPY is sufficiently active for ultrahigh-frequency data. Table 6.2 reports the descriptive statistics of these noise-modified estimators. The statistics include the quantile at 0.25, median, mean, quantile at 0.75, and standard deviation. The first panel reports the noise-modified volatility estimators. The second panel contains the noise-modified signed jump variation estimators NJ^{Δ}, NJ^{+} and NJ^{-} (with $NJ^{+} = NJ^{\Delta}I(NJ^{\Delta} > 0)$ and $NJ^{-} = NJ^{\Delta}I(NJ^{\Delta} < 0)$) while the third panel contains the modified noise-modified signed jump variation estimators $NJ^{*,\Delta}$, $NJ^{*,+}$ and $NJ^{*,-}$ (with $NJ^{+} = NJ^{*,\Delta}I(NJ^{*,\Delta} > 0)$ and $NJ^{-} = NJ^{*,\Delta}I(NJ^{*,\Delta} < 0)$). Comparing the estimators NJ^{Δ}, NJ^{+}, and NJ^{-} with the corresponding modified estimators $NJ^{*,\Delta}$, $NJ^{*,+}$, and $NJ^{*,-}$, the mean of the original estimator is

systematically more negative. This indicates that the noise-modified signed jump variation estimators are downward biased on average in the presence of non-zero drift with ultrahigh-frequency data.

Table 6.2. Data descriptive statistics

Note: This table contains the descriptive statistics of various estimators with ultrahigh-frequency SPDR S&P 500 ETF prices from 2010 January to 2021 September. The descriptive statistics include the quantile at 0.25, median, mean, quantile at 0.75, and standard deviation. The first panel contains the volatility estimators including the noise-modified Realized Variance (NRV), the noise-modified Bi-power Variation (NBV). The second panel contains the original signed jump variation estimators, including the noise-modified signed jump variation estimator (NJ^{Δ}), the noise-modified positive jump variation estimator $NJ^{+} = NJ^{\Delta} I(NJ^{\Delta} > 0)$, and the noise-modified negative jump variation estimators $NJ^{-} = NJ^{\Delta} I(NJ^{\Delta} < 0)$. The third panel contains the modified noise-modified signed jump variation estimators, including the modified noise-modified signed jump variation estimator ($NJ^{*,\Delta}$), the modified noise-modified positive jump variation estimator $NJ^{+} = NJ^{*,\Delta} I(NJ^{*,\Delta} > 0)$, and the modified noise-modified negative jump variation estimators $NJ^{-} = NJ^{*,\Delta} I(NJ^{*,\Delta} < 0)$. The fourth panel contains the noise-modified drift biases, including the noise-modified drift bias (ND^{Δ}), the noise-modified positive drift bias $ND^{+} = ND^{*,\Delta} I(ND^{*,\Delta} > 0)$, and the noise-modified negative drift bias $ND^{-} = ND^{*,\Delta} I(ND^{*,\Delta} < 0)$.

	$Q_{0.25}$	Median	Mean	$Q_{0.75}$	St. dev.
Volatility estimators					
NRV	0.159	0.295	0.624	0.591	1.546
NBV	0.159	0.298	0.641	0.604	1.613
Signed jump variation estimators					
NJ^{Δ}	-0.017	0.002	0.004	0.021	0.097
NJ^{+}	0.007	0.019	0.043	0.042	0.105
NJ^{-}	-0.046	-0.019	-0.040	-0.007	0.063
Modified signed jump variation estimators					
$NJ^{*,\Delta}$	-0.015	0.000	0.014	0.017	0.167
$NJ^{*,+}$	0.006	0.017	0.060	0.046	0.191
$NJ^{*,-}$	-0.033	-0.015	-0.032	-0.006	0.124
Drift biases					
ND^{Δ}	-0.019	0.003	-0.010	0.022	0.165
ND^{+}	0.008	0.018	0.041	0.040	0.138
ND^{-}	-0.071	-0.025	-0.074	-0.008	0.173

The bottom panel of Table 6.2 reports the descriptive statistics of noise-modified drift bias ND^{Δ}, with the noise-modified positive drift bias defined by $ND^{+} = ND^{\Delta} I(ND^{\Delta} > 0)$ and the noise-modified negative drift bias defined by $ND^{-} = ND^{\Delta} I(ND^{\Delta} < 0)$ also reported. First, the non-zero descriptive statistics of ND^{+} and ND^{-} show that the noise-modified drift bias

ND^{Δ} indeed has positive and negative signs, as found in the Monte Carlo study. Second, the results show that the drift bias in the noise-modified signed jump variation estimator is empirically not small. For example, the mean of the noise-modified negative drift bias ND^{-} is -0.074. This magnitude is greater than the level of the mean of both modified noise-modified signed jump variation estimators $J^{*,+}$ and $J^{*,-}$. The empirically large drift bias in the noise-modified signed jump variation estimator corroborates the finding in section 5.

7 Empirical applications

Forecasting price volatility allows us to predict how prices may vary in the future, which is important for investors and the financial industry involved with asset pricing (Black and Scholes, 1973), derivative pricing (Duffie et al., 2000), asset allocation (Merton, 1969), and risk management (Christoffersen and Diebold, 2000). One of the most popular predictors for forecasting volatility is probably the jump asymmetry or the signed jumps variation. The empirical finding by Patton and Sheppard (2015) shows that the signed jump variation has a negative and statistically significant effect on future volatility, and this finding is based on the use of the signed jump variation estimator J^{Δ}.

Based on a long-memory dependence volatility forecasting model framework introduced by Corsi (2009), We first show in this section that the predictability of the estimator J^{Δ} is mostly due to its drift bias D^{Δ}, while the predictability of the modified estimator $J^{*,\Delta}$, which estimate the signed jump variation in a much more accurate manner, is only limited. Then, We show that with ultra-high-frequency data, similar findings can be obtained by exploring the drift bias in the noise-modified signed jump variation estimator NJ^{Δ}. Finally, We show the economic importance of extracting the drift bias from the J^{Δ} and NJ^{Δ}estimators, in terms of the out-of-sample volatility forecasting accuracy.

7.1 Volatility forecasting application

7.1.1 Volatility forecasting models

A large volume of empirical studies has shown the predictive importance of long-memory dependence in financial market volatility. The long-memory dependence of volatility is typically characterized by an autocorrelation that is decaying slowly with the increase of the lag between volatility observations. Several different parametric autoregressive conditional

heteroskedasticity and stochastic volatility models have been proposed in the literature to capture this stylized fact (Robinson, 1991, Ding et al., 1993, Baillie et al., 1996, e.g., Andersen and Bollerslev, 1997, Breidt et al., 1998). The long-memory volatility dependence is also found in empirical observations for realized volatilities in subsequent research, and this similarly motivated the estimation of long-memory type ARFIMA models for realized volatilities in Areal and Taylor (2002), Andersen et al. (2003), Thomakos and Wang (2003), Pong et al. (2004), Koopman et al. (2005), and Deo et al. (2006) among others.

This paper eschews the above complicated fractionally integrated long-memory models and depends instead on the simple-to-estimate Heterogeneous autoregressive (HAR) class of volatility models proposed by Corsi (2009). The HAR formulation is based on a straightforward extension of the so-called Heterogeneous ARCH, or HARCH, class of models analysed by Müller et al. (1997), in which the conditional variance of the discretely sampled returns is parameterized as a linear function of the lagged squared returns over the identical return horizon together with the squared returns over longer and/or shorter return horizons. Although the HAR structure does not formally model long memory, the mixing of relatively few volatility components is capable of reproducing a remarkably slow volatility autocorrelation decay that is almost indistinguishable from that of a hyperbolic (long-memory) pattern over the most empirically relevant forecast horizon. Specifically, the HAR model presented by Corsi (2009) is defined as follows:

$$V_{t+1} = \beta_0 + \beta_d V_t + \beta_w \left(\frac{1}{5} \sum_{i=0}^{4} V_{t-i} \right) + \beta_m \left(\frac{1}{22} \sum_{i=0}^{21} V_{t-i} \right) + \epsilon_{t+1}, \qquad (7.1)$$

where V denotes the volatility measure being forecasted, typically RV. V_t denotes current volatility, which is the 1-day lagged value with respect to V_{t+1}. The remaining terms represent the average volatility over the past 5 days, $\frac{1}{5}\sum_{i=0}^{4} V_{t-i}$, and the past 22 days, $\frac{1}{22}\sum_{i=0}^{21} V_{t-i}$. ϵ_t is the disturbance. The predictors in this HAR model have some overlap. Specifically, the past 5-

day average volatility $\frac{1}{5}\sum_{i=0}^{4} V_{t-i}$ includes the past 1-day volatility V_t, and the past 22-day average volatility $\frac{1}{22}\sum_{i=0}^{21} V_{t-i}$ includes the average from the shorter 5-day lagged volatility $\sum_{i=0}^{4} V_{t-i}$. To reduce the overlap between these predictors, the HAR model used throughout this paper adopts the non-overlapping reparameterization of the HAR model suggested by Patton and Sheppard (2015),

$$V_{t+1} = \beta_0 + \beta_d V_t + \beta_w \left(\frac{1}{4}\sum_{i=1}^{4} V_{t-i}\right) + \beta_m \left(\frac{1}{17}\sum_{i=5}^{21} V_{t-i}\right) + \epsilon_{t+1}, \quad (7.2)$$

where $\frac{1}{4}\sum_{i=1}^{4} V_{t-i}$ does not include the average volatility over lags 2 to 5 and $\frac{1}{17}\sum_{i=5}^{21} V_{t-i}$ denotes the average volatility over lags 6 and 22. For the low-frequency data, e.g., 5 minutes, the volatility may be proxied by RV ($V = RV$). Using RV as the proxy for volatility, the HAR model becomes

$$RV_{t+1} = \beta_0 + \beta_d RV_t + \beta_w \overline{RV}_{w,t} + \beta_m \overline{RV}_{m,t} + \epsilon_{t+1}, \quad (7.3)$$

where $\overline{RV}_{w,t} = \frac{1}{4}\sum_{i=1}^{4} RV_{t-i}$ and $\overline{RV}_{m,t} = \frac{1}{17}\sum_{i=5}^{21} RV_{t-i}$.

The HAR model can be estimated by the Ordinary Least Square method (OLS) if the errors, ϵ_t, are independent, normally distributed, and have fixed volatility over sample days. However, the error term appears to change across the sample period in accordance with the level of the volatility (Patton and Sheppard, 2015), therefore, estimation by OLS has the disadvantage that the resulting estimates focus primarily on fitting periods of high volatility and place little weight on low volatility periods. To overcome this, Patton and Sheppard (2015) suggest a Weighted Least Squares (WLS) method to estimate the model. The WLS method attempts to provide a more efficient alternative to OLS by putting different weights on errors. Specifically, the WLS method puts relatively less weight on errors which are likely to have a large variance and more weight on errors which are likely to have a small variance. As for the weights, Patton and Sheppard (2015) suggest using the inverse of the fitted value of the HAR model estimated

by the OLS method. This idea is motivated by the positive relationship between the variance of residuals and the level of the fitted values of the HAR model estimated by the OLS method. Given the advantages of the WLS method, we will use the WLS method to estimate the models throughout this chapter. The statistical inference on the coefficient estimates is based on the Newey–West Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors proposed by Newey and West (1987). The full technical details describing the calculations of the WLS model estimation are provided in the Appendix.

Specification 1: To explore the impact of the signed jump variation on future volatility, Patton and Sheppard (2015) formulate a model by substituting the current realized variance RV in Equation (7.3) with the current signed jump variation estimator (J^{Δ}) and Bi-power variation estimator (B),

$$RV_{t+1} = \beta_0 + \beta_{J^{\Delta}} J_t^{\Delta} + \beta_C \, BV_t + \beta_w \, \overline{RV}_{w,t} + \beta_m \, \overline{RV}_{m,t} + \epsilon_{t+1}. \qquad (7.4)$$

We term this model the signed jump, J^{Δ}, model in this chapter.

Specification 2: To determine whether the coefficient on positive jump variation differs from that of negative jump variation, and thus whether the impact of jumps is driven more by the positive or negative jump variation, Patton and Sheppard (2015) extend the J^{Δ} model by replacing the signed jump variation estimator, J^{Δ}, with its decomposition, $J^{\Delta} = J^{+} + J^{-}$, with $J^{+} = J^{\Delta} I(J^{\Delta} > 0)$ and $J^{-} = J^{\Delta} I(J^{\Delta} < 0)$,

$$RV_{t+1} = \beta_0 + \beta_{J^{+}} J_t^{+} + \beta_{J^{-}} J_t^{-} + \beta_C \, BV_t + \beta_w \, \overline{RV}_{w,t} + \beta_m \, \overline{RV}_{m,t} + \epsilon_{t+1}. \qquad (7.5)$$

We term this model the $J^{\pm}$ model in this chapter.

Specification 3: Recall that the signed jump variation estimator can be decomposed into the modified signed jump variation estimator (the signed jump variation proxy) and a drift bias component ($J^{\Delta} = J^{*,\Delta} + D^{\Delta}$). To investigate the impact of the signed jump variation and the drift bias on future volatility, We formulate a new model by replacing the J^{Δ} component in the J^{Δ} model with $J^{*,\Delta} + D^{\Delta}$,

$$RV_{t+1} = \beta_0 + \beta_{J^{*,\Delta}} J_t^{*,\Delta} + \beta_{D^\Delta} D_t^\Delta + \beta_C \, BV_t + \beta_w \, \overline{RV}_{w,t} + \beta_m \, \overline{RV}_{m,t} + \epsilon_{t+1}. \quad (7.6)$$

We term this model the $J^{*,\Delta}D^\Delta$ model.

Specification 4: $J^{*,\Delta}$ can be decomposed into the positive and negative components: $J^{*,\Delta} = J^{*,+} + J^{*,-}$, where $J^{*,+} = J^{*,\Delta}(J^{*,\Delta} > 0)$ and $J^{*,-} = J^{*,\Delta}(J^{*,\Delta} < 0)$. And D^Δ can also be decomposed into positive and negative components: $D^\Delta = D^+ + D^-$, where $D^+ = D^\Delta(D^\Delta > 0)$ and $D^- = D^\Delta(D^\Delta < 0)$ (Recall that implication (b) in section 3 suggests that the drift bias D^Δ may have positive and negative signs). Therefore, J^Δ can be separated into four components,

$$J_t^\Delta = J_t^{*,+} + J_t^{*,-} + D_t^+ + D_t^-. \quad (7.7)$$

By replacing J^Δ in Equation (7.4) with this decomposition, We formulate a new model, defined as,

$$RV_{t+1} = \beta_0 + \beta_{J^{*,+}} J_t^{*,+} + \beta_{J^{*,-}} J_t^{*,-} + \beta_{D^+} D_t^+ + \beta_{D^-} D_t^- + \beta_C \, BV_t + \beta_w \, \overline{RV}_{w,t} + \beta_m \, \overline{RV}_{m,t} + \epsilon_{t+1}. \quad (7.8)$$

We term this model the $J^{*,\pm}D^\pm$ model.

7.1.2 In-sample estimation results

The first four columns of the upper panel of Table 7.1 report the in-sample estimation results of the J^Δ, $J^\pm$, $J^{*,\Delta}D^\Delta$, and $J^{*,\pm}D^\pm$ models, as detailed in Specifications (1) to (4) above, formalised by Equations () to (), respectively. The results show that there are three main patterns of the coefficient estimates of the J^Δ and $J^\pm$ models. The first pattern is that the estimates of β_C, β_w, and β_m , the coefficients on BV, RV_w, and RV_m, are also all positive and significant at the 5% level. This pattern evidences the stylized fact of volatility persistence, which has been found in a large volume of literature (e.g., Corsi (2009) and Andersen et al. (2007)). The second pattern is that the estimate of β_{J^Δ}, which relates to the variable measuring signed jumps, J^Δ, is negative and statistically significant at 5% level, which is consistent with

Patton and Sheppard (2015). This indicates that days dominated by negative jumps lead to higher future volatility, while days with positive jumps lead to lower future volatility, but this interpretations relies on J^{Δ} being an accurate measure for the signed jump variation. If J^{Δ} contains bias, we should be more careful when linking this finding to the predictability of the signed jump variation. The third pattern is that for the estimates of both β_{J^+} and β_{J^-}, which separate the effects of J^+ and J^-, are negative and statistically significant at the 5% level, with β_{J^-} larger in absolute magnitude than β_{J^+}, which is also consistent with Patton and Sheppard (2015). If J^+ and J^- are accurate measures of …, this finding indicates that the increase in future volatility is larger following a negative jump than the decrease in future volatility following a positive jump. If the J^+ and J^- estimators contain bias, one should not use this finding to indicate the predictability of the signed jumps.

Table 7.1. Model in-sample estimation and out-of-sample forecasting

Notes: The upper panel of the table provides the in-sample parameter estimates and measure of fit (R^2) for the J^{Δ}, $J^{\pm}$, $J^{*,\Delta}D^{\Delta}$, $J^{*,\pm}D^{\pm}$, D^{Δ}, and $D^{\pm}$ models (Specifications (1) to (6)). The bottom panel contains the MSE, MAE, and QLIKE losses of these models for evaluating their out-of-sample forecasting accuracy. The superscript (a) indicates that the model significantly outperforms the J^{Δ} model for the out-of-sample forecast accuracy based on the DMW test at the two-tail 5% significance level. The superscript (b) indicates that the model significantly outperforms the $J^{\pm}$ model for the out-of-sample forecast accuracy based on the DMW test at the two-tail 5% significance level.

	J^{Δ}	$J^{\pm}$	$J^{*,\Delta}D^{\Delta}$	$J^{*,\pm}D^{\pm}$	D^{Δ}	$D^{\pm}$
In-sample estimation						
$\beta_{J^{\Delta}}$	-0.705					
	(-2.47)					
β_{J^+}		-0.204				
		(-8.55)				
β_{J^-}		-0.941				
		(-4.86)				
$\beta_{J^{*,\Delta}}$			-0.716			
			(-0.99)			
$\beta_{D^{\Delta}}$			-0.903		-0.623	
			(-2.19)		(-5.96)	
$\beta_{J^{*,+}}$				-0.518		
				(-1.69)		
$\beta_{J^{*,-}}$				-0.149		
				(-0.83)		
β_{D^+}				-0.250		
				(-1.62)		

β_{D^-}				-1.183		-1.160
				(-5.86)		(-6.24)
β_C	0.595	0.533	0.572	0.550	0.555	0.481
	(16.12)	(13.62)	(16.38)	(10.14)	(16.97)	(13.46)
β_w	0.308	0.292	0.323	0.286	0.303	0.286
	(6.80)	(8.34)	(5.51)	(8.19)	(8.14)	(8.01)
β_m	0.084	0.090	0.086	0.099	0.098	0.106
	(3.41)	(3.87)	(2.82)	(3.99)	(4.13)	(4.41)
R^2	0.582	0.585	0.587	0.595	0.558	0.566
Out-of-sample forecast						
MSE $\times 10^7$	0.264	0.245	0.230	0.322	0.240	0.204
MAE $\times 10^4$	0.468	0.459	0.456	0.492	0.452	0.439[a,b]
QLIKE	0.194	0.185	0.203	0.232	0.179	0.170[a,b]

Since the simulation study of this paper found a large bias in J^{Δ}, J^{+}, and J^{-}, one should not link the findings of J^{Δ}, J^{+}, and J^{-} to the predictability of the signed jump variation. The estimation results for $J^{*,\Delta}D^{\Delta}$ and $J^{*,\pm}D^{\pm}$ models, where much more accurate estimators ($J^{*,\Delta,}$, $J^{*,+}$, and $J^{*,-}$) are modelled, and should be preferred to indicate the predictability of the signed jump variation.

The third column of the upper panel of Table 7.1 contains the results of the $J^{*,\Delta}D^{\Delta}$ model. The results show that the estimate of $\beta_{J^{*,\Delta}}$ (the coefficient of $J^{*,\Delta}$) is not statistically significant at the 5% level. This reveals that the days dominated by positive or negative jumps do not lead to a change in the level of future volatility, when the signed jumps are modified for drift bias. This result is quite different from that of Patton and Sheppard (2015), who found that days dominated by negative jumps lead to higher future volatility, and days with positive jumps lead to lower future volatility. But, this result is consistent with other studies which find only limited predictability of jumps in forecasting future volatility (Andersen et al., 2007, Sévi, 2014, Santos and Ziegelmann, 2014, Prokopczuk et al., 2016, Caporin, 2023, Bu et al., 2023). Via a signed jump variation estimator, modified to mitigate the influence of drift bias (the $J^{*,\Delta}$ estimator), this paper finds that signed jumps indeed do not help predict future volatility.

Besides, it can also be seen from the estimation results of the $J^{*,\Delta}D^{\Delta}$ model that the estimate of $\beta_{D^{\Delta}}$ (the coefficient of the drift bias) is negative and statistically significant at the 5% level. This result has three implications. First, the predictability of the signed jump variation estimator J^{Δ} is almost exclusively due to its drift bias. Second, days with negative drift bias lead to higher future volatility and days with positive drift bias lead to lower future volatility. Third, as drift is an increasing function of the drift bias, this result also implies that days dominated by negative drifts lead to higher future volatility, and days with positive drifts lead to lower future volatility. These three implications are novel findings in the literature and ones that cannot be detected without disentangling drift bias from the signed jump variation estimators.

The final column of Table 7.1 contains the estimation results of the $J^{*,\pm}D^{\pm}$ model. As the results show, the estimates of both β_{J^+} and β_{J^-} (the coefficients of modified positive jump variation estimators and modified negative jump variation estimators, respectively) are not statistically significant at the 5% level. This again indicates that future volatility cannot be predicted by either positive or negative jumps.

The results of the $J^{*,\pm}D^{\pm}$ model also shows that the estimate of β_{D^+} (the coefficient of positive drift bias) is negative but not statistically significant at the 5% level, and the estimate of β_{D^-} (the coefficient of negative drift bias) is statistically significant at 5% level and negative. The coefficient on the negative drift bias is larger in magnitude than on the positive drift bias, indicating that the increase in future volatility is larger in magnitude following a negative drift bias than the decrease in future volatility following a positive drift bias. The insignificant coefficient of the positive drift bias suggests that the impact of the positive drift bias on future volatility is limited.

The final row in the upper panel of Table 7.1 reports R^2 for measuring the goodness of fit of the models. The details for calculating R^2 are attached in the Appendix. As the results show,

the $R^2$of the $J^{*,\Delta}D^{\Delta}$ model is greater than that of the J^{Δ} model, and the $R^2$of the $J^{*,\pm}D^{\pm}$ model is greater than that of the $J^{\pm}$ model. This indicates that the models with drift bias disentangled from the original signed jump variation estimator lead to better in-sample goodness of fit.

Motivated by the limited in-sample evidence for the modified signed jump variation estimator, We formulate the following D^{Δ} model by removing the $J^{*,\Delta}$ component from the $J^{*,\Delta}D^{\Delta}$ model.

Specification 5: the D^{Δ} model:

$$RV_{t+1} = \beta_0 + \beta_{D^{\Delta}} D_t^{\Delta} + \beta_C\, BV_t + \beta_w\, \overline{RV}_{w,t} + \beta_m\, \overline{RV}_{m,t} + \epsilon_{t+1}. \qquad (7.9)$$

Motivated by the limited in-sample evidence of D^{+}, $J^{*,+}$, $J^{*,-}$ components, We formulate the following D^{-} model by removing the $D^{+}, J^{*,+}, J^{*,-}$ components from the $J^{*,\pm}D^{\pm}$ model.

Specification 6: the D^{-} model:

$$RV_{t+1} = \beta_0 + \beta_{D^{-}}\, D_t^{-} + \beta_C\, BV_t + \beta_w\, \overline{RV}_{w,t} + \beta_m\, \overline{RV}_{m,t} + \epsilon_{t+1}. \qquad (7.10)$$

The final two columns of the upper panel of Table 7.1 represent the in-sample estimation results for the D^{Δ} and D^{-} models. As shown by the results, the estimates of β_d, β_w, and β_m (for BV, RV_w, and RV_m) are also all positive and significant at the 5% level for both the D^{Δ} and D^{-} models, again indicating the stylized fact of the volatility persistence. Besides, both $\beta_{d^{\Delta}}$ of the D^{Δ} model and $\beta_{d^{-}}$ of the D^{-} model are negative and significant at the 5% level, again indicating the negative impact of the drift bias and the negative drift bias on future volatility.

7.1.3 Out-of-sample forecast results

This Section compares the six models defined in section 7.12 (Specifications (1)-(6)), in terms of their out-of-sample performances. The out-of-sample analysis is the only way to gauge

the forecasting performance (Giot and Laurent, 2007) and to interpret the additional predictive performance of new variables. In this chapter, the out-of-sample forecast is based on a rolling window of 500 observations over the full forecasting period. Thus, $n - 500$ forecasts, F_t, for $t = 501, \dots, n$, are obtained from each model. Due to estimation errors in the predictors (e.g., $RV, BV, J^{\Delta}, D^{\Delta}$), the models (Specifications (1)-(6)) may occasionally produce implausibly large or small forecasts. Thus, to make the forecast analysis in this paper more realistic, We apply an ‘‘insanity filter’’ (IF), as suggested by Swanson and White (1997), to the forecasts for all of these models. The IF substitutes the forecast with the unconditional mean within the estimation window if the forecasts are outside the interval between the minimum and maximum forecasting target (V) of this window. The insanity filter is commonly applied in the influential studies on volatility forecasting (e.g., Patton and Sheppard, 2015, Bollerslev et al., 2016, Bollerslev et al., 2018).

The evaluation criteria of forecasting accuracy rely on the loss functions. Not all functions are unbiased for evaluating the volatility forecasting accuracy due to the fact that volatility can only be measured by a proxy that is imperfect in nature (Patton, 2011). Among various loss functions, Patton (2011) finds that Squared Error (SE) and Quasi-likelihood (QLIKE) loss functions are unbiased in the presence of an imperfect volatility proxy. For this reason, this paper considers the SE and QLIKE loss functions for evaluating the forecasting accuracy of the models. Additionally, for robustness purposes, We also considers the Absolute Error (AE) loss function. The AE, SE, and QLIKE loss functions are standard methodologies in the volatility forecasting literature (e.g., Sévi, 2014, Patton and Sheppard, 2015, Liu et al., 2015, Bollerslev et al., 2016, Andersen et al., 2021).

The SE, AE, and QLIKE loss functions are defined by,

$$\mathrm{SE_t} = (RV_t - F_t)^2, \tag{7.11}$$

$$\mathrm{AE}_t = |RV_t - F_t|, \tag{7.12}$$

and

$$\mathrm{QLIKE}_t = \frac{RV_t}{F_t} - \ln\frac{RV_t}{F_t} - 1, \tag{7.13}$$

respectively, where $t = 501, \dots, n$. V_t indicates the volatility (e.g., measured by RV) at day t, and F_t indicate the volatility forecast from a particular model based on the 500 observations before day t. Smaller SE_t, AE_t, and QLIKE_t indicate more accurate forecasts. Based on the rolling window, SE_t, AE_t, and QLIKE_t for $t = 501, \dots, n$ are produced for each model. The mean value of SE_t, AE_t, and QLIKE_t for $t = 501, \dots, n$ are used to indicate the general forecasting accuracy of the model for $t = 501, \dots, n$. We term the mean value of SE_t, AE_t, and QLIKE_t by MSE, MAE, and QLIKE, respectively. A model with a smaller mean value of the loss function, relative to another model, indicates more precise forecasts on average.

However, the difference in the mean loss function of the two models may arise by chance, since the mean value of the loss function is calculated on a sample (e.g., the mean of SE_t is computed on the sample of $t = 501, \dots, n$) as opposed to the population. To reduce the uncertainty of the results, this paper uses hypothesis testing to make probabilistic statements about the population of the difference in the mean loss function of two models. The hypothesis testing is based on the Diebold-Mariano-West (DMW) test statistic developed by Diebold and Mariano (1995) and West (1996), with adjustment to the Newey-West Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors proposed by Newey and West (1987).[5] The DMW test has been widely applied to key forecast literature for testing the statistical significance of model forecast accuracy (Patton, 2011, Sévi, 2014, Patton and Sheppard, 2015, Andersen et al., 2021).

[5] The DMW results in this chapter were obtained using the robust_loss_1 function from Andrew Patton's Matlab code page, http://public.econ.duke.edu/~ap172/

To implement the DMW test, we first define d_t as the difference between the loss functions from two separate models,

$$d_t = \mathcal{L}_{\mathrm{t}}^{\mathrm{A}} - \mathcal{L}_{\mathrm{t}}^{\mathrm{B}}, \tag{7.14}$$

where $\mathcal{L}_{\mathrm{t}}^{\mathrm{A}}$ denotes a loss function calculated on the forecast by model A and $\mathcal{L}_{\mathrm{t}}^{\mathrm{B}}$ denotes the same loss function, but calculated on the forecast by model B. The null hypothesis is that the expectation of d_t equals zero, H_0: $E(d_t) = 0$. That is, the forecast accuracy of models A and B is identical. The alternative hypothesis is that the expectation of d_t does not equal zero, H_0: $E(d_t) \neq 0$. That is, the forecast accuracy of models A and B is different. To obtain the DMW test statistic for testing these hypotheses, d_t is first calculated on the sample ($t = 501, \dots, n$) then regressed on one, $d_t = \varphi_0 \cdot 1 + \varepsilon_t$. And the HAC t-statistics of the coefficient φ_0 is the DMW statistic.

If the DMW statistic exceeds a positive threshold (e.g., 1.96 for a two-sided 5% significance level), the null hypothesis is rejected, indicating that the forecast accuracy of model B is significantly better than that of model A. In other words, a positive test statistic reflects a larger loss function for model A than model B. Analogously, if the DMW statistic exceeds a negative threshold (e.g., -1.96 for a two-sided 5% significance level), the null hypothesis is rejected, indicating that the forecast accuracy of model B is significantly inferior to that of model A. If the DMW statistic does not exceed either the positive or negative threshold, the null hypothesis is not rejected, indicating that the forecast accuracy of model A is not significantly different from that of model B.

The lower panel of Table (7.1) reports the MSE, MAE, and QLIKE losses of the models as described by Specifications (1) to (6), for evaluating the out-of-sample forecasting accuracy of these models. The superscript (a) indicates that the model significantly outperforms the NJ^{Δ} model for the out-of-sample forecast accuracy, based on the DMW test at the two-tail 5% significance level. The superscript (b) indicates that the model significantly outperforms the

NJ$^{\pm}$ model for the out-of-sample forecast accuracy, based on the DMW test at the two-tail 5% significance level. As the results show, the mean loss functions of the D^{Δ} model and the $D^{\pm}$ model, with few exceptions, are generally lower than those of the $J^{*,\Delta}$D$^{\Delta}$ model and $J^{*,\pm}$D$^{\pm}$ model, respectively. This indicates that the signed jump variation and its sign decomposition do not help to forecast volatility according to out-of-sample evaluation. (recall that the only difference between the D^{Δ} model and the $J^{*,\Delta}$D$^{\Delta}$ model is the signed jump variation component and the only difference between the $D^{\pm}$ model and the $J^{*,\pm}$D$^{\pm}$ model is the positive and negative jump variation components). The limited out-of-sample evidence of signed jump variation and its sign decomposition corresponds to the limited in-sample evidence found in the previous section.

Additionally, the mean loss functions of both the D^{Δ} model and the $D^{\pm}$ model are smaller compared to their counterparts, the J^{Δ} model and the $J^{\pm}$ model. Moreover, based on the DMW test results, the $D^{\pm}$ model significantly outperforms both of the J^{Δ} model and the $J^{\pm}$ model according to the MAE and QLIKE loss functions. The superiority of the D^{Δ} and $D^{\pm}$ models over the J^{Δ} and $J^{\pm}$ models indicates the economic importance of extracting the drift bias component and its sign decomposition from the signed jump variation estimator, as measured by the out-of-sample volatility forecasting performance analysis.

7.2 Volatility forecasting with ultra-high-frequency data

7.2.1 Volatility forecasting models

To model volatility with ultra-high-frequency data, the volatility models defined by Specifications (1) to (4) are modified by replacing all variables with their noise-modified versions as defined in section 6. These noise-modified models are defined by the following Specifications (7) to (10).

Specification 7: This model is the noise-modified J^{Δ} model, which exploits the impact of noise-modified signed jump variation estimator on future volatility for ultra-high-frequency data,

$$NRV_{t+1} = \beta_0 + \beta_{NJ^{\Delta}} NJ_t^{\Delta} + \beta_C NBV_t + \beta_w \overline{NRV}_{w,t} + \beta_m \overline{NRV}_{m,t} + \epsilon_{t+1}. \quad (7.15)$$

We term this model the NJ^{Δ} model in this chapter.

Specification 8: This model is the noise-modified $J^{\pm}$ model, which exploits the impact of noise-modified positive and negative jump variation estimators on future volatility for ultra-high-frequency data,

$$\begin{aligned} NRV_{t+1} = \beta_0 + \beta_{NJ^{*,+}} NJ_t^{*,+} + \beta_{NJ^{*,-}} NJ_t^{*,-} + \beta_{ND^+} ND_t^{+} + \beta_{ND^-} ND_t^{-} \\ +\beta_C NBV_t + \beta_w \overline{NRV}_{w,t} + \beta_m \overline{NRV}_{m,t} + \epsilon_{t+1}. \end{aligned} \quad (7.16)$$

We term this model the $NJ^{\pm}$ model in this chapter.

Specification 9: Recall from Equation (4.20) that the noise-modified signed jump variation estimator can be decomposed into the drift-modified noise-modified signed jump variation estimator (the better signed jump variation proxy) and a drift bias component ($NJ^{\Delta} = NJ^{*,\Delta} + ND^{\Delta}$). To investigate the impact of the signed jump variation and the drift bias on future volatility for ultra-high-frequency data, We formulate a new model by replacing the NJ^{Δ} component in the NJ^{Δ} model with this decomposition,

$$NRV_{t+1} = \beta_0 + \beta_{NJ^{*,\Delta}} NJ_t^{*,\Delta} + \beta_{ND^{\Delta}} ND_t^{\Delta} + \beta_C NBV_t + \beta_w \overline{NRV}_{w,t} + \beta_m \overline{NRV}_{m,t} + \epsilon_{t+1}, \quad (7.17)$$

We term this model the $NJ^{*,\Delta}ND^{\Delta}$ model.

Specification 10: $NJ^{*,\Delta}$ can be decomposed into the positive and negative components: $NJ^{*,\Delta} = NJ^{*,+} + NJ^{*,-}$, where $NJ^{*,+} = NJ^{*,\Delta}(NJ^{*,\Delta} > 0)$ and $NJ^{*,-} = NJ^{*,\Delta}(NJ^{*,\Delta} < 0)$. Recall that the drift bias ND^{Δ} may have positive and negative signs. Thus, ND^{Δ} can also be decomposed into positive and negative components: $ND^{\Delta} = ND^{+} + ND^{-}$, where $ND^{+} =$

$ND^{\Delta}(ND^{\Delta} > 0)$ and $ND^{-} = ND^{\Delta}(ND^{\Delta} < 0)$. Therefore, NJ^{Δ} is separated into four components,

$$NJ_t^{\Delta} = NJ_t^{*,+} + NJ_t^{*,-} + ND_t^{+} + ND_t^{-}. \tag{7.18}$$

By replacing NJ_t^{Δ} in Equation (5.15) with this decomposition, We formulate a new model,

$$\begin{aligned} NRV_{t+1} = \beta_0 + \beta_{NJ^{*,+}}\, NJ_t^{*,+} + \beta_{NJ^{*,-}}\, NJ_t^{*,-} + \beta_{ND^{+}}\, ND_t^{+} + \beta_{ND^{-}}\, DN_t^{-} \\ +\beta_C\, NBV_t + \beta_w\, \overline{NRV}_{w,t} + \beta_m\, \overline{NRV}_{m,t} + \epsilon_{t+1}. \end{aligned} \tag{7.19}$$

We term this model the $\mathrm{NJ}^{*,\pm}\mathrm{ND}^{\pm}$ model.

7.2.2 In-sample estimation results

Table 7.2 reports in-sample estimation results of Specification (7) to (10) for the ultra-high-frequency SPY data. As the results show, across all of these models the coefficients β_d, β_w, and β_m (for NBV, NRV_w, and NRV_m, respectively) are all positive and significant at the 5% level. This corroborates the persistence of volatility previously found under the 5-minute frequency scheme. Still, for the NJ^{Δ} and $NJ^{\pm}$ models, the coefficients $\beta_{NJ^{\Delta}}$, $\beta_{NJ^{+}}$ and $\beta_{NJ^{-}}$ (of NJ^{Δ}, NJ^{+}, and NJ^{-}) are negative and significant at the 5% level ($\beta_{NJ^{+}}$ is somewhat not significant but still has a negative sign). This finding reflects that the signed jumps estimators exhibit a leverage effect on future volatility under ultra-high frequency.

For the $NJ^{*,\Delta}ND^{\Delta}$ model, the estimate of $\beta_{NJ^{*,\Delta}}$ (the coefficient on $NJ^{*,\Delta}$) is not statistically significant at the 5% level, while $\beta_{ND^{\Delta}}$ (the coefficient on ND^{Δ}) is negative and significant at the 5% level. This indicates that the signed jump variation component of NJ^{Δ} only has a limited impact, and the predictability of NJ^{Δ} is almost exclusively due to its drift bias (recall that $NJ^{\Delta} = NJ^{*,\Delta} + ND^{\Delta}$). This finding is consistent with the corresponding low-frequency $J^{*,\Delta}D^{\Delta}$ model, which showed that the predictability of the signed jump variation estimator J^{Δ} is driven by its drift bias.

For the $NJ^{*,\pm}ND^{\pm}$ model, $\beta_{NJ^{*,+}}$, $\beta_{NJ^{*,-}}$, and $\beta_{ND^{*,+}}$ (the coefficients on $NJ^{*,+}$, $NJ^{*,-}$ and $ND^{*,+}$, respectively) are not statistically significant at the 5% level, while only $\beta_{ND^{*,-}}$ (the coefficients on $ND^{*,-}$) is negative and significant at the 5% level. This indicates that the predictability of NJ^{Δ} is mainly attributed to its negative drift bias, as opposed to the positive or negative jump variations. This finding, which is revealed under noisy ultra-high frequency data, is consistent with that previously discovered from the corresponding low-frequency $J^{*,\pm}D^{\pm}$ model (recall that the results of $J^{*,\pm}D^{\pm}$ model shows that the predictability of J^{Δ} is due to its negative drift bias).

Table 7.2. Model in-sample estimation results and out-of-sample forecasting performance with ultrahigh-frequency data

Notes: The upper panel of the table provides the in-sample parameter estimates and measure of fit (R^2) of the NJ^{Δ}, $NJ^{\pm}$, $NJ^{*,\Delta}ND^{\Delta}$, $NJ^{*,\pm}ND^{\pm}$, ND^{Δ}, and $ND^{\pm}$ models (Specifications (7) to (12)). The bottom panel contains the MSE, MAE, and QLIKE losses of these models for evaluating their out-of-sample forecasting accuracy. The superscript (a) indicates that the model significantly outperforms the J^{Δ} model for the out-of-sample forecast accuracy based on the DMW test at the two-tail 5% significance level. The superscript (b) indicates that the model significantly outperforms the $J^{\pm}$ model for the out-of-sample forecast accuracy based on the DMW test at the two-tail 5% significance level.

	NJ^{Δ}	$NJ^{\pm}$	$NJ^{*,\Delta}ND^{\Delta}$	$NJ^{*,\pm}ND^{\pm}$	ND^{Δ}	$ND^{\pm}$
In-sample estimation results						
$\beta_{NJ^{\Delta}}$	-0.447					
	(-2.16)					
$\beta_{NJ^{+}}$		0.160				
		(0.57)				
$\beta_{NJ^{-}}$		-1.795				
		(-2.93)				
$\beta_{NJ^{*,\Delta}}$			-0.280			
			(-1.18)			
$\beta_{ND^{\Delta}}$			-0.900		-0.753	
			(-3.57)		(-3.53)	
$\beta_{NJ^{*,+}}$				-0.085		
				(-0.19)		
$\beta_{NJ^{*,-}}$				-0.613		
				(-1.57)		
$\beta_{ND^{+}}$				0.473		
				(1.23)		
$\beta_{ND^{-}}$				-2.035		-1.690
				(-4.14)		(-4.76)
β_{C}	0.684	0.623	0.641	0.532	0.638	0.559
	(15.24)	(10.90)	(14.86)	(9.16)	(14.52)	(11.02)
β_{w}	0.206	0.217	0.229	0.213	0.226	0.229
	(5.87)	(6.21)	(6.32)	(6.28)	(6.19)	(6.51)

β_m	0.024	0.024	0.028	0.030	0.029	0.031
	(1.17)	(1.22)	(1.37)	(1.55)	(1.41)	(1.51)
R^2	0.742	0.747	0.743	0.758	0.743	0.752
Out-of-sample forecast						
$\text{MSE(OOS)} \times 10^7$	0.080	0.093	0.094	0.094	0.072	0.080
$\text{MAE(OOS)} \times 10^4$	0.253	0.265	0.273	0.265	0.252	0.249
QLIKE(OOS)	0.185	0.218	0.217	0.194	$0.176^{a,b}$	$0.168^{a,b}$

Turning to the in-sample goodness of fit (R^2) of the NJ^{Δ}, $NJ^{\pm}$, $NJ^{*,\Delta}ND^{\Delta}$, and $NJ^{*,\pm}ND^{\pm}$ models reported in the final row of the upper panel of Table (7.2), the $NJ^{*,\Delta}ND^{\Delta}$ and $NJ^{*,\pm}ND^{\pm}$ models have greater R^2 than the NJ^{Δ} and $NJ^{*,\pm}ND^{\pm}$ models, respectively. This indicates that for ultra-high-frequency data, the model with the drift bias disentangled from the noise-modified signed jump variation estimator leads to better in-sample goodness of fit. This finding is consistent with that of the corresponding models using data sampled at the 5-minute frequency, specifically, the $J^{*,\Delta}D^{\Delta}$ and $J^{*,\pm}D^{\pm}$ models have greater R^2 than the J^{Δ} and $J^{*,\pm}D^{\pm}$ models, respectively.

Motivated by the limited in-sample evidence of the $NJ^{*,\Delta}$ estimator, We formulate the following ND^{Δ} model by removing the $NJ^{*,\Delta}$ component from the $NJ^{*,\Delta}ND^{\Delta}$ model.

Specification 11: the ND^{Δ} model:

$$NRV_{t+1} = \beta_0 + \beta_{ND^{\Delta}} ND_t^{\Delta} + \beta_C\, NBV_t + \beta_w\, \overline{NRV}_{w,t} + \beta_m\, \overline{NRV}_{m,t} + \epsilon_{t+1}. \qquad (7.20)$$

Motivated by the limited in-sample evidence of the ND^{+}, $NJ^{*,+}$, $NJ^{*,-}$ components, We formulate the following ND^{-} model by removing the ND^{+}, $NJ^{*,+}$, $NJ^{*,-}$ components from the $NJ^{*,\pm}ND^{\pm}$ model.

Specification 12: the ND^{-} model:

$$NRV_{t+1} = \beta_0 + \beta_{ND^{-}} ND_t^{-} + \beta_C\, NBV_t + \beta_w\, \overline{NRV}_{w,t} + \beta_m\, \overline{NRV}_{m,t} + \epsilon_{t+1}. \qquad (7.21)$$

The final two columns of the upper panel of Table 7.2 represents the in-sample estimation results for the ND^{Δ} and ND^{-} models. As shown by the results, the estimates of β_C, β_w, and β_m (for NBV, NRV_w, and NRV_m) are also all positive and significant at the 5% level for both the

ND^{Δ} and ND^{-} models, again indicating the stylized fact of the volatility persistence. Besides, both $\beta_{ND^{\Delta}}$ of the ND^{Δ} model and $\beta_{ND^{-}}$ of the ND^{-} model are negative and significant at the 5% level, again indicating the negative impact of the drift bias and the negative drift bias on future volatility.

7.2.3 Out-of-sample forecast results

The lower panel of Table 7.2 reports the MSE, MAE, and QLIKE losses of the models as described by Specifications (7) to (12), for evaluating out-of-sample forecasting accuracy. The superscript (a) indicates that the model significantly outperforms the NJ^{Δ} model for the out-of-sample forecast accuracy based on the DMW test at the two-tail 5% significance level. The superscript (b) indicates that the model significantly outperforms the $\mathrm{NJ}^{\pm}$ model for the out-of-sample forecast accuracy based on the DMW test at the two-tail 5% significance level. As the results show, the ND^{Δ} model and the $ND^{\pm}$ model performs better than the $NJ^{*,\Delta}ND^{\Delta}$ model and $NJ^{*,\pm}ND^{\pm}$ model, respectively, according to their out-of-sample accuracy. In detail, the mean loss functions of the ND^{Δ} model and the $ND^{\pm}$ model are lower than those of the $NJ^{*,\Delta}ND^{\Delta}$ model and $NJ^{*,\pm}ND^{\pm}$ model, respectively. The better performance of the ND^{Δ} model and the $ND^{\pm}$ model compared to the $NJ^{*,\Delta}ND^{\Delta}$ model and $NJ^{*,\pm}ND^{\pm}$ model indicate that the signed jump variation and its sign decomposition do not help forecast volatility, when evaluated by out-of-sample forecast accuracy, for ultra-high-frequency data (recall that the only difference between the ND^{Δ} model and the $NJ^{*,\Delta}ND^{\Delta}$ model is the modified noise-modified signed jump variation component and the only difference between the $ND^{\pm}$ model and the $NJ^{*,\pm}ND^{\pm}$ model is the modified noise-modified positive and the modified noise-modified negative jump variation components). The limited out-of-sample evidence of signed

jump variation and its sign decomposition corresponds to the limited in-sample evidence found in Section 7.2.2.

Additionally, the results also show that the ND^{Δ} and $ND^{\pm}$ models are superior to the NJ^{Δ} and NJ$^{\pm}$ model. Specifically, the mean loss function of either the ND^{Δ} model or the $ND^{\pm}$ model are smaller than those of both of the NJ^{Δ} model and the $NJ^{\pm}$ model. Moreover, based on the DMW test results, the $ND^{\pm}$ model is significantly superior to both of the NJ^{Δ} model and the $NJ^{\pm}$ model for the QLIKE loss function. The superiority of the ND^{Δ} and $ND^{\pm}$ models over the NJ^{Δ} and $NJ^{\pm}$ model indicates the economic importance of extracting the drift bias component and its sign decomposition from the noise-modified signed jump variation estimator under ultra-high-frequency data, in terms of the out-of-sample volatility forecasting precision.

8 Conclusion

The simulations of this paper reveal that with a rather low but realistic sampling frequency (e.g., 5 min), the bias due to a non-zero drift is also large in the realized semi-variances. When the signs of the realized semi-variance and the drift are equal, the realized semi-variances tend to overestimate the half volatility in the presence of a non-zero drift, and the bias increases with the magnitude of the drift. When the signs of the realized semi-variance and the drift are different, the realized semi-variances tend to underestimate the half volatility in the presence of a non-zero drift, and the bias is not sensitive to the variation of the drift. Moreover, despite the drift becoming extremely close to zero as the sampling frequency increases, my simulations show that the volatility estimators of Christensen et al. (2014), which are robust to microstructure noise and designed for ultra-high-frequency data, suffer from the same problem due to the use of pre-averaged returns. Consequently, the procedures derived from these volatility estimators such as the jump asymmetry estimation of Barndorff-Nielsen et al. (2008) have unsatisfactory performance in finite samples if log prices have a non-zero drift.

We propose an alternative construction of the jump asymmetry estimator, following the centring procedure of Laurent and Shi (2020). The simulations for the modified jump asymmetry estimator reveal dramatic improvement in the estimation accuracy. The newly proposed jump asymmetry estimator, along with its original version, is applied to high frequency log returns of the S&P 500 for the period from 1997 to 2021. From the results, We observe that the drift bias is empirically large and has positive or negative signs, which is consistent with the simulations. Furthermore, We apply these estimators to forecast volatility via the HAR volatility forecasting model framework by Corsi (2009). The results show that the predictive power of the jump asymmetry estimator previously found in Patton and Sheppard (2015) is almost exclusively due to drift bias in the presence of a non-zero drift. The modified

jump asymmetry estimators, which measure the jump asymmetry much more precisely, only exhibit limited predictability. We show that disentangling the drift bias from the jump asymmetry estimator also leads to significantly more accurate out-of-sample volatility forecasts.

References:

AÏT-SAHALIA, Y., JACOD, J. & LI, J. 2012. Testing for jumps in noisy high frequency data. *Journal of Econometrics,* 168**,** 207-222.

ANDERSEN, T. G. & BOLLERSLEV, T. 1997. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Journal of Finance,* 52**,** 975-1005.

ANDERSEN, T. G. & BOLLERSLEV, T. 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review,* 39**,** 885-905.

ANDERSEN, T. G., BOLLERSLEV, T. & DIEBOLD, F. X. 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics,* 89**,** 701-720.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. & EBENS, H. 2001. The distribution of realized stock return volatility. *Journal of Financial Economics,* 61**,** 43-76.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. & LABYS, P. 2000. Great realizations. *Risk,* 13**,** 105-108.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. & LABYS, P. 2003. Modeling and forecasting realized volatility. *Econometrica,* 71**,** 579-625.

ANDERSEN, T. G., LI, Y., TODOROV, V. & ZHOU, B. 2021. Volatility measurement with pockets of extreme return persistence. *Journal of Econometrics*.

AREAL, N. M. & TAYLOR, S. J. 2002. The realized volatility of FTSE-100 futures prices. *Journal of Futures Markets,* 22**,** 627-648.

BAILLIE, R. T., BOLLERSLEV, T. & MIKKELSEN, H. O. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics,* 74**,** 3-30.

BANDI, F. M. & RUSSELL, J. R. 2008. Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies,* 75**,** 339-369.

BARIGOZZI, M., BROWNLEES, C., GALLO, G. M. & VEREDAS, D. 2014. Disentangling systematic and idiosyncratic dynamics in panels of volatility measures. *Journal of Econometrics,* 182**,** 364-384.

BARNDORFF-NIELSEN, O. E., KINNEBROCK, S. & SHEPHARD, N. 2008. Measuring downside risk-realised semivariance. *CREATES Research Paper,* 42.

BARNDORFF-NIELSEN, O. E. & SHEPHARD, N. 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics,* 2**,** 1-37.

BARNDORFF-NIELSEN, O. E. & SHEPHARD, N. 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics,* 4**,** 1-30.

BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. & SHEPHARD, N. 2009. Realized kernels in practice: trades and quotes. *The Econometrics Journal,* 12**,** C1–C32.

BLACK, F. 1986. Noise. *The Journal of Finance,* 41**,** 528-543.

BLACK, F. & SCHOLES, M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy,* 81**,** 637-654.

BOLLERSLEV, T., HOOD, B., HUSS, J. & PEDERSEN, L. H. 2018. Risk everywhere: Modeling and managing volatility. *The Review of Financial Studies,* 31**,** 2729-2773.

BOLLERSLEV, T., MEDEIROS, M. C., PATTON, A. J. & QUAEDVLIEG, R. 2021. From zero to hero: Realized partial (co) variances. *Journal of Econometrics,* 231**,** 348-360.

BOLLERSLEV, T., PATTON, A. J. & QUAEDVLIEG, R. 2016. Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics,* 192**,** 1-18.

BREIDT, F. J., CRATO, N. & DE LIMA, P. 1998. The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics,* 83**,** 325-348.

BU, R., HIZMERI, R., IZZELDIN, M., MURPHY, A. & TSIONAS, M. 2023. The contribution of jump signs and activity to forecasting stock price volatility. *Journal of Empirical Finance,* 70**,** 144-164.

CAPORIN, M. 2023. The role of jumps in realized volatility modeling and forecasting. *Journal of Financial Econometrics,* 21**,** 1143-1168.

CHRISTENSEN, K., OOMEN, R. C. & PODOLSKIJ, M. 2014. Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics,* 114**,** 576-599.

CHRISTOFFERSEN, P. F. & DIEBOLD, F. X. 2000. How relevant is volatility forecasting for financial risk management? *Review of Economics and Statistics,* 82**,** 12-22.

CORSI, F. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics,* 7**,** 174-196.

CORSI, F., MITTNIK, S., PIGORSCH, C. & PIGORSCH, U. 2008. The volatility of realized volatility. *Econometric Reviews,* 27**,** 46-78.

CORSI, F., PIRINO, D. & RENO, R. 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics,* 159**,** 276-288.

CORSI, F. & RENÒ, R. 2012. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics,* 30**,** 368-380.

DEO, R., HURVICH, C. & LU, Y. 2006. Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. *Journal of Econometrics,* 131**,** 29-58.

DIEBOLD, F. X. & MARIANO, R. S. 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics,* 13.

DING, Z., GRANGER, C. W. & ENGLE, R. F. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance,* 1**,** 83-106.

DUFFIE, D., PAN, J. & SINGLETON, K. 2000. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica,* 68**,** 1343-1376.

DUONG, D. & SWANSON, N. R. 2015. Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. *Journal of Econometrics,* 187**,** 606-621.

FAN, J., IMERMAN, M. B. & DAI, W. 2016. What does the volatility risk premium say about liquidity provision and demand for hedging tail risk? *Journal of Business & Economic Statistics,* 34**,** 519-535.

GAO, L., HAN, Y., LI, S. Z. & ZHOU, G. 2018. Market intraday momentum. *Journal of Financial Economics,* 129**,** 394-414.

GIOT, P. & LAURENT, S. 2007. The information content of implied volatility in light of the jump/continuous decomposition of realized volatility. *Journal of Futures Markets,* 27**,** 337-359.

HANSEN, P. R. & LUNDE, A. 2006. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics,* 24**,** 127-161.

HAUTSCH, N. & PODOLSKIJ, M. 2013. Preaveraging-based estimation of quadratic variation in the presence of noise and jumps: theory, implementation, and empirical evidence. *Journal of Business & Economic Statistics,* 31**,** 165-183.

HUANG, X. & TAUCHEN, G. 2005. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics,* 3**,** 456-499.

JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKIJ, M. & VETTER, M. 2009. Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and Their Applications,* 119**,** 2249-2276.

KOOPMAN, S. J., JUNGBACKER, B. & HOL, E. 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance,* 12**,** 445-475.

LAURENT, S. & SHI, S. 2020. Volatility estimation and jump detection for drift–diffusion processes. *Journal of Econometrics,* 217**,** 259-290.

LEE, S. S. & MYKLAND, P. A. 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies,* 21**,** 2535-2563.

LEE, S. S. & MYKLAND, P. A. 2012. Jumps in equilibrium prices and market microstructure noise. *Journal of Econometrics,* 168**,** 396-406.

LIU, L. Y., PATTON, A. J. & SHEPPARD, K. 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics,* 187**,** 293-311.

MERTON, R. C. 1969. Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics,* 51**,** 247-257.

MÜLLER, U. A., DACOROGNA, M. M., DAVÉ, R. D., OLSEN, R. B., PICTET, O. V. & VON WEIZSÄCKER, J. E. 1997. Volatilities of different time resolutions—analyzing the dynamics of market components. *Journal of Empirical Finance,* 4**,** 213-239.

NELSON, D. B. 1990. ARCH models as diffusion approximations. *Journal of Econometrics,* 45**,** 7-38.

NEWEY, W. & WEST, K. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica,* 55**,** 703-708.

NIEDERHOFFER, V. & OSBORNE, M. F. M. 1966. Market making and reversal on the stock exchange. *Journal of the American Statistical Association,* 61**,** 897-916.

PARK, S. & LINTON, O. 2012. Realized volatility: Theory and applications. *Handbook of volatility models and their applications.* Wiley.

PATTON, A. J. 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics,* 160**,** 246-256.

PATTON, A. J. & SHEPPARD, K. 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics,* 97**,** 683-697.

PODOLSKIJ, M. & VETTER, M. 2009. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli,* 15**,** 634-658.

PONG, S., SHACKLETON, M. B., TAYLOR, S. J. & XU, X. 2004. Forecasting currency volatility: A comparison of implied volatilities and AR (FI) MA models. *Journal of Banking & Finance,* 28**,** 2541-2563.

PROKOPCZUK, M., SYMEONIDIS, L. & WESE SIMEN, C. 2016. Do jumps matter for volatility forecasting? Evidence from energy markets. *Journal of Futures Markets,* 36**,** 758-792.

ROBINSON, P. M. 1991. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics,* 47**,** 67-84.

ROLL, R. 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance,* 39**,** 1127-1139.

SANTOS, D. G. & ZIEGELMANN, F. A. 2014. Volatility forecasting via MIDAS, HAR and their combination: An empirical comparative study for IBOVESPA. *Journal of Forecasting,* 33**,** 284-299.

SÉVI, B. 2014. Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research,* 235**,** 643-659.

SWANSON, N. R. & WHITE, H. 1997. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting,* 13**,** 439-461.

THOMAKOS, D. D. & WANG, T. 2003. Realized volatility in the futures markets. *Journal of Empirical Finance,* 10**,** 321-353.

ZHANG, L., MYKLAND, P. A. & AÏT-SAHALIA, Y. 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association,* 100**,** 1394-1411.

Appendix

This section explains the motivation and calculation of using the Weighted Least Square (WLS) estimation method in this chapter. For the illustration, We consider the estimation of the J^{Δ} model as an example. The traditional method for estimating the J^{Δ} model relies on the Ordinary Least Square (OLS) method (e.g., Andersen et al., 2007, Corsi, 2009). To implement this method, for example, for the J^{Δ} model, first substitute the realized measures from the sample into this model,

$$RV_{23} = \beta_0 + \beta_{j^{\Delta}} J_{22}^{\Delta} + \beta_d \, RV_{22} + \beta_w \, \overline{RV}_{w,22} + \beta_m \, \overline{RV}_{m,22} + \epsilon_{23},$$

$$RV_{24} = \beta_0 + \beta_{j^{\Delta}} J_{23}^{\Delta} + \beta_d \, RV_{23} + \beta_w \, \overline{RV}_{w,23} + \beta_m \, \overline{RV}_{m,23} + \epsilon_{24},$$

$$\vdots$$

$$RV_n = \beta_0 + \beta_{j^{\Delta}} J_{n-1}^{\Delta} + \beta_d \, RV_{n-1} + \beta_w \, \overline{RV}_{w,n-1} + \beta_m \, \overline{RV}_{m,n-1} + \epsilon_n.$$

Note that the dependent variable starts from RV_{23} as the calculation of the respective independent variable $\overline{RV}_{m,t}$ requires a minimum of past 21 observations. The matrix presentation of above equations is,

$$\begin{bmatrix} RV_{23} \\ RV_{24} \\ \vdots \\ \vdots \\ RV_n \end{bmatrix} = \begin{bmatrix} 1 & J_{22}^{\Delta} & RV_{22} & \overline{RV}_{w,22} & \overline{RV}_{m,22} \\ 1 & J_{23}^{\Delta} & RV_{23} & \overline{RV}_{w,23} & \overline{RV}_{m,23} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & J_{n-1}^{\Delta} & RV_{n-1} & \overline{RV}_{w,n-1} & \overline{RV}_{m,n-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_{\mathrm{d}} \\ \beta_{\mathrm{w}} \\ \beta_{\mathrm{m}} \end{bmatrix} + \begin{bmatrix} \epsilon_{23} \\ \epsilon_{24} \\ \vdots \\ \vdots \\ \epsilon_n \end{bmatrix}. \quad (\mathrm{A}.1)$$

This can be rewritten more simply as:

$$y = X\beta + \epsilon, \quad (\mathrm{A}.2)$$

where $y = \begin{bmatrix} RV_{23} \\ RV_{24} \\ \vdots \\ \vdots \\ RV_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & J_{22}^{\Delta} & RV_{22} & \overline{RV}_{w,22} & \overline{RV}_{m,22} \\ 1 & J_{23}^{\Delta} & RV_{23} & \overline{RV}_{w,23} & \overline{RV}_{m,23} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & J_{n-1}^{\Delta} & RV_{n-1} & \overline{RV}_{w,n-1} & \overline{RV}_{m,n-1} \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_{\mathrm{d}} \\ \beta_{\mathrm{w}} \\ \beta_{\mathrm{m}} \end{bmatrix}$, and $\epsilon = \begin{bmatrix} \epsilon_{23} \\ \epsilon_{24} \\ \vdots \\ \vdots \\ \epsilon_n \end{bmatrix}$.

The OLS estimator of β ($\hat{\beta}_{OLS}$) is the solution to minimize the sum of squared residuals defined by,

$$(y - X\beta)'(y - X\beta),$$

and the solution of this minimization problem is

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y. \tag{A. 3}$$

The residuals based on above OLS solution of coefficients (e_{OLS}) are defined by,

$$e_{OLS} = y - X\hat{\beta}_{OLS}, \tag{A. 4}$$

The fitted value (or predicted value) of y is defined by,

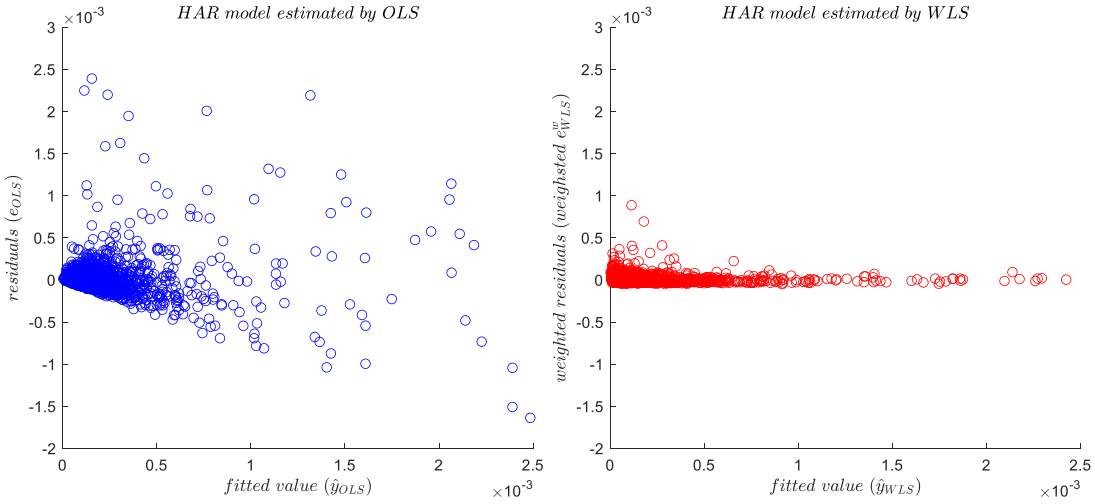
$$\hat{y}_{OLS} = \begin{bmatrix} \widehat{RV}_{23}^{OLS} \\ \widehat{RV}_{24}^{OLS} \\ \vdots \\ \vdots \\ \widehat{RV}_{n}^{OLS} \end{bmatrix} = X\hat{\beta}_{OLS}. \tag{A. 5}$$

One important (Gauss-Markov) assumption of Ordinary Least Square (OLS) is that the conditional variance of the disturbances should be fixed across the sample period (or homoskedasticity). For example, the OLS for the J^{Δ} model assumes the conditional variance of the disturbances of the J^{Δ} model to be fixed across the sample period, that is, $\mathrm{E}(\epsilon\epsilon'|X) = \sigma^2 I$, where σ^2 is the level of conditional variance and is fixed, and I indicates the identity matrix of size $n - 22$. However, as documented in the literature (e.g., Corsi et al., 2008, Patton and Sheppard, 2015), the assumption of homoskedasticity may be not realistic as the variance of the disturbances of the HAR model may vary substantially across the sample period (or heteroskedasticity) and is correlated its lagged value (or autocorrelation). For example, the

disturbances of the HAR model around 2008 financial crisis are generally greater than other tranquil periods (due to volatility changes much more dramatically during this period). To check if heteroskedasticity is present, the left panel of Figure A.1 reports the scatter plot of residuals (e_{OLS}) and fitted value ($\hat{y}_{OLS}$) for the J^{Δ} model estimated by the OLS method for the SPY sample of this chapter. [6] As the panel shows, the spread of the residuals is increasing as the fitted values changes. This implies that the level of the conditional volatility of the disturbances is likely to change dramatically across the sample period and increases with the level of the fitted values.

Figure A.1. The scatter plots for residuals and fitted value for the J^{Δ} model

Notes: The left panel reports the scatter plot of residuals (e_{OLS}) and fitted value ($\hat{y}_{OLS}$) for the J^{Δ} model estimated by the Ordinary Least Square (OLS) method. The residuals (e_{OLS}) and fitted value ($\hat{y}_{OLS}$) are obtained by Equation (A.4) and Equation (A.5), respectively. The right panel reports the scatter plot of weighted residuals (e^{w}_{WLS}) and fitted value ($\hat{y}_{WLS}$) for the J^{Δ} model estimated by the Weighted Least Square (WLS) method. The weighted residuals (e^{w}_{WLS}) and fitted value ($\hat{y}_{WLS}$) are obtained by Equation (A.9) and Equation (A.10), respectively. The data used for computing these residuals and fitted values is SPDR S&P 500 ETF from 1997 to 2021.



Since the level of the conditional volatility of the disturbances possibly changes across the sample period and is consistent with the level of the fitted values, estimation by OLS has the disadvantage that the resulting estimates focus primarily on fitting periods of high RV and place little weight on low RV periods. To overcome this, Weighted Least Squares (WLS) may be a

[6] Here, I only report the results for the J^{Δ} model in Figure A.1. But I confirm that the results for the remaining models lead to similar patterns.

solution. WLS attempts to provide a more efficient alternative to OLS by putting upon less weight to disturbances which are likely to have a large variance and more weight to disturbances which are likely to have a small variance. If each weight w is inversely proportional to the conditional variance of the corresponding disturbances, then the WLS estimator is more efficient than the OLS estimator. Motivated by the positive relationship between the volatility of the residuals and the level of OLS fitted values, Patton and Sheppard (2015) suggest using the inverse of the OLS fitted values as the weights. Specifically, for the J^{Δ} model, first obtain the OLS fitted value $\hat{y}_{OLS}$ from Equation (A.5). Then calculate the weights by the inverse value of each element in $\hat{y}_{OLS}$. The matrix form of the weights is defined by the matrix with the weights on the diagonal and zeroes everywhere else,

$$W = \begin{bmatrix} 1/\widehat{RV}_{23}^{OLS} & 0 & \cdots & 0 \\ 0 & 1/\widehat{RV}_{24}^{OLS} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\widehat{RV}_{n}^{OLS} \end{bmatrix}. \tag{A.6}$$

Finally, based on the weights W, the WLS estimator of β ($\hat{\beta}_{WLS}$) is the solution to the minimization problem of weighted sum of squared residuals,

$$(y - X\beta)'W(y - X\beta),$$

and the solution of this minimization problem is

$$\hat{\beta}_{WLS} = (X'WX)^{-1}X'Wy. \tag{A.7}$$

The residuals (e_{WLS}^{w}) based on above WLS estimated coefficients $\hat{\beta}_{WLS}$ are defined by,

$$e_{WLS} = \alpha W\left(y - X\hat{\beta}_{WLS}\right), \tag{A.8}$$

and the weighted residuals (e_{WLS}^{w}) are defined by,

$$e_{WLS}^{w} = \alpha W\left(y - X\hat{\beta}_{WLS}\right), \tag{A.9}$$

where the component $\alpha = 1/\left[1/(n-22)\sum_{j=1}^{n-22} w_j\right]$ (with w_j indicating the j th diagonal elements of the W matrix) is a scale which ensures that the average weight equals one ($1/(n-22)\sum_{j=1}^{n-22}\{\alpha w_j\} = 1$), and this facilitates comparing the WLS weighted residuals e_{WLS}^{w} with the above OLS (unweighted) residuals e_{OLS}.

The fitted value (or predicted value) of y is defined by,

$$\hat{y}_{WLS} = \begin{bmatrix} \widehat{RV}_{23}^{WLS} \\ \widehat{RV}_{24}^{WLS} \\ \vdots \\ \vdots \\ \widehat{RV}_{n}^{WLS} \end{bmatrix} = X\hat{\beta}_{WLS}. \qquad \text{(A. 10)}$$

To check if heteroskedasticity is reduced for weighted WLS residuals e_{WLS}^{w}, the right panel of Figure A.1 reports the scatter plot of weighted WLS residuals (e_{WLS}^{w}) and fitted value ($\hat{y}_{WLS}$) for the HAR model estimated by the WLS method.[7] Comparing the left panel with the right panel, the deviation of the weighted WLS residuals (e_{WLS}^{w}) is much smaller than that of the OLS residuals (e_{OLS}) as the fitted values changes. This implies that the heteroskedasticity of the weighted disturbances is reduced across the sample period. Motivated by the superiority of such WLS estimation method in terms of treating the heteroskedasticity of disturbances, We use apply this WLS estimation method for estimating all models throughout this chapter.

The population of coefficients β is unknown and the estimate of these coefficients (e.g., $\hat{\beta}_{WLS}$) is used for testing certain hypotheses about the population of coefficients β. To explore the impact of the predictors on the future realized volatility is weak or strong, this paper tests the null hypothesis that the population of coefficient β is zero and alternative hypothesis that the population of coefficient β is not zero. To test these hypotheses, the distribution (type of probability distribution, mean, variance) of $\hat{\beta}_{WLS}$ must be known. To obtain the type of

[7] Here, I only report the results for the J^A model in Figure A.1. But I confirm that the results for the remaining models in this chapter lead to similar patterns.

probability distribution of $\hat{\beta}_{WLS}$, first obtain a formulation of $\hat{\beta}_{WLS}$ by Equation (A.7) and Equation (A.2),

$$\hat{\beta}_{WLS} = \beta + (X'WX)^{-1}X'W\epsilon, \tag{A.11}$$

From Equation (8.11), it can then be seen that the type of probability distribution of $\hat{\beta}_{WLS}$ is consistent with that of disturbances ϵ. As often, the type of probability distribution of disturbances ϵ is assumed to be (multivariate) normal. Therefore, the type of probability distribution of the WLS estimated coefficients $\hat{\beta}_{WLS}$ is also a multivariate normal distribution. From Equation (A.11), it can also be easily calculated that the mean of the distribution of $\hat{\beta}_{WLS}$ ($E(\hat{\beta}_{WLS})$) equals β,

$$E(\hat{\beta}_{WLS}) = \beta \tag{A.12}$$

(Note that $E((X'WX)^{-1}X'W\epsilon) = (X'WX)^{-1}X'WE(\epsilon) = 0$ as one of the Gauss-Markov assumptions states that the disturbances average out to 0 for any values of X). From Equation (A.11), the variance-covariance of the coefficients $\hat{\beta}_{WLS}$ is $E\left((\hat{\beta}_{WLS} - \beta)(\hat{\beta}_{WLS} - \beta)'\right)$ and can be obtained by substituting $\hat{\beta}_{WLS}$ with $\beta + (X'WX)^{-1}X'W\epsilon$,

$$E\left((\hat{\beta}_{WLS} - \beta)(\hat{\beta}_{WLS} - \beta)'\right) = (X'WX)^{-1}X'WE(\epsilon\epsilon')W'X(X'WX)^{-1}. \tag{A.13}$$

The proportion $X'WE(\epsilon\epsilon')W'X$ in the middle of the right-hand side of above Equation (A.13) need to be estimated (Equation (A.13) is also known as the 'sandwich' estimator with two $(X'WX)^{-1}$ as upper and lower breads and $X'WE(\epsilon\epsilon')W'X$ as meat in the middle). To estimate the 'meat' proportion $X'WE(\epsilon\epsilon')W'X$, We use the heteroskedasticity and autocorrelation consistent (HAC) estimator by Newey and West (1987). The HAC estimator is robust to estimation bias due to the heteroskedasticity and autocorrelation of the variance of disturbances. Specifically, the HAC estimator for $X'WE(\epsilon\epsilon')W'X$ is given by,

$$X'WE(\epsilon\epsilon')W' = \frac{n}{n-k}\sum_{j=1}^{n-22}\left(w_j e_{WLS,j}\right)^2 x_j' x_j + \frac{n}{n-k}\sum_{l=1}^{L}\left(1-\frac{l}{L+1}\right)\sum_{j=l+1}^{n-22} w_j w_{j-l} e_{WLS,j} e_{WLS,j-l}\left(x_j' x_{j-l} + x_{j-l}' x_j\right) \quad \text{(A. 14)}$$

where x_j is the jth row of the X matrix, $e_{WLS,j}$ is the jth row of the e_{WLS} vector, w_j is the jth diagonal elements of the W matrix, and k is the number of predictors (including the constant) in the model (e.g., $k = 4$ for the J^{Δ} model), and L is the number of lags for compensating the estimation bias risen by the autocorrelation. Following Corsi and Renò (2012), We set the number of lags L equal to $2(h + 1)$, where h is the lead length (or the forecast horizon) of the dependent variable of the model (e.g., for the J^{Δ} model, $L = 4$ as $h = 1$). By substituting Equation (A.14) to Equation (A.13), the HAC variance-covariance matrix for $\hat{\beta}_{WLS}$ can be obtained. We use notation $\widehat{VCV}_{HAC}$ to indicate this HAC variance-covariance matrix.

In summary, the results of the distribution information on $\hat{\beta}_{WLS}$ show that $\hat{\beta}_{WLS}$ is distributed multivariate normal with mean equal to β and variance-covariance matrix equal to $\widehat{VCV}^{\mathrm{HAC}}$. Then, the test statistics T_{stat} for testing the hypotheses about the coefficients β are calculated by $T_{stat} = \hat{\beta}_{WLS} \oslash \mathrm{diag}\left(\widehat{VCV}_{HAC}\right)$, where "$\oslash$" indicates the Hadamard division (Hadamard division a binary operation that takes in two matrices of the same dimensions and returns a matrix of the divided corresponding elements), "diag" denotes the Matrix-to-vector diag operator (e.g., $\mathrm{diag}\left(\widehat{VCV}^{\mathrm{HAC}}\right)$ returns a vector of the diagonal entries of $\widehat{VCV}^{\mathrm{HAC}}$). For the HAR model, the population of the coefficients β and T_{stat} can be presented by,

$$\beta = \begin{bmatrix}\beta_0 \\ \beta_{\mathrm{d}} \\ \beta_{\mathrm{w}} \\ \beta_{\mathrm{m}}\end{bmatrix} \text{ and } T_{stat} = \begin{bmatrix}T_{stat,0} \\ T_{stat,d} \\ T_{stat,w} \\ T_{stat,m}\end{bmatrix}$$

Based on this formulation of β and T_{stat}, the hypotheses for each of the coefficients in the matrix β can be tested based on the corresponding test statistic in the matrix T_{stat}. For example, one can test the null & alternative hypotheses of β_{d} by the value of $T_{stat,d}$. If $T_{stat,d}$ is larger than 1.96, the false positive (or the probability of mistaken rejection of a null hypothesis) is smaller than 0.05 (two-tail). The probability of false positive is based on the p-value that corresponds to $t = 1.96$ with degree of freedom $\mathrm{df} = n - 22 - k = 6222 - 22 - 4 = 6196$ for the student t-distribution is approximately 0.05 (As often, to account for small sample bias, standard normal distribution is often replaced with the student t-distribution). As the risk of false positive is small, one can reject the null hypothesis that the population of β_{d} is zero thus accept the alternative hypothesis that the population of β_{d} is different from zero.

The goodness of fit of a statistical model describes how well it fits a set of observations. To measure the goodness of fit, this paper uses the coefficient of determination, denoted R^2. R^2 is defined by,

$$R^2 = 1 - \frac{\sum_{j=1}^{n-22}\left(y_j - \hat{y}_{WLS,j}\right)^2}{\sum_{j=1}^{n-22}\left(y_j - \frac{1}{n-22}\sum_{j=1}^{n-22} y_j\right)^2} \qquad \text{(A. 15)}$$

where y_j is the jth row of the y matrix and $\hat{y}_{WLS,j}$ is the jth row of the $\hat{y}_{WLS}$ matrix.