Journal of Financial Economics 109 (2013) 224–249

Contents lists available at SciVerse ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

# The leverage effect puzzle: Disentangling sources of bias at high frequency ☆

Yacine Aït-Sahalia [a,*,1], Jianqing Fan [b,2], Yingying Li [c,3]

[a] Princeton University and NBER, USA
[b] Princeton University, USA
[c] Hong Kong University of Science and Technology, Hong Kong Special Administrative Region

ARTICLE INFO

*Article history:*
Received 12 October 2011
Received in revised form
4 September 2012
Accepted 2 October 2012
Available online 27 February 2013

*JEL classifications:*
G12
C22
C14

*Keywords:*
High frequency data
Leverage effect
Market microstructure noise
Latent volatility
Correlation

ABSTRACT

The leverage effect refers to the generally negative correlation between an asset return and its changes of volatility. A natural estimate consists in using the empirical correlation between the daily returns and the changes of daily volatility estimated from high frequency data. The puzzle lies in the fact that such an intuitively natural estimate yields nearly zero correlation for most assets tested, despite the many economic reasons for expecting the estimated correlation to be negative. To better understand the sources of the puzzle, we analyze the different asymptotic biases that are involved in high frequency estimation of the leverage effect, including biases due to discretization errors, to smoothing errors in estimating spot volatilities, to estimation error, and to market microstructure noise. This decomposition enables us to propose novel bias correction methods for estimating the leverage effect.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

The leverage effect refers to the observed tendency of an asset's volatility to be negatively correlated with the asset's returns. Typically, rising asset prices are accompanied by declining volatility, and vice versa. The term "leverage" refers to one possible economic interpretation of this phenomenon, developed in Black (1976) and Christie (1982): as asset prices decline, companies become mechanically more leveraged since the relative value of their debt rises relative to that of their equity. As a result, it is natural to expect that their stock becomes riskier, hence more volatile. While this is only a hypothesis, this explanation is sufficiently prevalent in the literature that the term "leverage effect" has been adopted to describe the statistical regularity in question. It has also been documented that the effect is generally asymmetric: other things equal, declines in stock prices are accompanied by larger increases in volatility than the decline in volatility that accompanies rising stock markets (see, e.g., Nelson, 1991; Engle and Ng, 1993). Various discrete-time models with a leverage effect have been estimated by Yu (2005).

The magnitude of the effect, however, seems too large to be attributable solely to an increase in financial leverage:

☆ We are very grateful for the comments of the Editor and an anonymous referee.

* Corresponding author.
*E-mail addresses:* yacine@princeton.edu (Y. Aït-Sahalia), jqfan@princeton.edu (J. Fan), yyli@ust.hk (Y. Li).

[1] His research was supported by NSF Grant SES-0850533.
[2] His research was supported by NSF Grants DMS-0714554 and DMS-0704337.
[3] Her research was supported by the Bendheim Center for Finance at Princeton University and the RGC Grants DAG09/10.BM12 and GRF-602710 of the HKSAR.

0304-405X/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jfineco.2013.02.018

Figlewski and Wang (2000) noted among other findings that there is no apparent effect on volatility when leverage changes because of a change in debt or number of shares, only when stock prices change, which questions whether the effect is linked to financial leverage at all. As always, correlation does not imply causality. Alternative economic interpretations have been suggested: an anticipated increase in volatility requires a higher rate of return from the asset, which can only be produced by a fall in the asset price (see, e.g., French, Schwert, and Stambaugh, 1987; Campbell and Hentschel, 1992). The leverage explanation suggests that a negative return should make the firm more levered, hence riskier and therefore lead to higher volatility; the volatility feedback effect is consistent with the same correlation but reverses the causality: increases in volatility lead to future negative returns.

These different interpretations have been investigated and compared (see Bekaert and Wu, 2000), although at the daily and lower frequencies the direction of the causality may be difficult to ascertain since they both appear to be instantaneous at the level of daily data (see Bollerslev, Litvinova, and Tauchen, 2006). Using higher frequency data, namely, five-minute absolute returns, to construct a realized volatility proxy over longer horizons, Bollerslev, Litvinova, and Tauchen (2006) find a negative correlation between the volatility and the current and lagged returns, which lasts for several days, low correlations between the returns and the lagged volatility, and strong correlation between the high frequency returns and their absolute values. Their findings support the dual presence of a prolonged leverage effect at the intradaily level, and an almost instantaneous volatility feedback effect. Differences between the correlation measured using stock-level data and index-level data have been investigated by Duffee (1995). Bollerslev, Sizova, and Tauchen (2012) develop a representative agent model based on recursive preferences in order to generate a volatility process which exhibits clustering, fractional integration, and has a risk premium and a leverage effect.

Whatever the source(s) or explanation(s) for the presence of the leverage effect correlation, there is broad agreement in the literature that the effect should be present. So why is there a puzzle, as suggested by the title of this paper? As we will see, using high frequency data and standard estimation techniques, the data stubbornly refuse to conform to these otherwise appealing explanations. We find that, at high frequency and over short horizons, the estimated correlation $\rho$ between the asset returns and changes in its volatility is close to zero, instead of the strong negative value that we have come to expect. At longer horizons, or especially using option-implied volatilities in place of historical volatilities, the effect is present. If we accept that the true correlation is indeed negative, then this is especially striking since a correlation estimator relies on second moment, or quadratic (co)variation, and quantities like those should be estimated particularly well at high frequency, or instantaneously, using standard probability limit results. We call this disconnection the "leverage effect puzzle," and the purpose of this paper is to examine the reasons for it.

At first read, this behavior of the estimated correlation at high frequency can be reminiscent of the Epps Effect. Starting with Epps (1979), it has indeed been recognized that the empirical correlation between the returns of two assets tends to decrease as the sampling frequency of observation increases. One essential issue that arises in the context of high frequency estimation of the correlation coefficient between two assets is the asynchronicity of their trading, since two assets will generally trade, hence generate high frequency observations, at different times. Asynchronicity of the observations has been shown to have the potential to generate the Epps Effect.[4]

However, the asynchronicity problem is not an issue here since we are focusing on the estimation of the correlation between an asset's returns and its (own) volatility. Because the volatility estimator is constructed from the asset returns themselves, the two sets of observations are by construction synchrone. On the other hand, while asynchronicity is not a concern, one issue that is germane to the problem we consider in this paper is the fact that one of two variables entering the correlation calculation is latent, namely, the volatility of the asset returns. Relative to the Epps Effect, this gives rise to a different set of issues, specifically, the need to employ preliminary estimators or proxies for the volatility variable, such as realized volatility (RV), for example, in order to compute its correlation with asset returns. We will show that the latency of the volatility variable is partly responsible for the observed puzzle.

One further issue, which is in common at high frequency between the estimation of the correlation between two asset returns and the estimation of the correlation between an asset's return and its volatility, is that of market microstructure noise. When sampled at sufficiently high frequency, asset prices tend to incorporate noise that reflects the mechanics of the trading process, such as bid/ask bounces, the different price impact of different types of trades, limited liquidity, or other types of frictions. To address this issue, we will analyze the effect of using noise-robust high frequency volatility estimators for the purpose of estimating the leverage effect.[5]

[4] As a result, various data synchronization methods have been developed to address this issue: for instance, Hayashi and Yoshida (2005) have proposed a modification of the realized covariance which corrects for this effect; see also Large (2007), Griffin and Oomen (2008), Voev and Lunde (2007), Zhang (2011), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011), Kinnebrock and Podolskij (2008), and Aït-Sahalia, Fan, and Xiu (2010).

[5] In the univariate volatility case, many estimators have been developed to produce consistent estimators despite the presence of the noise. These include the Maximum-Likelihood Estimator (MLE) of Aït-Sahalia, Mykland, and Zhang (2005), shown to be robust to stochastic volatility by Xiu (2010), Two Scales Realized Volatility (TSRV) of Zhang, Mykland, Aït-Sahalia (2005), Multi-Scale Realized Volatility (MSRV), a modification of TSRV which achieves the best possible rate of convergence proposed by Zhang (2006), Realized Kernels (RK) by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), and the Pre-Averaging volatility (PAV) estimator by Jacod, Li, Mykland, Podolskij, and Vetter (2009). Related works include Bandi and Russell (2006), Delattre and Jacod (1997), Fan and Wang (2007), Gatheral and Oomen (2010), Hansen and Lunde (2006), Kalnina and Linton (2008), Li and Mykland (2007), Aït-Sahalia, Mykland, and Zhang (2011), and Li, Mykland, Renault, Zhang, and Zheng (2009). To estimate the correlation between two assets, or any two variables that are observable, Zhang (2011) proposed a consistent Two Scales Realized Covariance estimator (TSCV),

Related studies include the development of nonparametric estimators of the covariance between asset returns and changes in volatility in Bandi and Renò (2012) and Wang and Mykland (2009). Both papers propose nonparametric estimators of the leverage effect and develop the asymptotic theory for their respective estimators; our focus by contrast is on understanding the source of, and quantifying, the bias(es) that result from employing what is otherwise a natural approach to estimate that correlation.

Our main results are the following. We provide theoretical calculations that disentangle the biases involved in estimating the correlation between returns and changes in volatility, when a sequence of progressively more realistic estimators is employed. We proceed incrementally, in such a way that we can isolate the sources of the bias one by one. Starting with the spot volatility, an ideal but unavailable estimator since volatility is unobservable, we will see that the leverage effect parameter $\rho$ is already estimated with a bias that is due solely to discretization. This bias is small when the discretization step is small, but we will soon see that the optimal discretization step is not small when more realistic measures of volatilities are used. The unobservable spot volatility is frequently estimated by a local time-domain smoothing method which involves integrating the spot volatility over time, locally. Replacing the spot volatility by the (also unavailable) true integrated volatility, the bias for estimating $\rho$ is very large, but remains quantifiable. The incremental bias is due to smoothing. Replacing the true integrated volatility by an estimated integrated volatility, the bias for estimating $\rho$ becomes so large that, when calibrated on realistic parameter values, the estimated $\rho$ becomes essentially zero, which is indeed what we find empirically. The incremental bias represents the effect of the estimation error. We then examine the effect of using noise-robust estimators of the integrated volatility, and compute the resulting additional bias term, which can make the estimated leverage effect to go in the reverse direction. Based on the above results, we propose a regression approach to compute bias-corrected estimators of $\rho$. We mainly investigate these effects in the context of the Heston stochastic volatility model, which has the advantage of providing explicit expressions for all these bias terms. The effect of a jump component in the price process is also further analyzed.

The paper is organized as follows. Section 2 documents the presence of the leverage effect puzzle. The prototypical model for understanding the puzzle and nonparametric estimators for spot volatility are described in Section 3. Section 4 presents the main results of the paper, which unveil the biases of estimating the leverage effect parameter in all steps of approximations. Section 5 analyzes the role that price jumps can play when measuring the leverage effect. A possible solution to the puzzle is proposed in Section 6. Section 7 demonstrates the leverage effect puzzle, the effectiveness of the proposed solution, and the robustness to alternative models by Monte Carlo simulations. Section 8 presents empirical studies based on high frequency data from Standard and Poor's 500 Index (S&P 500) and Microsoft. Section 9 concludes. The Appendix contains the mathematical proofs.

## 2. Motivation: the leverage effect puzzle

To motivate the theoretical analysis that follows, we start with a straightforward empirical exercise to illustrate the leverage effect puzzle. Fig. 1 presents the time series of the log-returns of the S&P 500 index and the index itself. Large volatility periods accompanied by a decline of the index are symptomatic of the leverage effect.

Slightly more formally, a scatter plot of estimated changes of volatilities and returns provides a simple way to examine graphically the relationship between estimated changes in volatility and changes in log-prices (i.e., log-returns). Fig. 2 shows scatter plots of the differences of estimated daily volatilities $\hat{V}_t - \hat{V}_{t-m}$ against the corresponding returns of horizon $m$ days for several assets, where $\hat{V}_t$ is the integrated daily volatility estimated by the noise-robust TSRV estimator. If we start with long horizons, as shown in Fig. 2, we see that the effect is present albeit seriously underestimated in the data.

In addition to the evidence that comes from long horizons, the effect is even stronger empirically if we use a different measurement altogether of the asset volatility, based on market prices of derivatives. In the case of the S&P 500 index, we employ VIX, which is the square-root of the par variance swap rate with 30 days to maturity; that is, VIX measures the square-root of the risk-neutral expectation of the S&P 500 variance over the next 30 calendar days. Using this market-based volatility measure, the leverage effect is indeed very strong as demonstrated in Fig. 3.[6]

Yet, starting at the daily horizon, even when using high frequency volatility estimates, we see in Fig. 4 that the scatter plot of $\hat{D}_t = \hat{V}_t - \hat{V}_{t-1}$ against daily returns $R_t$ shows no apparent leverage effect for the different assets considered. As discussed in the Introduction, different economic explanations provide for different causation between returns and their volatility. To be robust against

(*footnote continued*)
Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) a Multivariate Realized Kernel (MRK), Kinnebrock and Podolskij (2008) a multivariate Pre-Averaging estimator, and Aït-Sahalia, Fan, and Xiu (2010) a multivariate Quasi-Maximum Likelihood Estimator (QMLE).

[6] VIX is subject to a risk premium which might make the observed correlation between VIX changes and asset returns even more negative than what results from the leverage effect, if the risk premium happens to increase when prices go down. So the point of Fig. 2 is not to suggest that VIX provides a solution to the measurement problem, but simply to investigate the magnitude of the effect when employing a data source for volatility that is not directly obtained from the price data itself. The leverage effect can be measured using options data by estimating a parametric model which also takes into account the risk premia: see, for example, Pan (2002) who estimates $\rho$ near $-0.5$, including jumps in prices, Bakshi, Cao, and Chen (1997) who estimate $\rho$ in the range $[-0.6,-0.8]$ without jumps, and around $-0.5$ when including jumps, and the likelihood-based estimates in Aït-Sahalia and Kimmel (2007) near $-0.75$ without jumps but across a wide range of parametric stochastic volatility specifications.

the timing differences that different causality explanations would generate, we next examine scatter plots of different time lags and leads such as $\{(\hat{D}_{t-1},R_t)\}$ and $\{(\hat{D}_t,R_{t-1})\}$. The evidence again reveals no leverage effect. Similar results are obtained if we employ different time periods and/or different noise-robust volatility estimators such as QMLE or PAV.

There are sound economic rationales to support a prior that a leverage effect is present in the data, and we do indeed find it in Figs. 1 and 2. So why are we unable to detect it on short horizons based on high frequency volatility estimates that should provide precise volatility proxies? This is the nature of the "leverage effect puzzle" that we seek to understand. Can it be the result of employing estimators that are natural at high frequency for the latent volatility variable, but somehow result in biasing the estimated correlation all the way down to zero? Why does this happen? The goal of this paper is to understand the sources of the puzzle and propose a solution.

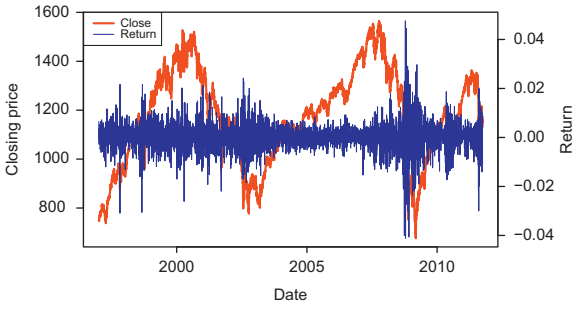


**Fig. 1.** S&P 500 index and returns. S&P 500 daily closing values plotted together with daily returns (January 1997–September 2011). A strong leverage effect is seen.

## 3. Data generating process and estimators

In order to study the leverage effect puzzle, we need two ingredients: nonparametric volatility estimators that are applicable at high frequency, and data generating processes for the log-returns and their volatility in the form of a stochastic volatility model. Employing a specific stochastic volatility model has the advantage that the properties of nonparametric estimators of the correlation between asset returns and their volatility become fully explicit. We can derive theoretically the asymptotic biases of different nonparametric estimators applied to this model, and verify their practical relevance via small-sample simulation experiments. Putting together, these ingredients lead to a solution to the leverage puzzle by

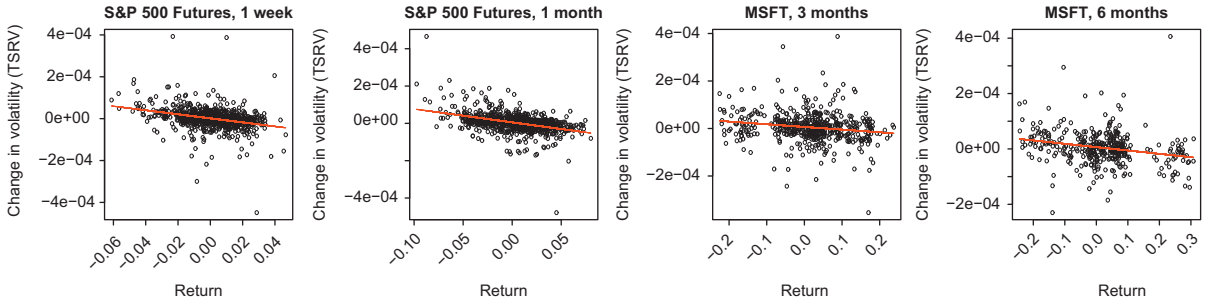


**Fig. 2.** Scatter plots of differences of estimated daily volatility $\hat{V}_t-\hat{V}_{t-m}$ versus returns over relatively long time span $m$ for S&P 500 futures data from January 2004 to December 2007 and Microsoft (MSFT) data from January 2005 to June 2007. Daily volatilities are estimated using TSRV based on high frequency minute-by-minute observations, and returns are calculated based on daily closing prices. From left to right: S&P 500 futures when time horizon $m$ is taken to be five days (a week), S&P 500 futures when time horizon $m$ is taken to be 21 days (a month), MSFT when time horizon $m$ is taken to be 63 days (three months), MSFT when time horizon $m$ is taken to be 126 days (six months). Solid lines are the least squares regression lines.

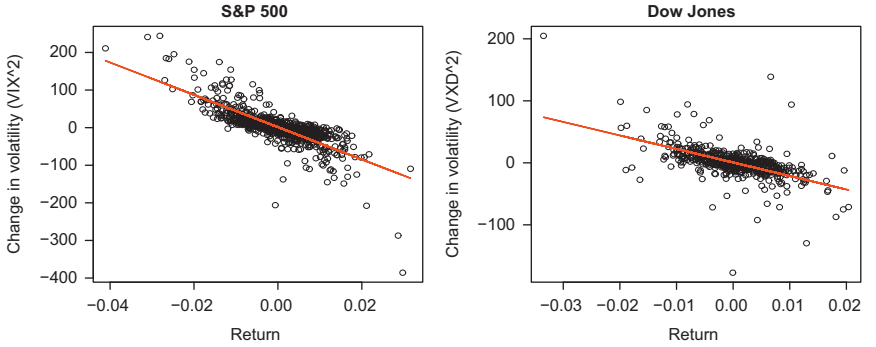


**Fig. 3.** Daily changes of squared volatility indices versus daily returns. Using the volatility indices as the proxy of volatility, the leverage effect can clearly be seen. Left: S&P 500 data from January 2004 to December 2007, in which the VIX is used as a proxy of the volatility; Right: Dow Jones Industrial Average data from January 2005 to March 2007 in which the Chicago Board Options Exchange (CBOE) DJIA Volatility Index (VXD) is used as the volatility measure.

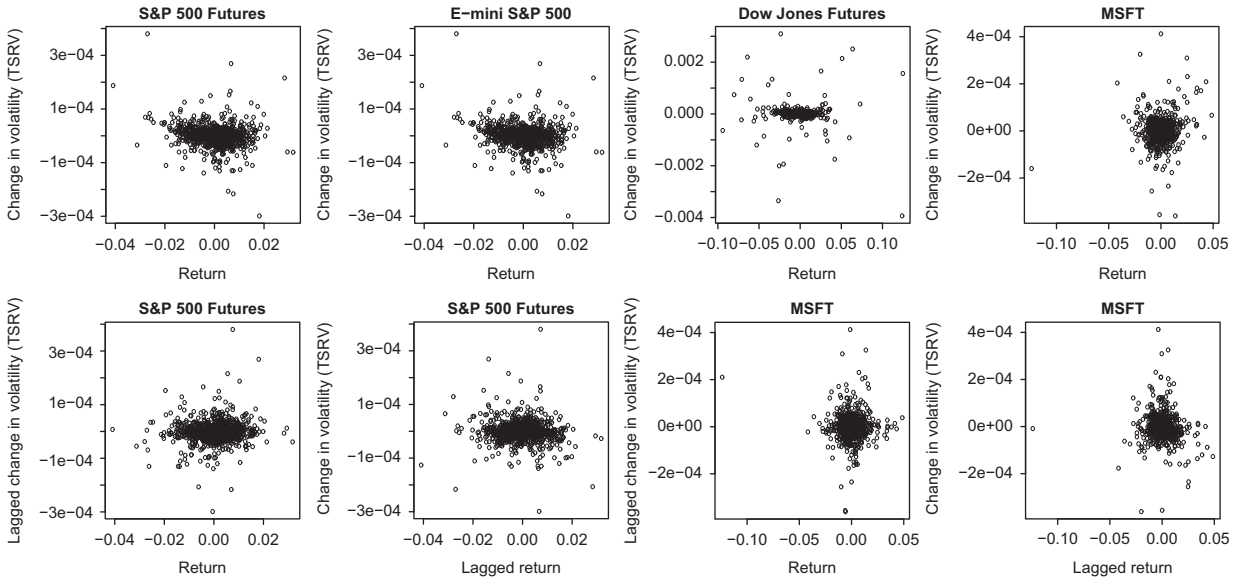


**Fig. 4.** Scatter plots of the changes of estimated daily volatilities versus daily returns, including leads and lags. Daily volatilities are estimated using TSRV based on high frequency minute-by-minute observations, and returns are calculated based on daily closing prices. Upper panel from left to right: S&P 500 futures data in 2004-2007, E-mini S&P 500 data in 2004-2007, Dow Jones futures data from January 2005 to March 2007, Microsoft data between January 2005 and June 2007. Lower: Scatter plots of differences of estimated daily volatility versus daily returns with leads and lags for S&P 500 futures 2004–2007 data (left two) and Microsoft data from January 2005 to June 2007 (right two).

introducing a tuning parameter (represented by $m$ below) that attempts to minimize the estimation bias. Of course, this solution assumes the constraints implicit in the estimation of the leverage effect in practice: in keeping with the spirit of the analysis of the paper, we attempt to fix the estimator employed in practice, rather than design a better estimator from scratch. In Sections 5 and 7, we discuss generalizations of the model considered here; in particular, we study what happens when jumps are present.

### 3.1. Stochastic volatility model

The specification we start with for this purpose is the stochastic volatility model of Heston (1993) for the log-price dynamics:

$$dX_t = (\mu - v_t/2)\,dt + v_t^{1/2}\,dB_t, \tag{1}$$

$$dv_t = \kappa(\alpha - v_t)\,dt + \gamma v_t^{1/2}\,dW_t, \tag{2}$$

where $B$ and $W$ are two standard Brownian motions with $E(dB_t\,dW_t) = \rho\,dt$, and the parameters $\mu$, $\alpha$, $\kappa$, $\gamma$, and $\rho$ are constants. We assume that the initial variance $v_0 > 0$ is a realization from the stationary (invariant) distribution of (2) so that $v_t$ is a stationary process. Under Feller's condition $2\kappa\alpha > \gamma^2$, the process $v_t$ stays positive, a condition that is always assumed in what follows. Note that

$$\rho = \lim_{s\to 0} \mathrm{Corr}(v_{t+s} - v_t, X_{t+s} - X_t) \tag{3}$$

so that the leverage effect is summarized by the parameter $\rho$ under the Heston model (1) and (2). We use $\sigma_t$ to denote $v_t^{1/2}$ in the following. In Sections 5 and 7, we will add jumps to the model to investigate their impact on the estimation of $\rho$.

Throughout the paper, we refer to the correlation (3) between changes in volatility and changes in asset log-prices, i.e., returns, as the "leverage effect." Other papers define it as the correlation between the level of volatility and returns, or the correlation between the level of absolute returns and returns (see, e.g., Bollerslev, Litvinova, and Tauchen, 2006.) The latter definition, however, would not predict that the parameter $\rho$ should be identified as the high frequency limit of that correlation; while that alternative definition is appropriate at lower frequencies, it yields a degenerate high frequency limit since it measures the correlation between two variables that are of different orders of magnitude in that limit. High frequency data can be employed to estimate the correlation between volatility levels and returns, but only over longer horizons, as it is indeed employed in Bollerslev, Litvinova, and Tauchen (2006).

We consider a different problem: the nature of the "leverage effect puzzle" we identify lies in the fact that it is difficult to translate the otherwise straightforward short horizon/high frequency limit (3) into a meaningful estimate of the parameter $\rho$.

### 3.2. Nonparametric estimation of volatility and sampling

Our first task will be to understand why natural approaches to estimate $\rho$ based on (3) do not yield a good estimator when nonparametric estimates of volatility based on high frequency data are employed. With a small time horizon $\Delta$ (e.g., one day or $\Delta = 1/252$ year), let

$$V_{t,\Delta} = \int_{t-\Delta}^{t} v_s\,ds \tag{4}$$

denote the integrated volatility from time $t-\Delta$ to $t$ and $\hat{V}_{t,\Delta}$ be an estimate of it based on the discretely observed

log-price process $X_t$, which additionally may be contaminated with the market microstructure noise. Recall that the quantity of interest is $\rho$ and is based on (3). However, the spot volatility process $v_t$ is not directly observable and has to be estimated by $\Delta^{-1}\hat{V}_{t,\Delta}$. Thus, corresponding to a given estimator $\hat{V}$, a natural and feasible estimator of $\rho$ is

$$\hat{\rho} = \mathrm{Corr}(\hat{V}_{t+s,\Delta} - \hat{V}_{t,\Delta}, X_{t+s} - X_t). \tag{5}$$

With $s = \Delta$, $\hat{V}_{t+s,\Delta}$ and $\hat{V}_{t,\Delta}$ are estimators of integrated volatilities over consecutive intervals. This is a natural choice for parameter $s$: changes of daily estimated integrated volatility are correlated with changes of daily prices in two consecutive days. However, as to be demonstrated later, the choice of $s = m\Delta$ (changes over multiple days apart) can be more advantageous.

We now specify the different nonparametric estimators of the integrated volatility that will be used for $\hat{V}_{t,\Delta}$. We assume that the log-price process $X_t$ is observed at higher frequency, corresponding to a time interval $\delta$ (e.g., one observation every ten seconds). In order for the nonparametric estimate $\hat{V}_{t,\Delta}$ to be sufficiently accurate, we need $\delta \ll \Delta$; asymptotically, we assume that $\Delta \to 0$ and $\delta \to 0$ in such a way that $\Delta/\delta \to \infty$.

In the absence of microstructure noise, the log-prices $X_{t-\Delta+i\delta}$ $(i = 0,1,\ldots,\Delta/\delta)$ are directly observable, and the most natural (and asymptotically optimal) estimator of $V_{t,\Delta}$ is the realized volatility

$$\hat{V}^{\mathrm{RV}}_{t,\Delta} = \sum_{i=0}^{\Delta/\delta-1} (X_{t-\Delta+(i+1)\delta} - X_{t-\Delta+i\delta})^2. \tag{6}$$

Here, for simplicity of exposition, we assume there is an observation at time $t-\Delta$, and that the ratio $\Delta/\delta$ is an integer (denoted by $n$ throughout the paper); otherwise $\Delta/\delta$ should be replaced by its integer part $[\Delta/\delta]$, without any asymptotic consequences.

In practice, high frequency observations of log-prices are likely to be contaminated with market microstructure noise. Instead of observing the log-prices $X_{t-\Delta+i\delta}$, we observe the noisy version

$$Z_{t-\Delta+i\delta} = X_{t-\Delta+i\delta} + \epsilon_{t-\Delta+i\delta}, \tag{7}$$

where the $\epsilon_{t-\Delta+i\delta}$'s are white noise random variables with mean zero and standard deviation $\sigma_\epsilon$. With this type of observation, we can use noise-robust methods to obtain consistent estimates of the integrated volatility. We will primarily employ PAV as it is one of the rate-efficient estimators. Results for TSRV are available in the Appendix for comparison purposes. The PAV estimator with weight function chosen as $g(x) = x \wedge (1-x)$ is defined as follows: with $\theta_{\mathrm{PAV}}$ a constant and $k_n = [\theta_{\mathrm{PAV}}\sqrt{n}]$ the window length over which the averaging takes place, let

$$\begin{aligned}\hat{V}^{\mathrm{PAV}}_{t,\Delta} \coloneqq \frac{12}{\theta_{\mathrm{PAV}}\sqrt{n}} \sum_{i=0}^{n-k_n+1} \Bigg( \frac{1}{k_n} \sum_{j=\lfloor k_n/2 \rfloor}^{k_n-1} Z_{t-\Delta+(i+j)\delta} \\ - \frac{1}{k_n} \sum_{j=0}^{\lfloor k_n/2 \rfloor - 1} Z_{t-\Delta+(i+j)\delta} \Bigg)^2 \\ - \frac{6}{\theta^2_{\mathrm{PAV}} n} \sum_{i=0}^{n-1} (Z_{t-\Delta+(i+1)\delta} - Z_{t-\Delta+i\delta})^2.\end{aligned} \tag{8}$$

A consistent estimator of the variance of this estimator is available, as well as a consistent estimator of the integrated quarticity $\int_{t-\Delta}^{t} \sigma_s^4\, ds$ (see (27) and (62)).

## 4. Biases in estimation of the leverage effect

We now present the results of the paper, consisting of the biases of estimators of the leverage effect parameter $\rho$ in four progressively more realistic scenarios, each employing a different nonparametric volatility estimator. We stress again that our purpose in analyzing these estimators is to match the empirical practice. These progressive scenarios help us document an incremental source for the bias: discretization, smoothing, estimation error, and market microstructure noise. The results in this section are based on the model (1) and (2): we apply nonparametric estimators, but study their properties when they are applied to a specific parametric model.

### 4.1. True spot volatility: discretization bias

First, we consider the unrealistic but idealized situation in which the spot volatility process $v_s$ is in fact directly observable. This helps us understand the error in estimating $\rho$ that is due to discretization alone. Theorem 1 reports the correlation between asset returns and changes of the instantaneous volatility, from which the bias can easily be computed.

*Theorem 1. Changes of the true spot volatility and changes of log-prices have the following correlation*:

$$\mathrm{Corr}(v_{s+t} - v_t, X_{s+t} - X_t) = \frac{\rho\sqrt{\frac{1-e^{-\kappa s}}{\kappa}}}{\sqrt{\left(s + \frac{e^{-\kappa s}-1}{\kappa}\right)\left(\frac{\gamma^2}{4\kappa^2} - \frac{\gamma\rho}{\kappa}\right) + s}}. \tag{9}$$

Let us denote the right-hand side of the expression in (9) as $C_1(s,\kappa,\gamma,\alpha,\rho)$. From Theorem 1, the bias due to the discrete approximation can be easily computed, in the form $C_1(s,\kappa,\gamma,\alpha,\rho) - \rho$. In particular, we have the following proposition expressing the bias as a function of the integration interval $\Delta$ and the interval length over which changes are evaluated, $m\Delta$, $m \geq 1$, under different asymptotic assumptions on the sampling scheme:

*Proposition 1. When $m\Delta \to 0$ (either with m fixed and $\Delta \to 0$, or $m \to \infty$ and $m\Delta \to 0$), we have*

$$\begin{aligned}&\mathrm{Corr}(v_{t+m\Delta} - v_t, X_{t+m\Delta} - X_t) \\ &\quad = \rho - \frac{\rho(\gamma^2 - 4\gamma\kappa\rho + 4\kappa^2)}{16\kappa} m\Delta + o(m\Delta).\end{aligned} \tag{10}$$

Since the value $\rho$ is negative, the first order of the bias is positive, which pulls the function $C_1(s,\kappa,\gamma,\alpha,\rho)$ towards zero, weakening the leverage effect. Fig. 5 shows how the function $C_1(m\Delta,\kappa,\gamma,\alpha,\rho)$ varies with $m$ for two sets of parameter values: $(\rho,\kappa,\gamma,\alpha,\mu) = (-0.8,5,0.5,\ 0.1,0.05)$ and $(\rho,\kappa,\gamma,\alpha,\mu) = (-0.3,5,0.05,0.04,0.02)$ when $\Delta$ is taken to be $1/252$ (one day). The former set of parameters was adapted from those in Aït-Sahalia and Kimmel (2007) and

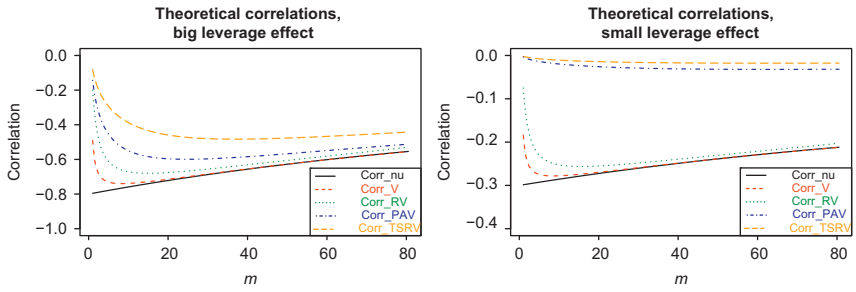


**Fig. 5.** Theoretical estimated leverage effect parameter $\rho$ as a function of the tuning parameter $m$ when $\Delta$ is taken to be $1/252$; using the spot volatility (`Corr_nu`), ideally estimated spot volatility (`Corr_V`), realized volatility estimator (`Corr_RV`), pre-averaging volatility estimator (`Corr_PAV`), and two-time scale volatility estimator (`Corr_TSRV`), respectively. They correspond, respectively, to the function $C_1(m\Delta,\kappa,\gamma,\alpha,\rho)$ in Theorem 1, $A_2/(B_2C_2)$ in Theorem 2, and the main terms in Theorems 3, 4, and 6, respectively (we used $C_m = m^2\Delta$ for the sake of accuracy. $C_m$=0 gives very similar graphs). Two sets of parameter values are considered. Left panel: $(\rho,\kappa,\gamma,\alpha,\mu) = (-0.8,5,0.5,0.1,0.05)$; right panel: $(\rho,\kappa,\gamma,\alpha,\mu) = (-0.3,5,0.05,0.04,0.02)$.

the latter set was taken to weaken the leverage effect but still maintain Feller's condition: $2\kappa\alpha > \gamma^2$. As expected, the smaller the $m$, the smaller the discretization bias.

### 4.2. True integrated volatility: smoothing bias

The spot volatilities are latent. They can be (and usually are) estimated by a local average of integrated volatility, which is basically a smoothing operation, over a small time horizon $\Delta$. How big are the biases for estimating $\rho$ even in the idealized situation where the integrated volatility is known precisely? The following theorem gives an analytic expression for the resulting smoothing bias:

*Theorem 2. Changes of the true integrated volatility and changes of log-prices have the following correlation*:

$$\mathrm{Corr}(V_{t+m\Delta,\Delta}-V_{t,\Delta},X_{t+m\Delta}-X_t) = A_2/(B_2C_2), \tag{11}$$

*where*

$$A_2 = 2\gamma(1-\Delta\kappa)+4\Delta\kappa^2\rho-2\gamma e^{-\Delta\kappa} + e^{-\Delta\kappa(m+1)}(e^{2\Delta\kappa}(\gamma-4\kappa\rho)-2e^{\Delta\kappa}(\gamma-2\kappa\rho)+\gamma),$$

$$B_2 = 2\sqrt{e^{-\Delta\kappa(m+1)}(2e^{\Delta\kappa m}-(e^{\Delta\kappa}-1)^2)+2\Delta\kappa-2},$$

*and*

$$C_2 = \sqrt{\gamma^2(\Delta\kappa m+e^{-\Delta\kappa m}-1)+4\gamma\kappa\rho(-\Delta\kappa m-e^{-\Delta\kappa m}+1)+4\Delta\kappa^3 m}.$$

While the expressions in Theorem 2 are exact, further insights can be gained when we consider the resulting asymptotic expansion as $\Delta\to 0$. We consider both situations where $m$ is fixed and $m\to\infty$ while still $m\Delta\to 0$.

*Proposition 2. The following asymptotic expansions show the incremental bias due to smoothing induced by the local integration of spot volatilities*:

$$\begin{aligned}&\mathrm{Corr}(V_{t+m\Delta,\Delta}-V_{t,\Delta},X_{t+m\Delta}-X_t)\\ &\quad= \mathrm{Corr}(v_{t+m\Delta}-v_t,X_{t+m\Delta}-X_t)\frac{(2m-1)}{2\sqrt{m^2-m/3}}\\ &\qquad+\begin{cases}\mathrm{O}(\Delta) & \text{when } \Delta\to 0 \text{ for any } m,\\ \mathrm{o}(m\Delta) & \text{when } m\to\infty \text{ and } m\Delta\to 0.\end{cases}\end{aligned} \tag{12}$$

The first factor on the right-hand side of (12) is the same as if the true spot volatility were observable. For the second factor, it is the asymptotic bias, which is $\sqrt{3/8}\approx 0.612$ when $m=1$. Note that the asymptote of the bias factor

$$\frac{(2m-1)}{2\sqrt{m^2-m/3}} = 1+\mathrm{O}\left(\frac{1}{m}\right) \quad \text{when } m \text{ is large.} \tag{13}$$

Hence, when $m$ is large, the bias of estimated $\rho$ based on integrated volatilities is asymptotically the same as that of the estimated $\rho$ based on spot volatilities.

Fig. 5 shows the resulting numerical values for the same sets of parameter values considered above. They are plotted along with the correlations of the other estimators to facilitate comparisons. First, as expected, the bias is larger than that when spot volatilities are available. Fig. 5 also reveals an interesting shape of biases of the idealized estimate of spot volatility. When $m$ is small, the bias is large and so is when $m$ is large. There is an optimal choice of $m$ that minimizes the bias. For the case $\Delta = 1/252$, with the chosen parameters as in the left panel of Fig. 5 [$(\rho,\kappa,\gamma,\alpha,\mu) = (-0.8,5,0.5,0.1,0.05)$], the optimum is $m_0 = 8$ with the optimal value $-0.74$, leading to a bias of 0.06. On the other hand, using the natural choice $m=1$, the estimated correlation is about $-0.5$, meaning that the bias is about 40% of the true value.

### 4.3. Estimated integrated volatility: shrinkage bias due to estimation error

Theorems 1 and 2 provide a partial solution to the puzzle. If the spot volatility were observable, the ideal estimate of leverage effect is to use the change of volatility over two consecutive intervals against the changes of the prices over the same time interval, i.e., $m=1$. However, when the spot volatility has to be estimated, even with the ideally estimated integrated volatility $V_{t,\Delta}$, the choice of $m=1$ is far from being optimal. Indeed, the resulting bias is quite large: for $\rho=-0.8$, with the same set of parameters as above, the estimated $\rho$ is about $-0.5$ even when employing the idealized true integrated volatility $V_{t,\Delta}$. When the sample version of integrated volatility is used, we should expect that the leverage effect is further masked by estimation error. This is due to the well-known shrinkage bias of computing correlation when variables are

measured with errors. In fact, we already know that it becomes so large that it masks completely the leverage effect when $m=1$ is used as in Fig. 4. We now derive the theoretical bias expressions corresponding to this more realistic case.

The following theorem calculates the bias of using a data-driven estimator of the integrated volatility in the absence of microstructure noise. In other words, we use the realized volatility estimator. As introduced in Section 3.2, we use $n$ to denote the number of observations during each interval $\Delta$, and assume for simplicity that the observation intervals are equally spaced at a distance $\delta=\Delta/n$.

*Theorem 3. When $\Delta\to 0$ and $n\Delta\to C\in(0,\infty)$, the following expansion shows the incremental bias due to estimation error induced by the use of RV*:

$$\begin{aligned}&\mathrm{Corr}(\hat{V}^{RV}_{t+m\Delta,\Delta}-\hat{V}^{RV}_{t,\Delta},X_{t+m\Delta}-X_t)\\&=\mathrm{Corr}(v_{t+m\Delta}-v_t,X_{t+m\Delta}-X_t)\frac{(2m-1)}{2\sqrt{m^2-\frac{m}{3}}}\\&\times\left(1+\frac{12\alpha\kappa+6\gamma^2}{(3\gamma^2m-\gamma^2)\kappa C-\frac{3}{2}\gamma^2\kappa^2CC_m}\right)^{-1/2}+R^{RV}_r,\end{aligned}\tag{14}$$

*where $R^{RV}_r$ is $O(\Delta)$ for any fixed $m$, with $C_m$ above replaced by* 0; *and* $o(m\Delta)$ *when* $m^2\Delta\to C_m\in(0,\infty)$.

The above theorem documents the bias when there is no market microstructure noise. Interestingly, it is decomposed into two factors. The first factor is the smoothing bias and the second factor is the shrinkage bias due to the estimation errors. The second factor reflects the cost of estimating the latency of volatility process. The larger the $C$, the smaller the shrinkage bias. Similarly, within a reasonable range such that $m\Delta$ is not too big, the larger the $m$, the smaller the shrinkage bias.

To appreciate the bias due to the use of RV, the main term in Theorem 3 as a function of $m$ is depicted in Fig. 5 for the same sets of the aforementioned parameters. The daily sampling frequency is taken to be $n=390$ (one observation per minute) so that $C=390/252$. The choice of $m=1$ corresponds to the natural estimator but it results in a very large bias.

Even in the absence of market microstructure noise, the estimated correlation based on the natural estimator

$$\hat{\rho}^{\mathrm{RV}}=\mathrm{Corr}(\hat{V}^{\mathrm{RV}}_{t+\Delta,\Delta}-\hat{V}^{\mathrm{RV}}_{t,\Delta},X_{t+\Delta}-X_t)\tag{15}$$

is very close to zero. This provides a mathematical explanation for why the leverage effect cannot be detected empirically using a natural approach. On the other hand, Theorem 3 also hints at a solution to the leverage effect puzzle: with an appropriate choice of $m$, there is hope to make the leverage effect detectable. For the left panel of Fig. 5, if the optimal $m=15$ is used, the estimated correlation is about $-0.68$, when the true value is $-0.8$.

### 4.4. Estimated noise-robust integrated volatility: shrinkage bias due to estimation error and noise correction error

Under the more realistic case where the presence of market microstructure noise under (7) is recognized, the integrated volatility $V_t$ is estimated based on noisy log-returns, using bias-corrected high frequency volatility estimators such as TSRV or PAV. In this case, as we will see, detecting the leverage effect based on the natural estimator is even harder. It may in fact even result in an estimated correlation coefficient with the wrong sign. Again, the tuning parameter $m$ can help with the issue.

We consider the PAV estimator as defined in (8), the corresponding result for TSRV can be found in the Appendix.

*Theorem 4. When $\Delta\to 0$, $n^{1/2}\Delta\to C_{\mathrm{PAV}}$, and $\sigma^2_\epsilon/\Delta\to C_\epsilon$ with $C_{\mathrm{PAV}}$ and $C_\epsilon\in(0,\infty)$, the following expansion shows the incremental bias due to estimation error and noise correction induced by the use of PAV*:

$$\begin{aligned}&\mathrm{Corr}(\hat{V}^{\mathrm{PAV}}_{t+m\Delta,\Delta}-\hat{V}^{\mathrm{PAV}}_{t,\Delta},Z_{t+m\Delta}-Z_t)\\&=\mathrm{Corr}(v_{t+m\Delta}-v_t,X_{t+m\Delta}-X_t)\frac{(2m-1)}{2\sqrt{m^2-m/3}}\\&\times(1+A_4+B_4+C_4)^{-1/2}+R^{\mathrm{PAV}}_r,\end{aligned}\tag{16}$$

*where $R^{\mathrm{PAV}}_r$ is $O(\Delta)$ for any fixed $m$, with $C_m$ below replaced by* 0; *and* $o(m\Delta)$ *when* $m^2\Delta\to C_m\in(0,\infty)$, *and*

$$A_4=\frac{24\Phi_{22}\theta_{\mathrm{PAV}}(2\alpha\kappa+\gamma^2)}{\psi_2^2C_{\mathrm{PAV}}\kappa\gamma^2(6m-2-3\kappa C_m)},$$

$$B_4=\frac{96\Phi_{12}C_\epsilon}{\theta_{\mathrm{PAV}}\psi_2^2C_{\mathrm{PAV}}\gamma^2(6m-2-3\kappa C_m)},$$

$$C_4=\frac{48\Phi_{11}C_\epsilon^2}{\theta^3_{\mathrm{PAV}}\psi_2^2C_{\mathrm{PAV}}\alpha\gamma^2(6m-2-3\kappa C_m)},$$

*where* $\psi_2=\frac{1}{12},\Phi_{11}=\frac{1}{6},\Phi_{12}=\frac{1}{96},\Phi_{22}=\frac{151}{80\,640}$.

For the same reasons behind the above theorem, using the parameter $m$ helps in resolving the leverage effect problems. When $\theta_{\mathrm{PAV}}$ is taken to be 0.5, with $m=1$ and the same set of parameters $(\rho,\kappa,\gamma,\alpha,\mu,\Delta,n)=(-0.8,5,0.5,0.1,0.05,1/252,390)$, the leverage effect is barely noticeable whereas using $m=25$ yields a correlation of about $-0.60$. Even though the bias is large, the leverage effect is clearly noticeable.

Again, the estimation biases can be decomposed into two factors. The first factor is the smoothing bias, the same as that in the RV. The second factor reflects the shrinkage biases due to estimation errors and noise correction errors. The rate of convergence of PAV is slower than that of RV (but RV is biased). This is reflected in the factor $C_{\mathrm{PAV}}$ which is of order $n^{1/2}\Delta$, rather than $C=n\Delta$ in RV.

Theorem 4 and a parallel result to Theorem 4 for TSRV (see Theorem 6 in the Appendix) are illustrated in Fig. 5 in which the main terms of the correlations are graphed. For TSRV, when $\theta_{\mathrm{TSRV}}$ is taken to be 0.5 (see details about the TSRV setting and notation in the Appendix), with $m=1$ and the same set of parameters $(\rho,\kappa,\gamma,\alpha,\Delta,n)=(-0.8,5,0.5,0.1,1/252,390)$, the leverage effect is nearly zero whereas using $m=37$ yields a correlation of about $-0.48$.

### 4.5. Another view on sources of biases

The correlation between the returns of an asset and their volatilities is their covariance divided by their standard deviations. A natural question is which of these three quantities are understated/overstated in the process. We focus on the case where $m^2\Delta \to C_m \in (0,\infty)$ for ease of presentation below.

Under the conditions of Theorem 4, we have (see details in the Appendix)

$$\begin{aligned}&\mathrm{Cov}(\hat{V}^{\mathrm{PAV}}_{t+m\Delta,\Delta}-\hat{V}^{\mathrm{PAV}}_{t,\Delta},Z_{t+m\Delta}-Z_t)\\&\quad=\mathrm{Cov}(V_{t+m\Delta,\Delta}-V_{t,\Delta},X_{t+m\Delta}-X_t)(1+\mathrm{o}(m\Delta)),\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(\hat{V}^{\mathrm{PAV}}_{t+m\Delta,\Delta}-\hat{V}^{\mathrm{PAV}}_{t,\Delta})&=\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})\\&\quad\times\left[1+\mathrm{O}\left(\frac{1}{m}\right)+\mathrm{o}(m\Delta)\right],\end{aligned}$$

$$\mathrm{Var}(Z_{t+m\Delta}-Z_t)=\mathrm{Var}(X_{t+m\Delta}-X_t)[1+\mathrm{o}(m\Delta)].$$

That means that the shrinkage bias due to estimation error mainly comes from the denominator, more specifically, the variance of change in volatilities.

By contrast, both the covariance and variance contribute to the source of the smoothing bias. Indeed, under the same conditions as above, we have

$$\begin{aligned}&\mathrm{Cov}(V_{t+m\Delta,\Delta}-V_{t,\Delta},X_{t+m\Delta}-X_t)\\&\quad=\Delta\mathrm{Cov}(v_{t+m\Delta}-v_t,X_{t+m\Delta}-X_t)\left(1-\frac{1}{2m}+\mathrm{o}(m\Delta)\right),\end{aligned}$$

$$\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})=\Delta^2\,\mathrm{Var}(v_{t+m\Delta}-v_t)\left(1-\frac{1}{3m}+\mathrm{o}(m\Delta)\right).$$

Hence, both the numerator and the denominator, more specifically, the covariance and the variance of change in volatilities, contribute to the smoothing bias.

For the discretization bias, we have the following relations:

$$\frac{1}{m\Delta}\,\mathrm{Cov}(v_{t+m\Delta}-v_t,X_{t+m\Delta}-X_t)=\alpha\gamma\rho-\frac{1}{2}\alpha\gamma\kappa\rho m\Delta+\mathrm{o}(m\Delta),$$

$$\frac{1}{m\Delta}\,\mathrm{Var}(v_{t+m\Delta}-v_t)=\alpha\gamma^2-\frac{1}{2}\alpha\gamma^2\kappa m\Delta+\mathrm{o}(m\Delta),$$

$$\frac{1}{m\Delta}\,\mathrm{Var}(X_{t+m\Delta}-X_t)=\alpha+\frac{\alpha\gamma m\Delta(\gamma-4\kappa\rho)}{8\kappa}+\mathrm{o}(m\Delta).$$

They imply that all three components in the calculation of the correlation contribute to the discretization bias.

## 5. The effect of jumps

Jumps are an important feature of asset returns. The extent to which their presence impacts the measurement of the leverage effect depends primarily on two factors: first, price jumps that are not accounted for when estimating volatility do bias upwards the volatility estimates and affect the correlation measurement; second, if there are jumps in volatility and the volatility process tends to jump at the same time as the price process, then the bias can go in either direction depending upon whether those co-jumps tend to be of the same sign or not.

If we allow for co-jumps in price and volatility, then the notion of a leverage effect needs to be extended to incorporate not only the correlation arising between the two Brownian shocks, but also the correlation arising between the two jump terms. This results in a total covariation between the price and volatility processes that includes both a continuous and a discontinuous part, namely,

$$\rho\gamma\int_0^t v_s\,ds+\sum_{0\le s\le t}(\Delta X_s)(\Delta v_s),$$

where $\Delta X_s$ and $\Delta v_s$ denote jumps in log-price and volatility, respectively. The total quadratic variations of prices and volatility are respectively given by

$$\int_0^t v_s\,ds+\sum_{0\le s\le t}(\Delta X_s)^2\quad\text{and}\quad\gamma^2\int_0^t v_s\,ds+\sum_{0\le s\le t}(\Delta v_s)^2.$$

Now there are different possible definitions for the leverage effect: one is $\rho$, as before, which we can estimate using the continuous part of the covariation. But another part is due to the co-jumps, $\sum_{0\le s\le t}(\Delta X_s)(\Delta v_s)$.

The two parts can in principle be estimated consistently, although any estimates of the latter are likely to be unreliable in practice since they would necessarily rely on observing jumps that are rare to begin with for each series, but in fact rely on jumps in both price and volatility series that happen at the same time. We will therefore focus on analyzing the first effect, namely, the impact of price jumps on estimating the correlation between the Brownian shocks to price and volatility, involving a direct effect on the covariation measurement and an indirect effect on the volatility level measurement. The impact of co-jumps will be studied via simulation in Section 7.2.

We consider for this purpose a natural extension of the Heston model that allows for jumps in the price process, as follows:

$$dX_t=(\mu-v_t/2)\,dt+v_t^{1/2}\,dB_t+J_t\,dN_t,\tag{17}$$

$$dv_t=\kappa(\alpha-v_t)\,dt+\gamma v_t^{1/2}\,dW_t,\tag{18}$$

where $B_t$ and $W_t$ are Brownian motions with correlation $\rho$, $N_t$ is a Poisson process with intensity $\lambda$, and $J_t$ denotes the jump size which is assumed independent of everything else. We assume that $J_t$ follows a distribution with mean zero and variance $\sigma_J^2$. We label this model Heston(J).

We analyze what happens to natural estimators of the leverage effect parameter $\rho$. In the absence of market microstructure noise, we employ the truncated realized volatility estimator

$$\hat{V}^{\mathrm{RV,TR}}_{t,\Delta}=\sum_{i=0}^{\Delta/\delta-1}(X_{t-\Delta+(i+1)\delta}-X_{t-\Delta+i\delta})^2\mathbf{1}_{\{|X_{t-\Delta+(i+1)\delta}-X_{t-\Delta+i\delta}|\le a\delta^{\varpi}\}}\tag{19}$$

for some $\varpi\in(0,1/2)$ and $a>0$. $\hat{V}^{\mathrm{RV,TR}}_{t,\Delta}$ is a consistent estimator of the integrated volatility $V_{t,\Delta}$ that filters out the large increments for the purpose of computing the continuous part of the quadratic variation. By including only increments that are of an order of magnitude smaller than what the jumps can generate, this estimator computes the sum of squared log-returns only for log-returns that are likely to have been generated by the Brownian part of the model, and is known to effectively address the problem of the upward bias in volatility that is caused

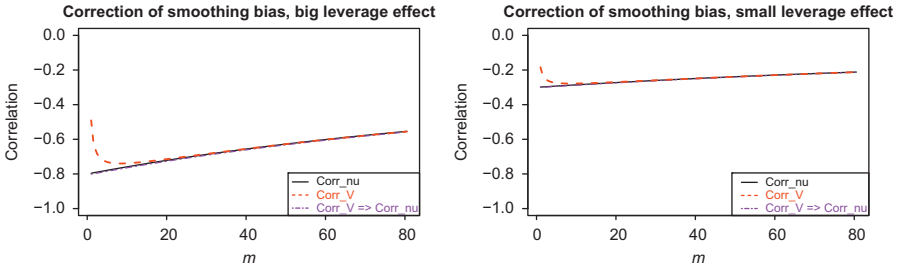


**Fig. 6.** Effectiveness of the multiplicative correction of smoothing bias, based on the main term in (21). The solid curve (labeled as `Corr_nu`) depicts $\mathrm{Corr}(v_{t+m\Delta}-v_t, X_{t+m\Delta}-X_t)$ as a function of $m$; the dash curve (labeled as `Corr_V`) shows $\mathrm{Corr}(V_{t+m\Delta,\Delta}-V_{t,\Delta}, X_{t+m\Delta}-X_t)$, and the dot-dashed curve (labeled as `Corr_V` ⇒ `Corr_nu`) plots $(2\sqrt{m^2-m/3}/(2m-1))\mathrm{Corr}(V_{t+m\Delta,\Delta}-V_{t,\Delta}, X_{t+m\Delta}-X_t)$. After correction, the estimate of $\rho$ based on the integrated volatility $V$ is approximately the same as that based on the spot volatility. Left panel: $(\rho,\kappa,\gamma,\alpha,\mu,\Delta)=(-0.8,5,0.5,0.1,0.05,1/252)$; right panel: $(\rho,\kappa,\gamma,\alpha,\mu,\Delta)=(-0.3,5,0.05,0.04,0.02,1/252)$.

by the jumps (see Mancini, 2009; Aït-Sahalia and Jacod, 2009). Without truncation, price jumps could be a substantial source of attenuation of the leverage effect.

We have the following result about the correlation when the truncated realized volatility is used as a proxy for the continuous part of the volatility:

*Theorem 5. When* $\Delta\to 0$, $n\Delta\to C$, *and* $m^2\Delta\to C_m$ *with* $C$, $C_m\in(0,\infty)$, *and* $5/16<\varpi<1/2$,

$$\begin{aligned}&\mathrm{Corr}(\hat{V}^{\mathrm{RV,TR}}_{t+m\Delta,\Delta}-\hat{V}^{\mathrm{RV,TR}}_{t,\Delta}, X_{t+m\Delta}-X_t)\\&=\mathrm{Corr}(v_{t+m\Delta}-v_t, X_{t+m\Delta}-X_t)\frac{(2m-1)}{2\sqrt{m^2-m/3}}\\&\cdot\left(1+\frac{12\alpha\kappa+6\gamma^2}{(3\gamma^2 m-\gamma^2)\kappa C-\frac{3}{2}\gamma^2\kappa^2 CC_m}\right)^{-1/2}\\&\cdot\left(\frac{\alpha+\left(\frac{\gamma^2\alpha}{8\kappa}-\frac{\gamma\alpha\rho}{2}\right)m\Delta}{\alpha+\sigma_f^2\lambda+\left(\frac{\gamma^2\alpha}{8\kappa}-\frac{\gamma\alpha\rho}{2}\right)m\Delta}\right)^{1/2}+\mathrm{o}(m\Delta).\end{aligned}\tag{20}$$

Theorem 5 shows that, once we filter out the jumps using truncated realized volatility, then up to an additional bias factor which is due to the addition of jump variance in the variance of $X_{t+m\Delta}-X_t$, the estimated leverage effect $\rho$ is subject to the same bias terms as when jumps are absent.

## 6. A solution to the puzzle

Sections 4 and 5 documented the various biases arising when estimating the leverage effect parameter $\rho$ in four progressively more realistic scenarios. The message was decidedly gloomy: even in idealized situations, the bias is large, and attempts to correct for the latency of the volatility, or for the presence of market microstructure noise, do not improve matters. In fact, they often make matters worse. But, fortunately, they also point towards potential solutions to the bias problem, even if one insists upon using estimators that are constructed by plugging into the correlation log-returns and realized volatility-type estimators.

### 6.1. Back to the latent spot volatility

We focus on model (1) and (2) first. We show that all the additional biases that are introduced by the latency of the spot volatility can be corrected, and the problem is reduced to the discretization bias left in Theorem 1.

Recall the asymptotic expression given in Proposition 2, which can be inverted to yield

$$\begin{aligned}&\mathrm{Corr}(v_{t+m\Delta}-v_t, X_{t+m\Delta}-X_t)\\&=\frac{2\sqrt{m^2-m/3}}{(2m-1)}\mathrm{Corr}(V_{t+m\Delta,\Delta}-V_{t,\Delta}, X_{t+m\Delta}-X_t)\\&+\begin{cases}\mathrm{O}(\Delta) & \text{when } \Delta\to 0 \text{ for any } m,\\ \mathrm{o}(m\Delta) & \text{when } m\to\infty,\ m\Delta\to 0.\end{cases}\end{aligned}\tag{21}$$

Thus, up to a multiplicative correction factor that is independent of the model's parameters, the integrated volatility $V$ can work as well as the spot volatility $v$. The effectiveness of this simple bias correction is demonstrated in Fig. 6.

In the absence of microstructure noise, using the realized volatility (6), the asymptotic relative bias in comparison with the use of the true spot volatility is given by Theorem 3. Using the expressions given there, we can correct the bias due to the estimate by realized volatility back to that based on the spot volatility. However, such a correction involves unknown parameters in the Heston model, which is nontrivial to estimate due to the stochastic volatility, which relies on a nonparametric correction. An alternative approach is to use the following result, demonstrated in the Appendix. This avoids the challenge of directly estimating the unknown parameters in the model. For ease of presentation we again focus on the case where $m^2\Delta\to C_m\in(0,\infty)$ in the following.

*Proposition 3. When* $\Delta\to 0$, $n\Delta\to C$ *and* $m^2\Delta\to C_m$ *with* $C$ *and* $C_m\in(0,\infty)$,

$$\begin{aligned}\mathrm{Corr}(v_{t+m\Delta}-v_t, X_{t+m\Delta}-X_t)=c_3\frac{2\sqrt{m^2-m/3}}{(2m-1)}\mathrm{Corr}(\hat{V}^{\mathrm{RV}}_{t+m\Delta,\Delta}\\-\hat{V}^{\mathrm{RV}}_{t,\Delta}, X_{t+m\Delta}-X_t)+\mathrm{o}(m\Delta),\end{aligned}\tag{22}$$

*where $c_3$ is given by*

$$c_3 = \left(1 - \frac{4\mathrm{E}[\sigma_t^4]\Delta^2}{n\,\mathrm{Var}(\hat{V}^{\mathrm{RV}}_{t+m\Delta,\,\Delta} - \hat{V}^{\mathrm{RV}}_{t,\Delta})}\right)^{-1/2}. \tag{23}$$

Note that in (23), the stationarity of the process of $v_t$ is used so that the correction factor does not depend on $t$.

In practice, we can estimate $\mathrm{E}[\sigma_t^4]$ nonparametrically based on the fact that the realized quarticity

$$RQ_t^n := \frac{n}{3}\sum_{i=0}^{n-1}(X_{t+(i+1)\delta} - X_{t+i\delta})^4$$

satisfies

$$\mathrm{E}[RQ_t^n] = \Delta^2\mathrm{E}[\sigma_t^4](1+\mathrm{o}(1)) \tag{24}$$

for any fixed $\Delta$ as $n\to\infty$. A long-run average of scaled realized quarticity can be used to estimate $\mathrm{E}[\sigma_t^4]$. The variance in (23) can be estimated by its sample version.

For the PAV estimator, the bias correction admits the same form as (22) with a different correction factor.

*Proposition 4. When $\Delta\to 0$, $n^{1/2}\Delta\to C_{\mathrm{PAV}}$, $\sigma_\epsilon^2/\Delta\to C_\epsilon$, and $m^2\Delta\to C_m$ for constants $C_{\mathrm{PAV}}$, $C_\epsilon$, and $C_m\in(0,\infty)$,*

$$\mathrm{Corr}(v_{t+m\Delta}-v_t, X_{t+m\Delta}-X_t) = c_4\frac{2\sqrt{m^2-m/3}}{(2m-1)}\mathrm{Corr}(\hat{V}^{\mathrm{PAV}}_{t+m\Delta,\Delta} - \hat{V}^{\mathrm{PAV}}_{t,\Delta}, Z_{t+m\Delta}-Z_t) + \mathrm{o}(m\Delta), \tag{25}$$

*where*

$$c_4 = \left(1 - \frac{2(A_4'+B_4'+C_4')}{n^{1/2}\mathrm{Var}(\hat{V}^{\mathrm{PAV}}_{t+m\Delta,\Delta} - \hat{V}^{\mathrm{PAV}}_{t,\Delta})}\right)^{-1/2}, \tag{26}$$

*with*

$$A_4' = \frac{4\Phi_{22}\theta_{\mathrm{PAV}}\mathrm{E}[\sigma_t^4]\Delta^2}{\psi_2^2},\quad B_4' = \frac{8\Phi_{12}\mathrm{E}[\sigma_t^2]\sigma_\epsilon^2\Delta}{\theta_{\mathrm{PAV}}\psi_2^2},$$
$$C_4' = \frac{4\Phi_{11}\sigma_\epsilon^4}{\theta_{\mathrm{PAV}}^3\psi_2^2},$$

*where $\psi_2, \Phi_{11}, \Phi_{12}, \Phi_{22}$ are the constants given in Theorem 4.*

One can make use of the long-run average of the quantity $\Gamma_t^n$ defined in Jacod, Li, Mykland, Podolskij, and Vetter (2009) to estimate $A_4'+B_4'+C_4'$:

$$\begin{aligned}\Gamma_t^n = {}& \frac{4\Phi_{22}}{3\theta_{\mathrm{PAV}}\Delta^{1/2}\psi_2^4}\sum_{i=0}^{n-k_n+1}\\ &\times\left(\frac{1}{k_n}\sum_{j=\lfloor k_n/2\rfloor}^{k_n-1} Z_{t-\Delta+(i+j)\delta} - \frac{1}{k_n}\sum_{j=0}^{\lfloor k_n/2\rfloor-1} Z_{t-\Delta+(i+j)\delta}\right)^4\\ &- \frac{4\delta}{\theta_{\mathrm{PAV}}^3\Delta^{3/2}}\left(\frac{\Phi_{12}}{\psi_2^3} - \frac{\Phi_{22}}{\psi_2^4}\right)\sum_{i=0}^{n-2k_n+1}\left(\left(\frac{1}{k_n}\sum_{j=\lfloor k_n/2\rfloor}^{k_n-1} Z_{t-\Delta+(i+j)\delta}\right.\right.\\ &\left.\left. - \frac{1}{k_n}\sum_{j=0}^{\lfloor k_n/2\rfloor-1} Z_{t-\Delta+(i+j)\delta}\right)^2\sum_{j=i+k_n}^{i+2k_n-1}(Z_{t-\Delta+(j+1)\delta} - Z_{t-\Delta+j\delta})^2\right)\\ &+ \frac{\delta}{\theta_{\mathrm{PAV}}^3\Delta^{3/2}}\left(\frac{\Phi_{11}}{\psi_2^2} - 2\frac{\Phi_{12}}{\psi_2^3} + \frac{\Phi_{22}}{\psi_2^4}\right)\\ &\times\sum_{i=1}^{n-2}(Z_{t-\Delta+(i+1)\delta} - Z_{t-\Delta+i\delta})^2(Z_{t-\Delta+(i+3)\delta} - Z_{t-\Delta+(i+2)\delta})^2.\end{aligned} \tag{27}$$

Then, we have

$$E(\Gamma_t^n) = \Delta^{-1/2}(A_4'+B_4'+C_4')(1+\mathrm{o}(1)) \tag{28}$$

for any fixed $\Delta$ as $n\to\infty$. A long-run average of scaled $\Gamma_t^n$ can be used to estimate $A_4'+B_4'+C_4'$.

A result parallel to Proposition 4 for TSRV (see Proposition 6) can be found in the Appendix.

### 6.2. Correction in the presence of jumps

In the presence of jumps, for model (17) and (18), we have the following result.

*Proposition 5. When $\Delta\to 0$, $n\Delta\to C$, and $m^2\Delta\to C_m$ with $C$ and $C_m\in(0,\infty)$, and $5/16<\varpi<1/2$,*

$$\mathrm{Corr}(v_{t+m\Delta}-v_t, X_{t+m\Delta}-X_t) = c_5\frac{2\sqrt{m^2-m/3}}{(2m-1)}\mathrm{Corr}(\hat{V}^{\mathrm{RV,TR}}_{t+m\Delta,\Delta} - \hat{V}^{\mathrm{RV,TR}}_{t,\Delta}, X_{t+m\Delta}-X_t) + \mathrm{o}(m\Delta), \tag{29}$$

*where $c_5$ is given by*

$$c_5 = \left(1 - \frac{4\mathrm{E}[\sigma_t^4]\Delta^2}{n\,\mathrm{Var}(\hat{V}^{\mathrm{RV,TR}}_{t+m\Delta,\Delta} - \hat{V}^{\mathrm{RV,TR}}_{t,\Delta})}\right)^{-1/2} \cdot\left(\frac{\mathrm{Var}(X_{t+m\Delta}-X_t)}{\mathrm{Var}(X_{t+m\Delta}-X_t) - \mathrm{Var}(J_{(t,t+m\Delta)})}\right)^{1/2}, \tag{30}$$

*and $J_{(s,t)} = \sum_{r\in[s,t]}(X_r - X_{r-})$.*

To implement the result above in practice, we can estimate $\mathrm{E}[\sigma_t^4]$ consistently using a long-run average of truncated realized quarticity $(n/3)\sum_{i=0}^{n-1}(X_{t+(i+1)\delta} - X_{t+i\delta})^4 1_{\{|X_{t+(i+1)\delta}-X_{t+i\delta}|\le a\delta^{\varpi}\}}$ (scaled by $\Delta^2$), and $\mathrm{Var}(J_{(t,t+m\Delta)})$ by $m$ times the long-run variance of $\sum_{i=0}^{n}(X_{t+(i+1)\delta} - X_{t+i\delta})1_{\{|X_{t+(i+1)\delta}-X_{t+i\delta}|>a\delta^{\varpi}\}}$, for some $a>0$ and $5/16<\varpi<1/2$.

### 6.3. Correcting the discretization bias from spot volatilities

The above results reveal that the biases due to the various estimates are correctable back to the case where the spot volatility can be viewed as observable. However, Theorem 1 implies that the estimate of $\rho$ based on $v_t$ itself is also biased. If the model were known, then the bias in (10) can be computed and corrected. However, this depends on the Heston model and its unknown parameters.

A parameter-independent method is as follows. Let $\rho_m = \mathrm{Corr}(v_{t+m\Delta}-v_t, X_{t+m\Delta}-X_t)$. Then, by Theorem 1 we see that

$$\rho_m = \rho + bm + \mathrm{o}(m\Delta). \tag{31}$$

This suggests that the parameter of interest $\rho$ (as well as the slope $b$ but this is not needed) can be estimated by running a linear regression of the data $\{(m,\rho_m)\}$. The bias-corrected estimate of $\rho$ is simply the intercept of that linear regression. The scatter plot of $\{(m,\rho_m)\}$ can also suggest a region of $m$ to run the above simple linear

regression (31). A data-driven choice of the range of $m$ to be used for the regression is given in Section 7.1.2.

The above discussion suggests a rather general strategy for bias correction. First, compute the simple correlation between estimated changes of volatilities and changes of prices. Second, conduct a preliminary bias correction according to (21)–(30) or (61), depending on which estimated volatilities are used. Third, run the simple regression Eq. (31) for the preliminary bias-corrected estimated correlations. Fourth, take the intercept of the simple linear regression as the final estimate. The method turns out to be very effective in practice, as we now see.

## 7. Monte-Carlo simulations

In this section, we use simulations to reproduce the leverage effect puzzle and its proposed solution, and to verify the practical validity of the results presented in the previous sections at finite sample. We devised the bias terms and their correction using the tractable mathematics of the Heston model. In these simulations, we first seek both to validate the theoretical results presented in the previous section under the Heston model, and then examine the conjecture that these corrections can be useful in practice under different data generating processes than what was assumed in the theory part of the paper.

### 7.1. Data generating processes under prototypical models

We first validate the theoretical results above using the Heston model (1)–(2) without price jumps, and then with price jumps (17)–(18), as assumed in our derivations. We employ broadly realistic parameter values: $\alpha=0.1$, $\gamma=0.5$, $\kappa=5$, $\rho=-0.8$, and $\mu=0.05$ over $252*5$ trading days in five years ($\Delta=1/252$). The sampling frequency is one minute per sample, giving an intraday number of observations of $n=390$. Therefore, the total number of observations over five years is $N=252*390*5=491{,}400$. The true price is latent. Instead, the observed data $\{Z_{i\delta}\}_{i=1}^{491{,}400}$ are contaminated with market microstructure as in (7): the noise is independent and identically distributed (i.i.d.) $\mathcal{N}(0,\sigma_\epsilon^2)$ with $\sigma_\epsilon=0.0005$.

#### 7.1.1. Visualizing the leverage effect puzzle

With the latent spot volatility $v_t$ and latent price $X_t$ known in simulated data, we can easily examine the correlation of $\{(X_{t\Delta}-X_{(t-1)\Delta},v_{t\Delta}-v_{(t-1)\Delta})\}$ over $N$ observations. As expected, the leverage effect is strong, with the sample correlation being $-0.79$ for a given realization. This is in line with the result of Theorem 1.

Next, consider the more realistic situation that the spot volatility needs to be estimated by a smoothing method such as a local integrated average $V_{t,\Delta}=\int_{t-\Delta}^{t}\sigma_t^2\,dt$. A natural estimate is the average of daily spot volatility $\hat{V}_{t,\Delta}=n^{-1}\sum_{j=1}^{n}\hat{\sigma}^2_{t-\Delta+j\Delta/n}$. In this ideal situation, $\sigma^2_{t-\Delta+j\Delta/n}$ is known, resulting in $V_{t,\Delta}=n^{-1}\sum_{j=1}^{n}\sigma^2_{t-\Delta+j\Delta/n}$. The correlation of $\{(X_{(t+1)\Delta}-X_{t\Delta},V_{(t+1)\Delta,\Delta}-V_{t\Delta,\Delta})\}_{t=1}^{1259}$ is $-0.49$ for the given realization examined above. This is in line with the result of Theorem 2. The magnitude of the leverage effect parameter $\rho$ is significantly under estimated. To appreciate the effect of the tuning parameter $m$, the upper panel of Fig. 7 plots the correlation $\{(X_{(t+m)\Delta}-X_{t\Delta},v_{(t+m)\Delta,\Delta}-v_{t\Delta,\Delta})\}_{t=1}^{1260-m}$ and $\{(X_{(t+m)\Delta}-X_{t\Delta},V_{(t+m)\Delta,\Delta}-V_{t\Delta,\Delta})\}_{t=1}^{1260-m}$ against $m$. To examine the sampling variabilities, the simulation is conducted 100 times. The averages of the sample correlations are plotted along with its standard deviation (SD) in the figure. The impact of $m$ can easily be seen and the natural estimate based on $V_{t,\Delta}$ with $m=1$ is far from optimal.

In practice, the integrated volatility is not observable. It has to be estimated using the discretely observed data. In absence of the market microstructure noise, the realized volatility provides a good estimate of the integrated volatility. Using RV based on the simulated latent prices $X_i^n$, we have a sample correlation of $-0.26$ for the same realization discussed above based on $\{(X_{(t+1)\Delta}-X_{t\Delta},\hat{V}^{\mathrm{RV}}_{(t+1)\Delta,\Delta}-\hat{V}^{\mathrm{RV}}_{t\Delta,\Delta})\}_{t=1}^{1259}$. More generally, the correlation of $\{(X_{(t+m)\Delta}-X_{t\Delta},\hat{V}^{\mathrm{RV}}_{(t+m)\Delta,\Delta}-\hat{V}^{\mathrm{RV}}_{t\Delta,\Delta})\}_{t=1}^{1260-m}$ as a function of $m$ is depicted in the lower left panel of Fig. 7. As above, this is repeated 100 times so that the average correlations along with their errors at each $m$ are computed.

For a more realistic situation, the integrated volatility has to be estimated based on the contaminated log-prices $Z_t$ in (7). The volatility parameter is now estimated by the correlation $\{(Z_{(t+m)\Delta}-Z_{t\Delta},\hat{V}^{\mathrm{PAV}}_{(t+m)\Delta,\Delta}-\hat{V}^{\mathrm{PAV}}_{t\Delta,\Delta})\}_{t=1}^{1260-m}$ or $\{(Z_{(t+m)\Delta}-Z_{t\Delta},\hat{V}^{\mathrm{TSRV}}_{(t+m)\Delta,\Delta}-\hat{V}^{\mathrm{TSRV}}_{t\Delta,\Delta})\}_{t=1}^{1260-m}$ with a suitable choice of $m$. The lower middle and lower right plots of Fig. 7 show the correlation as a function of $m$. In particular, when $m=1$, the sample correlation is merely $-0.13$ for PAV and $-0.06$ for TSRV for the same simulated path as mentioned above, which would be interpreted in practice as showing little support for the leverage effect. But we know that this is due to the statistical bias of the procedure as demonstrated in Theorems 4 and 6. For this realization, using PAV with $m=31$, the sample correlation is $-0.63$; and using TSRV with $m=50$, the sample correlation is $-0.53$. While this is still a biased estimate, the leverage effect can be clearly seen.

The averages of these correlations, based on the 100 simulations, against $m$ are plotted together in Fig. 8. These are in line with the theory (see the left panel of Fig. 5).

#### 7.1.2. Effectiveness of the bias correction method

We now illustrate the effectiveness of the bias correction method proposed in Section 6. We simulate sample paths with the same parameters as above. $\theta_{\mathrm{TSRV}}$ and $\theta_{\mathrm{PAV}}$ are both taken to be 0.5.

For each volatility or volatility proxy $v$, $V$, $\hat{V}^{\mathrm{RV}}$, $\hat{V}^{\mathrm{PAV}}$, and $\hat{V}^{\mathrm{TSRV}}$, let $\hat{\rho}_m$ be the sample version or bias-corrected estimate of $\rho_m=\mathrm{Corr}(v_{t+m\Delta}-v_t,X_{t+m\Delta}-X_t)$. We call this preliminary bias-correction. In practice, the model parameters are unknown. We use the non parametric methods as described in Section 6.1 to obtain the preliminary corrections. More specifically, $\mathrm{E}[\sigma_t^4]$ in RV is estimated by the long-run average of scaled realized quarticity based on Eq. (24);

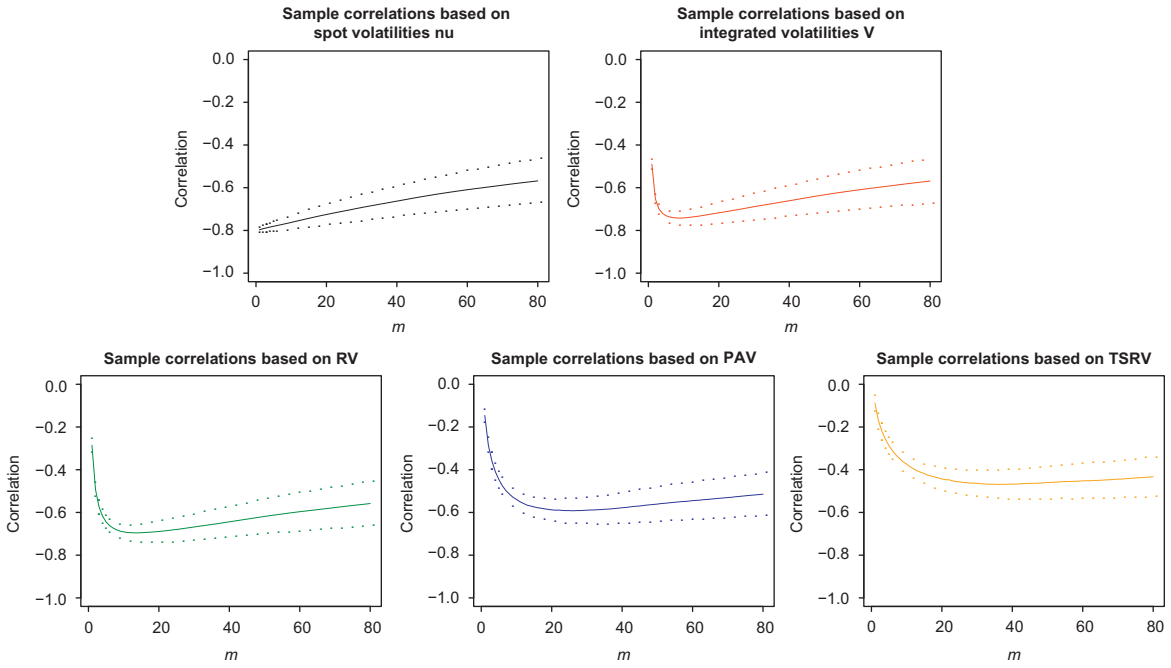


**Fig. 7.** Sample correlations between the log-returns and the changes of spot volatilities (upper left), the changes of integrated volatilities (upper right), the changes of realized volatility (lower left), the changes of PAV (lower middle), and the changes of TSRV (lower right) over a period of $m$ days. The results are based on 100 simulations. The solid curve is the average over one hundred simulations. The dots are one standard deviations away from the averages. Parameters: $(\rho,\kappa,\gamma,\alpha,\mu,\sigma_\epsilon,\Delta,n)=(-0.8,5,0.5,0.1,0.05,0.0005,1/252,390)$.

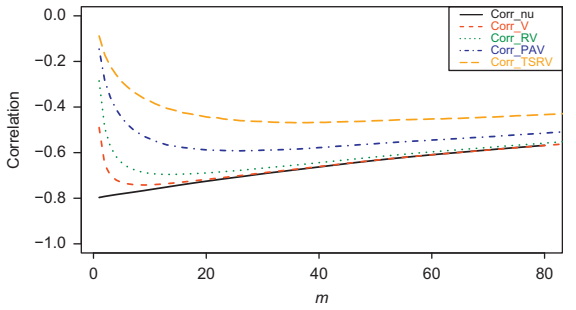


**Fig. 8.** Average sample correlations based on simulated data. The average sample correlations between the changes of log-prices over a period of $m$ days and the difference of the spot volatility $v$, the difference of the integrated volatility $V$, the difference of the RV estimates, the difference of the PAV estimates, and the difference of the TSRV estimates over the same period. Comparing this with the left panel of Fig. 5, we see how the simulation results are in line with the theory. Parameters: $(\rho,\kappa,\gamma,\alpha,\mu,\sigma_\epsilon,\Delta,n)=(-0.8,5,0.5,0.1,0.05,0.0005,1/252,390)$.

$A'_4+B'_4+C'_4$ in PAV is estimated by the long-run average of scaled $\Gamma_t^n$ based on Eq. (28). For the unknown values of $\mathrm{E}[\sigma_t^4]$ and $\sigma_\epsilon^2$ in TSRV, we use long-run average of $\hat{Q}_t^n$ as defined in (62) and long-run average of $(\hat{V}^{\mathrm{RV}}_{t,\Delta}-\hat{V}^{\mathrm{TSRV}}_{t,\Delta})/2n$ as discussed above Eq. (62) in the Appendix.

In our simulation and empirical studies, we employ the following automated method to select the range of $m$ for the linear extrapolation in Section 6.3. Note that the average of $\hat{\rho}_m$ over many simulations should behave like the black solid curve in Figs. 5, 6 or 8. For each given sample path, $\hat{\rho}_m$ can deviate from the theoretical curve as demonstrated in (the left panel of) Fig. 9, and the deviation can be large when $m$ is small. Therefore, choosing an appropriate range for linear extrapolation is important and challenging. Our data-driven procedure goes as follows:

1. Compute $\hat{\rho}_m$ for every $m$ in $[1,l]$ ($l=252$, say). Let $m_1$, $m_2$, $m_3$ be the positions corresponding to the smallest, second smallest, third smallest among $\{\hat{\rho}_m\}_{m=a_0}^{l/2}$ ($a_0=6$, say, which avoids instable estimates for small $m$). The notation $\{\hat{\rho}_m\}_{m=a_0}^{l/2}$ means the sequence $\hat{\rho}_m$ when $m$ runs from $a_0$ to $l/2$. Set $m^*=\max\{m_1,m_2,m_3\}$. $m^*$ basically corresponds to the minimum value of $\{\hat{\rho}_m\}_{m=a_0}^{l/2}$, but is computed more robustly. It is the lower end point of the range of $m$ to be used for regression.
2. Run the simple linear regression based on the pair of the data $\{m,\hat{\rho}_m\}_{m=m^*}^{m^*+k}$, for $k$ in $[k_0,l-m^*]$, where $k_0$ ($=11$, say) is the minimum number of data points needed to run such a regression. Let $k^*$ be the value of $k$ that corresponds to the largest multiple-$R^2$. It is taken to determine the upper end point of the range of values of $m$ to be used.
3. Run a regression based on $\{m,\hat{\rho}_m\}_{m=m^*}^{m^*+k^*}$; the intercept of the regression is taken as our final estimate of $\rho$.

In the simulation studies, we take $l=252$, $a_0=6$, and $k_0=11$. The results are not sensitive to the choices of these parameters. Fig. 9 demonstrates how this works on a simulated sample path. More extensive results are given

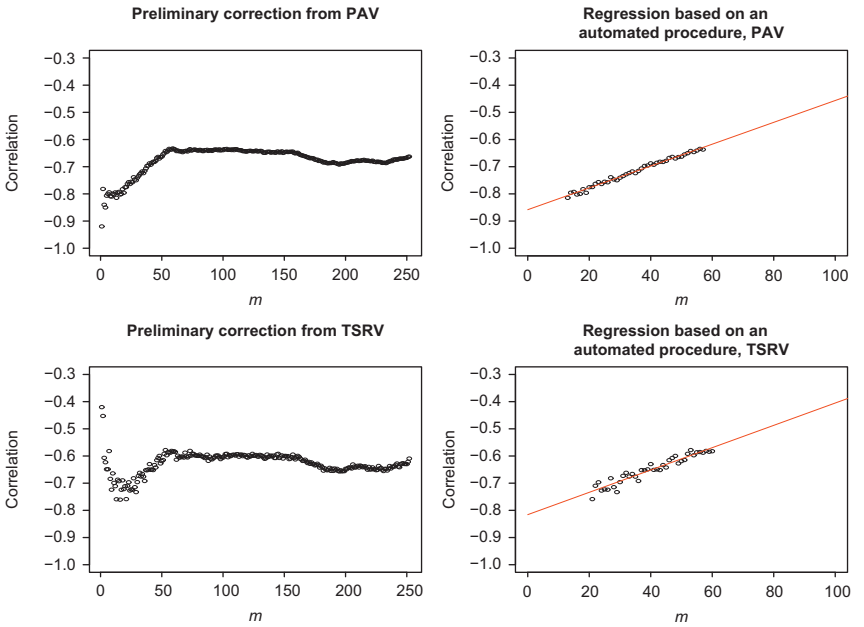


**Fig. 9.** Scatter plots of preliminary bias-corrected estimates of the leverage effect parameter $\hat{\rho}_m$ against $m$ for one simulated realization are plotted on the left (upper: PAV, lower: TSRV); the regression based on the automated procedure described in Section 7.1.2 is illustrated on the right. The final estimate based on PAV is $-0.86$, that based on TSRV is $-0.82$ for this realization. The true leverage effect parameter $\rho$ is $-0.8$.

**Table 1**
Performance of the (feasible) bias correction method based on nonparametrically estimated asymptotic quantities and an automated linear regression for Heston model. The 100 estimates of $\rho$ are summarized by their minimum, first quartile, median, third quartile, maximum, mean, and SD. Parameters: $(\rho,\kappa,\gamma,\alpha,\mu,\sigma_\epsilon,\Delta,n)=(-0.8,5,0.5,0.1,0.05,0.0005,1/252,390)$.

| Estimator | Min. | 1st Qu. | Median | 3rd Qu. | Max. | Mean | SD |
|---|---|---|---|---|---|---|---|
| $\hat{\rho}_{\mathrm{RV}}$ | −0.95 | −0.84 | −0.81 | −0.78 | −0.67 | −0.82 | 0.047 |
| $\hat{\rho}_{\mathrm{PAV}}$ | −0.99 | −0.85 | −0.82 | −0.77 | −0.64 | −0.81 | 0.064 |
| $\hat{\rho}_{\mathrm{TSRV}}$ | −1.00 | −0.87 | −0.83 | −0.74 | −0.63 | −0.81 | 0.089 |

in Table 1, which shows that the automated method works very well among 100 simulations.

For the simulation studies, we have the data without microstructural noise available and hence, $\hat{\rho}_m$ can be computed based on the realized volatility via (22). Let us denote the bias-corrected estimate of $\rho$ as $\hat{\rho}_{\mathrm{RV}}$ after running the automated linear extrapolation algorithm. In the presence of microstructure noise, $\hat{\rho}_m$ can be computed based on PAV (see (25)) or TSRV (see (61)). After running the automated linear extrapolation algorithm, the results are denoted, respectively, as $\hat{\rho}_{\mathrm{PAV}}$ and $\hat{\rho}_{\mathrm{TSRV}}$. Table 1 summarizes the results of 100 simulations of minute-by-minute ($n=390$) data over a five-year period ($T=5$) for the model (1) and (2) with $\alpha=0.1$, $\gamma=0.5$, $\kappa=5$, $\rho=-0.8$, and $\mu=0.05$.

The means of these corrected estimates are all close to the true value $\rho=-0.8$, which implies that these estimates are nearly unbiased. The fact that the problems become progressively harder can easily be seen from the SD of the estimates.

In summary, Table 1 provides stark evidence that the methods in Section 6 solve the leverage effect puzzle. It also quantifies the extent to which the problem gets progressively harder. When the sampling frequency is more frequent than one sample per minute, the estimation error can be reduced. We omit the details here.

### 7.1.3. Models with jumps

We now consider the model Heston(J) (17) and (18), which incorporates jumps. For the diffusion part, we use the same parameter values as above: $\alpha=0.1$, $\gamma=0.5$, $\kappa=5$, $\rho=-0.8$, $\mu=0.05$, $n=390$ over 252∗5 trading days in five years. The jumps are of intensity $\lambda=5$ with sizes distributed as $\mathcal{N}(0,0.015^2)$ ($\sigma_J=0.015$). The empirical findings in Andersen, Benzoni, and Lund (2002) are taken as reference when selecting the parameter values. We consider the case without market microstructure noise. The observed data are $\{X_{i\delta}\}_{i=1}^{491,400}$. For the truncated realized volatility and truncated realized quarticity, we use a truncation level $a\delta^{\varpi}=4\sqrt{\hat{\alpha}}\delta^{1/2}$ following the suggested value and the rationale behind Aït-Sahalia and Jacod (2012, Section 9.1), where $\hat{\alpha}$ is the scaled long-run average of daily realized volatilities, $\hat{\alpha}=\Delta^{-1}$∗average of $\{\hat{V}^{RV}_{t,\Delta}\}_{t=1}^{1260}$, which is taken as a rough estimate of $\alpha$.

Fig. 10 shows the mean and standard deviation of the sample correlations between the log-returns and the changes of the truncated realized volatility based on 100

simulated sample paths. This is in line with Theorem 5 and shows the same feature as the no-jump case.

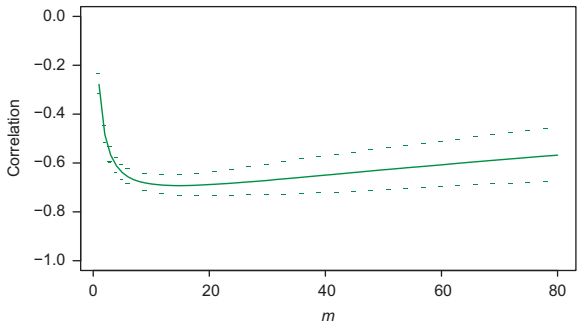


**Fig. 10.** Sample correlations based on truncated RV for the model Heston(J). Sample correlations between the log-returns and the changes of truncated realized volatilities over a period of $m$ days for 100 sample paths simulated based on the model Heston(J) (17) and (18). The solid curve is the average of one hundred simulations; the dots are one standard deviations away from the averages. Parameters: $(\rho,\kappa,\gamma,\alpha,\mu,\sigma_\epsilon,\Delta,n,\lambda,\sigma_J)=(-0.8,5,0.5,0.1,0.05,0.0005,1/252,390,5,0.015)$.

We next consider the correction based on Proposition 5. To estimate the quantities in $c_5$, we use the long-run variances to estimate $\mathrm{Var}(\hat{V}^{\mathrm{RV,TR}}_{t+m\Delta,\Delta}-\hat{V}^{\mathrm{RV,TR}}_{t,\Delta})$ and $\mathrm{Var}(X_{t+m\Delta}-X_t)$; apply

$$m\sum_{i=0}^{n}(X_{t+(i+1)\delta}-X_{t+i\delta})^2 1_{\{|X_{t+(i+1)\delta}-X_{t+i\delta}|>a\delta^{\varpi}\}}$$

to estimate $\mathrm{Var}(J_{(t,t+m\Delta)})$; and employ a long-run scaled average of truncated realized quarticity

$$\Delta^{-2}\frac{n}{3}\sum_{i=0}^{n-1}(X_{t+(i+1)\delta}-X_{t+i\delta})^4 1_{\{|X_{t+(i+1)\delta}-X_{t+i\delta}|\le a\delta^{\varpi}\}}$$

to estimate $\mathrm{E}[\sigma_t^4]$. Then, based on Eq. (29), we can correct the raw correlation $\mathrm{Corr}(\hat{V}^{\mathrm{RV,TR}}_{t+m\Delta,\Delta}-\hat{V}^{\mathrm{RV,TR}}_{t,\Delta},X_{t+m\Delta}-X_t)$ back to $\mathrm{Corr}(v_{t+m\Delta}-v_t,X_{t+m\Delta}-X_t)$. The same automated linear extrapolation procedure as described in Section 7.1.2 is used to estimate $\rho$. The results are collected in Table 2 which shows that the correction is effective.

This illustrates that the leverage effect puzzle exists beyond the Heston model. In particular, when jumps are

**Table 2**
Performance of the bias correction based on nonparametrically estimated asymptotic quantities and linear regression for model Heston(J), starting with the raw correlation based on truncated realized volatility. The 100 estimates of $\rho$ ($\hat{\rho}_{RV,TR}$) are summarized by their minimum, first quartile, median, third quartile, maximum, mean, and SD. Parameters: $(\rho,\kappa,\gamma,\alpha,\mu,\Delta,n,\sigma_J,\lambda)=(-0.8,5,0.5,0.1,0.05,1/252,390,0.015,5)$.

| Estimator | Min. | 1st Qu. | Median | 3rd Qu. | Max. | Mean | SD |
|---|---|---|---|---|---|---|---|
| $\hat{\rho}_{\mathrm{RV,TR}}$ | −0.94 | −0.85 | −0.82 | −0.79 | −0.68 | −0.82 | 0.048 |

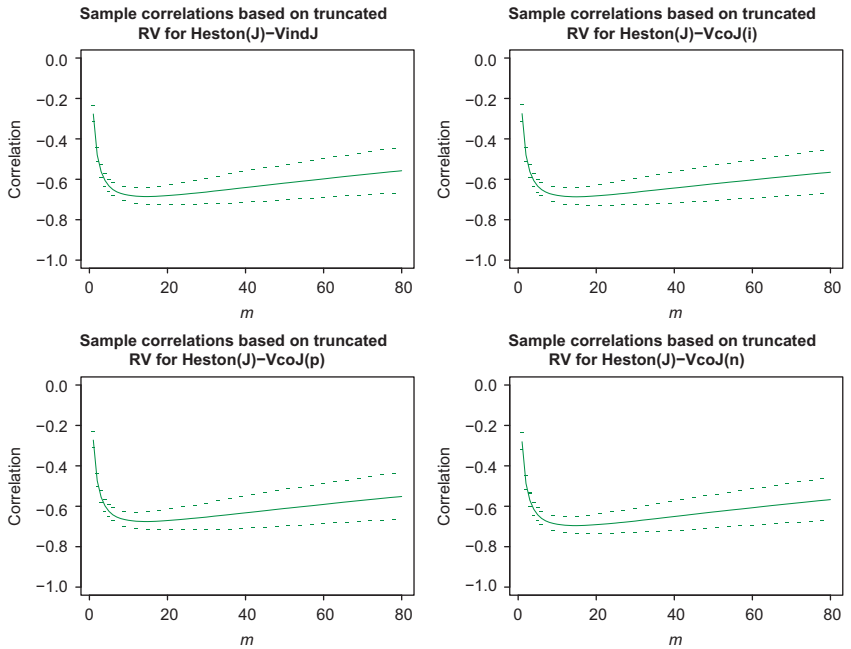


**Fig. 11.** Sample correlations between the log-returns and the changes of truncated realized volatilities over a period of $m$ days for one hundred sample paths simulated based on models Heston(J)–VindJ, Heston(J)–VcoJ(i), Heston(J)–VcoJ(p), and Heston(J)–VcoJ(n). The solid curves are the average of 100 simulations; the dots are one standard deviations away from the averages. Parameters: $(\rho,\kappa,\gamma,\alpha,\mu,\Delta,n,\lambda,\sigma_J,\lambda^v,\sigma_{J^v},\rho_{Jp},\rho_{Jn})=(-0.8,5,0.5,0.1,\ 0.05,1/252,\ 390,5,0.015,5,0.01,0.75,-0.75)$.

present, the nature of the puzzle is the same as in the continuous case when truncated realized measures are used. The proposed nonparametric correction remains effective.

### 7.2. Alternative data generating processes

The bias correction methods were derived for the Heston model. Yet, the method for the correction itself is nonparametric, independent of the prototypical model. The question that arises naturally is whether the nonparametric corrections still work with different models. The results that we now present from nine alternative models show that the proposed correction method works about as well for those models as it does for the Heston model.

#### 7.2.1. Jumps in prices and volatilities

To investigate the performance of the proposed bias correction methods in a model that is different from the one assumed for the theoretical analysis, we first add volatility jumps to the model (17) and (18).

When adding volatility jumps, we consider two cases. First, we add the term $J_t^v\, dN_t^v$ to the volatility stochastic differential equations (18), where $N_t^v$ is a Poisson process with intensity $\lambda^v$ independent of $N_t$, and $J_t^v$ is independent of everything else with mean zero and variance $\sigma_{J^v}^2$. We label this model "Heston(J)–VindJ."; We then consider the cases where volatility and price jump together, i.e., $N_t^v = N_t$. We study three cases: "Heston(J)–VcoJ(i)" is the model where $J_t^v$ is independent of $J_t$; "Heston(J)–VcoJ(p)" is the model where $J_t^v$ and $J_t$ are positively correlated with correlation $\rho_{Jp}$, and "Heston(J)–VcoJ(n)" is the model where $J_t^v$ and $J_t$ are negatively correlated with correlation $\rho_{Jn}$.

We consider the case where there is no market microstructure noise. The parameter values are: $\alpha = 0.1$, $\gamma = 0.5$, $\kappa = 5$, $\rho = -0.8$, $\mu = 0.05$, $n = 390$, $\lambda = \lambda^v = 5$, $\rho_{Jp} = 0.75$, and $\rho_{Jn} = -0.75$. $J_t$ and $J_t^v$ are normally distributed with $\sigma_J = 0.015$ and $\sigma_{J^v} = 0.01$. Based on the 100 sample paths simulated over 252∗5 trading days in five years, we found that the correlations exhibit the same features as above

**Table 3**
Performance of the bias correction based on nonparametrically estimated asymptotic quantities and linear regression for models Heston(J)–VindJ, Heston(J)–VcoJ(i), Heston(J)–VcoJ(p), and Heston(J)–VcoJ(n), starting with the raw correlation based on truncated realized volatility. For each of these models, the 100 estimates of $\rho$ ($\hat{\rho}_{RV,TR}$) are summarized by their minimum, first quartile, median, third quartile, maximum, mean, and SD. Parameters: $(\rho,\kappa,\gamma,\alpha,\mu,\Delta,n,\sigma_J,\lambda,\sigma_{J^v},\lambda^v,\rho_{Jp},\rho_{Jn}) = (-0.8,5,0.5,0.1,0.05,1/252,390,0.015,5,0.01,5,0.75,-0.75)$.

| Estimator/Model | Min. | 1st Qu. | Median | 3rd Qu. | Max. | Mean | SD |
|---|---|---|---|---|---|---|---|
| $\hat{\rho}_{RV,TR}$, Heston(J)–VindJ | −0.94 | −0.85 | −0.81 | −0.79 | −0.67 | −0.82 | 0.047 |
| $\hat{\rho}_{RV,TR}$, Heston(J)–VcoJ(i) | −0.97 | −0.84 | −0.81 | −0.79 | −0.67 | −0.81 | 0.051 |
| $\hat{\rho}_{RV,TR}$, Heston(J)–VcoJ(p) | −0.97 | −0.83 | −0.80 | −0.78 | −0.68 | −0.81 | 0.052 |
| $\hat{\rho}_{RV,TR}$, Heston(J)–VcoJ(n) | −0.95 | −0.85 | −0.82 | −0.80 | −0.66 | −0.82 | 0.048 |

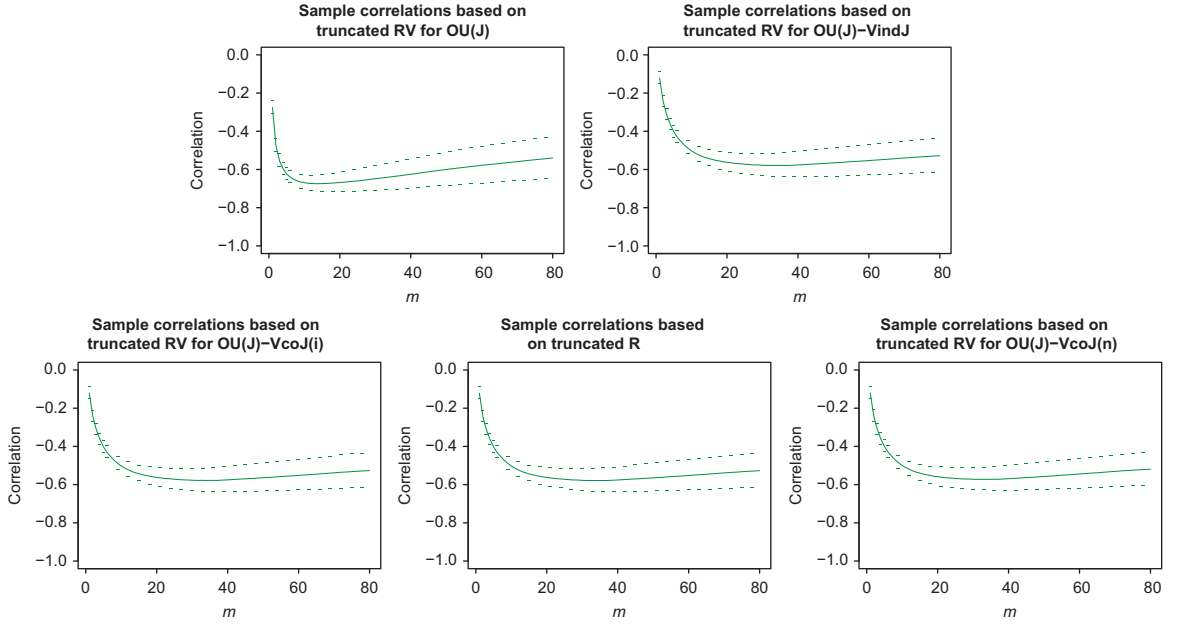


**Fig. 12.** Sample correlations between the log-returns and the changes of truncated realized volatilities over a period of $m$ days for one hundred sample paths simulated based on models OU(J), OU(J)–VindJ, OU(J)–VcoJ(i), OU(J)–VcoJ(p), and OU(J)–VcoJ(n). The solid curves are the average of 100 simulations; the dots are one standard deviations away from the averages. Parameters: $(\rho,\alpha,\gamma,\kappa,\mu,\Delta,n,\lambda,\sigma_J,\lambda^v,\sigma_{J^v},\rho_{Jp},\rho_{Jn}) = (-0.8,-1,0.5,5,0.05,1/252,\ 390,5,0.015,5,0.01,0.75,-0.75)$.

(see Fig. 11). The correction method as discussed in Section 7.1.3 works well for these non-prototypical models: see Table 3.

*7.2.2. Non-Heston stochastic volatility models*

We now employ a different stochastic volatility model where the log-variance follows an Ornstein-Uhlenbeck process. We first consider the $SV_1(J)$ model as in Andersen, Benzoni, and Lund (2002). We label this model as "OU(J)"

$$\frac{dS_t}{S_t} = (\mu - \lambda_t \bar{J})\,dt + \sqrt{v_t}\,dB_t + J_t\,dN_t, \tag{32}$$

$$d\ln v_t = \kappa(\alpha - \ln v_t)\,dt + \gamma\,dW_t, \tag{33}$$

where $\lambda_t = \lambda_0 + \lambda_1 v_t$, $\bar{J}$ is a parameter, and $J_t$ is such that $\ln(1+J_t) \sim \mathrm{N}(\ln(1+\bar{J}) - 0.5\sigma_J^2, \sigma_J^2)$. We choose the following (annualized) parameters which are in line with the empirical results in Andersen, Benzoni, and Lund (2002): $\alpha = -1$, $\mu = 0.05$, $\gamma = 0.5$, $\kappa = 5$, $\rho = -0.8$, $\bar{J} = 0$, $\lambda_1 = 0$, $\lambda_0 = 5$, $\sigma_J = 0.015$.

We further consider the case where there are additional jumps in volatilities, both independent jumps and co-jumps. More specifically, model "OU(J)–VindJ" refers to the case when $J_t^v\,dN_t^v$ is added to (33), where $N_t^v$ is a Poisson process with intensity $\lambda^v$ independent of $N_t$, and $J_t^v$ is independent of everything else following $\ln(1+J_t^v) \sim \mathrm{N}(0, \sigma_{J^v}^2)$. The model "OU(J)–VcoJ(i)" is when $N_t^v = N_t$ and $J_t^v$ is independent of $J_t$; "OU(J)–VcoJ(p)" is when $N_t^v = N_t$ and $J_t^v$ is positively correlated with $J_t$; "OU(J)–VcoJ(n)" is when $N_t^v = N_t$ and $J_t^v$ is negatively correlated with $J_t$. $J_t^v$ is again distributed such that $\ln(1+J_t^v) \sim \mathrm{N}(0, \sigma_{J^v}^2)$. We simulated OU(J)–VcoJ(p) and OU(J)–VcoJ(n) such that $\ln(1+J_t^v)$ and $\ln(1+J_t)$ are correlated with correlations $\rho_{Jp} = 0.75$ and $\rho_{Jn} = -0.75$, respectively. We again take $\sigma_{J^v}$ to be 0.01. Fig. 12 summarizes the raw correlations. The leftmost points on the plots are the averages of the naive estimates. They are about $-0.12$ for all the models OU(J), OU(J)–VindJ, OU(J)–VcoJ(i), OU(J)–VcoJ(p), and OU(J)–VcoJ(n), and are seriously biased. As $m$ increases, the smoothing biases decrease up to the point at which the discretization biases dominate.

Fig. 12 illustrates further that the leverage effect puzzle exists beyond the Heston model and the Heston model with jumps. Table 4 collects the results of the correction as used in Section 7.1.3. The results show that the bias correction, while derived from the Heston model, is also effective in the above alternative models.

## 8. Empirical evidence on the leverage effect at high frequency

We now apply our bias-corrected methods to examine the presence of the leverage effect using high frequency data. We have seen in Section 2 that, due to the latency of the volatility process, it is nearly impossible to use only returns data and no extraneous volatility proxy to get as nice a plot as what was shown in Fig. 4. Nevertheless, we will demonstrate that, after applying bias correction, we can uncover the presence of a strong leverage effect in high frequency data. We only focus on the data of S&P 500

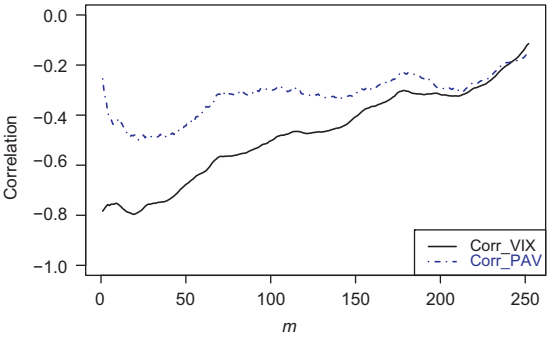


**Fig. 13.** S&P 500 leverage effects based on PAV and VIX. The raw sample correlations based on PAV with minute-by-minute data and VIX (squared) are plotted, for different horizons $m$, for S&P 500 in the time period 2004–2007.

**Table 4**
Performance of the bias correction based on nonparametrically estimated asymptotic quantities and linear regression for models OU(J), OU(J)–VindJ, OU(J)–VcoJ(i), OU(J)–VcoJ(p), and OU(J)–VcoJ(n), starting with the raw correlation based on truncated realized volatility. For each of these models, the 100 estimates of $\rho$ ($\hat{\rho}_{RV,TR}$) are summarized by their minimum, first quartile, median, third quartile, maximum, mean, and SD. Parameters: $(\rho,\alpha,\gamma,\kappa,\mu,\Delta,n,\sigma_J,\lambda,\sigma_{J^v},\lambda^v,\rho_{Jp},\rho_{Jn}) = (-0.8,-1,0.5,5,0.051/252,390,0.015,5,0.01,5,0.75,-0.75)$.

| Estimator/Model | Min. | 1st Qu. | Median | 3rd Qu. | Max. | Mean | SD |
|---|---|---|---|---|---|---|---|
| $\hat{\rho}_{RV,TR}$, OU(J) | −0.98 | −0.85 | −0.82 | −0.79 | −0.68 | −0.82 | 0.055 |
| $\hat{\rho}_{RV,TR}$, OU(J)–VindJ | −0.99 | −0.85 | −0.82 | −0.79 | −0.68 | −0.82 | 0.055 |
| $\hat{\rho}_{RV,TR}$, OU(J)–VcoJ(i) | −0.98 | −0.85 | −0.81 | −0.78 | −0.67 | −0.81 | 0.057 |
| $\hat{\rho}_{RV,TR}$, OU(J)–VcoJ(p) | −0.99 | −0.85 | −0.81 | −0.79 | −0.68 | −0.82 | 0.056 |
| $\hat{\rho}_{RV,TR}$, OU(J)–VcoJ(n) | −0.99 | −0.85 | −0.81 | −0.79 | −0.68 | −0.82 | 0.055 |

**Table 5**
The sample correlations at different horizons $m$ between the returns of S&P 500 (January 2004–December 2007) and the estimated changes of volatilities, using PAV with sampling frequencies at one per minute and VIX (squared).

| $m$ | 1 | 2 | 5 | 21 | 63 | 126 | 252 |
|---|---|---|---|---|---|---|---|
| Corr_PAV | −0.26 | −0.32 | −0.40 | −0.49 | −0.37 | −0.32 | −0.15 |
| Corr_VIX | −0.78 | −0.77 | −0.76 | −0.79 | −0.61 | −0.47 | −0.11 |

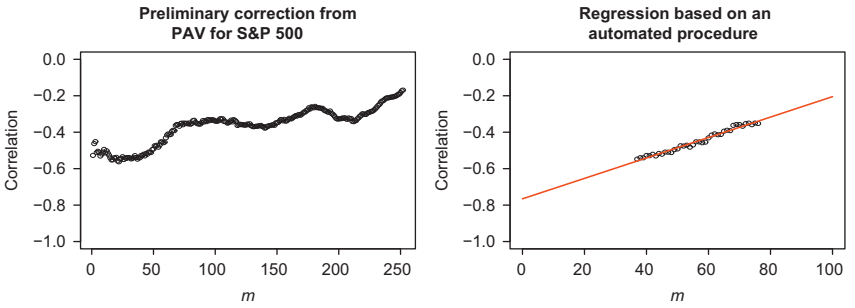


**Fig. 14.** Bias correction procedure applied to S&P 500 returns. The plot of the left is a scatter plot of preliminary bias-corrected estimates based on PAV of the leverage effect parameter $\rho$ against $m$ for S&P 500 2004–2007 minute-by-minute data. The plot on the right shows how estimates in the range identified by an automated procedure are further aggregated by using a simple linear regression to obtain a final estimate of the leverage effect.

**Table 6**
The sample correlations at different horizons $m$ between the returns (January 2005–June 2007) of Microsoft and the estimated changes of volatilities, using PAV with sampling frequencies at one per minute.

| $m$ | 1 | 2 | 5 | 10 | 21 | 63 | 126 | 252 |
|---|---|---|---|---|---|---|---|---|
| Corr_PAV | 0.03 | 0.00 | −0.02 | −0.04 | −0.17 | −0.34 | −0.40 | −0.28 |

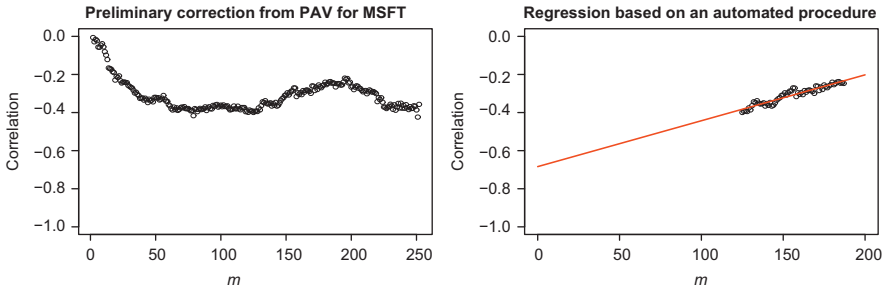


**Fig. 15.** Bias correction procedure applied to Microsoft returns. The plot on the left is a scatter plot of preliminary bias-corrected estimates based on PAV of the leverage effect parameter $\rho$ against $m$ based on the minute-by-minute data of Microsoft returns in the time period January 2005–June 2007. The plot on the right shows how estimates in the range identified by an automated procedure are further aggregated by using a simple linear regression to obtain a final estimate of the leverage effect.

and Microsoft Corp.; we have applied the methods to various other data sets and the conclusions are similar.

### 8.1. S&P 500

Based on the high frequency returns (one sample per minute) on S&P 500 futures from January 2004 to December 2007, the naive or natural estimates give the results reported in Table 5. The leverage effect at the natural choice of $m=1$ is nearly zero. This is the main message of the paper: the natural choice of $m=1$ leads to estimates that are seriously biased.

Even with the data-optimized choice of $m$, the correlation with PAV is around $-0.5$, significantly smaller than that computed based on VIX. Fig. 13 summarizes the sample correlations based on PAV and VIX, respectively, as a function of the horizon $m$. We note the loose resemblance between the two empirical curves in Fig. 13 and the predictions of the theory from Fig. 5.

We now apply our bias-corrected methods. First, we compute the preliminarily bias-corrected estimates $\hat{\rho}_m$ using PAV for $m$ in [1,252]. The scatter plot is presented in Fig. 14, which is somewhat curvy. As illustrated in Fig. 9, although on average we would expect to see a smooth curve like the black solid curve in Fig. 8, variations should be expected for individual sample paths. The automated data-driven procedure described in Section 7.1.2 (with $a_0=25$, $k_0=10$)[7] identifies a range of $m$ for the regression: $m=[37,76]$ (see Fig. 14) and our final estimate is $\hat{\rho}_{\mathrm{PAV}}=-0.77$.

The estimation with extrapolation has its own issues, including the dependence on the choice of the range of $m$ selected for the regression, and could overstate the true magnitude of $\rho$, but it is likely closer to it than what the a priori natural choice of $m=1$ produces.

[7] In practice, to best use the automated procedure, we should take into account the behavior of the preliminary corrections (see left panel of Figs. 14 and 15 when choosing $a_0$. We should use an $a_0$ close to the right boundary of the first "decreasing zone" to get rid of random effects for small $m$. Note that our automated procedure is fairly robust to the choice of $a_0$. Roughly, for $a_0$ between 20 and 40 for S&P 500 data, or between 5 and 100 for Microsoft data, the estimated leverage effect remains very close.

### 8.2. Microsoft

We now use our method to examine how strong the leverage effect is for Microsoft. The high frequency returns at sample frequencies of one data point per minute from January 2005 to June 2007 are used for estimating the leverage effect parameter. PAV is employed as the volatility estimator. Again, we apply both the naive method, the simple sample correlation, and the more sophisticated volatility estimation method, based on preliminary correction and linear regression. Table 6 summarizes the results of the simple sample correlations. As before, the leverage effect is barely noticeable for the natural choices of $m$ consisting of small values.

The preliminary corrections based on PAV are summarized on the left panel of Fig. 15 and the regression is illustrated on the right. Our automated procedure (with $a_0=25$, $k_0=10$) identifies the range $m=[125,187]$ for the linear extrapolation, which leads to an estimated leverage effect parameter $\hat{\rho}_{\mathrm{PAV}}=-0.68$.

## 9. Conclusions

We showed that there are different sources of error when estimating the leverage effect using high frequency data, a discretization error due to not observing the full instantaneous stochastic processes, a smoothing error due to using integrated volatility in place of spot volatilities, an estimation error due to the need to estimate the integrated volatility using the price process, and a noise correction error introduced by the need to correct the integrated volatility estimates for the presence of market microstructure noise. Perhaps paradoxically, attempts to improve the estimation by employing statistically better volatility estimators (such as noise-robust estimators) can actually make matters worse as far as the estimation of the leverage effect is concerned.

These errors tend to be large even when the window size is small and lead to significant bias in the leverage effect estimation. They are typically convex as a function of the length of time, controlled by $m$ which is used to compute changes in the volatilities and prices. These errors can have an adverse effect on the assessment of the leverage effect.

Fortunately, these errors are correctable to some extent. There is still a substantial discretization bias that remains when using the spot volatility over a longer time horizon, yet a reasonable large choice of $m$ is necessary so that biases based on integrated volatility become correctable. This led us to further correcting the biases by aggregating the information in various preliminary estimates of the leverage effect over different values of $m$. This is achieved by using a simple linear regression technique. The effectiveness of the method is demonstrated using both simulated examples and an empirical study of real asset returns.

Of course, to demonstrate the effect, our analysis necessarily proceeded by analyzing estimators based on the realized correlation between log-returns and realized volatility-type quantities. To be consistent with the theoretical analysis in the first part of the paper, what we proposed as a bias correction procedure is what's feasible given the constraint that we only consider estimators of that type. But, once the puzzle is identified and understood, a real solution must include the development from scratch of estimators that are not hindered by that constraint.

## Appendix A. Preliminary results

We first compute some moments that are related to the Heston model. They will be useful for proofs of the theorems. Throughout the Appendix, we use the notation $E_v$ and $\mathrm{Var}_v$ to denote the conditional mean and conditional variance given the latent volatility process $\{v_t\}$, and $E_t$ to denote the conditional expectation given the filtration up to time $t$. Other similar notations will be adopted.

### A.1. Conditional moments of returns

Rewrite the process as

$$dX_t=(\mu-v_t/2)\,dt+\rho v_t^{1/2}\,dW_t+\sqrt{1-\rho^2}v_t^{1/2}\,dZ_t,$$

where $Z_t$ is another Brownian motion process independent of $W$. Let $Y_t=\gamma X_t-\rho v_t$, which eliminates the $dW_t$ term. Then, it follows that

$$dY_t=[\gamma\mu-\rho\kappa\alpha+(\rho\kappa-\gamma/2)v_t]\,dt+\gamma\sqrt{1-\rho^2}v_t^{1/2}\,dZ_t.$$

Denoting by $a=\mu-\rho\kappa\alpha/\gamma$, $b=\rho\kappa/\gamma-1/2$, and $c=\rho/\gamma$, we have from the above expression that

$$X_u-X_s=\int_s^u\{(a+bv_t)\,dt+\sqrt{1-\rho^2}v_t^{1/2}\,dZ_t\}+c(v_u-v_s). \tag{34}$$

Hence, conditioning on the process $\{v_t\}$, $X_u-X_s$ is normally distributed with mean

$$E_v(X_u-X_s)=\int_s^u(a+bv_t)\,dt+c(v_u-v_s)\equiv\mu_v \tag{35}$$

and variance

$$\mathrm{Var}_v(X_u-X_s)=(1-\rho^2)\int_s^u v_t\,dt\equiv\sigma_v^2. \tag{36}$$

Using the moment formulas of the normal distribution, we can easily obtain the first four moments for the changes of the prices:

$$E_v(X_u-X_s)^2=\mu_v^2+\sigma_v^2,$$
$$E_v(X_u-X_s)^3=\mu_v^3+3\mu_v\sigma_v^2,$$
$$E_v(X_u-X_s)^4=\mu_v^4+3\sigma_v^4+6\mu_v^2\sigma_v^2.$$

### A.2. Cross-moments of the Feller process

We now compute the cross-moments of the Feller process $\{v_t\}$. First of all, it is well known that

$$E(v_t)=\alpha\quad\text{and}\quad\mathrm{Var}(v_t)=\frac{\gamma^2\alpha}{2\kappa}. \tag{37}$$

Using again Itô's formula, we have

$$d(e^{\kappa t}v_t)=\kappa\alpha e^{\kappa t}\,dt+\gamma e^{\kappa t}v_t^{1/2}\,dW_t,$$

which implies for $s>t$,

$$E(v_s|v_t)=e^{-\kappa(s-t)}v_t+\alpha(1-e^{-\kappa(s-t)}). \tag{38}$$

Similarly, by using Itô's formula again

$$d(e^{\kappa t}v_t)^2=(2\kappa\alpha+\gamma^2)e^{2\kappa t}v_t\,dt+2\gamma e^{2\kappa t}v_t^{3/2}\,dW_t.$$

This together with (38) implies that for $s>t$,

$$\begin{aligned}E(v_s^2|v_t)&=e^{-2\kappa(s-t)}v_t^2+e^{-2\kappa s}\int_t^s(2\kappa\alpha+\gamma^2)e^{2\kappa u}E(v_u|v_t)\,du\\&=e^{-2\kappa(s-t)}v_t^2+\frac{2\kappa\alpha+\gamma^2}{\kappa}(v_t-\alpha)(e^{-\kappa(s-t)}-e^{-2\kappa(s-t)})\\&\quad+\frac{2\kappa\alpha^2+\gamma^2\alpha}{2\kappa}(1-e^{-2\kappa(s-t)}).\end{aligned} \tag{39}$$

Therefore, for $r\le s$,

$$\begin{aligned}E(v_rv_s)&=E(v_rE(v_s|v_r))\\&=\mathrm{E}[v_r^2e^{-\kappa(s-r)}+\alpha(1-e^{-\kappa(s-r)})v_r]\\&=\alpha^2+\gamma^2\alpha e^{-\kappa(s-r)}/(2\kappa).\end{aligned} \tag{40}$$

Using the same technique, we can calculate higher moments and cross-moments. From

$$\begin{aligned}d(e^{\kappa t}v_t)^3&=3e^{2\kappa t}v_t^2(\kappa\alpha e^{\kappa t}\,dt+\gamma e^{\kappa t}v_t^{1/2}\,dW_t)\\&\quad+3(e^{\kappa t}v_t)\gamma^2e^{2\kappa t}v_t\,dt,\end{aligned} \tag{41}$$

we have

$$E(e^{\kappa t}v_t)^3=Ev_0^3+3\int_0^t(\kappa\alpha+\gamma^2)e^{3\kappa u}Ev_u^2\,du.$$

Using the fact that $Ev_t^3=Ev_0^3$ and $Ev_u^2=\alpha^2+\gamma^2\alpha/(2\kappa)$, we deduce that

$$Ev_t^3=\left(\alpha+\frac{\gamma^2}{\kappa}\right)\left(\alpha^2+\frac{\gamma^2\alpha}{2\kappa}\right).$$

Recall that $E_t$ denotes the conditional expectation given the filtration up to time $t$. For $s>t$, we deduce from (41) that

$$E_t(e^{\kappa s}v_s)^3=e^{\kappa t}v_t^3+\int_t^s(3\kappa\alpha+3\gamma^2)e^{3\kappa u}E_tv_u^2\,du.$$

Now, substituting (39) into the above expression, we obtain after some calculation that

$$\begin{aligned}E_tv_s^3&=e^{-3\kappa(s-t)}[v_t^3+3\beta_1(e^{\kappa(s-t)}-1)v_t^2+1.5\beta_1\beta_2(e^{2\kappa(s-t)}\\&\quad-2e^{\kappa(s-t)}+1)v_t+0.5\alpha\beta_1\beta_2(3e^{\kappa(s-t)}\\&\quad-3e^{2\kappa(s-t)}+e^{3\kappa(s-t)}-1)],\end{aligned} \tag{42}$$

where $\beta_1=\alpha+\gamma^2/\kappa$ and $\beta_2=2\alpha+\gamma^2/\kappa$. For $r<s<u$, by using conditional expectation and (38), we have

$$\begin{aligned}E(v_rv_sv_u)&=Ev_rv_s[\alpha+e^{-\kappa(u-s)}(v_s-\alpha)]\\&=\alpha E(v_rv_s)+e^{-\kappa(u-s)}\mathrm{E}[v_rE_r(v_s^2-\alpha v_s)].\end{aligned}$$

Substituting (38)–(40) into the above formula, the resulting expression involves only the first three moments of $v_r$, which has already been derived. Therefore, after some calculation, it follows that

$$\begin{aligned}E(v_rv_sv_u)&=\alpha^3+\frac{\gamma^2\alpha^2}{2\kappa}[e^{-\kappa(s-r)}+e^{-\kappa(u-r)}+e^{-\kappa(u-s)}\\&\quad+\gamma^2\kappa^{-1}\alpha^{-1}e^{-\kappa(u-r)}].\end{aligned} \tag{43}$$

The fourth-order cross-moment can be derived analogously using what has already been derived along with Itô's formula:

$$\begin{aligned}d(e^{\kappa t}v_t)^4&=4e^{3\kappa t}v_t^3(\kappa\alpha e^{\kappa t}\,dt+\gamma e^{\kappa t}v_t^{1/2}\,dW_t)\\&\quad+6(e^{2\kappa t}v_t^2)\gamma^2e^{2\kappa t}v_t\,dt.\end{aligned}$$

We omit the detailed derivations, but state the following results:

$$E(v_t^4)=\left(\alpha+\frac{3\gamma^2}{2\kappa}\right)\left(\alpha+\frac{\gamma^2}{\kappa}\right)\left(\alpha^2+\frac{\gamma^2\alpha}{2\kappa}\right),$$

and for $r<s<u<t$,

$$\begin{aligned}E(v_rv_sv_uv_t)&=\alpha^4+\frac{\alpha^3\gamma^2}{2\kappa}[e^{-\kappa(u-r)}+e^{-\kappa(t-r)}\\&\quad+e^{-\kappa(s-r)}+e^{-\kappa(u-s)}+e^{-\kappa(t-s)}+e^{-\kappa(t-u)}]\\&\quad+\frac{\alpha^2\gamma^4}{2\kappa^2}[e^{-\kappa(t+u-r-s)}+e^{-\kappa(u-r)}+2e^{-\kappa(t-r)}\\&\quad+e^{-\kappa(s+t-u-r)}/2+e^{-\kappa(t-s)}]\\&\quad+\frac{\alpha\gamma^6}{4\kappa^3}[e^{-\kappa(t+u-r-s)}+2e^{-\kappa(t-r)}].\end{aligned} \tag{44}$$

## Appendix B. Proofs

*Proof of Theorem 1 and Proposition 1.* Let us first compute the covariance. It follows from (34) that

$$\begin{aligned}&\mathrm{Cov}(v_{t+s}-v_t,X_{t+s}-X_t)\\&\quad=E(v_{t+s}-v_t)\left[\int_t^{t+s}(a+bv_u)\,du+c(v_{t+s}-v_t)\right]\\&\quad=b\int_t^{t+s}[E(v_uv_{t+s})-E(v_uv_t)]\,du+cE(v_{t+s}-v_t)^2.\end{aligned}$$

Now, using the moment formulas (37) and (40) and some simple calculus, we have

$$\mathrm{Cov}(v_{t+s}-v_t,X_{t+s}-X_t)=\alpha\gamma\rho[1-\exp(-s\kappa)]/\kappa.$$

By using (37) and (40) again, we easily obtain

$$\mathrm{Var}(v_{t+s}-v_s)=\gamma^2\alpha[1-\exp(-\kappa s)]/\kappa. \tag{45}$$

Hence, it remains to compute $\mathrm{Var}(X_{t+s}-X_t)$. By (35) and (36),

$$\begin{aligned}\mathrm{Var}(X_{t+s}-X_t)&=\mathrm{Var}(\mu_v)+E(\sigma_v^2)=E(\mu_v^2+\sigma_v^2)-(E(\mu_v))^2\\&=\int_t^{t+s}\int_t^{t+s}E(a+bv_r)(a+bv_u)\,dr\,du\\&\quad+2bc\int_t^{t+s}Ev_r(v_{t+s}-v_t)\,dr+c^2E(v_{t+s}-v_t)^2\\&\quad+(1-\rho^2)\alpha s-\left(\int_t^{t+s}a+bE(v_r)\,dr\right)^2.\end{aligned}$$

Using the moments for $v_t$ computed in Appendix A.2, after some calculus, we obtain

$$\mathrm{Var}(X_{t+s}-X_t)=\left(s+\frac{e^{-\kappa s}-1}{\kappa}\right)\left(\frac{\gamma^2\alpha}{4\kappa^2}-\frac{\gamma\alpha\rho}{\kappa}\right)+\alpha s. \tag{46}$$

Finally, combinations of the covariance and variance expressions lead to the correlation formula in Theorem 1.

Expanding the result of Theorem 1 around $s=0$, we obtain Proposition 1. □

*Proof of Theorem 2 and Proposition 2.* Recall $V_{t,\Delta} = \int_{t-\Delta}^{t} v_s\,ds$. Let us compute the variance of the change of the ideally estimated spot volatility. Note that $E(V_{t+m\Delta,\Delta} - V_{t,\Delta}) = 0$. Using the stationarity of the process $\{v_t\}$, we have

$$\begin{aligned}\mathrm{Var}(V_{t+m\Delta,\Delta} - V_{t,\Delta}) &= E(V_{t+m\Delta,\Delta} - V_{t,\Delta})^2 \\ &= 2\int_{t-\Delta}^{t}\int_{t-\Delta}^{t} Ev_s v_u\,ds\,du - 2\int_{t+(m-1)\Delta}^{t+m\Delta}\int_{t-\Delta}^{t} Ev_s v_u\,ds\,du.\end{aligned}$$

Now, by (40), the above variance is given by

$$\begin{aligned}&4\int_{t-\Delta}^{t}\int_{t-\Delta}^{u}\left[\alpha^2 + \frac{\gamma^2\alpha}{2\kappa}e^{-\kappa(u-s)}\right]ds\,du \\ &-2\int_{t+(m-1)\Delta}^{t+m\Delta}\int_{t-\Delta}^{t}\left[\alpha^2 + \frac{\gamma^2\alpha}{2\kappa}e^{-\kappa(u-s)}\right]ds\,du.\end{aligned}$$

Simple calculus leads to

$$\mathrm{Var}(V_{t+m\Delta,\Delta} - V_{t,\Delta}) = \alpha\gamma^2 B_2^2/4\kappa^3,$$

where $B_2$ is as given in Theorem 2. Comparing this with the variance of differenced spot volatilities, we have

$$\frac{\mathrm{Var}(V_{t+m\Delta,\Delta} - V_{t,\Delta})}{\Delta^2\,\mathrm{Var}(v_{t+m\Delta} - v_t)(1-1/3m)} = 1 + R_v^V, \tag{47}$$

where $R_v^V$ is $O(\Delta)$ for any fixed $m$ as $\Delta\to 0$, and $o(m\Delta)$ if $m\to\infty$ and $m\Delta\to 0$.

Next, we compute the covariance. By (35) and the double expectation formula, we have

$$\begin{aligned}&\mathrm{Cov}(V_{t+m\Delta,\Delta} - V_{t,\Delta}, X_{t+m\Delta} - X_t) \\ &= E\left[\int_{t+(m-1)\Delta}^{t+m\Delta} v_s\,ds - \int_{t-\Delta}^{t} v_s\,ds\right] \\ &\quad\times\left[\int_{t}^{t+m\Delta}(a+bv_r)\,dr + c(v_{t+m\Delta} - v_t)\right] \\ &= b\int_{t+(m-1)\Delta}^{t+m\Delta}\int_{t}^{t+m\Delta} E(v_s v_r)\,dr\,ds - b\int_{t-1}^{t}\int_{t}^{t+m\Delta} E(v_s v_r)\,dr\,ds \\ &\quad + c\int_{t+(m-1)\Delta}^{t+m\Delta} Ev_s(v_{t+m\Delta} - v_t)\,ds - c\int_{t-1}^{t} Ev_s(v_{t+m\Delta} - v_t)\,ds.\end{aligned}$$

Using (40), after some calculus, we obtain that

$$\mathrm{Cov}(V_{t+m\Delta,\Delta} - V_{t,\Delta}, X_{t+m\Delta} - X_t) = \alpha\gamma A_2/(4\kappa^3),$$

where $A_2$ is again as given in Theorem 2. The conclusion of Theorem 2 follows from (46) and the above results. Comparing this with the covariance based on the spot volatilities, we have

$$\frac{\mathrm{Cov}(V_{t+m\Delta,\Delta} - V_{t,\Delta}, X_{t+m\Delta} - X_t)}{\Delta\mathrm{Cov}(v_{t+m\Delta} - v_t, X_{t+m\Delta} - X_t)(1-1/2m)} = 1 + R_c^V, \tag{48}$$

where $R_c^V$ is $O(\Delta)$ for any fixed $m$ as $\Delta\to 0$, and $o(m\Delta)$ if $m\to\infty$ and $m\Delta\to 0$.

By (47) and (48), the following asymptotic expressions are easily obtained:

$$\begin{aligned}&\mathrm{Corr}(V_{t+m\Delta,\Delta} - V_{t,\Delta}, X_{t+m\Delta} - X_t) \\ &= \mathrm{Corr}(v_{t+m\Delta} - v_t, X_{t+m\Delta} - X_t)\frac{(2m-1)}{2\sqrt{m^2 - \frac{m}{3}}} \\ &\quad + \begin{cases} O(\Delta) & \text{when } \Delta\to 0 \text{ for any fixed } m, \\ o(m\Delta) & \text{when } m\to\infty, m\Delta\to 0,\end{cases}\end{aligned} \tag{49}$$

which proves Proposition 2. □

*Proof of Theorem 3 and Proposition 3.* The calculation is very involved, and we separate the steps into several subsections. Recall that $n = \Delta/\delta$. Without loss of generality, we assume that $t = \Delta$ and rewrite $\hat{V}^{\mathrm{RV}}_{\Delta,\Delta} = \hat{V}^{\mathrm{RV}}_{\Delta}$. Note that it is easy to verify that $E(\hat{V}^{RV}_{(m+1)\Delta,\Delta} - \hat{V}^{RV}_{\Delta}) = 0$.

(a) *Calculation of* $E[\hat{V}^{RV}_{\Delta}(X_{(m+1)\Delta} - X_\Delta)]$: Note that $\hat{V}^{\mathrm{RV}}_{\Delta}$ and $X_{(m+1)\Delta} - X_\Delta$ involve two different time intervals. By conditioning on the latent process $\{v_t\}$, $\hat{V}^{\mathrm{RV}}_{\Delta}$ and $X_{(m+1)\Delta} - X_\Delta$ are independent by (34). Thus,

$$E[\hat{V}^{\mathrm{RV}}_{\Delta}(X_{(m+1)\Delta} - X_\Delta)] = E[E_v\hat{V}^{\mathrm{RV}}_{\Delta}\;E_v(X_{(m+1)\Delta} - X_\Delta)].$$

Using (34)–(36), the above expectation is given by

$$\begin{aligned}&\sum_{i=0}^{n-1} E\left\{\left[\int_{i\delta}^{(i+1)\delta}(a+bv_r)\,dr + cv_{(i+1)\delta} - cv_{i\delta}\right]^2\right. \\ &\quad\left.+(1-\rho^2)\int_{i\delta}^{(i+1)\delta} v_r\,dr\right\} \\ &\quad\cdot\left\{\int_{\Delta}^{(m+1)\Delta}(a+bv_r)\,dr + cv_{(m+1)\Delta} - cv_\Delta\right\}.\end{aligned} \tag{50}$$

Expanding the first curly bracket into four terms, we have four product terms with the second curly bracket in (50). Denote those four terms by $I_1,\ldots,I_4$, respectively.

We now deal with each of the four terms. The first term is given by

$$\begin{aligned}I_1 &\equiv \sum_{i=0}^{n-1}\left\{\int_{\Delta}^{(m+1)\Delta}(a+bv_r)\,dr + cv_{(m+1)\Delta} - cv_\Delta\right\} \\ &\quad\left[\int_{i\delta}^{(i+1)\delta}(a+bv_r)\,dr\right]^2.\end{aligned}$$

Expressing the square-term above as the double integral, $I_1$ involves only the third cross-moment of the process $\{v_t\}$. By using (40) and (43), it follows that as $\Delta\to 0$ and $n\to\infty$ or as $m\to\infty$, $m\Delta\to 0$, and $n\to\infty$,

$$\begin{aligned}I_1 &= \sum_{i=0}^{n-1}\int_{\Delta}^{(m+1)\Delta}\int_{i\delta}^{(i+1)n\delta}\int_{i\delta}^{r} -2\frac{\gamma^4 a^3}{2\kappa^2\alpha^2}e^{-\kappa(s-u)}\,du\,dr\,ds \\ &\quad + \sum_{i=0}^{n-1}\int_{i\delta}^{(i+1)\delta}\int_{i\delta}^{r} 2ca^2\frac{\gamma^4}{2\kappa^2\alpha}[e^{-\kappa((m+1)\Delta-u)} - e^{-\kappa(\Delta-u)}]\,du\,dr \\ &= -\frac{a^2\gamma^4(a+\alpha c\kappa)}{2\alpha^2\kappa^2}\frac{m\Delta^3}{n} + o\left(\frac{m\Delta^3}{n}\right),\end{aligned}$$

in particular,

$$\frac{I_1}{m\Delta^2} = O\left(\frac{\Delta}{n}\right).$$

Using the same argument, the second term can be calculated as follows:

$$\begin{aligned}I_2 &\equiv 2c\sum_{i=0}^{n-1}\left\{\int_{\Delta}^{(m+1)\Delta}(a+bv_r)\,dr + cv_{(m+1)\Delta} - cv_\Delta\right\} \\ &\quad\times\left[\int_{i\delta}^{(i+1)\delta}(a+bv_r)\,dr\right](v_{(i+1)\delta} - v_{i\delta}) \\ &= \sum_{i=0}^{n-1}\int_{\Delta}^{(m+1)\Delta}\int_{i\delta}^{(i+1)\delta} 2ca^2\frac{\gamma^4}{2\kappa^2\alpha}[e^{-\kappa(s-r)} - e^{-\kappa(s-i\delta)}]\,dr\,ds\end{aligned}$$

$$+\int_{i\delta}^{(i+1)\delta} 2ac^2\frac{\gamma^4}{2\kappa^2}[e^{-\kappa((m+1)-i\delta)}-e^{-\kappa((m+1)\Delta-r)}-e^{-\kappa(\Delta-i\delta)} +e^{-\kappa(\Delta-r)}]\,dr,$$

where the cross-moment function of the process $\{v_t\}$ is used. We have as $\Delta\to 0$ and $n\to\infty$ or as $m\to\infty$, $m\Delta\to 0$, and $n\to\infty$,

$$I_2=\frac{ac\gamma^4(a+\alpha c\kappa)}{2\alpha\kappa}\frac{m\Delta^3}{n}+o\left(\frac{m\Delta^3}{n}\right),$$

hence,

$$\frac{I_2}{m\Delta^2}=O\left(\frac{\Delta}{n}\right).$$

Similarly, we can calculate the third term and the fourth term based on the cross-moments of the process $\{v_t\}$. They are given by

$$I_3=-c^2\gamma^4(e^{\Delta\kappa}-1)e^{-\Delta\kappa(m+1)}(e^{\Delta\kappa m}-1)(a+\alpha c\kappa)/(2\kappa^3),$$
$$I_4=\gamma^2(\rho^2-1)(e^{\Delta\kappa}-1)e^{-\Delta\kappa(m+1)}(e^{\Delta\kappa m}-1)(a+\alpha c\kappa)/(2\kappa^3).$$

(b) *Calculation of* $E\hat{V}^{RV}_{(m+1)\Delta,\Delta}(X_{(m+1)\Delta}-X_\Delta)$ *and the covariance*: By the definition of $\hat{V}^{RV}_{(m+1)\Delta,\Delta}$, it follows that

$$\begin{aligned}E\hat{V}^{RV}_{(m+1)\Delta,\Delta}&(X_{(m+1)\Delta}-X_\Delta)\\ &=E\sum_{i=0}^{n-1}[X_{m\Delta+(i+1)\delta}-X_{m\Delta+i\delta}]^2\\ &\quad\times\{(X_{m\Delta+i\delta}-X_\Delta)+(X_{m\Delta+(i+1)\delta}-X_{m\Delta+i\delta})\\ &\quad+(X_{(m+1)\Delta}-X_{m\Delta+(i+1)\delta})\}.\end{aligned}$$

Let $J_1$, $J_2$, and $J_3$ be, respectively, the product of the first, second, and third term in the curly bracket with that in the square bracket. Each of these terms can be treated similarly as those in Appendix B. That is, by conditioning on the process $\{v_t\}$, they can be reduced to the calculation of the cross-moments of $\{v_t\}$, by using the conditional moments in Appendix A.1. After tedious calculations involving the cross-moments discussed in Appendix A.2, we can obtain asymptotic expressions for $J_1$, $J_2$, and $J_3$. Using these together with what we get for $I_1,\ldots,I_4$, we can easily obtain an asymptotic expression of $\mathrm{Cov}(\hat{V}^{RV}_{(m+1)\Delta,\Delta}-\hat{V}^{RV}_{\Delta},X_{(m+1)\Delta}-X_\Delta)$. Comparing this asymptotic expression with what we have obtained in Theorem 2, we conclude that, as $\Delta\to 0$, and $n\Delta\to C\in(0,\infty)$,

$$\begin{aligned}m^{-1}\Delta^{-2}&\,\mathrm{Cov}(\hat{V}^{RV}_{(m+1)\Delta,\Delta}-\hat{V}^{RV}_{\Delta},X_{(m+1)\Delta}-X_\Delta)\\ &=m^{-1}\Delta^{-2}\,\mathrm{Cov}(V_{(m+1)\Delta,\Delta}-V_{\Delta,\Delta},X_{(m+1)\Delta}-X_\Delta)\\ &\quad+\begin{cases}O(\Delta) & \text{for any fixed } m,\\ o(m\Delta) & \text{when } m\to\infty \text{ and } m\Delta\to 0.\end{cases}\end{aligned}\tag{51}$$

(c) *Calculation of the variance of changes of estimated RV*: Let $Y_i=X_{(i+1)\delta}-X_{i\delta}$. Then,

$$E(\hat{V}^{RV}_{\Delta})^2=\sum_{i=0}^{n-1}EY_i^4+2\sum_{i=1}^{n-1}\sum_{j=0}^{i-1}EY_i^2Y_j^2.\tag{52}$$

By using the expression at the end of Appendix A, we have

$$\begin{aligned}EY_i^4=E&\left(\int_{i\delta}^{(i+1)\delta}(a+bv_r)\,dr+cv_{(i+1)\delta}-cv_{i\delta}\right)^4\\ &+3E\left((1-\rho^2)\int_{i\delta}^{(i+1)\delta}v_r\,dr\right)^2\\ &+6(1-\rho^2)E\int_{i\delta}^{(i+1)\delta}v_r\,dr\\ &\cdot\left(\int_{i\delta}^{(i+1)\delta}(a+bv_r)\,dr+cv_{(i+1)\delta}-cv_{i\delta}\right)^2.\end{aligned}$$

By conditioning on the process $\{v_t\}$, $Y_i^2$ and $Y_j^2$ are conditionally independent for $j<i$. Appealing to (35) and (36), we have that for $j<i$,

$$\begin{aligned}EY_i^2Y_j^2=E\Bigg\{&\left[\left(\int_{i\delta}^{(i+1)\delta}(a+bv_r)\,dr+cv_{(i+1)\delta}-cv_{i\delta}\right)^2\right.\\ &\left.+(1-\rho^2)\int_{i\delta}^{(i+1)\delta}v_r\,dr\right]\\ &\cdot\left[\left(\int_{j\delta}^{(j+1)\delta}(a+bv_r)\,dr+cv_{(j+1)\delta}-cv_{j\delta}\right)^2\right.\\ &\left.+(1-\rho^2)\int_{j\delta}^{(j+1)\delta}v_r\,dr\right]\Bigg\}.\end{aligned}$$

Both terms above only involve the cross-moments of the process $\{v_t\}$. After tedious calculations, we obtain

$$\begin{aligned}E(\hat{V}^{RV}_{\Delta})^2=&\left(\alpha^2+\frac{\alpha\gamma^2}{2\kappa}\right)\Delta^2-\frac{\alpha\gamma^2}{6}\Delta^3\\ &+\frac{\alpha(3\rho^4-6\rho^2+4)(2\alpha\kappa+\gamma^2)}{\kappa}\frac{\Delta^2}{n}+O(\Delta^4),\end{aligned}\tag{53}$$

when $\Delta\to 0$ and $n\Delta\to C\in(0,\infty)$. This is the same for $E(\hat{V}^{RV}_{(m+1)\Delta,\Delta})^2$.

By conditioning on the process $\{v_t\}$, using the conditional independence, we have

$$\begin{aligned}E\hat{V}^{RV}_{\Delta}\hat{V}^{RV}_{(m+1)\Delta,\Delta}=&\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}E\left[\left(\int_{i\delta}^{(i+1)\delta}(a+bv_r)\,dr+cv_{(i+1)\delta}-cv_{i\delta}\right)^2\right.\\ &\left.+(1-\rho^2)\int_{i\delta}^{(i+1)\delta}v_r\,dr\right]\\ &\cdot\left[\left(\int_{j\delta}^{(j+1)\delta}(a+bv_{m\Delta+r})\,dr+cv_{m\Delta+(j+1)\delta}\right.\right.\\ &\left.\left.-cv_{m\Delta+j\delta}\right)^2.+(1-\rho^2)\int_{j\delta}^{(j+1)\delta}v_{m\Delta+r}\,dr\right].\end{aligned}$$

Additional calculations involving the cross-moments of the process $\{v_t\}$ yield

$$E\hat{V}^{RV}_{\Delta}\hat{V}^{RV}_{(m+1)\Delta,\Delta}=\frac{\alpha(2\alpha\kappa+\gamma^2)}{\kappa}\Delta^2-\alpha\gamma^2m\Delta^3+O(m\Delta^4),\tag{54}$$

as $\Delta\to 0$ and $n\Delta\to C\in(0,\infty)$, for fixed $m$ or when $m\to\infty$ and $m\Delta\to 0$. Combination of (53) and (54) results in

$$\begin{aligned}\mathrm{Var}(\hat{V}^{RV}_{(m+1)\Delta,\Delta}-\hat{V}^{RV}_{\Delta})&=E(\hat{V}^{RV}_{\Delta})^2+E(\hat{V}^{RV}_{(m+1)\Delta,\Delta})^2-2E(\hat{V}^{RV}_{\Delta}\hat{V}^{RV}_{(m+1)\Delta,\Delta})\\ &=2\left[\left(\alpha^2+\frac{\alpha\gamma^2}{2\kappa}\right)\Delta^2-\frac{\alpha\gamma^2}{6}\Delta^3+\frac{\alpha(2\alpha\kappa+\gamma^2)}{\kappa}\frac{\Delta^2}{n}\right]\\ &\quad-2\left[\frac{\alpha(2\alpha\kappa+\gamma^2)}{\kappa}\Delta^2-\alpha\gamma^2m\Delta^3\right]+O(m\Delta^4),\end{aligned}\tag{55}$$

as $\Delta\to 0$ and $n\Delta\to C\in(0,\infty)$, for fixed $m$ or when $m\to\infty$ and $m\Delta\to 0$. Comparing this with the variance expression obtained in the proof of Theorem 2, we have

$$\mathrm{Var}(\hat{V}^{\mathrm{RV}}_{(m+1)\Delta,\Delta}-\hat{V}^{\mathrm{RV}}_{\Delta})=\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})+\frac{2\alpha(2\alpha\kappa+\gamma^2)}{\kappa}\frac{\Delta^2}{n}+\mathrm{O}(m\Delta^4),$$

or, equivalently,

$$m^{-1}\Delta^{-3}\,\mathrm{Var}(\hat{V}^{\mathrm{RV}}_{(m+1)\Delta,\Delta}-\hat{V}^{\mathrm{RV}}_{\Delta})=m^{-1}\Delta^{-3}\left(\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})+\frac{4E\sigma_t^4\Delta^2}{n}\right)+\mathrm{O}(\Delta),\tag{56}$$

$$m^{-1}\Delta^{-3}\,\mathrm{Var}(\hat{V}^{\mathrm{RV}}_{(m+1)\Delta,\Delta}-\hat{V}^{\mathrm{RV}}_{\Delta})=m^{-1}\Delta^{-3}\,\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})+\frac{2\alpha(2\alpha\kappa+\gamma^2)}{\kappa Cm}+\mathrm{O}(\Delta),\tag{57}$$

as $\Delta\to 0$ and $n\Delta\to C\in(0,\infty)$, for fixed $m$ or when $m\to\infty$ and $m\Delta\to 0$.

(d) *Adjustment to the leverage parameter*: Further, from (45) and (47), we have

$$m^{-1}\Delta^{-3}\,\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})=\gamma^2\alpha-\frac{\gamma^2\alpha}{3m}+\begin{cases}\mathrm{O}(\Delta) & \text{for any fixed } m,\\ -\frac{1}{2}\alpha\gamma^2\kappa m\Delta+\mathrm{o}(m\Delta) & \text{when } m^2\Delta\to C_m\in(0,\infty),\end{cases}\tag{58}$$

and (56) becomes

$$m^{-1}\Delta^{-3}\,\mathrm{Var}(\hat{V}^{\mathrm{RV}}_{(m+1)\Delta,\Delta}-\hat{V}^{\mathrm{RV}}_{\Delta})=m^{-1}\Delta^{-3}\,\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})\times\left(1+\frac{6(2\alpha\kappa+\gamma^2)}{(3\gamma^2m-\gamma^2)\kappa C-\frac{3}{2}\gamma^2\kappa^2CC_m}\right)+R_v^{RV},\tag{59}$$

where $R_v^{RV}$ is $\mathrm{O}(\Delta)$ for fixed $m$, with $C_m$ above replaced by 0; and $\mathrm{o}(m\Delta)$ when $m^2\Delta\to C_m\in(0,\infty)$.

By using (51), (59), and (49), we can easily obtain the following relationship:

$$\begin{aligned}&\mathrm{Corr}(\hat{V}^{\mathrm{RV}}_{(m+1)\Delta,\Delta}-\hat{V}^{\mathrm{RV}}_{\Delta},X_{(m+1)\Delta}-X_{\Delta})\\&=\mathrm{Corr}(V_{(m+1)\Delta}-V_{\Delta},X_{(m+1)\Delta}-X_{\Delta}))\cdot\frac{1}{\sqrt{1+\dfrac{12\alpha\kappa+6\gamma^2}{(3\gamma^2m-\gamma^2)\kappa C-\frac{3}{2}\gamma^2\kappa^2CC_m}}}+R_c^{RV}\\&=\mathrm{Corr}(v_{m+t}-v_t,X_{m+t}-X_t)\cdot\frac{1-1/2m}{\sqrt{\left(1+\dfrac{12\alpha\kappa+6\gamma^2}{(3\gamma^2m-\gamma^2)\kappa C-\frac{3}{2}\gamma^2\kappa^2CC_m}\right)\left(1-\dfrac{1}{3m}\right)}}+R_c^{RV},\end{aligned}$$

where $R_c^{RV}$ is $\mathrm{O}(\Delta)$ for fixed $m$, with $C_m$ above replaced by 0; and $\mathrm{o}(m\Delta)$ when $m^2\Delta\to C_m\in(0,\infty)$.

Proposition 3 follows from (49), (51), and (56). □

We now state the result parallel to Theorem 4 for TSRV. Let $\theta_{\mathrm{TSRV}}$ be a constant, $L=[\theta_{\mathrm{TSRV}}n^{2/3}]$ be the number of grids over which the subsampling is performed, and $\overline{n}=(n-L+1)/n$. The TSRV estimator is defined as

$$\hat{V}^{\mathrm{TSRV}}_{t,\Delta}=\frac{1}{L}\sum_{i=0}^{n-L}(Z_{t-\Delta+(i+L)\delta}-Z_{t-\Delta+i\delta})^2-\frac{\overline{n}}{n}\sum_{i=0}^{n-1}(Z_{t-\Delta+(i+1)\delta}-Z_{t-\Delta+i\delta})^2.\tag{60}$$

*Theorem 6. When $\Delta\to 0$, $n^{1/3}\Delta\to C_{\mathrm{TSRV}}$, and $\sigma_\epsilon^2/\Delta\to C_\epsilon$ with $C_{\mathrm{TSRV}}$ and $C_\epsilon\in(0,\infty)$, the following expansion shows the incremental bias due to estimation error and noise correction induced by the use of TSRV*:

$$\begin{aligned}&\mathrm{Corr}(\hat{V}^{\mathrm{TSRV}}_{t+m\Delta,\Delta}-\hat{V}^{\mathrm{TSRV}}_{t,\Delta},Z_{t+m\Delta}-Z_t)\\&=\mathrm{Corr}(v_{t+m\Delta}-v_t,X_{t+m\Delta}-X_t)\frac{(2m-1)}{2\sqrt{m^2-m/3}}\\&\quad\times(1+A_6+B_6)^{-1/2}+R_r^{\mathrm{TSRV}},\end{aligned}$$

*where*

$$A_6=\frac{96\theta_{\mathrm{TSRV}}^{-2}C_\epsilon^2}{C_{\mathrm{TSRV}}\alpha\gamma^2(6m-2-3\kappa C_m)}$$

*and*

$$B_6=\frac{8\theta_{\mathrm{TSRV}}(2\alpha\kappa+\gamma^2)}{\kappa C_{\mathrm{TSRV}}\gamma^2(6m-2-3\kappa C_m)}$$

*and $R_r^{\mathrm{TSRV}}$ is $\mathrm{O}(\Delta)$ for any fixed $m$, with $C_m$ above replaced by*0; *and $\mathrm{o}(m\Delta)$ when $m^2\Delta\to C_m\in(0,\infty)$.*

A result parallel to Proposition 4 for TSRV is the following.

*Proposition 6. When $\Delta\to 0$, $n^{1/3}\Delta\to C_{\mathrm{TSRV}}$, $\sigma_\epsilon^2/\Delta\to C_\epsilon$, and $m^2\Delta\to C_m$ with $C_{\mathrm{TSRV}}$, $C_\epsilon$, and $C_m\in(0,\infty)$,*

$$\begin{aligned}&\mathrm{Corr}(v_{t+m\Delta}-v_t,X_{t+m\Delta}-X_t)\\&=c_6\frac{2\sqrt{m^2-m/3}}{(2m-1)}\mathrm{Corr}(\hat{V}^{\mathrm{TSRV}}_{t+m\Delta,\Delta}-\hat{V}^{\mathrm{TSRV}}_{t,\Delta},Z_{t+m\Delta}-Z_t)+\mathrm{o}(m\Delta),\end{aligned}$$

*where*

$$c_6=\left(1-\frac{48\theta_{\mathrm{TSRV}}^{-2}\sigma_\epsilon^4+8\theta_{\mathrm{TSRV}}\mathrm{E}[\sigma_t^4]\Delta^2}{3n^{1/3}\,\mathrm{Var}(\hat{V}^{\mathrm{TSRV}}_{t+m\Delta,\Delta}-\hat{V}^{\mathrm{TSRV}}_{t,\Delta})}\right)^{-1/2}.\tag{61}$$

Two unknown quantities are involved and can be estimated nonparametrically here. For $\sigma_\epsilon$, we have under our model that $\mathrm{E}(\hat{V}^{\mathrm{RV}}_{t,\Delta}-\hat{V}^{\mathrm{TSRV}}_{t,\Delta})/2n=\sigma_\epsilon^2(1+\mathrm{o}(1))$. A long-run average of $(\hat{V}^{\mathrm{RV}}_{t,\Delta}-\hat{V}^{\mathrm{TSRV}}_{t,\Delta})/2n$ can be used as a good estimate of $\sigma_\epsilon^2$. This is similar to the way the average of the subsampled RV estimators is bias-corrected to construct TSRV. For $\mathrm{E}[\sigma_t^4]$, consistent noise-robust estimators of $\int_{t-\Delta}^t\sigma_s^4\,ds$ are proposed in Zhang, Mykland, Aït-Sahalia (2005) and Jacod, Li, Mykland, Podolskij, and Vetter (2009). We can use, for instance, the estimator called $\hat{Q}_t^n$ in the latter paper:

$$\hat{Q}_t^n=\frac{1}{3\theta_{\mathrm{PAV}}^2\psi_2^2\Delta}\sum_{i=0}^{n-k_n+1}\left(\frac{1}{k_n}\sum_{j=\lfloor k_n/2\rfloor}^{k_n-1}Z_{t-\Delta+(i+j)\delta}\right.$$

$$-\frac{1}{k_n}\sum_{j=0}^{\lfloor k_n/2\rfloor-1} Z_{t-\Delta+(i+j)\delta}\Bigg)^4 - \frac{\delta}{\theta_{\mathrm{PAV}}^4\psi_2^2\Delta^2}\sum_{i=0}^{n-2k_n+1}\Bigg(\Bigg(\frac{1}{k_n}\sum_{j=\lfloor k_n/2\rfloor}^{k_n-1} Z_{t-\Delta+(i+j)\delta} - \frac{1}{k_n}\sum_{j=0}^{\lfloor k_n/2\rfloor-1} Z_{t-\Delta+(i+j)\delta}\Bigg)^2 \times \sum_{j=i+k_n}^{i+2k_n-1}(Z_{t-\Delta+(j+1)\delta}-Z_{t-\Delta+j\delta})^2\Bigg) + \frac{\delta}{4\theta_{\mathrm{PAV}}^4\psi_2^2\Delta^2}\sum_{i=1}^{n-2}(Z_{t-\Delta+(i+1)\delta}-Z_{t-\Delta+i\delta})^2 \times (Z_{t-\Delta+(i+3)\delta}-Z_{t-\Delta+(i+2)\delta})^2, \tag{62}$$

where $\psi_2=\frac{1}{12}$, $k_n=[\theta_{\mathrm{PAV}}\sqrt{n}]$ for an appropriately chosen $\theta_{\mathrm{PAV}}$ for any $\Delta$. A scaled long-run average of this estimator can be used to estimate $\mathrm{E}[\sigma_t^4]$, based on the fact that $E(\hat{Q}_t^n)=\Delta^2 E\sigma_t^4(1+\mathrm{o}(1))$.

*Proof of Theorem 6, Proposition 6 and Theorem 4, Proposition 4.* Under the assumptions that $n^{1/3}\Delta\to C_{\mathrm{TSRV}}\in(0,\infty)$ and $\sigma_\epsilon^2/\Delta\to C_\epsilon\in(0,\infty)$, we have

$$m^{-1}\Delta^{-2}\,\mathrm{Cov}(\hat{V}^{\mathrm{TSRV}}_{(m+1)\Delta,\Delta}-\hat{V}^{\mathrm{TSRV}}_{\Delta},Z_{(m+1)\Delta}-Z_\Delta) = m^{-1}\Delta^{-2}\,\mathrm{Cov}(V_{(m+1)\Delta,\Delta}-V_{\Delta,\Delta},X_{(m+1)\Delta}-X_\Delta)+R_c^{\mathrm{TSRV}} \tag{63}$$

and

$$m^{-1}\Delta^{-3}\,\mathrm{Var}(\hat{V}^{\mathrm{TSRV}}_{(m+1)\Delta,\Delta}-\hat{V}^{\mathrm{TSRV}}_{\Delta}) = m^{-1}\Delta^{-3}\left(\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})+\frac{16\theta_{\mathrm{TSRV}}^{-2}\sigma_\epsilon^4}{n^{1/3}} + \frac{8\theta_{\mathrm{TSRV}}E\sigma_t^4\Delta^2}{3n^{1/3}}\right)+R_\nu^{\mathrm{TSRV}} \tag{64}$$

$$m^{-1}\Delta^{-3}\,\mathrm{Var}(\hat{V}^{\mathrm{TSRV}}_{(m+1)\Delta,\Delta}-\hat{V}^{\mathrm{TSRV}}_{\Delta}) = m^{-1}\Delta^{-3}\,\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})+\frac{16\theta_{\mathrm{TSRV}}^{-2}C_\epsilon^2}{mC_{\mathrm{TSRV}}}+\frac{8\theta_{\mathrm{TSRV}}E\sigma_t^4}{3mC_{\mathrm{TSRV}}}+R_\nu^{\mathrm{TSRV}} = m^{-1}\Delta^{-3}\,\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})[1+A_6+B_6]+R_\nu^{\mathrm{TSRV}}, \tag{65}$$

where by (58) $A_6=96\theta_{\mathrm{TSRV}}^{-2}C_\epsilon^2/C_{\mathrm{TSRV}}\alpha\gamma^2(6m-2-3\kappa C_m)$, $B_6=8\theta_{\mathrm{TSRV}}(2\alpha\kappa+\gamma^2)/\kappa C_{\mathrm{TSRV}}\gamma^2(6m-2-3\kappa C_m)$, and $R_c^{\mathrm{TSRV}}$ and $R_\nu^{\mathrm{TSRV}}$ are $\mathrm{O}(\Delta)$ for any fixed $m$, with $C_m$ above replaced by 0; and $\mathrm{o}(m\Delta)$ when $m^2\Delta\to C_m\in(0,\infty)$.

Under the assumptions that $n^{1/2}\Delta\to C_{\mathrm{PAV}}\in(0,\infty)$ and $\sigma_\epsilon^2/\Delta\to C_\epsilon\in(0,\infty)$, with the constants $\psi_2$, $\Phi_{11}$, $\Phi_{12}$, $\Phi_{22}$ as specified in Theorem 4, we have

$$m^{-1}\Delta^{-2}\,\mathrm{Cov}(\hat{V}^{\mathrm{PAV}}_{(m+1)\Delta,\Delta}-\hat{V}^{\mathrm{PAV}}_{\Delta},Z_{(m+1)\Delta}-Z_\Delta) = m^{-1}\Delta^{-2}\,\mathrm{Cov}(V_{(m+1)\Delta,\Delta}-V_{\Delta,\Delta},X_{(m+1)\Delta}-X_\Delta)+R_c^{\mathrm{PAV}} \tag{66}$$

and

$$m^{-1}\Delta^{-3}\,\mathrm{Var}(\hat{V}^{\mathrm{PAV}}_{(m+1)\Delta,\Delta}-\hat{V}^{\mathrm{PAV}}_{\Delta}) = m^{-1}\Delta^{-3}\left(\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})+\frac{8\Phi_{22}\theta_{\mathrm{PAV}}E\sigma_t^4\Delta^2}{\psi_2^2 n^{1/2}} + \frac{16\Phi_{12}E\sigma_t^2\sigma_\epsilon^2\Delta}{\theta_{\mathrm{PAV}}\psi_2^2 n^{1/2}}+\frac{8\Phi_{11}\sigma_\epsilon^4}{\theta_{\mathrm{PAV}}^2\psi_2^2 n^{1/2}}\right)+R_\nu^{\mathrm{PAV}} \tag{67}$$

$$m^{-1}\Delta^{-3}\,\mathrm{Var}(\hat{V}^{\mathrm{PAV}}_{(m+1)\Delta,\Delta}-\hat{V}^{\mathrm{PAV}}_{\Delta}) = m^{-1}\Delta^{-3}\,\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})+\frac{8\Phi_{22}\theta_{\mathrm{PAV}}E\sigma_t^4}{m\psi_2^2C_{\mathrm{PAV}}} + \frac{16\Phi_{12}E\sigma_t^2C_\epsilon}{m\theta_{\mathrm{PAV}}\psi_2^2C_{\mathrm{PAV}}}+\frac{8\Phi_{11}C_\epsilon^2}{m\theta_{\mathrm{PAV}}^2\psi_2^2C_{\mathrm{PAV}}}+R_\nu^{\mathrm{PAV}} = m^{-1}\Delta^{-3}\,\mathrm{Var}(V_{t+m\Delta,\Delta}-V_{t,\Delta})[1+A_4+B_4+C_4]+R_\nu^{\mathrm{PAV}}, \tag{68}$$

where

$A_4=24\Phi_{22}\theta_{\mathrm{PAV}}(2\alpha\kappa+\gamma^2)/\psi_2^2C_{\mathrm{PAV}}\kappa\gamma^2(6m-2-3\kappa C_m)$,

$B_4=96\Phi_{12}C_\epsilon/\theta_{\mathrm{PAV}}\psi_2^2C_{\mathrm{PAV}}\gamma^2(6m-2-3\kappa C_m)$,

$C_4=48\Phi_{11}C_\epsilon^2/\theta_{\mathrm{PAV}}^2\psi_2^2C_{\mathrm{PAV}}\alpha\gamma^2(6m-2-3\kappa C_m)$, and $R_c^{\mathrm{PAV}}$ and $R_\nu^{\mathrm{PAV}}$ are $\mathrm{O}(\Delta)$ for any fixed $m$, with $C_m$ above replaced by 0; and $\mathrm{o}(m\Delta)$ when $m^2\Delta\to C_m\in(0,\infty)$.

Theorem 6 follows from (63) and (65), and Theorem 4 from (66) and (68). Proposition 6 follows from (63) and (64), and Proposition 4 from (66) and (67). □

*Proof of Theorem 5 and Proposition 5.* Let $X^c$ be the continuous part of the log-price process: $dX^c=(\mu-v_t/2)\,dt+v_t^{1/2}\,dB_t$. The proof is based on comparing the covariance and variances with jump components to those without the jump component so that the previous calculations can be used. We first introduce some notation to facilitate the technical arguments. Let

$$I_t^j=\{i:0\le i\le n-1,\ X \text{ process has jumps between } t+i\delta \text{ and } t+(i+1)\delta\},$$

and

$$I_t^c=\{i:0\le i\le n-1,\ X \text{ process has no jump between } t+i\delta \text{ and } t+(i+1)\delta\}$$

be, respectively, the set of time indices for the process $X_t$ with jumps and without jumps, and

$$\Delta X_{t,i}^c=X^c_{t-\Delta+(i+1)\delta}-X^c_{t-\Delta+i\delta},$$

$$\Delta X_{t,i}=X_{t-\Delta+(i+1)\delta}-X_{t-\Delta+i\delta}.$$

Define $\hat{V}^{\bar{RV}}_{t,\Delta}=\sum_{i=0}^{\Delta/\delta-1}(\Delta X^c_{t,i})^2$ the quadratic variation of the continuous part,

$$A_{1,t}=\sum_{i\in I_t^j}(\Delta X^c_{t,i})^2,\quad A_{2,t}=\sum_{i\in I_t^j}(\Delta X_{t,i})^2 1_{\{|\Delta X_{t,i}|\le a\delta^{\varpi}\}};$$

and

$$A_{3,t}=\sum_{i\in I_t^c}(\Delta X_{t,i})^2 1_{\{|\Delta X_{t,i}|>a\delta^{\varpi}\}}.$$

Then, we have the following simple relation:

$$\hat{V}^{\mathrm{RV,TR}}_{t,\Delta}=\hat{V}^{\bar{RV}}_{t,\Delta}-A_{1,t}+A_{2,t}-A_{3,t}.$$

Our aim is to show $A_{i,t}$ for $i=1,2,3$ are negligible by evaluating their second moments.

To this end, let $P_{J,\delta}=P\{\text{at least one jump between } (0,\delta)\}$, which is of order $O(\delta)$. By independence of the jump and continuous parts of the $X$ process, we have

$$\begin{aligned}E(A_{1,t})^2 &= E\left(\sum_{i=0}^{n-1}(\Delta X^c_{t,i})^4 1_{i\in I^J_t}+2\sum_{0\le i<j\le n-1}(\Delta X^c_{t,i})^2(\Delta X^c_{t,j})^2 1_{i\in I^J_t}1_{j\in I^J_t}\right)\\ &= E\left(\sum_{i=0}^{n-1}(\Delta X^c_{t,i})^4 P_{J,\delta}+2\sum_{0\le i<j\le n-1}(\Delta X^c_{t,i})^2(\Delta X^c_{t,j})^2 P^2_{J,\delta}\right)\\ &= P^2_{J,\delta}E((\hat V^{R\tilde V}_{t,\Delta})^2)+\sum_{i=1}^{n}E(\Delta X^c_{t,i})^4(P_{J,\delta}-P^2_{J,\delta})\\ &= O(\delta^2\cdot\Delta^2)+O(n\delta^2\cdot\delta)=O\left(\frac{\Delta^3}{n^2}\right).\end{aligned}$$

Following a similar calculation:

$$\begin{aligned}E(A_{2,t})^2 &\le a^4\delta^{4\varpi}nP_{J,\delta}+2\sum_{0\le i<j\le n}a^4\delta^{4\varpi}P^2_{J,\delta}\\ &= O(\delta^{4\varpi}(n^2\delta^2+n\delta))=O\left(\frac{\Delta^{1+4\varpi}}{n^{4\varpi}}\right).\end{aligned}$$

To analyze $A_{3,t}$, we first apply the Cauchy-Schwarz inequality to obtain

$$A^2_{3,t}\le n\sum_{i\in I^c_t}(\Delta X_{t,i})^4 1_{\{|\Delta X_{t,i}|>a\delta^{\varpi}\}}.$$

Taking expectation on both sides and utilizing the Cauchy-Schwarz inequality again, we obtain

$$E(A^2_{3,t})\le n\sum_{i=0}^{n-1}\sqrt{E(\Delta X^c_{t,i})^8\cdot P_{c,\delta}},$$

where $P_{c,\delta}=P\{|X^c_\delta-X^c_0|>a\delta^{\varpi}\}$. By similar derivation as in Fan, Li, and Yu (2012), we have that $P_{c,\delta}$ is exponentially small as $\delta\to0$ for any $0<\varpi<1/2$. Yet, $E(\Delta X^c_{t,i})^8=O((\Delta/n)^4)$. Therefore, $E(A^2_{3,t})=o(\Delta^k)$ for any $k$.

Now, we are ready to compute the variances and covariance involved in the theorem. First, notice that

$$\hat V^{RV,TR}_{t+m\Delta,\Delta}-\hat V^{RV,TR}_{t,\Delta}=\hat V^{R\tilde V}_{t+m\Delta,\Delta}-\hat V^{R\tilde V}_{t,\Delta}+D_{m\Delta,t},$$

where

$$D_{m\Delta,t}=-(A_{1,t+m\Delta}-A_{1,t})+(A_{2,t+m\Delta}-A_{2,t})-(A_{3,t+m\Delta}-A_{3,t}).$$

Thus, by the covariance formula and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}&|\mathrm{Var}(\hat V^{RV,TR}_{t+m\Delta,\Delta}-\hat V^{RV,TR}_{t,\Delta})-\mathrm{Var}(\hat V^{R\tilde V}_{t+m\Delta,\Delta}-\hat V^{R\tilde V}_{t,\Delta})|\\ &\quad\le 2\sqrt{\mathrm{Var}(\hat V^{R\tilde V}_{t+m\Delta,\Delta}-\hat V^{R\tilde V}_{t,\Delta})\mathrm{Var}(D_{m\Delta,t})}+\mathrm{Var}(D_{m\Delta,t}).\end{aligned}$$

Using the order of magnitude of $E(A_{i,t})^2$ for $i=1,2,3$, it is easy to see that

$$\mathrm{Var}(D_{m\Delta,t})\le E(D^2_{m\Delta,t})=O(\Delta^{1+8\varpi})\quad\text{for } 0<\varpi<1/2.$$

Therefore, when $\frac{5}{16}<\varpi<\frac12$ and $m^2\Delta\to C_m\in(0,\infty)$,

$$|\mathrm{Var}(\hat V^{RV,TR}_{t+m\Delta,\Delta}-\hat V^{RV,TR}_{t,\Delta})-\mathrm{Var}(\hat V^{R\tilde V}_{t+m\Delta,\Delta}-\hat V^{R\tilde V}_{t,\Delta})|=o(\Delta^3).\tag{69}$$

The relation between the observed components and continuous component is simply

$$\mathrm{Var}(X_{t+m\Delta}-X_t)=\mathrm{Var}(X^c_{t+m\Delta}-X^c_t)+\sigma^2_J\lambda m\Delta.\tag{70}$$

We now relate the covariance component. By independence of the jump part and the continuous part of the $X$ process, we have when $\frac{5}{16}<\varpi<\frac12$ and $m^2\Delta\to C_m\in(0,\infty)$,

$$\begin{aligned}&\mathrm{Cov}(\hat V^{RV,TR}_{t+m\Delta,\Delta}-\hat V^{RV,TR}_{t,\Delta},X_{t+m\Delta}-X_t)\\ &\quad-\mathrm{Cov}(\hat V^{R\tilde V}_{t+m\Delta,\Delta}-\hat V^{R\tilde V}_{t,\Delta},X^c_{t+m\Delta}-X^c_t)\\ &\quad=\mathrm{Cov}(D_{m\Delta,t},X_{t+m\Delta}-X_t)\\ &\quad=O(\sqrt{\Delta^{1+8\varpi}\cdot m\Delta})=o(\Delta^2).\end{aligned}\tag{71}$$

By (69)–(71), it is easy to see when $\frac{5}{16}<\varpi<\frac12$,

$$\begin{aligned}&\mathrm{Corr}(\hat V^{RV,TR}_{t+m\Delta,\Delta}-\hat V^{RV,TR}_{t,\Delta},X_{t+m\Delta}-X_t)\\ &\quad=\mathrm{Corr}(\hat V^{R\tilde V}_{t+m\Delta,\Delta}-\hat V^{R\tilde V}_{t,\Delta},X^c_{t+m\Delta}-X^c_t)\\ &\qquad\cdot\sqrt{\frac{\mathrm{Var}(X^c_{t+m\Delta}-X^c_t)}{\mathrm{Var}(X^c_{t+m\Delta}-X^c_t)+\sigma^2_J\lambda m\Delta}}(1+o(m\Delta)).\end{aligned}\tag{72}$$

Proposition 5 follows from (72) and Theorem 5 follows by substituting

$$\mathrm{Var}(X^c_{t+m\Delta}-X^c_t)=\alpha m\Delta+\left(\frac{\gamma^2\alpha}{8\kappa}-\frac{\gamma\alpha\rho}{2}\right)m^2\Delta^2+o(m^2\Delta^2).\qquad\square$$

## References


Aït-Sahalia, Y., Fan, J., Xiu, D., 2010. High-frequency covariance estimates with noisy and asynchronous data. Journal of the American Statistical Association 105, 1504–1517.

Aït-Sahalia, Y., Jacod, J., 2009. Testing for jumps in a discretely observed process. Annals of Statistics 37, 184–222.

Aït-Sahalia, Y., Jacod, J., 2012. Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High Frequency Data. Journal of Economic Literature 50, 1007–1050.

Aït-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. Journal of Financial Economics 83, 413–452.

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351–416.

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2011. Ultra high frequency volatility estimation with dependent microstructure noise. Journal of Econometrics 160, 160–175.

Andersen, T.G., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. Journal of Finance 57, 1239–1284.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. Journal of Finance 52, 2003–2049.

Bandi, F.M., Renò, R., 2012. Time-varying leverage effects. Journal of Econometrics 169, 94–113.

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

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. Econometrica 76, 1481–1536.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2011. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. Journal of Econometrics 162, 149–169.

Bekaert, G., Wu, G., 2000. Asymmetric volatility and risk in equity markets. Review of Financial Studies 13, 1–42.

Black, F., 1976. Studies of stock price volatility changes. In: Proceedings of the 1976 Meetings of the American Statistical Association, pp. 171–181.

Bollerslev, T., Litvinova, J., Tauchen, G., 2006. Leverage and volatility feedback effects in high-frequency data. Journal of Financial Econometrics 4, 353–384.

Bollerslev, T., Sizova, N., Tauchen, G.T., 2012. Volatility in equilibrium: a symmetries and dynamic dependencies. Review of Finance 16, 31–80.

Campbell, J.Y., Hentschel, L., 1992. No news is good news: an asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281–318.

Christie, A.A., 1982. The stochastic behavior of common stock variances: value, leverage and interest rate effects. Journal of Financial Economics 10, 407–432.

Delattre, S., Jacod, J., 1997. A central limit theorem for normalized functions of the increments of a diffusion process in the presence, of round-off errors. Bernoulli 3, 1–28.
Duffee, G.R., 1995. Stock returns and volatility: a firm-level analysis. Journal of Financial Economics 37, 399–420.
Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.
Epps, T.W., 1979. Comovements in stock prices in the very short run. Journal of the American Statistical Association 74, 291–296.
Fan, J., Li, Y., Yu, K., 2012. Vast volatility matrix estimation using high frequency data for portfolio selection. Journal of the American Statistical Association 107, 412–428.
Fan, J., Wang, Y., 2007. Multi-scale jump and volatility analysis for high-frequency financial data. Journal of the American Statistical Association 102, 1349–1362.
Figlewski, S., Wang, X., 2000. Is the ''leverage effect'' a leverage effect? Unpublished working paper. New York University, Fudan University.
French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.
Gatheral, J., Oomen, R.C., 2010. Zero-intelligence realized variance estimation. Finance and Stochastics 14, 249–283.
Griffin, J.E., Oomen, R.C., 2008. Sampling returns for realized variance calculations: Tick time or transaction time?. Econometric Reviews 27, 230–253.
Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. Journal of Business and Economic Statistics 24, 127–161.
Hayashi, T., Yoshida, N., 2005. On covariance estimation of non-synchronously observed diffusion processes. Bernoulli 11, 359–379.
Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bonds and currency options. Review of Financial Studies 6, 327–343.
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.
Kalnina, I., Linton, O., 2008. Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. Journal of Econometrics 147, 47–59.
Kinnebrock, S., Podolskij, M., 2008. Estimation of the quadratic covariation matrix in noisy diffusion models. Unpublished working paper. University of Oxford, Heidelberg University.
Large, J., 2007. Accounting for the Epps effect: realized covariation, cointegration and common factors. Unpublished working paper. University of Oxford.
Li, Y., Mykland, P.A., 2007. Are volatility estimators robust with respect to modeling assumptions? Bernoulli 13, 601–622.
Li, Y., Mykland, P.A., Renault, E., Zhang, L., Zheng, X., 2009. Realized volatility when sampling times are possibly endogenous. Unpublished working paper. Hong Kong University of Science and Technology, Brown University, University of Chicago, University of Illinois at Chicago.
Mancini, C., 2009. Nonparametric threshold estimation for models with stochastic diffusion coefficient and jumps. Scandinavian Journal of Statistics 36, 270–296.
Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.
Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. Journal of Financial Economics 63, 3–50.
Voev, V., Lunde, A., 2007. Integrated covariance estimation using high-frequency data in the presence of noise. Journal of Financial Econometrics 5, 68–104.
Wang, D.C., Mykland, P.A., 2009. The estimation of the leverage effect with high frequency data. Unpublished working paper. Princeton University, University of Chicago.
Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. Journal of Econometrics 159, 235–250.
Yu, J., 2005. On leverage in a stochastic volatility model. Journal of Econometrics 127, 165–178.
Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli 12, 1019–1043.
Zhang, L., 2011. Estimating covariation: Epps effect and microstructure noise. Journal of Econometrics 160, 33–47.
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.