

NBER WORKING PAPER SERIES

THE LEVERAGE EFFECT PUZZLE:  
DISENTANGLING SOURCES OF BIAS AT HIGH FREQUENCY

Yacine Ait-Sahalia  
Jianqing Fan  
Yingying Li

Working Paper 17592  
<http://www.nber.org/papers/w17592>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
November 2011

Ait-Sahalia's research was supported by NSF grant SES-0850533. Fan's research was supported by NSF grants DMS-0714554 and DMS-0704337. Li's research was supported by the Bendheim Center for Finance at Princeton University and the RGC grant DAG09/10.BM12 at Hong Kong University of Science and Technology. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2011 by Yacine Ait-Sahalia, Jianqing Fan, and Yingying Li. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Leverage Effect Puzzle: Disentangling Sources of Bias at High Frequency  
Yacine Ait-Sahalia, Jianqing Fan, and Yingying Li  
NBER Working Paper No. 17592  
November 2011  
JEL No. C22,G12

**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.

Yacine Ait-Sahalia  
Department of Economics  
Fisher Hall  
Princeton University  
Princeton, NJ 08544-1021  
and NBER  
yacine@princeton.edu

Yingying Li  
Department of Information Systems,  
Business Statistics  
Hong Kong University of Science and Technology  
yyli@ust.hk

Jianqing Fan  
Bendheim Center for Finance  
26 Prospect Ave  
Princeton NJ 08540  
jqfan@princeton.edu

# 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) and 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: 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 et al. (1987) and 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 et al. (2006)). Using higher frequency data, namely five-minute absolute returns to construct a realized volatility proxy over longer horizons, Bollerslev et al. (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 et al. (2011) 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, 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, quantities and as such should be estimated particularly well at high frequency, or instantaneously, using standard probability limit results. We call this disconnect 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.<sup>1</sup>

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

---

<sup>1</sup>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 et al. (2008b), Kinnebrock and Podolskij (2008) and Ait-Sahalia et al. (2010).

volatility estimator is constructed from the asset returns themselves, the two sets of observations are by construction synchronous. 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.<sup>2</sup>

Our main results are the following. We provide theoretical results to disentangle the biases involved in estimating the correlation between the returns and volatilities with a sequence of progressively more realistic estimators. 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

---

<sup>2</sup>In the univariate volatility case, many estimators have been developed to produce consistent estimators despite the presence of the noise. These include the Two Scales Realized Volatility (TSRV) of Zhang et al. (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 et al. (2008a), the Pre-Averaging volatility estimator (PAV) by Jacod et al. (2009), and the Quasi-Maximum Likelihood Estimator (QMLE) of Xiu (2010) which extends the parametric Maximum-Likelihood Estimator of Aït-Sahalia et al. (2005) to the setting of stochastic volatility. Related work 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 et al. (2011) and Li et al. (2010). 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), Barndorff-Nielsen et al. (2008b) a Multivariate Realized Kernel (MRK), Kinnebrock and Podolskij (2008) a multivariate Pre-Averaging estimator and Aït-Sahalia et al. (2010) a multivariate Quasi-Maximum Likelihood Estimator.

already estimated with a bias that is due solely to discretization. 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 even larger, 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 of course go in the reverse direction. Based on the above results, we propose a regression approach to compute bias-corrected estimators of  $\rho$ . We 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 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 leverage effect parameter in all steps of approximations. A novel solution to the puzzle is proposed in Section 5, which is convincingly demonstrated by Monte Carlo simulations in Section 6 and by empirical studies in Section 7 using the high-frequency data from S&P500 and Microsoft. Section 8 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. 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). Figure 1 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 Figure 1, we see that the effect is present in the data.

+++ Insert Figure 1 Here +++

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 thirty day to maturity; that is, VIX measures the square-root of the risk neutral expectation of the S&P 500 variance over the next thirty calendar days. Using this market-based volatility measure, the leverage effect is indeed very strong as demonstrated in Figure 2.

+++ Insert Figure 2 Here +++

Yet, starting at the daily horizon, even when using high frequency volatility estimates, we see in Figure 3 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 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.

+++ Insert Figure 3 Here +++

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 Figures 1 and 2. So why are we unable to detect it on short horizon 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.

### 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. Put together, these ingredients lead to a novel solution to the leverage puzzle by introducing a tuning parameter (represented by  $m$  below) that attempts to minimize the estimation bias.

#### 3.1 Stochastic Volatility Model

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

$$dX_t = (\mu - \nu_t/2)dt + \sigma_t dB_t \tag{1}$$

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

where  $\nu_t = \sigma_t^2$ ,  $B$  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  $\nu_0 > 0$  is a realization from the stationary (invariant) distribution of (2) so that  $\nu_t$  is a stationary process. Under Feller's condition  $2\kappa\alpha > \gamma^2$ , the process  $\nu_t$  stays positive, a condition that is always assumed in what follows. Note that

$$\rho = \lim_{s \rightarrow 0} \text{Corr}(\nu_{t+s} - \nu_t, X_{t+s} - X_t) \tag{3}$$

so that the leverage effect is summarized by the parameter  $\rho$  under the Heston model (1)-(2).

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 et al. (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 et al. (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 statistical 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

$$V_{t,\Delta} = \int_{t-\Delta}^t \nu_s ds \quad (4)$$

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  $\nu_t$  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

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

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 10 seconds). In order for the nonparametric estimate  $\hat{V}_{t,\Delta}$  to be sufficiently accurate, we need  $\delta \ll \Delta$ ; asymptotically, we assume that  $\Delta \rightarrow 0$  and  $\delta \rightarrow 0$  in such a way that  $\Delta/\delta \rightarrow \infty$ .

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

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

Here, for simplicity of exposition, we assume there is an observation at time  $t - \Delta$ , and that the ratio  $\Delta/\delta$  is an integer; 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  $X_{t+i\delta}$ , we observe the noisy version

$$Z_{t+i\delta} = X_{t+i\delta} + \epsilon_{t+i\delta}, \quad (7)$$

where the  $\epsilon_{t+i\delta}$ 's are white noise random variables with mean zero and standard deviation  $\sigma_\epsilon$ . With this type of observations, we can use noise-robust methods such as TSRV, PAV, QMLE or RK to obtain consistent estimates of the integrated volatility. We will first use the TSRV estimator, as it is relatively simple to analyze. Specifically, letting  $n = \Delta/\delta$ ,  $\theta_{\text{TSRV}}$  be a constant,  $L = [\theta_{\text{TSRV}} n^{2/3}]$  the number of grids over which the subsampling is performed and  $\bar{n} = (n - L + 1)/n$ , the TSRV estimator is defined as

$$\hat{V}_{t,\Delta}^{\text{TSRV}} = \frac{1}{L} \sum_{i=0}^{n-L} (Z_{t-\Delta+(i+L)\delta} - Z_{t-\Delta+i\delta})^2 - \frac{\bar{n}}{n} \sum_{i=0}^{n-1} (Z_{t-\Delta+(i+1)\delta} - Z_{t-\Delta+i\delta})^2. \quad (8)$$

The TSRV estimator is simple to analyze but is not rate-optimal, converging at rate  $n^{1/6}$  instead of the optimal rate  $n^{1/4}$ . Thus, it is expected to incur a slightly larger estimation error. We therefore consider the rate-efficient pre-averaging volatility estimator (PAV) as proposed by Jacod et al. (2009) with the weight function chosen as  $g(x) = x \wedge (1 - x)$ . More specifically, let  $\theta_{\text{PAV}}$  be a constant,  $k_n = [\theta_{\text{PAV}} \sqrt{n}]$ , we consider

$$\begin{aligned} \hat{V}_{t,\Delta}^{\text{PAV}} = & \frac{12}{\theta_{\text{PAV}} \sqrt{n}} \sum_{i=0}^{n-k_n+1} \left( \frac{1}{k_n} \sum_{j=[k_n/2]}^{k_n-1} Z_{t-\Delta+(i+j)\delta} - \frac{1}{k_n} \sum_{j=0}^{[k_n/2]-1} Z_{t-\Delta+(i+j)\delta} \right)^2 \\ & - \frac{6}{\theta_{\text{PAV}}^2 n} \sum_{i=0}^{n-1} (Z_{t-\Delta+(i+1)\delta} - Z_{t-\Delta+i\delta})^2. \end{aligned} \quad (9)$$

A consistent estimator of the variance is provided in Jacod et al. (2009), as well as a consistent estimator of the integrated quarticity  $\int_{t-\Delta}^t \sigma_s^4 ds$  (see (21)).

## 4 Biases in Estimation of the Leverage Effect

We now present the first 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. These progressive scenarios help us document an incremental source for the bias: discretization, smoothing, estimation error and market microstructure noise.

### 4.1 True Spot Volatility: Discretization Bias

First, we consider the unrealistic but idealized situation in which the spot volatility process  $\nu_s$  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:*

$$\text{Corr}(\nu_{s+t} - \nu_t, X_{s+t} - X_t) = \frac{\rho \sqrt{\frac{1-e^{-\kappa s}}{\kappa}}}{\sqrt{\left(s + \frac{e^{-\kappa s}-1}{\kappa}\right) \left(\frac{\gamma^2}{4\kappa^2} - \frac{\gamma\rho}{\kappa}\right) + s}}. \quad (10)$$

Let us denote the right hand side of the expression in Theorem 1 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  $m\Delta \rightarrow 0$ , we have*

$$\text{Corr}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t) = \rho - \frac{\rho(\gamma^2 - 4\gamma\kappa\rho + 4\kappa^2)}{16\kappa} m\Delta + o(m\Delta). \quad (11)$$

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. Figure 4 shows precisely how the function  $C_1(m\Delta, \kappa, \gamma, \alpha, \rho)$  varies with  $m$  for two sets of parameter values:  $(\rho, \kappa, \gamma, \alpha) = (-0.8, 5, 0.5, 0.1)$  and  $(\rho, \kappa, \gamma, \alpha) = (-0.3, 5, 0.05, 0.04)$  when  $\Delta$  is taken to be  $1/252$ . The former set of parameters was adapted from those in Ait-Sahalia and Kimmel (2010) and the latter set

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

+++ Insert Figure 4 Here +++

## 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:*

$$\text{Corr}(V_{t+m\Delta,\Delta} - V_{t,\Delta}, X_{t+m\Delta} - X_t) = A_2/(B_2C_2) \quad (12)$$

where

$$\begin{aligned} A_2 &= 2\gamma(1 - \Delta\kappa) + 4\Delta\kappa^2\rho - 2\gamma e^{-\Delta\kappa} \\ &\quad + e^{-\Delta\kappa(m+1)} (e^{2\Delta\kappa}(\gamma - 4\kappa\rho) - 2e^{\Delta\kappa}(\gamma - 2\kappa\rho) + \gamma), \\ B_2 &= 2\sqrt{e^{-\Delta\kappa(m+1)} (2e^{\Delta\kappa m} - (e^{\Delta\kappa} - 1)^2) + 2\Delta\kappa - 2}, \end{aligned}$$

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  $\Delta \rightarrow 0$ . We focus again on both situations where  $m$  is fixed and  $m \rightarrow \infty$  while still  $m\Delta \rightarrow 0$ .

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

$$\begin{aligned} \text{Corr}(V_{t+m\Delta,\Delta} - V_{t,\Delta}, X_{t+m\Delta} - X_t) &= \text{Corr}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t) \frac{(2m-1)}{2\sqrt{m^2 - m/3}} \\ &\quad + \begin{cases} O(\Delta) \text{ when } \Delta \rightarrow 0 \text{ for any } m \\ o(m\Delta) \text{ when } m \rightarrow \infty, m\Delta \rightarrow 0 \end{cases}. \end{aligned} \quad (13)$$

The first expression is true when  $m$  is any fixed integer. For the second expression, note that the asymptote of the correction factor

$$\frac{(2m-1)}{2\sqrt{m^2 - m/3}} = 1 + O\left(\frac{1}{m}\right). \quad (14)$$

Hence, when  $m$  is large, unlike what the initial intuition might have suggested, the bias of estimated  $\rho$  based on integrated volatilities is asymptotically the same as that of the estimated  $\rho$  based on spot volatilities.

Figure 4 shows the resulting numerical values (dotted curves) for the same sets of parameters. 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 employed. Figure 4 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 Figure 4  $[(\rho, \kappa, \gamma, \alpha) = (-0.8, 5, 0.5, 0.1)]$ , the optimum is  $m_0 = 16$  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 Figures 1 and 2. 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. Let  $n$  be the number of observations during each interval  $\Delta$ . Assume for simplicity that the observation intervals are equally spaced at a distance  $\delta = \Delta/n$ .

**Theorem 3.** *When  $n\Delta \rightarrow C$  and  $m^2\Delta \rightarrow C_m$  for  $C, C_m \in (0, \infty)$ , the following expansion shows the incremental bias due to estimation error induced by the use of RV:*

$$\begin{aligned} \text{Corr}(\hat{V}_{t+m\Delta, \Delta}^{RV} - \hat{V}_{t, \Delta}^{RV}, X_{t+m\Delta} - X_t) &= \text{Corr}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t) \frac{(2m-1)}{2\sqrt{m^2 - m/3}} \\ &\times \left( 1 + \frac{12\alpha\kappa + 6\gamma^2}{(3\gamma^2m - \gamma^2)\kappa C - \frac{3}{2}\gamma^2\kappa^2 C C_m} \right)^{-1/2} [1 + o(m\Delta)]. \end{aligned} \quad (15)$$

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, 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 Figure 4 for the same sets of parameters as mentioned above. The daily sampling frequency is taken to be  $n = 390$  (one observation per minute) so that  $C = 390/252$ . In particular, 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

$$\hat{\rho}^{RV} = \text{Corr}(\hat{V}_{t+\Delta, \Delta}^{RV} - \hat{V}_{t, \Delta}^{RV}, X_{t+\Delta} - X_t) \quad (16)$$

is very close to 0. 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 Figure 4, if the optimal  $m = 27$  is used, the estimated correlation is now  $-0.694$ , 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 allowance is made for the presence of market microstructure noise under (7), the integrated volatility  $V_t$  is estimated based on noisy log-returns, using bias-corrected high-frequency volatility estimators such as TSRV, PAV, QMLE or RK. 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 resolve the issue.

We start with TSRV and then consider PAV as well. Other methods can be employed too, but the computations become increasingly tedious – more so than they already are! Recall the definition of  $\theta_{\text{TSRV}}$  in the TSRV estimator, which determines the constant factor of the large scale RV.

**Theorem 4.** *When  $n^{1/3}\Delta \rightarrow C_{\text{TSRV}}$ ,  $\sigma_\epsilon^2/\Delta \rightarrow C_\epsilon$  and  $m^2\Delta \rightarrow C_m$  with  $C_{\text{TSRV}}$ ,  $C_\epsilon$  and  $C_m \in (0, \infty)$ , the following expansion shows the incremental bias due to estimation error and noise correction induced by the use of TSRV:*

$$\begin{aligned} \text{Corr}(\hat{V}_{t+m\Delta, \Delta}^{\text{TSRV}} - \hat{V}_{t, \Delta}^{\text{TSRV}}, Z_{t+m\Delta} - Z_t) &= \text{Corr}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t) \frac{(2m-1)}{2\sqrt{m^2 - m/3}} \\ &\times (1 + A_4 + B_4)^{-1/2} [1 + o(m\Delta)], \end{aligned}$$

where

$$\begin{aligned} A_4 &= \frac{96\theta_{\text{TSRV}}^{-2} C_\epsilon^2}{C_{\text{TSRV}} \alpha \gamma^2 (6m - 2 - 3\kappa C_m)} \\ B_4 &= \frac{8\theta_{\text{TSRV}} (2\alpha\kappa + \gamma^2)}{\kappa C_{\text{TSRV}} \gamma^2 (6m - 2 - 3\kappa C_m)}. \end{aligned}$$

For the same reasons behind the above theorem, using the parameter  $m$  helps resolving the leverage effect problems. When  $\theta_{\text{TSRV}}$  is taken to be 0.5, 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 barely noticeable whereas using  $m = 73$  yields a correlation of  $-0.483$ . Even though the bias is large, the leverage effect is clearly noticeable.

Again, the estimating biases can be decomposed into two factors. The first factor is the smoothing bias, the same as that in the RV and PAV below. The second factor reflects the shrinkage biases due to estimation errors and noise correction errors. The rate of convergence

of TSRV is slower than that of RV. This is reflected in the factor  $C_{\text{TSRV}}$  which is of order  $n^{1/3}\Delta$ , rather than  $C = n\Delta$  in RV. Similarly, since PAV below has a faster rate of convergence than TSRV, its corresponding shrinkage bias is smaller than TSRV but larger than RV. This is reflected in  $C_{\text{PAV}} = n^{1/2}\Delta$  in Theorem 5 below.

A parallel result to Theorem 4 for PAV is the following.

**Theorem 5.** *When  $n^{1/2}\Delta \rightarrow C_{\text{PAV}}$ ,  $\sigma_\epsilon^2/\Delta \rightarrow C_\epsilon$ , and  $m^2\Delta \rightarrow C_m$  with  $C_{\text{PAV}}$ ,  $C_\epsilon$  and  $C_m \in (0, \infty)$ , the following expansion shows the incremental bias due to estimation error and noise correction induced by the use of PAV:*

$$\begin{aligned} \text{Corr}(\hat{V}_{t+m\Delta, \Delta}^{\text{PAV}} - \hat{V}_{t, \Delta}^{\text{PAV}}, Z_{t+m\Delta} - Z_t) &= \text{Corr}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t) \frac{(2m-1)}{2\sqrt{m^2 - m/3}} \\ &\quad \times (1 + A_5 + B_5 + C_5)^{-1/2} [1 + o(m\Delta)], \end{aligned}$$

where

$$\begin{aligned} A_5 &= \frac{24\Phi_{22}\theta_{\text{PAV}}(2\alpha\kappa + \gamma^2)}{\psi_2^2 C_{\text{PAV}} \kappa \gamma^2 (6m - 2 - 3\kappa C_m)} \\ B_5 &= \frac{96\Phi_{12}C_\epsilon}{\theta_{\text{PAV}} \psi_2^2 C_{\text{PAV}} \gamma^2 (6m - 2 - 3\kappa C_m)} \\ C_5 &= \frac{48\Phi_{11}C_\epsilon^2}{\theta_{\text{PAV}}^3 \psi_2^2 C_{\text{PAV}} \alpha \gamma^2 (6m - 2 - 3\kappa C_m)}, \end{aligned}$$

where  $\psi_2 = \frac{1}{12}$ ,  $\Phi_{11} = \frac{1}{6}$ ,  $\Phi_{12} = \frac{1}{96}$ ,  $\Phi_{22} = \frac{151}{80640}$ .

Theorem 4 and Theorem 5 are also illustrated in Figure 4 in which the main terms of the correlations are plotted. Again when  $m = 1$ , the correlation is nearly zero, whereas with the ideal choice of  $m = 48$ , the ideal correlation is  $-0.599$ , still significantly smaller than the true one of  $-0.8$ .

## 5 A Solution to the Puzzle: Model-Independent Bias Corrections

The previous section 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 correcting 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.

## 5.1 Back to the Latent Spot Volatility

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 Theorem 2, which can be inverted to yield:

$$\text{Corr}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t) = \frac{2\sqrt{m^2 - m/3}}{(2m - 1)} \text{Corr}(V_{t+m\Delta, \Delta} - V_{t, \Delta}, X_{t+m\Delta} - X_t) + O(\Delta). \quad (17)$$

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  $\nu$ . The effectiveness of this simple bias correction is demonstrated in Figure 5.

+++ Insert Figure 5 Here +++

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 of realized volatility back to that based on the spot volatility. However, such a correction involves unknown parameters in the Heston model and depends on the parametric assumption. However, the method is applicable to a wider array of data generating processes. A nonparametric correction consists in using the following result, demonstrated in the Appendix, in the proof of Theorem 3:

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

$$\text{Corr}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t) = c_3 \frac{2\sqrt{m^2 - m/3}}{(2m - 1)} \text{Corr}(\hat{V}_{t+m\Delta, \Delta}^{RV} - \hat{V}_{t, \Delta}^{RV}, X_{t+m\Delta} - X_t) + o(m\Delta), \quad (18)$$

where  $c_3$  is given by

$$c_3 = \left( 1 - \frac{4E[\sigma_t^4] \Delta^2}{n \text{Var}(\hat{V}_{t+m\Delta, \Delta}^{RV} - \hat{V}_{t, \Delta}^{RV})} \right)^{-1/2}. \quad (19)$$

Note that in (19), the stationarity of the process of  $\nu_t$  is used so that the correction factor does not depend on  $t$ .

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

$$\frac{n}{3} \sum_{i=0}^{n-1} (X_{t+(i+1)\delta} - X_{t+i\delta})^4 \xrightarrow{\mathbb{P}} \Delta \int_t^{t+\Delta} \sigma_s^4 ds \text{ as } n \rightarrow \infty$$

for any fixed  $\Delta$ . Hence a long run average of scaled quarticity can be used to estimate  $E[\sigma_t^4]$ . The variance in (19) can be estimated by its sample version.

For the TSRV estimator, the bias correction admits the same form as (18) with a different correction.

**Proposition 4.** *When  $n^{1/3}\Delta \rightarrow C_{TSRV}$ ,  $\sigma_\epsilon^2/\Delta \rightarrow C_\epsilon$  and  $m^2\Delta \rightarrow C_m$  for constants  $C_{TSRV}$ ,  $C_\epsilon$  and  $C_m \in (0, \infty)$ ,*

$$\text{Corr}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t) = c_4 \frac{2\sqrt{m^2 - m/3}}{(2m - 1)} \text{Corr}(\hat{V}_{t+m\Delta, \Delta}^{TSRV} - \hat{V}_{t, \Delta}^{TSRV}, Z_{t+m\Delta} - Z_t) + o(m\Delta)$$

where

$$c_4 = \left( 1 - \frac{48\theta_{TSRV}^{-2}\sigma_\epsilon^4 + 8\theta_{TSRV}E[\sigma_t^4]\Delta^2}{3n^{1/3}\text{Var}(\hat{V}_{t+m\Delta, \Delta}^{TSRV} - \hat{V}_{t, \Delta}^{TSRV})} \right)^{-1/2}. \quad (20)$$

Two unknown quantities are involved and can be estimated nonparametrically here. For  $\sigma_\epsilon$ , since  $\hat{V}_{t, \Delta}^{RV} \approx 2n\sigma_\epsilon^2 + \int_{t-\Delta}^t \sigma_s^2 ds$ ,  $\hat{V}_{t, \Delta}^{TSRV} \approx \int_{t-\Delta}^t \sigma_s^2 ds$  for fixed  $\Delta$  and big  $n$ , we can conclude that a long run average of  $(\hat{V}_{t, \Delta}^{RV} - \hat{V}_{t, \Delta}^{TSRV})/2n$  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  $E[\sigma_t^4]$ , since market microstructure noise is involved, the situation is a bit more complicated than before. Consistent noise-robust estimators of  $\int_{t-\Delta}^t \sigma_s^4 ds$  are proposed in Zhang et al. (2005)

and Jacod et al. (2009); we can use for instance the estimator called  $\hat{Q}_t^n$  in the latter paper:

$$\begin{aligned}
\hat{Q}_t^n &= \frac{1}{3\theta_{PAV}^2\psi_2^2} \sum_{i=0}^{n-k_n+1} \left( \frac{1}{k_n} \sum_{j=\lfloor k_n/2 \rfloor}^{k_n-1} Z_{t-\Delta+(i+j)\delta} - \frac{1}{k_n} \sum_{j=0}^{\lfloor k_n/2 \rfloor-1} Z_{t-\Delta+(i+j)\delta} \right)^4 \\
&\quad - \frac{\delta}{\theta_{PAV}^4\psi_2^2} \sum_{i=0}^{n-2k_n+1} \left( \left( \frac{1}{k_n} \sum_{j=\lfloor k_n/2 \rfloor}^{k_n-1} Z_{t-\Delta+(i+j)\delta} - \frac{1}{k_n} \sum_{j=0}^{\lfloor k_n/2 \rfloor-1} Z_{t-\Delta+(i+j)\delta} \right)^2 \times \right. \\
&\quad \quad \left. \sum_{j=i+k_n}^{i+2k_n-1} (Z_{t-\Delta+(j+1)\delta} - Z_{t-\Delta+j\delta})^2 \right) \\
&\quad + \frac{\delta}{4\theta_{PAV}^4\psi_2^2} \sum_{i=1}^{n-2} (Z_{t-\Delta+(i+1)\delta} - Z_{t-\Delta+i\delta})^2 (Z_{t-\Delta+(i+3)\delta} - Z_{t-\Delta+(i+2)\delta})^2. \\
&\rightarrow_{\mathbb{P}} \int_{t-\Delta}^t \sigma_t^4 dt,
\end{aligned} \tag{21}$$

where  $\psi_2 = \frac{1}{\sqrt{2}}$ ,  $k_n = \lfloor \theta_{PAV} \sqrt{n} \rfloor$  for an appropriately chosen  $\theta_{PAV}$ . A scaled long run average of this estimator can be used to estimate  $E[\sigma_t^4]$ .

For PAV, we have

**Proposition 5.** *When  $n^{1/2}\Delta \rightarrow C_{PAV}$ ,  $\sigma_\epsilon^2/\Delta \rightarrow C_\epsilon$ , and  $m^2\Delta \rightarrow C_m$  for constants  $C_{PAV}$ ,  $C_\epsilon$  and  $C_m \in (0, \infty)$ ,*

$$\text{Corr}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t) = c_5 \frac{2\sqrt{m^2 - m/3}}{(2m - 1)} \text{Corr}(\hat{V}_{t+m\Delta, \Delta}^{PAV} - \hat{V}_{t, \Delta}^{PAV}, Z_{t+m\Delta} - Z_t) + o(m\Delta)$$

where

$$c_5 = \left( 1 - \frac{2(A'_5 + B'_5 + C'_5)}{n^{1/2} \text{Var}(\hat{V}_{t+m\Delta, \Delta}^{PAV} - \hat{V}_{t, \Delta}^{PAV})} \right)^{-1/2}, \tag{22}$$

with

$$\begin{aligned}
A'_5 &= \frac{4\Phi_{22}\theta_{PAV}E[\sigma_t^4]}{\psi_2^2}, \\
B'_5 &= \frac{8\Phi_{12}E[\sigma_t^2]\sigma_\epsilon^2}{\theta_{PAV}\psi_2^2}, \\
C'_5 &= \frac{4\Phi_{11}\sigma_\epsilon^4}{\theta_{PAV}^3\psi_2^2},
\end{aligned}$$

where  $\psi_2, \Phi_{11}, \Phi_{12}, \Phi_{22}$  are as in Theorem 5.

A more direct way is to use the long run average of the quantity  $\Gamma_t^n$  defined in (3.7) of Jacod et al. (2009) to estimate  $A'_5 + B'_5 + C'_5$ .

## 5.2 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  $\nu_t$  itself is also biased. If the model were known, then the bias (11) can be computed and corrected. However, this depends on the Heston model and its unknown parameters.

A parameter-independent method is as follows. Let  $\hat{\rho}_m = \text{Corr}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t)$ . Then, by Theorem 1 we see that

$$\hat{\rho}_m = \rho + bm + o(m\Delta). \quad (23)$$

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  $\{(m, \hat{\rho}_m)\}$ . The bias-corrected estimate of  $\rho$  is simply the intercept of that linear regression. The scatter plot of  $\{(m, \hat{\rho}_m)\}$  can also suggest a region of  $m$  to run the above simple linear regression (23).

The above discussion suggests a rather general strategy for bias correction. The method does not depend on the Heston model parameters. First, compute the simple correlation between estimated changes of volatilities and estimated changes of prices. Second, conduct a preliminary bias correction according to (17) – (22), depending on which estimated volatilities are used. Third, run the simple regression equation (17) 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 will now see.

## 6 Monte-Carlo Simulations

In this section, we use simulation studies to reproduce the leverage effect puzzle and its proposed solution and to verify the practical validity of the results presented in the previous section.

### 6.1 The data generating process

The true log-price is simulated from the Heston model (1)-(2) with 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 ( $\Delta = 1/252$ ). Each day, the sampling frequency is one minute per sample, giving an

intra-day number of observations of  $n = 390$ . Therefore the total number of observations over 5 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 the market microstructure as in (7) with i.i.d  $\mathcal{N}(0, \sigma_\epsilon^2)$  noise, and  $\sigma_\epsilon = 0.0005$ . (The case when the observation frequency is higher,  $n = 1560$  is also studied. The results are collected at the end of this section.)

## 6.2 Vizualizing the leverage effect puzzle

With the latent spot volatility  $\nu_t$  and latent price  $X_t$  known in simulated data, we can easily examine the correlation of  $\{(X_{t\Delta} - X_{(t-1)\Delta}, \nu_{t\Delta} - \nu_{(t-1)\Delta})\}$  over  $N$  observations. As expected, the leverage effect is strong, with the sample correlation being  $-0.787$  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  $\hat{V}_{t,\Delta} = n^{-1} \sum_{j=1}^n \hat{\sigma}_{t-\Delta+j/N}^2$ . In this ideal situation,  $\sigma_{t-\Delta+j/N}^2$  is known, resulting in  $V_{t,\Delta} = n^{-1} \sum_{j=1}^n \sigma_{t-\Delta+j/N}^2$ . The correlation of  $\{(X_{(t+1)\Delta} - X_{t\Delta}, V_{(t+1)\Delta,\Delta} - V_{t\Delta,\Delta})\}_{t=1}^{1259}$  is  $-0.462$  for the given realization. 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 panels of Figure 6 plots the correlation  $\{(X_{(t+m)\Delta} - X_{t\Delta}, \nu_{(t+m)\Delta,\Delta} - \nu_{t\Delta,\Delta})\}_{t=1}^{1260-m}$  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 in the figure. The impact of  $m$  can easily be seen and the natural estimate based on  $V_{t\Delta,\Delta}$  with  $m = 1$  is far from optimal.

+++ Insert Figure 6 Here +++

In practice, the integrated volatility is not observable. It has to be estimated using the discretely observed data. In absence of 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  $-0.25$  for the same realization discussed above. More generally, the correlation of  $\{(X_{(t+m)\Delta} - X_{t\Delta}, \hat{V}_{(t+m)\Delta,\Delta}^{\text{RV}} - \hat{V}_{t\Delta,\Delta}^{\text{RV}})\}_{t=1}^{1260-m}$  as a function of  $m$  is depicted in the lower left panel of Figure 6. 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  $\{(Z_{(t+m)\Delta} - Z_{t\Delta}, \hat{V}_{(t+m)\Delta, \Delta}^{\text{PAV}} - \hat{V}_{t\Delta, \Delta}^{\text{PAV}})\}_{t=1}^{1260-m}$  or  $\{(Z_{(t+m)\Delta} - Z_{t\Delta}, \hat{V}_{(t+m)\Delta, \Delta}^{\text{TSRV}} - \hat{V}_{t\Delta, \Delta}^{\text{TSRV}})\}_{t=1}^{1260-m}$  with a suitable choice of  $m$ . The lower middle and right panels of Figure 6 shows the correlation as a function of  $m$ . In particular, when  $m = 1$ , the sample correlation is merely  $-0.113$  for PAV and  $-0.107$  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 Theorem 4. Using PAV with  $m = 26$ , the sample correlation is  $-0.682$ ; and using TSRV with  $m = 62$ , the sample correlation is  $-0.553$ . While this is still a biased estimate, the leverage effect is now clearly seen.

The averages of these correlations against  $m$  are plotted together in Figure 7. These are in line with what the theory predicts (see the left panel of Figure 4).

+++ Insert Figure 7 Here +++

### 6.3 Effectiveness of the Bias Correction Method

We now illustrate the effectiveness of the bias correction method proposed in Section 5. We simulate sample paths with the same parameters as above.  $\theta_{\text{TSRV}}$  and  $\theta_{\text{PAV}}$  are both taken to be 0.5. For the linear regression method, we use the set of values of  $m$  as  $\{6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17\}$ . Let us denote the bias corrected estimate of  $\rho$  as  $\text{corNutorho}$  with the correlation based on the spot volatility as the input. Similarly,  $\text{corVtorho}$ ,  $\text{corRVtorho}$ ,  $\text{corPAVtorho}$  and  $\text{corTSRVtorho}$  are the bias corrected  $\rho$  using the linear regression respectively based on the corrected curves from  $V$ ,  $\hat{V}^{\text{RV}}$ ,  $\hat{V}^{\text{PAV}}$  and  $\hat{V}^{\text{TSRV}}$ , using equations (17), (18), (22) and (20). The true values of  $E\sigma_t^4$ ,  $E\sigma_t^2$  and  $\sigma_\epsilon$  are plugged in. Table 1 summarizes the results based on 100 simulations of ( $T = 5$ ) minute-by-minute ( $n=390$ ) data over a five-year period.

+++ Insert Table 1 Here +++

The mean of all of these corrected estimates are all close to the true value  $\rho = -0.8$ . This means that these estimates are unbiased. The progressive harder of the problems can easily be seen from the SD of the estimates. As demonstrated in Figure 8, the TSRV with the available sample size incurs large estimation errors. This translates into the large error for estimating

$\rho$ . When the sampling frequency is more frequent than one sample per minute, the estimation error can be reduced.

In summary, Table 1 provides a stark evidence that the methods in Section 5 solve the leverage effect puzzle. It also quantifies the extent to which the problem gets progressively harder.

In practice, the parameters are unknown and hence the correction based directly on the model parameters as in Table 1 is not feasible. In the following, we demonstrate the effectiveness of the non-parametric methods as described in section 5.1 to obtain corrections. `corRVtorhoE`, `corPAVtorhoE` and `corTSRVtorhoE` are the bias corrected  $\rho$  using the linear regression based on the non-parametrically corrected curves from  $\hat{V}^{RV}$ ,  $\hat{V}^{PAV}$  and  $\hat{V}^{TSRV}$  respectively. Estimates are based on the same simulated sample paths as above. The results are collected in Table 2.

+++ Insert Table 2 Here +++

In this section, we used a fixed range of  $m$  to run the regression regardless of the method and realizations. As discussed in Section 5, the data-dependent and estimator-dependent choice of the range can improve the results further. The scatterplot helps us empirically determine an appropriate range. To illustrate the point, we randomly select a sample path from the simulation, and then plot the corrected correlations up to the curve corresponding to Theorem 1: see Figure 8. After that, we identify the range of  $m$  values such that the corrected curve is roughly linear. For this sample path, a range of  $m = (30, 31, \dots, 70)$  gives us a `corPAVtorhoE` of -0.83, and a range of  $m = (40, 41, \dots, 80)$  gives us a `corTSRVtorhoE` of -0.85. These are better than the corrections based on a generic choice of  $m = (6, 7, \dots, 17)$ , which yields an estimated correlation of -0.71 and -0.90 respectively. This tailor-made method will be used in the empirical study below in order to improve the estimation.

+++ Insert Figure 8 Here +++

Note that as shown in the upper right panel of Figure 8, occasionally we may obtain extreme results that the corrected correlation has an absolute value bigger than 1, especially when  $m$  is small. One can truncate these observations back to the correct range as one likes. Indeed, typically these values won't affect our final results because we typically choose reasonably big values of  $m$  in the final step of correction using regression.

Without bias correction, as demonstrated in Section 2, sampling data at higher frequency does not give us a better assessment of the leverage effect. With the bias correction, we would expect better results. Table 3 and Figure 9 demonstrate this further by the results with the sampling frequency of one observation every 15 seconds, from which we see similar results but with reduced estimation errors.

+++ Insert Table 3 Here +++

+++ Insert Figure 9 Here +++

## 7 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 Figure 3. Nevertheless, we will demonstrate that the new tool is able to reveal the presence of a strong leverage effect contained in high-frequency data. We only focus on the S&P 500 and MSFT returns; we have applied the methods to various data sets and the conclusions are similar.

### 7.1 S&P 500 data

Based on the high-frequency returns (1 minute per sample) on S&P 500 futures from January 2004 to December 2007, the naive or natural estimates give the results reported in Table 4. The leverage effect at the natural choice of  $m = 1$  is nearly 0. Even with the data-optimized choice of  $m$ , the correlation with TSRV is estimated to be around  $-0.44$  and that with the PAV is around  $-0.50$ , much smaller than that computed based on the VIX. The upper panel of Figure 10 summarizes the sample correlations based on TSRV, PAV and VIX respectively, versus horizon  $m$ .

+++ Insert Table 4 Here +++

We now apply our bias corrected methods. First, we compute the preliminarily bias-corrected estimates using both TSRV and PAV for a wide range of choice of  $m$ . Their scatter plots are presented in Figure 10, which are quite curly. We then took the first increasing region (because the biases are expected to get larger with  $m$  after correction so that the slope should be positive) that appears approximately linear. For TSRV, we take  $m = (55, 56 \dots, 100)$  to run the regression and obtain the intercept of  $-0.85$ . This is our bias-corrected estimated leverage effect parameter implied in the high frequency data. Similarly, for PAV, we selected the range  $m = (55, 56, \dots, 85)$ , the bias-corrected correlation based on PAV is estimated as  $-0.68$ .

+++ Insert Figure 10 Here +++

## 7.2 Microsoft

We now use our method to examine how strong the leverage effect for the Microsoft corporation. The high-frequency returns at sample frequencies of one data point per minute and one per 5 seconds from January 2005 to June 2007 are used for estimating the leverage effect parameter. Again, we apply both the naive method, the simple sample correlation, and more sophisticated volatility estimation methods, based on preliminary correction and linear regression. Table 5 summarizes the results. Again, the leverage effects are barely noticeable at both sampling frequencies for natural choices of  $m$  (small values of  $m$ ).

+++ Insert Table 5 Here +++

For the data sampled at one observation per minute, the preliminary corrections are summarized in the left panel of Figure 11. This helps us determine a region where the linear regression should be run. Based on regression (taking  $m = (130, 131, \dots, 170)$ ), the leverage effect parameter is estimated as  $-0.72$  with TSRV and  $-0.68$  with PAV. Again, we find that the leverage effect is much stronger than one would obtain without bias-correction.

+++ Insert Figure 11 Here +++

The analysis for the data sampling at 5-second frequency produces similar results. The range  $m = (130, 131, \dots, 170)$  is determined based on preliminary corrections depicted in Figure 12 (left panel). Using this range of estimates to run a simple linear regression yields an estimated leverage effect parameter. It is  $-0.76$  based on TSRV and  $-0.78$  based on PAV.

+++ Insert Figure 12 Here +++

## 8 Conclusions

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.

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 also concave as a function of the length of time, controlled by  $m$ , used to compute changes in the variables. We have shown that these errors can have an adverse effect on the assessment of the leverage effect.

Fortunately, these errors are correctable to the extent where spot volatility is used for a certain range of  $m$ . There is still a substantial discretization bias that remains when using the spot volatility over a large time horizon, yet a reasonable large choice of  $m$  is necessary so that biases based on integrated volatility becomes correctable. This leads us to further correct 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 methods' effectiveness is demonstrated using both simulated examples and empirical study of real asset returns data.

In summary, a seemingly natural application of integrated volatility estimators to assess the leverage effect can lead to severe bias. 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. We show instead that to assess the leverage effect using high-frequency data, it is necessary to first do a preliminary bias correction and then further aggregate these preliminary estimates by running a local linear regression.

# Appendix

## A Preliminary Results

We first compute some moments that are related to the Heston model. They will be useful for proofs of Theorems 1–4. Throughout the appendix, we use the notation  $E_\nu$  and  $\text{Var}_\nu$  to denote the conditional mean and conditional variance given the latent volatility process  $\{\nu_t\}$ , and  $E_t$  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

$$dX_t = (\mu - \nu_t/2)dt + \rho\nu_t^{1/2}dW_t + \sqrt{1 - \rho^2}\nu_t^{1/2}dZ_t, \quad (24)$$

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

$$dY_r = [\gamma\mu - \rho\kappa\alpha + (\rho\kappa - \gamma/2)\nu_r] dr + \sqrt{1 - \rho^2}\nu_r^{1/2}dZ_r.$$

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

$$X_u - X_s = \int_s^u \left\{ (a + b\nu_t) dt + \sqrt{1 - \rho^2}\nu_t^{1/2}dZ_t \right\} + c(\nu_u - \nu_s). \quad (25)$$

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

$$E_\nu(X_u - X_s) = \int_s^u (a + b\nu_t)dt + c(\nu_u - \nu_s) \equiv \mu_\nu \quad (26)$$

and variance

$$\text{Var}_\nu(X_u - X_s) = (1 - \rho^2) \int_s^u \nu_t dt \equiv \sigma_\nu^2. \quad (27)$$

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

$$\begin{aligned} E_\nu(X_u - X_s)^2 &= \mu_\nu^2 + \sigma_\nu^2, \\ E_\nu(X_u - X_s)^3 &= \mu_\nu^3 + 3\mu_\nu\sigma_\nu^2, \\ E_\nu(X_u - X_s)^4 &= \mu_\nu^4 + 3\sigma_\nu^4 + 6\mu_\nu^2\sigma_\nu^2. \end{aligned}$$

## A.2 Cross-Moments of the Feller Process

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

$$E(\nu_t) = \alpha \text{ and } \text{Var}(\nu_t) = \frac{\gamma^2 \alpha}{2\kappa}. \quad (28)$$

Using again the Ito's formula, we have

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

which implies for  $s > t$ ,

$$E(\nu_s | \nu_t) = e^{-\kappa(s-t)} \nu_t + \alpha(1 - e^{-\kappa(s-t)}). \quad (29)$$

Similarly, by using the Ito's formula again,

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

This together with (29) imply that for  $s > t$ ,

$$\begin{aligned} E(\nu_s^2 | \nu_t) &= e^{-2\kappa(s-t)} \nu_t^2 + e^{-2\kappa s} \int_t^s (2\kappa \alpha + \gamma^2) e^{2\kappa u} E(\nu_u | \nu_t) du \\ &= e^{-2\kappa(s-t)} \nu_t^2 + \frac{2\kappa \alpha + \gamma^2}{\kappa} (\nu_t - \alpha) (e^{-\kappa(s-t)} - e^{-2\kappa(s-t)}) \\ &\quad + \frac{2\kappa \alpha^2 + \gamma^2 \alpha}{2\kappa} (1 - e^{-2\kappa(s-t)}). \end{aligned} \quad (30)$$

Therefore, for  $r \leq s$ ,

$$\begin{aligned} E(\nu_r \nu_s) &= E(\nu_r E(\nu_s | \nu_r)) \\ &= E[\nu_r^2 e^{-\kappa(s-r)} + \alpha(1 - e^{-\kappa(s-r)}) \nu_r] \\ &= \alpha^2 + \gamma^2 \alpha e^{-\kappa(s-r)} / (2\kappa). \end{aligned} \quad (31)$$

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

$$d(e^{\kappa t} \nu_t)^3 = 3e^{2\kappa t} \nu_t^2 (\kappa \alpha e^{\kappa t} dt + \gamma e^{\kappa t} \nu_t^{1/2} dW_t) + 3(e^{\kappa t} \nu_t) \gamma^2 e^{2\kappa t} \nu_t dt, \quad (32)$$

we have

$$E(e^{\kappa t} \nu_t)^3 = E\nu_0^3 + 3 \int_0^t (\kappa \alpha + \gamma^2) e^{3\kappa u} E\nu_u^2 du.$$

Using the fact that  $E\nu_t^3 = E\nu_0^3$  and  $E\nu_u^2 = \alpha^2 + \gamma^2\alpha/(2\kappa)$ , we deduce that

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

Recall that  $E_t$  denote the conditional expectation given the filtration up to time  $t$ . For  $s > t$ , we deduce from (32) that

$$E_t(e^{\kappa s}\nu_s)^3 = e^{\kappa t}\nu_t^3 + \int_t^s (3\kappa\alpha + 3\gamma^2)e^{3\kappa u}E_t\nu_u^2 du.$$

Now, substituting (30) into the above expression, we obtain after some calculation that

$$\begin{aligned} E_t\nu_s^3 = & e^{-3\kappa(s-t)} \left[ \nu_t^3 + 3\beta_1(e^{\kappa(s-t)} - 1)\nu_t^2 + 1.5\beta_1\beta_2(e^{2\kappa(s-t)} - 2e^{\kappa(s-t)} + 1)\nu_t \right. \\ & \left. + 0.5\alpha\beta_1\beta_2(3e^{\kappa(s-t)} - 3e^{2\kappa(s-t)} + e^{3\kappa(s-t)} - 1) \right], \end{aligned} \quad (33)$$

where  $\beta_1 = \alpha + \gamma^2/\kappa$  and  $\beta_2 = 2\alpha + \gamma^2/\kappa$ . For  $r < s < u$ , by using conditional expectation and (29), we have

$$\begin{aligned} E(\nu_r\nu_s\nu_u) &= E\nu_r\nu_s[\alpha + e^{-\kappa(u-s)}(\nu_s - \alpha)] \\ &= \alpha E(\nu_r\nu_s) + e^{-\kappa(u-s)}E[\nu_r E_r(\nu_s^2 - \alpha\nu_s)]. \end{aligned}$$

Substituting (29)-(31) into the above formula, the resulting expression involves only the first three moments of  $\nu_r$ , which has already been derived. Therefore, after some calculation, it follows that

$$E(\nu_r\nu_s\nu_u) = \alpha^3 + \frac{\gamma^2\alpha^2}{2\kappa} [e^{-\kappa(s-r)} + e^{-\kappa(u-r)} + e^{-\kappa(u-s)} + \gamma^2\kappa^{-1}\alpha^{-1}e^{-\kappa(u-r)}]. \quad (34)$$

The fourth order cross-moment can be derived analogously using what has already been derived along with the Ito formula:

$$d(e^{\kappa t}\nu_t)^4 = 4e^{3\kappa t}\nu_t^3(\kappa\alpha e^{\kappa t}dt + \gamma e^{\kappa t}\nu_t^{1/2}dW_t) + 6(e^{2\kappa t}\nu_t^2)\gamma^2 e^{2\kappa t}\nu_t dt.$$

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

$$E(\nu_t^4) = \left(\alpha + \frac{3\gamma^2}{2\kappa}\right)\left(\alpha + \frac{\gamma^2}{\kappa}\right)\left(\alpha^2 + \frac{\gamma^2\alpha}{2\kappa}\right), \quad (35)$$

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

$$\begin{aligned} E(\nu_r\nu_s\nu_u\nu_t) = & \alpha^4 + \frac{\alpha^3\gamma^2}{2\kappa} \left[ e^{-\kappa(u-r)} + e^{-\kappa(t-r)} + e^{-\kappa(s-r)} + e^{-\kappa(u-s)} + e^{-\kappa(t-s)} + e^{-\kappa(t-u)} \right] \\ & + \frac{\alpha^2\gamma^4}{2\kappa^2} \left[ e^{-\kappa(t+u-r-s)} + e^{-\kappa(u-r)} + 2e^{-\kappa(t-r)} + e^{-\kappa(s+t-u-r)}/2 + e^{-\kappa(t-s)} \right] \\ & + \frac{\alpha\gamma^6}{4\kappa^3} \left[ e^{-\kappa(t+u-r-s)} + 2e^{-\kappa(t-r)} \right]. \end{aligned} \quad (36)$$

## B Proofs of the Theorems

### B.1 Proof of Theorem 1

Let us first compute the covariance. It follows from (25) that

$$\begin{aligned} & \text{Cov}(\nu_{t+s} - \nu_t, X_{t+s} - X_t) \\ &= E(\nu_{t+s} - \nu_t) \left[ \int_t^{t+s} (a + b\nu_u) du + c(\nu_{t+s} - \nu_t) \right] \\ &= b \int_t^{t+s} [E(\nu_u \nu_{t+s}) - E(\nu_u \nu_t)] du + cE(\nu_{t+s} - \nu_t)^2. \end{aligned}$$

Now, using the moment formulas (28) and (31) and some simple calculus, we have

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

By using (28) and (31) again, we easily obtain

$$\text{Var}(\nu_{t+s} - \nu_s) = \gamma^2\alpha[1 - \exp(-\kappa s)]/\kappa. \quad (37)$$

Hence, it remains to compute  $\text{Var}(X_{t+s} - X_t)$ . By (26) and (27),

$$\begin{aligned} \text{Var}(X_{t+s} - X_t) &= \text{Var}(\mu_\nu) + E(\sigma_\nu^2) = E(\mu_\nu^2 + \sigma_\nu^2) - (E(\mu_\nu))^2 \\ &= \int_t^{t+s} \int_t^{t+s} E(a + b\nu_r)(a + b\nu_u) dr du + 2bc \int_t^{t+s} E\nu_r(\nu_{t+s} - \nu_t) dr \\ &\quad + c^2 E(\nu_{t+s} - \nu_t)^2 + (1 - \rho^2)\alpha s - \left( \int_t^{t+s} a + bE(\nu_r) dr \right)^2. \end{aligned}$$

Using the moments for  $\nu_t$  computed in section A.2, after some calculus, we obtain

$$\text{Var}(X_{t+s} - X_t) = \left( s + \frac{e^{-\kappa s} - 1}{\kappa} \right) \left( \frac{\gamma^2\alpha}{4\kappa^2} - \frac{\gamma\alpha\rho}{\kappa} \right) + \alpha s. \quad (38)$$

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 the Proposition 1.

### B.2 Proof of Theorem 2 and Proposition 2

Recall  $V_{t,\Delta} = \int_{t-\Delta}^t \nu_s ds$ . 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  $\{\nu_t\}$ ,

we have

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

Now, by (31), the above variance is given by

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

Simple calculus leads to

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

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

$$\frac{\text{Var}(V_{t+m\Delta,\Delta} - V_{t,\Delta})}{\Delta^2 \text{Var}(\nu_{t+m\Delta} - \nu_t)(1 - 1/3m)} = 1 + R_v(m, \Delta), \quad (39)$$

where  $R_v(m, \Delta)$  is such that

$$\limsup_{m\Delta \rightarrow 0} \frac{|R_v(m, \Delta)|}{\Delta} < \infty.$$

In particular,  $R_v(m, \Delta) = O(\Delta)$  for any  $m$  as  $\Delta \rightarrow 0$ , and  $R_v(m, \Delta) = o(m\Delta)$  if  $m \rightarrow \infty$  and  $m\Delta \rightarrow 0$ .

Next, we compute the covariance. By (26) and the double expectation formula, we have

$$\begin{aligned}& \text{Cov}(V_{t+m\Delta,\Delta} - V_{t,\Delta}, X_{t+m\Delta} - X_t) \\ &= E \left[ \int_{t+(m-1)\Delta}^{t+m\Delta} \nu_s ds - \int_{t-\Delta}^t \nu_s ds \right] \left[ \int_t^{t+m\Delta} (a + b\nu_r) dr + c(\nu_{t+m\Delta} - \nu_t) \right] \\ &= b \int_{t+(m-1)\Delta}^{t+m\Delta} \int_t^{t+m\Delta} E(\nu_s \nu_r) dr ds - b \int_{t-1}^t \int_t^{t+m\Delta} E(\nu_s \nu_r) dr ds \\ & \quad + c \int_{t+(m-1)\Delta}^{t+m\Delta} E\nu_s(\nu_{t+m\Delta} - \nu_t) ds - c \int_{t-1}^t E\nu_s(\nu_{t+m\Delta} - \nu_t) ds.\end{aligned}$$

Using (31), after some calculus, we obtain that

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

where  $A_2$  is again as given in Theorem 2. The conclusion of Theorem 2 follows from (38) and the above results. Comparing this with the covariance based on the spot volatilities, we have

$$\frac{\text{Cov}(V_{t+m\Delta,\Delta} - V_{t,\Delta}, X_{t+m\Delta} - X_t)}{\Delta \text{Cov}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t)(1 - 1/2m)} = 1 + R_c(m, \Delta), \quad (40)$$

where  $R_c(m, \Delta)$  satisfies  $\limsup_{m\Delta \rightarrow 0} \frac{|R_c(m, \Delta)|}{\Delta} < \infty$ .

By (39) and (40) the following asymptotic expressions are easily obtained:

$$\begin{aligned} \text{Corr}(V_{t+m\Delta, \Delta} - V_{t, \Delta}, X_{t+m\Delta} - X_t) &= \text{Corr}(\nu_{t+m\Delta} - \nu_t, X_{t+m\Delta} - X_t) \frac{(2m-1)}{2\sqrt{m^2 - \frac{m}{3}}} \\ &+ \begin{cases} O(\Delta), & \text{when } \Delta \rightarrow 0 \text{ for any } m \\ o(m\Delta), & \text{when } m \rightarrow \infty, m\Delta \rightarrow 0, \end{cases} \end{aligned} \quad (41)$$

which proves the Proposition 2.

### B.3 Proof of Theorem 3 and Proposition 3

The calculation is very involved. We separate them into several subsections so that the structure of computation can be better recognized. Recall that  $n = \Delta/\delta$ . Without loss of generality, we assume that  $t = \Delta$  and rewrite  $\hat{V}_{\Delta, \Delta}^{RV} = \hat{V}_{\Delta}^{RV}$ . Note that it is easy to verify that  $E(\hat{V}_{(m+1)\Delta, \Delta}^{RV} - \hat{V}_{\Delta}^{RV}) = 0$ .

#### B.3.1 Calculation of $E[\hat{V}_{\Delta}^{RV}(X_{(m+1)\Delta} - X_{\Delta})]$

Note that  $\hat{V}_{\Delta}^{RV}$  and  $X_{(m+1)\Delta} - X_{\Delta}$  involve two different time intervals. By conditioning on the latent process  $\{\nu_t\}$ ,  $\hat{V}_{\Delta}^{RV}$  and  $X_{(m+1)\Delta} - X_{\Delta}$  are independent by (25). Thus,

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

Using (25)–(27), the above expectation is given by

$$\begin{aligned} &\sum_{i=0}^{n-1} E \left\{ \left[ \int_{i\delta}^{(i+1)\delta} (a + b\nu_r) dr + c\nu_{(i+1)\delta} - c\nu_{i\delta} \right]^2 + (1 - \rho^2) \int_{i\delta}^{(i+1)\delta} \nu_r dr \right\} \\ &\cdot \left\{ \int_{\Delta}^{(m+1)\Delta} (a + b\nu_r) dr + c\nu_{(m+1)\Delta} - c\nu_{\Delta} \right\}. \end{aligned} \quad (42)$$

Expanding the first curly bracket into four terms, we have 4 product terms with the second curly bracket in (42). Denote those four terms by  $I_1, \dots, I_4$ , respectively.

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

$$I_1 \equiv \sum_{i=0}^{n-1} \left\{ \int_{\Delta}^{(m+1)\Delta} (a + b\nu_r) dr + c\nu_{(m+1)\Delta} - c\nu_{\Delta} \right\} \left[ \int_{i\delta}^{(i+1)\delta} (a + b\nu_r) dr \right]^2.$$

Expressing the square-term above as the double integral,  $I_1$  involves only the third cross moment of the process  $\{\nu_t\}$ . By using (31) and (34), it follows that

$$\begin{aligned} I_1 &= \sum_{i=0}^{n-1} \int_{\Delta}^{(m+1)\Delta} \int_{i\delta}^{(i+1)n\delta} \int_{i\delta}^r -2 \frac{\gamma^4 a^3}{2\kappa^2 \alpha^2} e^{-\kappa(s-u)} dudrds \\ &\quad + \sum_{i=0}^{n-1} \int_{i\delta}^{(i+1)\delta} \int_{i\delta}^r 2ca^2 \frac{\gamma^4}{2\kappa^2 \alpha} \left[ e^{-\kappa((m+1)\Delta-u)} - e^{-\kappa(\Delta-u)} \right] dudr \\ &= - \frac{a^2 \gamma^4 (a + \alpha c \kappa)}{2\alpha^2 \kappa^2} \frac{m\Delta^3}{n} + R_1, \end{aligned}$$

where  $R_1$  satisfies that  $\lim_{n \rightarrow \infty} \sup_{m \geq 1} \sup_{\Delta \leq 1} \frac{|R_1|n}{m\Delta^3} = 0$ . In particular, we have,

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

as  $\Delta \rightarrow 0$  and  $n \rightarrow \infty$  or as  $m \rightarrow \infty$ ,  $m\Delta \rightarrow 0$ , and  $n \rightarrow \infty$ .

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

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

where the cross moment function of the process  $\{\nu_t\}$  is used. We have

$$I_2 = \frac{ac\gamma^4(a + \alpha c \kappa)}{2\alpha \kappa} \frac{m\Delta^3}{n} + R_2,$$

where  $R_2$  satisfies that  $\lim_{n \rightarrow \infty} \sup_{m \geq 1} \sup_{\Delta \leq 1} \frac{|R_2|n}{m\Delta^3} = 0$ . Hence,

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

as  $\Delta \rightarrow 0$  and  $n \rightarrow \infty$  or as  $m \rightarrow \infty$ ,  $m\Delta \rightarrow 0$ , and  $n \rightarrow \infty$ .

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

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

### B.3.2 Calculation of $E\hat{V}_{(m+1)\Delta,\Delta}^{RV}(X_{(m+1)\Delta} - X_\Delta)$ and covariance

By the definition of  $\hat{V}_{(m+1)\Delta,\Delta}^{RV}$ , it follows that

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

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 square bracket. Each of these terms can be treated similarly as those in Section B.3.1. That is, by conditioning on the process  $\{\nu_t\}$ , they can be reduced to the calculation of the cross moments of  $\{\nu_t\}$ , by using the conditional moments in section A.1. After tedious calculations involving the cross moments discussed in section A.2, we can obtain asymptotic expressions for  $J_1$ ,  $J_2$  and  $J_3$ . Using these together with what we get for  $I_1, \dots, I_4$ , we can easily obtain an asymptotic expression of  $\text{Cov}(\hat{V}_{(m+1)\Delta,\Delta}^{RV} - \hat{V}_\Delta^{RV}, X_{(m+1)\Delta} - X_\Delta)$ . Comparing this asymptotic expression with what we have obtained in Theorem 2, we conclude that

$$\begin{aligned} &m^{-1}\Delta^{-2} \text{Cov}(\hat{V}_{(m+1)\Delta,\Delta}^{RV} - \hat{V}_\Delta^{RV}, X_{(m+1)\Delta} - X_\Delta) \\ &= m^{-1}\Delta^{-2} \text{Cov}(V_{(m+1)\Delta,\Delta} - V_{\Delta,\Delta}, X_{(m+1)\Delta} - X_\Delta) + O\left(\frac{1}{n}\right) \text{ as } n \rightarrow \infty \\ &= m^{-1}\Delta^{-2} \text{Cov}(V_{(m+1)\Delta,\Delta} - V_{\Delta,\Delta}, X_{(m+1)\Delta} - X_\Delta) + o(m\Delta), \end{aligned} \quad (43)$$

as  $\Delta \rightarrow 0$ ,  $n\Delta \rightarrow C$  and  $m\Delta \rightarrow 0$ .

### B.3.3 Calculation of the variance of changes of estimated RV

Let  $Y_i = X_{(i+1)\delta} - X_{i\delta}$ . Then,

$$E(\hat{V}_\Delta^{RV})^2 = \sum_{i=0}^{n-1} EY_i^4 + 2 \sum_{i=1}^{n-1} \sum_{j=1}^{i-1} EY_i^2 Y_j^2. \quad (44)$$

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

$$\begin{aligned} EY_i^4 &= E \left( \int_{i\delta}^{(i+1)\delta} (a + b\nu_r) dr + c\nu_{(i+1)\delta} - c\nu_{i\delta} \right)^4 + 3E \left( (1 - \rho^2) \int_{i\delta}^{(i+1)\delta} \nu_r dr \right)^2 \\ &\quad + 6(1 - \rho^2)E \int_{i\delta}^{(i+1)\delta} \nu_r dr \cdot \left( \int_{i\delta}^{(i+1)\delta} (a + b\nu_r) dr + c\nu_{(i+1)\delta} - c\nu_{i\delta} \right)^2. \end{aligned}$$

By conditioning on the process  $\{\nu_t\}$ ,  $Y_i^2$  and  $Y_j^2$  are conditionally independent for  $j < i$ . Appealing to (26) and (27), we have that for  $j < i$

$$EY_i^2Y_j^2 = E \left\{ \left[ \left( \int_{i\delta}^{(i+1)\delta} (a + b\nu_r) dr + c\nu_{(i+1)\delta} - c\nu_{i\delta} \right)^2 + (1 - \rho^2) \int_{i\delta}^{(i+1)\delta} \nu_r dr \right] \cdot \left[ \left( \int_{j\delta}^{(j+1)\delta} (a + b\nu_r) dr + c\nu_{(j+1)\delta} - c\nu_{j\delta} \right)^2 + (1 - \rho^2) \int_{j\delta}^{(j+1)\delta} \nu_r dr \right] \right\}.$$

Both terms above only involve the cross moments of the process  $\{\nu_t\}$ . After tedious calculations, we can show that

$$E\left(\hat{V}_\Delta^{\text{RV}}\right)^2 = \left(\alpha^2 + \frac{\alpha\gamma^2}{2\kappa}\right)\Delta^2 - \frac{\alpha\gamma^2}{6}\Delta^3 + \frac{\alpha(3\rho^4 - 6\rho^2 + 4)(2\alpha\kappa + \gamma^2)}{\kappa} \frac{\Delta^2}{n} + o\left(\frac{\Delta}{n}\right), \quad (45)$$

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

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

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

Again, tedious calculations involving the cross moments of the process  $\{\nu_t\}$  yield

$$E\hat{V}_\Delta^{\text{RV}}\hat{V}_{(m+1)\Delta,\Delta}^{\text{RV}} = \frac{\alpha(2\alpha\kappa + \gamma^2)}{\kappa} \Delta^2 - \alpha\gamma^2 m \Delta^3 + o(\Delta^3), \quad (46)$$

as  $\Delta \rightarrow 0$  and  $n\Delta \rightarrow C$ . Combination of (45) and (46) results in

$$\begin{aligned} \text{Var}(\hat{V}_{(m+1)\Delta,\Delta}^{\text{RV}} - \hat{V}_\Delta^{\text{RV}}) &= E(\hat{V}_\Delta^{\text{RV}})^2 + E(\hat{V}_{(m+1)\Delta,\Delta}^{\text{RV}})^2 - 2E(\hat{V}_\Delta^{\text{RV}}\hat{V}_{(m+1)\Delta,\Delta}^{\text{RV}}) \\ &= 2 \left[ \left(\alpha^2 + \frac{\alpha\gamma^2}{2\kappa}\right)\Delta^2 - \frac{\alpha\gamma^2}{6}\Delta^3 + \frac{\alpha(2\alpha\kappa + \gamma^2)}{\kappa} \frac{\Delta^2}{n} \right] \\ & \quad - 2 \left[ \frac{\alpha(2\alpha\kappa + \gamma^2)}{\kappa} \Delta^2 - \alpha\gamma^2 m \Delta^3 \right] + o(\Delta^3). \end{aligned} \quad (47)$$

as  $\Delta \rightarrow 0$  and  $n\Delta \rightarrow C$ . Comparing this with the variance expression obtained in the proof of Theorem 2, we have

$$\text{Var}(\hat{V}_{(m+1)\Delta,\Delta}^{\text{RV}} - \hat{V}_\Delta^{\text{RV}}) = \text{Var}(V_{t+m\Delta,\Delta} - V_{t,\Delta}) + \frac{2\alpha(2\alpha\kappa + \gamma^2)}{\kappa} \frac{\Delta^2}{n} + o(\Delta^3),$$

Or, equivalently,

$$\begin{aligned} & m^{-1}\Delta^{-3} \text{Var}(\hat{V}_{(m+1)\Delta,\Delta}^{\text{RV}} - \hat{V}_{\Delta}^{\text{RV}}) \\ &= m^{-1}\Delta^{-3} \left( \text{Var}(V_{t+m\Delta,\Delta} - V_{t,\Delta}) + \frac{4E\sigma_t^4\Delta^2}{n} \right) + o\left(\frac{1}{m}\right) \end{aligned} \quad (48)$$

$$= m^{-1}\Delta^{-3} \text{Var}(V_{t+m\Delta,\Delta} - V_{t,\Delta}) + \frac{2\alpha(2\alpha\kappa + \gamma^2)}{\kappa C m} + o(m\Delta), \quad (49)$$

when  $m^2\Delta \rightarrow C_m$ .

### B.3.4 Adjustment to Leverage Parameter

Further from (37) and (39), we have

$$m^{-1}\Delta^{-3} \text{Var}(V_{t+m\Delta,\Delta} - V_{t,\Delta}) = \gamma^2\alpha - \frac{\gamma^2\alpha}{3m} - \frac{1}{2}\alpha\gamma^2\kappa m\Delta + o(m\Delta), \quad (50)$$

and (48) becomes

$$m^{-1}\Delta^{-3} \text{Var}(\hat{V}_{(m+1)\Delta,\Delta}^{\text{RV}} - \hat{V}_{\Delta}^{\text{RV}}) \quad (51)$$

$$= m^{-1}\Delta^{-3} \text{Var}(V_{t+m\Delta,\Delta} - V_{t,\Delta}) \left( 1 + \frac{6(2\alpha\kappa + \gamma^2)}{(3\gamma^2m - \gamma^2)\kappa C - \frac{3}{2}\gamma^2\kappa^2 C C_m} \right) + o(m\Delta). \quad (52)$$

By using (43), (51) and (41), we can easily obtain the following relationship:

$$\begin{aligned} & \text{Corr}(\hat{V}_{(m+1)\Delta,\Delta}^{\text{RV}} - \hat{V}_{\Delta}^{\text{RV}}, X_{(m+1)\Delta} - X_{\Delta}) \\ &= \text{Corr}(V_{(m+1)\Delta} - V_{\Delta}, X_{(m+1)\Delta} - X_{\Delta}) \cdot \frac{1}{\sqrt{1 + \frac{12\alpha\kappa + 6\gamma^2}{(3\gamma^2m - \gamma^2)\kappa C - \frac{3}{2}\gamma^2\kappa^2 C C_m}}} [1 + o(m\Delta)] \\ &= \text{Corr}(\nu_{m+t} - \nu_t, X_{m+t} - X_t) \cdot \frac{1 - 1/2m}{\sqrt{(1 + \frac{12\alpha\kappa + 6\gamma^2}{(3\gamma^2m - \gamma^2)\kappa C - \frac{3}{2}\gamma^2\kappa^2 C C_m})(1 - \frac{1}{3m})}} [1 + o(m\Delta)], \end{aligned}$$

as  $\Delta \rightarrow 0$ ,  $n\Delta \rightarrow C$  and  $m^2\Delta \rightarrow C_m$ .

Proposition 3 follows from (41), (43) and (48).

## B.4 Proof of Theorem 4 and Theorem 5

Under the assumptions that  $n^{1/3}\Delta \rightarrow C_{TSRV}$  and  $\sigma_\epsilon^2/\Delta \rightarrow C_\epsilon$ , we have

$$\begin{aligned} & m^{-1}\Delta^{-2} \text{Cov}(\hat{V}_{(m+1)\Delta,\Delta}^{\text{TSRV}} - \hat{V}_{\Delta}^{\text{TSRV}}, Z_{(m+1)\Delta} - Z_{\Delta}) \\ &= m^{-1}\Delta^{-2} \text{Cov}(V_{(m+1)\Delta,\Delta} - V_{\Delta,\Delta}, X_{(m+1)\Delta} - X_{\Delta}) + o(m\Delta) \end{aligned} \quad (53)$$

and

$$\begin{aligned}
& m^{-1} \Delta^{-3} \text{Var}(\hat{V}_{(m+1)\Delta, \Delta}^{\text{TSRV}} - \hat{V}_{\Delta}^{\text{TSRV}}) \\
&= m^{-1} \Delta^{-3} \left( \text{Var}(V_{t+m\Delta, \Delta} - V_{t, \Delta}) + \frac{16\theta_{\text{TSRV}}^{-2} \sigma_{\epsilon}^4}{n^{1/3}} + \frac{8\theta_{\text{TSRV}} E \sigma_t^4 \Delta^2}{3n^{1/3}} \right) + o(m\Delta) \tag{54}
\end{aligned}$$

$$\begin{aligned}
&= m^{-1} \Delta^{-3} \text{Var}(V_{t+m\Delta, \Delta} - V_{t, \Delta}) + \frac{16\theta_{\text{TSRV}}^{-2} C_{\epsilon}^2}{m C_{\text{TSRV}}} + \frac{8\theta_{\text{TSRV}} E \sigma_t^4}{3m C_{\text{TSRV}}} + o(m\Delta) \\
&= m^{-1} \Delta^{-3} \text{Var}(V_{t+m\Delta, \Delta} - V_{t, \Delta}) [1 + A_4 + B_4 + o(m\Delta)], \tag{55}
\end{aligned}$$

where  $A_4 = \frac{96\theta_{\text{TSRV}}^{-2} C_{\epsilon}^2}{C_{\text{TSRV}} \alpha \gamma^2 (6m-2-3\kappa C_m)}$ ,  $B_4 = \frac{8\theta_{\text{TSRV}} (2\alpha\kappa + \gamma^2)}{\kappa C_{\text{TSRV}} \gamma^2 (6m-2-3\kappa C_m)}$ , by (50).

Under the assumptions that  $n^{1/2} \Delta \rightarrow C_{\text{PAV}}$  and  $\sigma_{\epsilon}^2 / \Delta \rightarrow C_{\epsilon}$ , with the constants  $\psi_2$ ,  $\Phi_{11}$ ,  $\Phi_{12}$ ,  $\Phi_{22}$  as specified in Theorem 5, we have

$$\begin{aligned}
& m^{-1} \Delta^{-2} \text{Cov}(\hat{V}_{(m+1)\Delta, \Delta}^{\text{PAV}} - \hat{V}_{\Delta}^{\text{PAV}}, Z_{(m+1)\Delta} - Z_{\Delta}) \\
&= m^{-1} \Delta^{-2} \text{Cov}(V_{(m+1)\Delta, \Delta} - V_{\Delta, \Delta}, X_{(m+1)\Delta} - X_{\Delta}) + o(m\Delta), \tag{56}
\end{aligned}$$

and

$$\begin{aligned}
& m^{-1} \Delta^{-3} \text{Var}(\hat{V}_{(m+1)\Delta, \Delta}^{\text{PAV}} - \hat{V}_{\Delta}^{\text{PAV}}) \\
&= m^{-1} \Delta^{-3} \left( \text{Var}(V_{t+m\Delta, \Delta} - V_{t, \Delta}) + \frac{8\Phi_{22} \theta_{\text{PAV}} E \sigma_t^4 \Delta^2}{\psi_2^2 n^{1/2}} + \frac{16\Phi_{12} E \sigma_t^2 \sigma_{\epsilon}^2 \Delta}{\theta_{\text{PAV}} \psi_2^2 n^{1/2}} + \frac{8\Phi_{11} \sigma_{\epsilon}^4}{\theta_{\text{PAV}}^2 \psi_2^2 n^{1/2}} \right) + o(m\Delta) \tag{57}
\end{aligned}$$

$$\begin{aligned}
&= m^{-1} \Delta^{-3} \text{Var}(V_{t+m\Delta, \Delta} - V_{t, \Delta}) + \frac{8\Phi_{22} \theta_{\text{PAV}} E \sigma_t^4}{m \psi_2^2 C_{\text{PAV}}} + \frac{16\Phi_{12} E \sigma_t^2 C_{\epsilon}}{m \theta_{\text{PAV}} \psi_2^2 C_{\text{PAV}}} + \frac{8\Phi_{11} C_{\epsilon}^2}{m \theta_{\text{PAV}}^2 \psi_2^2 C_{\text{PAV}}} + o(m\Delta) \\
&= m^{-1} \Delta^{-3} \text{Var}(V_{t+m\Delta, \Delta} - V_{t, \Delta}) [1 + A_5 + B_5 + C_5 + o(m\Delta)], \tag{58}
\end{aligned}$$

where  $A_5 = \frac{24\Phi_{22} \theta_{\text{PAV}} (2\alpha\kappa + \gamma^2)}{\psi_2^2 C_{\text{PAV}} \kappa \gamma^2 (6m-2-3\kappa C_m)}$ ,  $B_5 = \frac{96\Phi_{12} C_{\epsilon}}{\theta_{\text{PAV}} \psi_2^2 C_{\text{PAV}} \gamma^2 (6m-2-3\kappa C_m)}$  and  $C_5 = \frac{48\Phi_{11} C_{\epsilon}^2}{\theta_{\text{PAV}}^2 \psi_2^2 C_{\text{PAV}} \alpha \gamma^2 (6m-2-3\kappa C_m)}$ .

Similar as in section B.3.4, Theorem 4 follows by (53) and (55), and Theorem 5 by (56) and (58). Proposition 4 follows by (53) and (54), and Proposition 5 by (56) and (57).

## References

- Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010), “High-Frequency Covariance Estimates with Noisy and Asynchronous Data,” *Journal of the American Statistical Association*, 105, 1504–1517.
- Aït-Sahalia, Y. and Kimmel, R. (2010), “Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions,” *Journal of Financial Economics*, 98, 113–144.
- Aït-Sahalia, Y., Mykland, P. A., and 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.
- (2011), “Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise,” *Journal of Econometrics*, 160, 190–203.
- Bandi, F. M. and 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., and Shephard, N. (2008a), “Designing realized kernels to measure ex-post variation of equity prices in the presence of noise,” *Econometrica*, 76, 1481–1536.
- (2008b), “Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading,” Tech. rep., Department of Mathematical Sciences, University of Aarhus.
- Bekaert, G. and 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., and Tauchen, G. (2006), “Leverage and Volatility Feedback Effects in High-Frequency Data,” *Journal of Financial Econometrics*, 4, 353–384.
- Bollerslev, T., Sizova, N., and Tauchen, G. T. (2011), “Volatility in Equilibrium: Asymmetries and Dynamic Dependencies,” *Review of Finance*, forthcoming.

- Campbell, J. Y. and 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. and 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. and Ng, V. K. (1993), “Measuring and Testing the Impact of News on Volatility,” *The 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. and 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. and Wang, X. (2000), “Is the ”Leverage Effect” a Leverage Effect?” Tech. rep., New York University.
- French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987), “Expected stock returns and volatility,” *Journal of Financial Economics*, 19, 3–29.
- Gatheral, J. and Oomen, R. C. (2010), “Zero-intelligence realized variance estimation,” *Finance and Stochastics*, 14, 249–283.
- Griffin, J. and Oomen, R. (2008), “Sampling returns for realized variance calculations: Tick time or transaction time?” *Econometric Reviews*, 27, 230–253.
- Hansen, P. R. and Lunde, A. (2006), “Realized Variance and Market Microstructure Noise,” *Journal of Business and Economic Statistics*, 24, 127–161.
- Hayashi, T. and 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., and Vetter, M. (2009), “Microstructure Noise in the Continuous Case: The Pre-Averaging Approach,” *Stochastic Processes and Their Applications*, 119, 2249–2276.
- Kalnina, I. and Linton, O. (2008), “Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error,” *Journal of Econometrics*, 147, 47–59.
- Kinnebrock, S. and Podolskij, M. (2008), “Estimation of the Quadratic Covariation Matrix in Noisy Diffusion Models,” Tech. rep., Oxford-Man Institute, University of Oxford.
- Large, J. (2007), “Accounting for the Epps Effect: Realized Covariation, Cointegration and Common Factors,” Tech. rep., Oxford-Man Institute, University of Oxford.
- Li, Y., Mykland, P., Renault, E., Zhang, L., and Zheng, X. (2010), “Realized Volatility When Sampling Times are Possibly Endogenous,” Tech. rep., Hong Kong University of Science and Technology.
- Li, Y. and Mykland, P. A. (2007), “Are Volatility Estimators Robust with Respect to Modeling Assumptions?” *Bernoulli*, 13, 601–622.
- Nelson, D. B. (1991), “Conditional Heteroskedasticity in Asset Returns: A New Approach,” *Econometrica*, 59, 347–370.
- Voev, V. and Lunde, A. (2007), “Integrated Covariance Estimation Using High-Frequency Data in the Presence of Noise,” *Journal of Financial Econometrics*, 5, 68–104.
- 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.

— (2011), “Estimating Covariation: Epps Effect and Microstructure Noise,” *Journal of Econometrics*, 160, 33–47.

Zhang, L., Mykland, P. A., and Ait-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.

Table 1: Performance of the bias correction method based on asymptotic formulas and linear regression. The 100 estimates of  $\rho$  are summarized by its minimum, first quartile, median, third quartile, maximum, mean and SD.

	Min.	1st Qu.	Median	3rd Qu.	Max.	Mean	SD
corNutorho	-0.8555	-0.8172	-0.8005	-0.7808	-0.7511	-0.7999	0.026
corVtorho	-0.8827	-0.8238	-0.8019	-0.7814	-0.7415	-0.8041	0.031
corRVtorho	-0.9282	-0.8323	-0.7993	-0.7709	-0.7075	-0.8007	0.043
corPAVtorho	-1.0330	-0.8320	-0.7921	-0.7174	-0.6138	-0.7794	0.079
corTSRVtorho	-1.0760	-0.8351	-0.7776	-0.7020	-0.4765	-0.7744	0.117

Table 2: Performance of the bias correction method based on nonparametrically estimated asymptotic quantities and linear regression. The 100 estimates of  $\rho$  are summarized by its minimum, first quartile, median, third quartile, maximum, mean and SD.

	Min.	1st Qu.	Median	3rd Qu.	Max.	Mean	SD
corRVtorhoE	-0.8916	-0.8250	-0.8007	-0.7779	-0.7312	-0.8029	0.034
corPAVtorhoE	-0.9530	-0.8380	-0.7941	-0.7513	-0.6545	-0.7941	0.062
corTSRVtorhoE	-1.9230	-0.9647	-0.8241	-0.7452	-0.5587	-0.8742	0.211

Table 3: The results based on the sample frequency  $n = 1560$ , one observation per 15 seconds. They show the performance of the bias correction method based on asymptotic formulas and linear regression (“corNutorho”, “corVtorho”, “corRVtorho”, “corPAVtorho”, “corTSRVtorho”); and that of bias correction method based nonparametrically estimated asymptotic quantities and linear regression (“corRVtorhoE”, “corPAVtorhoE”, “corTSRVtorhoE”). The 100 estimates of  $\rho$  are summarized by its minimum, first quartile, median, third quartile, maximum, mean and SD. The range for regression is  $m = (6, 7, \dots, 17)$ .

	Min.	1st Qu.	Median	3rd Qu.	Max.	Mean	SD
corNutorho	-0.8527	-0.8141	-0.8023	-0.7812	-0.7316	-0.7988	0.024
corVtorho	-0.8727	-0.8253	-0.8099	-0.7869	-0.7223	-0.8053	0.029
corRVtorho	-0.8777	-0.8285	-0.8104	-0.7875	-0.6944	-0.8051	0.033
corPAVtorho	-0.8968	-0.8313	-0.7989	-0.7558	-0.6119	-0.7942	0.058
corTSRVtorho	-0.9921	-0.8709	-0.7953	-0.7313	-0.5364	-0.7999	0.094
corRVtorhoE	-0.8745	-0.8270	-0.8093	-0.7872	-0.7067	-0.8056	0.030
corPAVtorhoE	-0.8962	-0.8287	-0.7973	-0.7682	-0.6743	-0.7970	0.045
corTSRVtorhoE	-1.7800	-0.9657	-0.8850	-0.8184	-0.6027	-0.9147	0.175

Table 4: The sample correlation between the returns of S&P500 (2004-2007) and its estimated changes of volatilities, using TSRV, PAV and VIX (squared).

m	1	2	5	21	63	126	252
TSRV	-0.196	-0.257	-0.333	-0.434	-0.309	-0.267	-0.122
PAV	-0.255	-0.318	-0.403	-0.494	-0.368	-0.317	-0.152
VIX(sq)	-0.784	-0.774	-0.761	-0.792	-0.614	-0.469	-0.114

Table 5: The sample correlation between the returns of Microsoft and its estimated changes of volatilities, using TSRV and PAV with sampling frequencies at one per minute and one per 5 seconds.

m	1	2	5	10	21	63	126	252
One observation per minute								
TSRV	0.087	0.049	0.016	-0.027	-0.120	-0.290	-0.363	-0.221
PAV	0.030	-0.002	-0.017	-0.039	-0.169	-0.339	-0.405	-0.280
One observation per 5 seconds								
TSRV	-0.006	-0.034	-0.047	-0.069	-0.207	-0.364	-0.362	-0.345
PAV	-0.051	-0.074	-0.102	-0.115	-0.257	-0.412	-0.400	-0.404

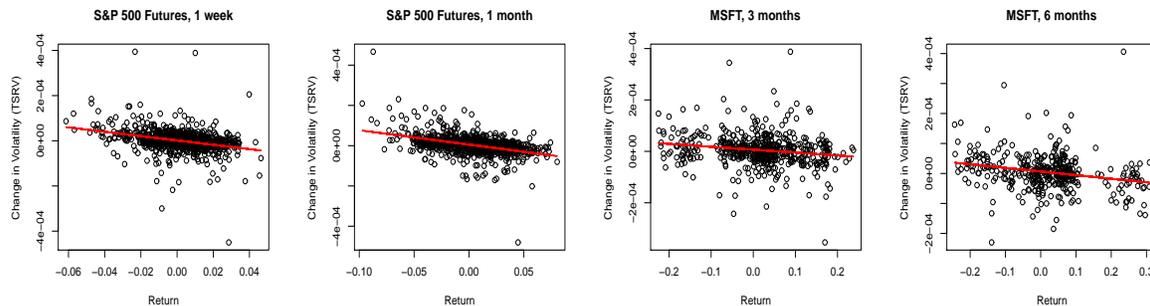


Figure 1: Scatter plots of differences of estimated daily volatility  $\hat{V}_t - \hat{V}_{t-m}$  versus returns over relatively long time span  $m$  for S&P 500 futures 2004-2007 data and Microsoft data from January 2005 to June 2007. Daily volatilities are estimated using TSRV based on high frequency 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 5 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 red lines are the least squares regression lines.

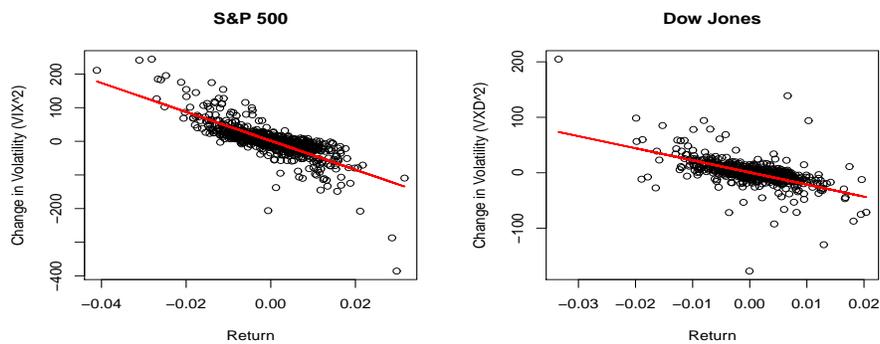


Figure 2: Changes of squared volatility indices versus returns. Using the volatility indices as the proxy of volatility, the leverage effect can clearly be seen. Left panel: S&P 500 data from January 03, 2004 to December 05, 2007, in which the VIX is used as a proxy of the volatility; Right panel: Dow Jones Industrial Average data from January 03, 2005 to March 30, 2007 in which the CBOE DJIA Volatility Index (VXD) is used as the volatility measure.

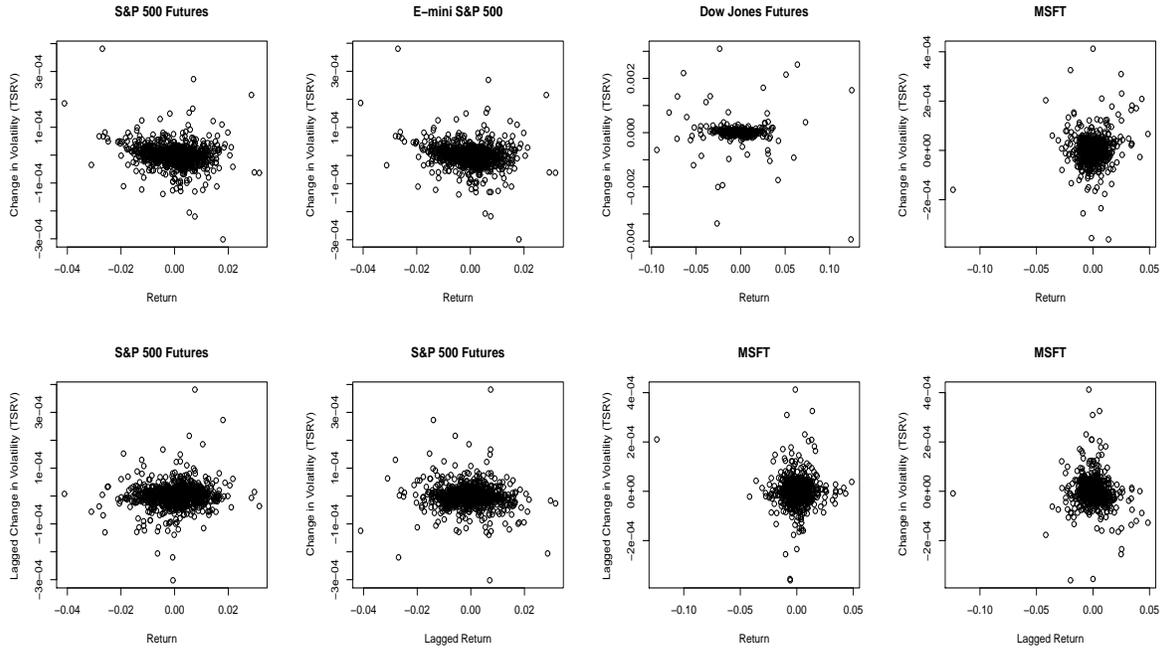


Figure 3: Upper: Scatter plots of the changes of estimated daily volatilities versus daily returns. Daily volatilities are estimated using TSRV based on high frequency minute-by-minute observations, and returns are calculated based on daily closing prices. From left to right: S&P 500 futures 2004-2007 data, E-mini S&P 500 2004-2007 data, Dow Jones futures January 2005 – March 2007 data, Microsoft January 2005 – June 2007 data. 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).

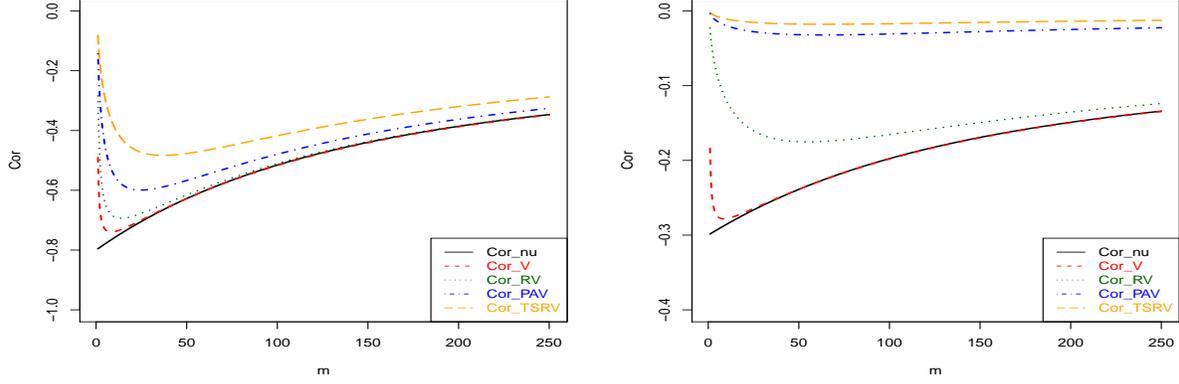


Figure 4: The theoretical estimated leverage effect parameter  $\rho$  as a function of the tuning parameter  $m$  when  $\Delta$  is taken to be  $1/252$ ; using the spot volatility (**Cor\_nu**), ideally estimated spot volatility (**Cor\_V**), realized volatility estimator (**Cor\_RV**), pre-averaging volatility estimator, (**Cor\_PAV**) and two-time scale volatility estimator (**Cor\_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 5 respectively. Two sets of parameters are used. Left panel:  $(\rho, \kappa, \gamma, \alpha) = (-0.8, 5, 0.5, 0.1)$ ; right panel:  $(\rho, \kappa, \gamma, \alpha) = (-0.3, 5, 0.05, 0.04)$ .

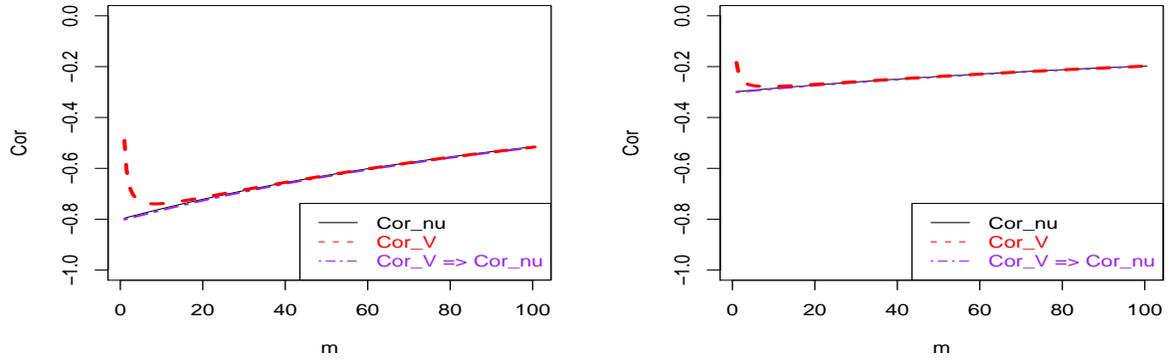


Figure 5: The effectiveness of multiplicative correction of smoothing bias, based on the main term in (17). After correction, the estimate of  $\rho$  is approximately the same as that based on the observable spot volatility. Left panel:  $(\rho, \kappa, \gamma, \alpha, \Delta) = (-0.8, 5, 0.5, 0.1, 1/252)$ ; right panel:  $(\rho, \kappa, \gamma, \alpha, \Delta) = (-0.3, 5, 0.05, 0.04, 1/252)$ .

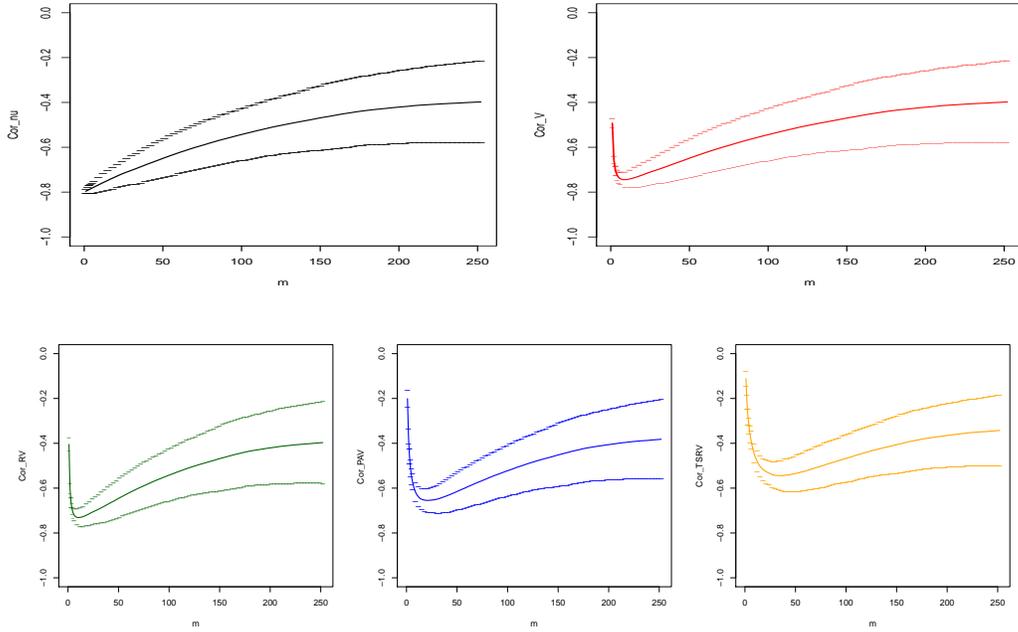


Figure 6: Sample correlation between the log-returns and the changes of spot volatility (upper left), the changes of integrated volatility (upper right), the differences of realized volatility (lower left), the differences of PAV (lower middle), and the differences of TSRV (lower right) over a period of  $m$  days. The results are based on 100 simulations. The solid curve is the average over 100 simulations. The dots are one standard deviations away from the averages. Parameters:  $(\rho, \kappa, \gamma, \alpha, \sigma_\epsilon, \Delta, n) = (-0.8, 5, 0.5, 0.1, 0.0005, 1/252, 390)$ .

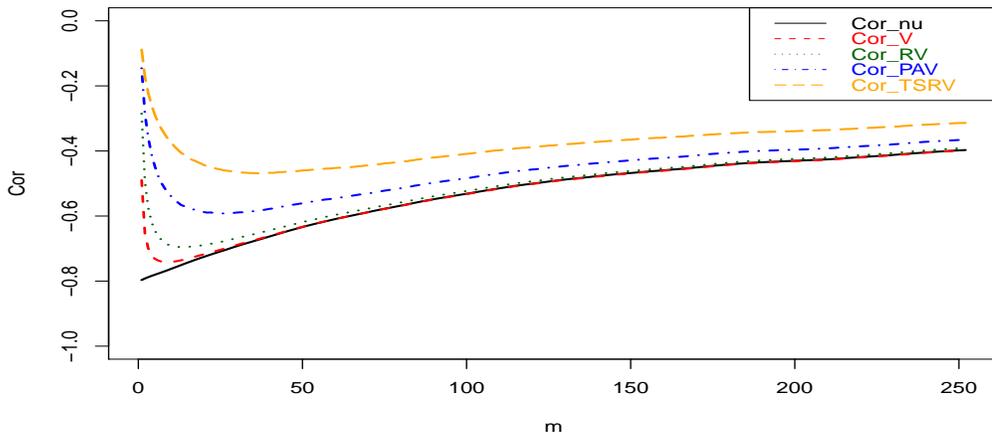


Figure 7: The average sample correlations between the changes of log-prices over a period of  $m$  days and the difference of the spot volatility  $\nu$ , the difference of the integrated volatility  $V$ , the difference of the RV estimates, and the difference of the TSRV estimates over the same period. Comparing this with the left panel of Figure 4, we see how the simulation results are in line with the theory. Parameters:  $(\rho, \kappa, \gamma, \alpha, \sigma_\epsilon, \Delta, n) = (-0.8, 5, 0.5, 0.1, 0.0005, 1/252, 390)$

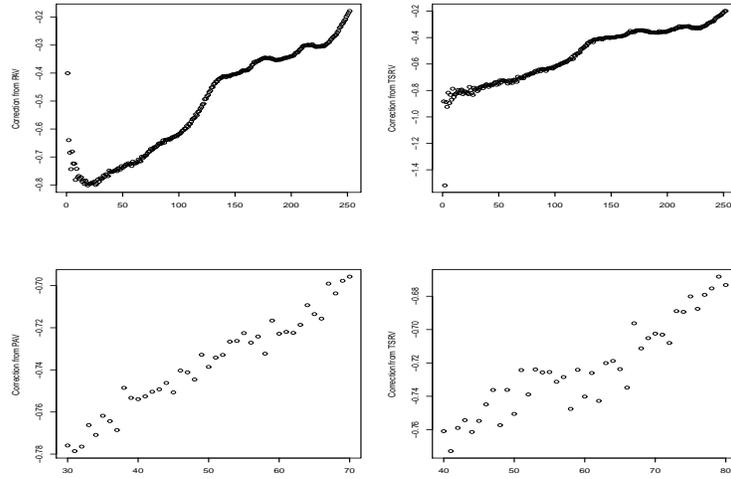


Figure 8: The scatter plot of preliminary bias corrected estimates of the leverage effect parameter  $\rho$  against  $m$  for one simulated realization. Left panel is based on the PAV and the right panel is based on TSRV. The bottom panel is the zoomed version of the plots on the top panel, in which the first linearly increasing interval is depicted. The estimates in this range are further aggregated by using a simple linear regression to obtain a final estimate of the leverage effect.

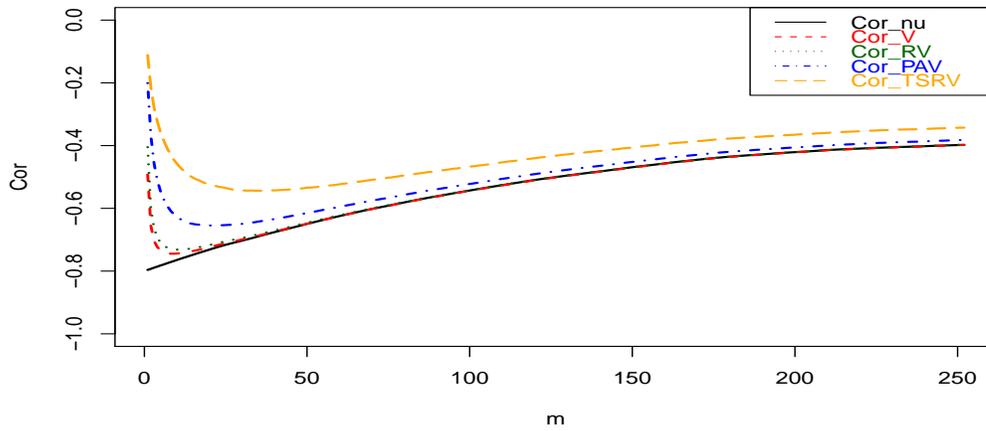


Figure 9: The same as Figure 7 except that the data are sampled at every 15 seconds ( $n=1560$ ).

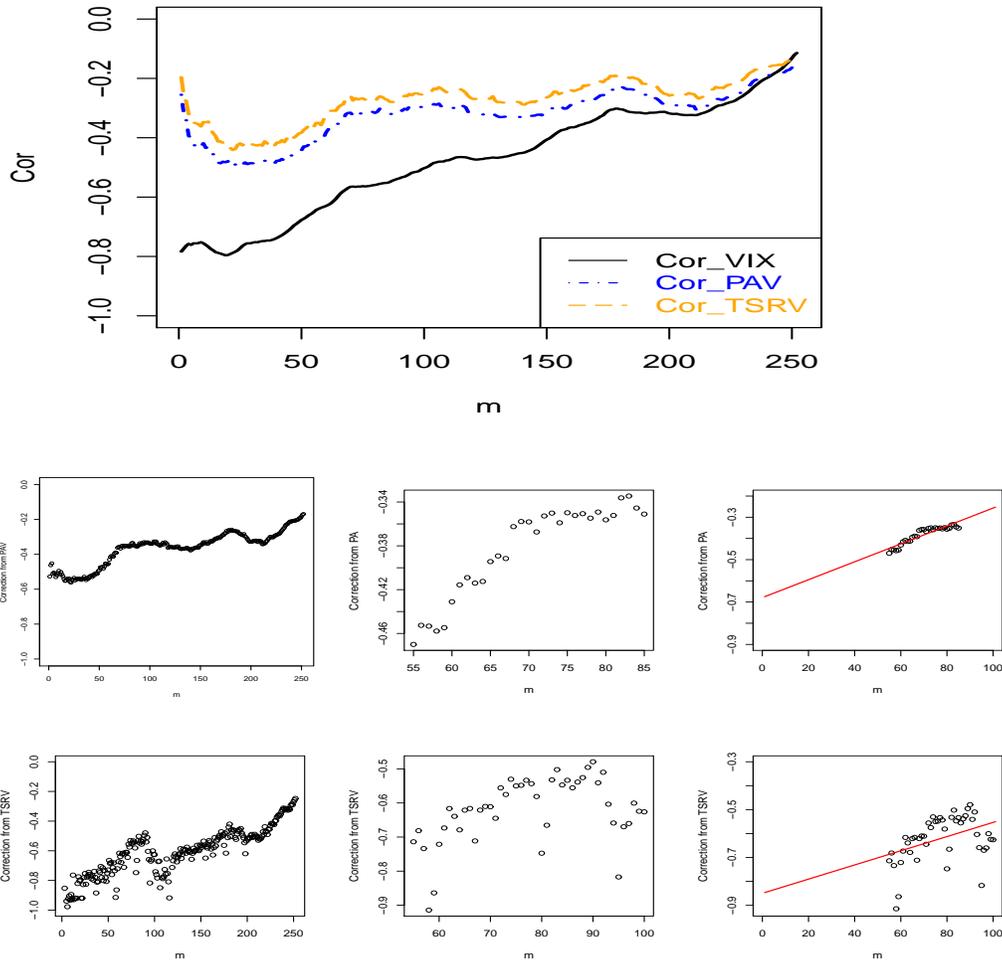


Figure 10: Upper: The raw sample correlations based on VIX (squared), PAV and TSRV, for different horizons  $m$  based on the minute-by-minute data of S&P500 returns in the time period 2004-2007. Lower: the scatter plots of preliminary bias corrected estimates of the leverage effect parameter  $\rho$  against  $m$  for the same data set. The middle plots are the zoomed versions of the plots on the left, in which the first linearly increasing interval is depicted. The plots on the right show how estimates in the range are further aggregated by using a simple linear regression to obtain a final estimate of the leverage effect.

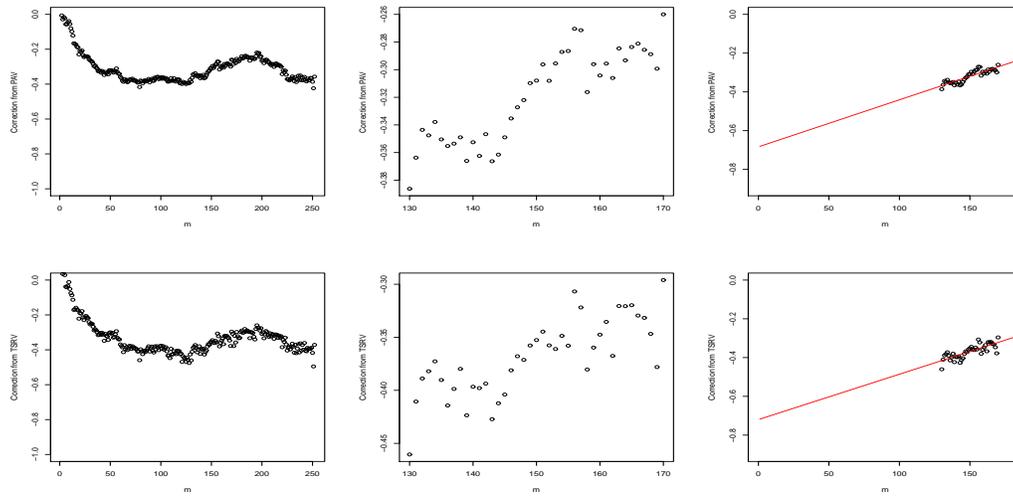


Figure 11: The scatter plots of preliminary bias corrected estimates of the leverage effect parameter  $\rho$  against  $m$  based on the minute-by-minute data of Microsoft returns in the time period Jan/2005-Jun/2007. Upper panel is based on the PAV and the lower panel is based on TSRV. The middle are the zoomed versions of the plots on the left, in which the first linearly increasing interval is depicted. The plots on the right show how estimates in the range are further aggregated by using a simple linear regression to obtain a final estimate of the leverage effect.

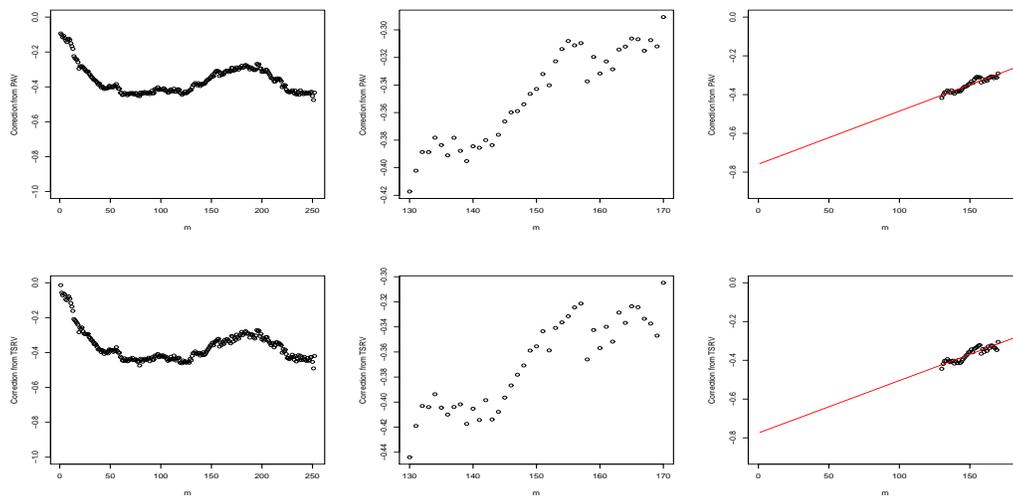


Figure 12: The same as in Figure 11 except that the data are sampled at the frequency of one observation per 5 seconds.