Ann Inst Stat Math (2016) 68:673–703
DOI 10.1007/s10463-015-0507-z

Optimal restricted quadratic estimator of integrated volatility

Liang-Ching Lin · Meihui Guo

Received: 1 March 2013 / Revised: 24 December 2014 / Published online: 3 March 2015

Abstract Estimation of the integrated volatility is an important problem in high-frequency financial data analysis. In this study, we propose a quadratic unbiased estimator of the integrated volatility for stochastic volatility models with microstructure noise. The proposed estimator minimizes the finite sample variance in the class of quadratic estimators based on symmetric Toeplitz matrices. We show the proposed estimator has an asymptotic mixed normal distribution with optimal convergence rate $n^{-1/4}$ and achieves the maximum likelihood estimator efficiency for constant volatility case. Simulation results show that our proposed estimator attains better finite sample efficiency than state-of-the-art methods. Finally, a real data analysis is conducted for illustration.

Keywords High-frequency data · Integrated volatility · Microstructure noise · Signal-to-noise ratio · Stochastic volatility model

1 Introduction

In security markets, financial data taken at a finer time scale such as tick-by-tick data have become readily available due to advances in data acquisition and processing techniques. These high-frequency data provide a rich source for volatility analysis,

L.-C. Lin · M. Guo (✉)
Department of Applied Mathematics, National Sun Yat-Sen University, Kaohsiung, Taiwan
e-mail: guomh@math.nsysu.edu.tw

L.-C. Lin
e-mail: lclin@mail.ncku.edu.tw

Present address:
L.-C. Lin
Department of Statistics, National Cheng Kung University, Tainan, Taiwan

which plays an important role in derivative pricing, portfolio allocation, and risk management. It is usually assumed that the high-frequency efficient log price process $\{p_t\}$ satisfies the following continuous-time stochastic volatility model (SVM),

$$\mathrm{d}p_t = u_t \mathrm{d}t + \sigma_t \mathrm{d}W_t \tag{1}$$

where u_t is the drift term, σ_t (the spot volatility at time t) is a continuous-time stochastic process and W_t is a Brownian motion. See for instance, Andersen et al. (2001), Aït-Sahalia et al. (2005), Zhang (2006), Barndorff-Nielsen et al. (2009), Reiss (2011), Lee and Guo (2012) and Lin et al. (2013). A leverage effect between the return and conditional volatility, see for example Model (21) in Sect. 4.1, may be considered in the models. The integrated volatility of $\{p_t\}$ in the unit interval [0, 1] is defined as the cumulated volatility in the interval, $\int_0^1 \sigma_s^2 ds$, which under Model (1) is equal to the quadratic variation of $\{p_t\}$, i.e.,

$$[p, p]_t = \lim_{\max \Delta_{t_i} \to 0} \sum_{i=1}^{n} \left(p_{t_i} - p_{t_{i-1}}\right)^2 = \int_0^1 \sigma_s^2 \mathrm{d}s, \tag{2}$$

where $(t_1, t_2, \cdots, t_n)$ denotes a partition of [0, 1] and $\Delta_{t_i} = t_i - t_{i-1}$. By (2), the sum of the squared efficient returns converges to the integrated volatility as the partition length Δ_{t_i} shrinks to zero. However, empirical evidence shows that sum of the observed high-frequency squared returns shoots up as the sampling time decreases to zero, see for example Fan and Wang (2007). To accommodate this empirical fact, microstructure noise component is encompassed in the price model of high-frequency data. Specifically, the true (efficient) price is usually assumed to be contaminated by the market microstructure effects, see for example, Aït-Sahalia et al. (2005) and Bandi and Russell (2006). Accordingly, in this paper, we assume that the observed log price process $\{\widetilde{p}_t\}$ satisfies the following model,

$$\widetilde{p}_t = p_t + \eta_t, \tag{3}$$

where p_t is the efficient log price at time t satisfying Model (1), the instantaneous volatility process σ_t satisfies

$$E|\sigma_s - \sigma_t|^{2q} \leq C_q |t - s|^q, \quad q \leq 2, \tag{4}$$

where $C_q > 0$ are constants and $\{\eta_t\}$ is a white noise process independent of $\{p_t\}$ with $E(\eta_t) = 0$, $\mathrm{Var}(\eta_t) = \sigma_\eta^2$, $E\left(\eta_t^4\right)/\left(E\left(\eta_t^2\right)\right)^2 = \lambda$ and $E\left(\eta_t^8\right) < \infty$. Barndorff-Nielsen et al. (2008) gave further discussion on the independence assumption between $\{p_t\}$ and $\{\eta_t\}$ and the white noise assumption of $\{\eta_t\}$. The microstructure noise results from either the information or the non-information-related factors, which include the bid-ask spread, the differences in trade sizes, informational asymmetries of traders, inventory control effects, the discreteness of price changes, and others.

Assume $\widetilde{p}_t$s are observed at the equispaced time points $(t_1, t_2, \cdots, t_n)$ where $t_i = i/n$. For the sake of simplicity, in the sequel we denote $p_{t_i} = p_i$, $\widetilde{p}_{t_i} = \widetilde{p}_i$ and

$\eta_{t_i} = \eta_i$. Hence, the observed log return at time t_i is

$$\widetilde{r}_i = \widetilde{p}_i - \widetilde{p}_{i-1} = r_i + \varepsilon_i,$$

where $r_i = p_i - p_{i-1}$ denotes the nominal return and $\varepsilon_i = \eta_i - \eta_{i-1}$ is regarded as the microstructure noise at t_i, respectively. Since $\{\eta_t\}$ is a white noise process, the microstructure noise process $\{\varepsilon_t\}$ is an MA(1) process with variance

$$\mathrm{Var}(\epsilon_t) = \sigma_\varepsilon^2 = 2\sigma_\eta^2. \tag{5}$$

Define the realized volatility,

$$\mathrm{RV} = \sum_{j=1}^{n} \widetilde{r}_j^2,$$

which represents the aggregate squared observed returns. Under the assumption of Model (3), the realized volatility

$$\mathrm{RV} = \sum_{j=1}^{n} r_j^2 + \sum_{j=1}^{n} \varepsilon_j^2 + 2\sum_{j=1}^{n} r_j \varepsilon_j = \int_0^1 \sigma_s^2 \mathrm{d}s + nE\left(\varepsilon^2\right) + o_p(n) \tag{6}$$

increases with the sample size (hence also with the sampling frequency), which echoes the aforementioned empirical fact. In view of (6), the RV of high-frequency data is mostly composed of the latent market microstructure noise, which hinders the estimation of the integrated volatility in practice.

In the literature, a lot of effort has been devoted to improve the estimation of integrated volatility. Some of the important contributions include but are not limited to Barndorff-Nielsen and Shephard (2002), Aït-Sahalia et al. (2005), Zhang et al. (2005), Zhang (2006), Bandi and Russell (2006), Fan and Wang (2007), Bandi and Russell (2008), Barndorff-Nielsen et al. (2008), Barndorff-Nielsen et al. (2009), and Reiss (2011). See also Zhou (1996), Andersen et al. (2000), Hansen and Lunde (2006), Kalnina and Linton (2008), Jacod et al. (2009), Sun (2006) and Xiu (2010) for related research in integrated volatility. Bibinger and Mykland (2013) discuss some relation of the aforementioned methods and show their asymptotic equivalence. In these studies, quadratic estimators of the form $\widetilde{R}' W \widetilde{R}$ is a popular choice, where $\widetilde{R} = (\widetilde{r}_1, \ldots, \widetilde{r}_n)'$ and $W = (w_{ij})_{1 \le i,j \le n}$ is a symmetric matrix. Examples include Zhang et al. (2005), Zhang (2006), Barndorff-Nielsen et al. (2008) and Jacod et al. (2009) which will be introduced below.

Zhang et al. (2005) proposed a "two-scale estimator" defined as

$$T_2^{(K)} = \frac{1}{K} \sum_{j>K} (\widetilde{p}_j - \widetilde{p}_{j-K})^2 - \frac{n-K+1}{nK} \mathrm{RV}$$

which is a bias-corrected average of the squared K-period returns and the RV is the realized volatility based on all the observed returns. The optimal period K is obtained by minimizing the asymptotic variance of the estimator $T_2^{(K)}$. Zhang et al. (2005) proved that the two-scale estimator has the convergence rate $n^{-1/6}$. The two-scale estimator is a quadratic type estimator and the weights matrix is $W = W^* - (n - K + 1)/(nK)I_n$ where I_n is an $n \times n$ identity matrix and $W^* = (w^*_{ij})_{1\leq i,j\leq n}$ is a symmetric matrix. When $j - i = d \geq K$ $w^*_{ij} = 0$ and for $0 \leq j - i = d \leq K - 1$

$$w^*_{ij} = \begin{cases} \dfrac{i}{K}, & i = 1, \ldots, K - 1 - d, \\ \dfrac{K-d}{K}, & i = K - d, \ldots, n - K + 1, \\ \dfrac{n+1-d-i}{K}, & i = n - K + 2, \ldots, n - d. \end{cases}$$

Zhang (2006) generalized the two-scale estimator to the following multi-scale realized volatility using m multiple sampling frequencies $(K_1, \ldots, K_m)$,

$$T_m = \sum_{i=1}^{m} a_i T_2^{(K_i)}, \quad a_i = 12\frac{1}{m^2}\frac{i/m - 1/2 - 1/(2m)}{1 - 1/m^2}.$$

The multi-scale estimator has a weak convergence rate $n^{-1/4}$. Barndorff-Nielsen et al. (2008) proposed the following realized kernel estimator

$$T_k = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right) L_h,$$

where the kernel function $k(x)$ is a weight function defined on [0, 1], and

$$L_h = L_{-h} = \sum_{j=1}^{n-h} \tilde{r}_j \tilde{r}_{j+h}, \tag{7}$$

is the lag $h(\geq 0)$ sample autocovariance. Barndorff-Nielsen et al. (2008) showed that the optimal kernel function minimizing the asymptotic variance is $k(x) = (1 + |x|)e^{-|x|}$ (see Proposition 1 of page 1498 in Barndorff-Nielsen et al. 2008) with $H = \xi n^{1/2}$ where $\xi^2 = \sigma_\eta^2 / \sqrt{\int_0^1 \sigma_s^4 ds}$. The realized kernel estimator has a weak convergence rate $n^{-1/4}$. Jacod et al. (2009) presented the following generalized pre-averaging approach

$$T_p = \frac{1}{k_n \psi_2} \sum_{i=0}^{n-k_n+1} (\bar{r}_i^n)^2 - \frac{\psi_1}{2k_n^2 \psi_2} \sum_{i=1}^{n} \tilde{r}_i^2, \tag{8}$$

where $k_n = c\sqrt{n} + o\left(n^{-1/4}\right)$,

$$\bar{r}_i^n = \sum_{j=1}^{k_n - 1} g(j/k_n)\tilde{r}_{i+j},$$

$\psi_1 = \int_0^1 \left(g'(u)\right)^2 \mathrm{d}u$ and $\psi_2 = \int_0^1 g^2(u)\mathrm{d}u$ for a given function g. The function g defined on [0, 1] is continuous, piecewise C^1 with a piecewise Lipschitz derivative g' and satisfies $g(0) = g(1) = 0$ and $\int_0^1 g(s)^2\mathrm{d}s > 0$. A simple choice is $g(x) = \min\{x, 1-x\}$ which implies $\psi_1 = 1$ and $\psi_2 = 1/12$. Koike (2014) proposed an optimal weight function for pre-averaging covariance estimation. The pre-averaging estimator has a weak convergence rate $n^{-1/4}$. Since the multi-scale estimator T_m, the realized kernel estimator T_k and the pre-averaging estimator T_p all are linear combinations of $\tilde{r}_i\tilde{r}_j$, by appropriately rearranging the coefficients, we can show that they are also a type of quadratic estimators. Moreover, the quasi-maximum likelihood estimation of Xiu (2010) also behaves like an iterative exponential realized kernel asymptotically in light of quadratic representation.

In this study, we proposed a quadratic estimator based on a symmetric Toeplitz matrix, i.e., each descending diagonal from left to right is constant. The proposed estimator is unbiased and minimizes the finite sample variance in the class of quadratic estimators based on symmetric Toeplitz matrices. Define the signal-to-noise ratio as

$$S_{nr} = \frac{E\left(\int_0^1 \sigma_s^2 \mathrm{d}s\right)}{n\sigma_\varepsilon^2},$$

which represents the ratio of the integrated volatility to the n-folds variance of the microstructure noise. The optimal weights are functions of S_{nr} solved from fifth order difference equations. A recursive algorithm, using the Newton method and the Gauss–Seidel method was developed to solve S_{nr} and the optimal weights. The proposed estimator converges weakly to a mixed normal distribution at the rate $n^{-1/4}$ and achieves the MLE efficiency in the constant volatility case. Note that $n^{-1/4}$ is the optimal convergence rate for integrated volatility estimators, see for example Aït-Sahalia et al. (2005) and Reiss (2011).

The main differences between the proposed estimator and $T_2^{(K)}$, T_m, T_k and T_p are listed below.

(i) The optimal weights of the proposed estimator are only constrained by unbiasedness condition. These data-driven weights dynamically adjust with the size of S_{nr}. The weights of $T_2^{(K)}$, T_m, T_k and T_p are constrained to be functions of parameters or kernels, such as $K, K_1, \ldots, K_m, k(x)$ or g.

(ii) The proposed estimator minimizes the finite sample variance in the class of Toeplitz type quadratic estimators, while $T_2^{(K)}$, T_m and T_k aim to minimize the asymptotic variances in the class of quadratic estimators with aforementioned constraint weights. In view of the slow $n^{-1/4}$ convergence rate of these estimators, we expect the proposed estimator attains better efficiency in practical application with smallish sample size. Our simulation results in Sect. 4 also support the advantage of the proposed estimator in both smallish and large samples.

Sun (2006) also considers the quadratic type estimator $\tilde{R}'W\tilde{R}$. The main differences between the estimators of ours and Sun (2006) are listed below.

(iii) Both Sun's (2006) and our estimators are obtained to minimize the finite sample variance based on unbiasedness condition. Sun (2006) derives the optimal weights under the distributional assumptions that the volatility does not change much within the sampling period, i.e., $\int_{t_{i-1}}^{t_i} \sigma_s^2 ds = \int_0^1 \sigma_s^2 \mathrm{d}s/n$, $\forall i = 1, \ldots, n$, and the kurtosis of η_t is 3. Our optimal weights are obtained by assuming that the weighting matrix is a symmetric Toeplitz matrix; no distributional assumptions of Sun (2006) are required. Hence, our approach is suitable for estimating the integrated volatility in either volatile or non-volatile market.

(iv) The optimal weights of Sun (2006) are functions of the ratio $\lambda_{\mathrm{Sun}} = \int_0^1 \sigma_s^2 \mathrm{d}s/(n\sigma_\eta^2)$, which is estimated by a consistent estimator. And our optimal weights are functions of S_{nr}, which will be recursively estimated via an algorithm.

(v) Since Sun's (2006) approach is based on a quadratic form of $n(n+1)/2$ non-zero weights, it requires $O(n^2)$ arithmetical operations to obtain the integrated volatility estimator. However, our proposed estimator only depends on $\ell(\ll n)$ weights [cf. (9)], which requires only $O(n)$ computational complexity.

The remainder of this paper is organized as follows. In Sect. 2, we describe our estimator of the integrated volatility, based on a linear function of sample autocovariances. The asymptotic distribution of the proposed estimator is derived. In Sect. 3, a recursive algorithm is developed to compute the optimal weights. A simulation and an empirical study are provided in Sect. 4. Finally, in Sect. 5, we draw our conclusions. Part of the proofs is provided in Appendix A and the figures and the tables are shown in Appendix B and Appendix C, respectively.

2 Optimal restricted quadratic estimator

Throughout, we assume the efficient log price process $\{p_t : t \geq 0\}$ satisfies the SVM (1) and the observed log price $\tilde{p}_t$ satisfies (3). The notations p_i, $\tilde{p}_i$, r_i and $\tilde{r}_i$ are defined as in the previous section. Consider the symmetric Toeplitz type quadratic estimator $S_L(\boldsymbol{\theta}) = \widetilde{R}'W\widetilde{R}$ where $W = (w_{ij})_{1\leq i,j\leq n}$ with $w_{ii} = \theta_0$, $i = 1, \ldots, n$, $w_{ij} = \theta_{|i-j|}/2$ if $1 \leq |i-j| \leq \ell$ and $w_{ij} = 0$ if $|i-j| > \ell$. Equivalently, we can express the proposed estimator S_L as a linear function of the sample autocovariances,

$$
\begin{aligned}
S_L(\boldsymbol{\theta}) &= \theta_0 \sum_{i=1}^{n+1} \widetilde{r}_i^2 + \theta_1 \sum_{i=1}^{n} \widetilde{r}_i\widetilde{r}_{i+1} + \theta_2 \sum_{i=1}^{n-1} \widetilde{r}_i\widetilde{r}_{i+2} + \ldots + \theta_\ell \sum_{i=1}^{n+1-\ell} \widetilde{r}_i\widetilde{r}_{i+\ell} \\
&= \theta_0 L_0 + \theta_1 L_1 + \theta_2 L_2 + \ldots + \theta_\ell L_\ell, \qquad (9)
\end{aligned}
$$

where L_h is the lag h sample autocovariance defined in (7), $\widetilde{r}_1 \equiv \widetilde{p}_1 = p_1 + \eta_1$ and $\widetilde{r}_{n+1} \equiv -\widetilde{p}_n = -p_n - \eta_n$. We assume $\ell/n \to 0$ as $n \to \infty$. The optimal weights $\boldsymbol{\theta}^* = (\theta_0^*, \theta_1^*, \theta_2^*, \ldots, \theta_\ell^*)'$ are chosen to satisfy the unbiasedness

$$
E\left(S_L(\boldsymbol{\theta}^*) - \int_0^1 \sigma_s^2 ds\right) = 0
$$

and the minimum variance condition

$$\operatorname{Var}\left(S_L(\boldsymbol{\theta}^*) - \int_0^1 \sigma_s^2 ds\right) = \min_{\boldsymbol{\theta}} \operatorname{Var}\left(S_L(\boldsymbol{\theta}) - \int_0^1 \sigma_s^2 ds\right).$$

Throughout, we let $\mathcal{G}$ denote the σ-field generated by $\{\sigma_t,\ t \geq 0\}$ and $E_{\mathcal{G}}(\cdot) = E[\cdot|\mathcal{G}]$ denote the conditional expectation with respect to $\mathcal{G}$. The conditional expectation is defined up to almost-sure equivalence. In the following Lemma 1, we derive the moments of the sample autocovariance function L_h and the estimator S_L when the drift term $u_t = 0$.

Lemma 1 (i) $E_{\mathcal{G}}(L_0) = \int_0^1 \sigma_s^2 \mathrm{d}s + n\sigma_\varepsilon^2$, $E_{\mathcal{G}}(L_1) = -\frac{n}{2}\sigma_\varepsilon^2$, *and* $E_{\mathcal{G}}(L_h) = 0$, $\forall\, h \geq 2$, *where* σ_ε^2 *is defined in* (5).

(ii) *If* $\theta_0 = 1$ *and* $\theta_1 = 2$, *then* $E_{\mathcal{G}}\left(S_L - \int_0^1 \sigma_s^2 \mathrm{d}s\right) = 0$ *and hence* S_L *is an unbiased estimator of* $\int_0^1 \sigma_s^2 \mathrm{d}s$, *that is* $E\left(S_L - \int_0^1 \sigma_s^2 \mathrm{d}s\right) = 0$.

Remark 1 When the drift term $u_t \neq 0$, the nominal return

$$r_i^u = r_i + \int_{t_{i-1}}^{t_i} u_s \mathrm{d}s,$$

where r_i denotes the nominal return when the drift term $u_t = 0$. Then, the observed log return $\tilde{r}_i = r_i^u + \varepsilon_i$ and $L_h = \sum_{i=1}^{n+1-h} \tilde{r}_i \tilde{r}_{i+h}$, $h = 0, \ldots, \ell$. The results of Lemma 1 are modified as follows:

(i) $E_{\mathcal{G}}(L_0) = \int_0^1 \sigma_s^2 \mathrm{d}s + n\sigma_\varepsilon^2 + O(1/n)$, $E_{\mathcal{G}}(L_1) = -\frac{n}{2}\sigma_\varepsilon^2 + O(1/n)$, $E_{\mathcal{G}}(L_h) = O(1/n)$, $h = 2, \ldots, \ell$.

(ii) If $\theta_0 = 1$ and $\theta_1 = 2$, then $E_{\mathcal{G}}\left(S_L - \int_0^1 \sigma_s^2 \mathrm{d}s\right) = O(\ell/n)$ and hence S_L is asymptotically unbiased.

In view of Lemma 1, hereinafter we consider

$$S_L(\boldsymbol{\theta}) = L_0 + 2L_1 + \sum_{i=2}^{\ell} \theta_i L_i,$$

i.e., $\theta_0 = 1$ and $\theta_1 = 2$, to insure the unbiasedness of S_L. We have

$$\begin{aligned}\operatorname{Var}\left(S_L - \int_0^1 \sigma_s^2 \mathrm{d}s\right) &= E\left[\operatorname{Var}_{\mathcal{G}}\left(S_L - \int_0^1 \sigma_s^2 \mathrm{d}s\right)\right] \\ &\quad + \operatorname{Var}\left[E_{\mathcal{G}}\left(S_L - \int_0^1 \sigma_s^2 \mathrm{d}s\right)\right] \\ &= E\left[\operatorname{Var}_{\mathcal{G}}(S_L)\right] \qquad (10) \\ &= \sum_{i=0}^{\ell}\sum_{j=0}^{\ell} \theta_i\theta_j E\left[\operatorname{Cov}_{\mathcal{G}}(L_i,\ L_j)\right], \qquad (11)\end{aligned}$$

where (10) is due to Lemma 1 (ii). Since the microstructure noise $\epsilon_i = \eta_i - \eta_{i-1}$ is a zero-mean MA(1) process and is independent of $\{r_j\}$, thus for $k \geq 3$ we have

$$\begin{aligned} &\mathrm{Cov}_{\mathcal{G}}\left(L_h,\ L_{h+k}\right) \\ &= \mathrm{Cov}_{\mathcal{G}}\left(\sum_{j=1}^{n-h+1}(r_j+\epsilon_j)(r_{j+h}+\epsilon_{j+h}),\ \sum_{i=1}^{n-h-k+1}(r_i+\epsilon_i)(r_{i+h+k}+\epsilon_{i+h+k})\right) \\ &= \mathrm{Cov}_{\mathcal{G}}\left(\sum_{j=1}^{n-h+1}\epsilon_j\epsilon_{j+h},\ \sum_{i=1}^{n-h-k+1}\epsilon_i\epsilon_{i+h+k}\right) = 0. \end{aligned} \tag{12}$$

Therefore, given the sigma-field $\mathcal{G}$, the conditional covariance matrix of the vector variable

$$(L_0, 2L_1, \theta_2 L_2, \ldots, \theta_\ell L_\ell)'$$

is an $(\ell+1)\times(\ell+1)$ symmetric pentadiagonal matrix, with only the main diagonal and the first two diagonals above and below it are non-zero. To simplify the notation, we set

$$\mu_h = E\left[\mathrm{Var}_{\mathcal{G}}\left(L_h\right)\right],\ \rho_h = E\left[\mathrm{Cov}_{\mathcal{G}}\left(L_{h-1},\ L_h\right)\right],\ \nu_h = E\left[\mathrm{Cov}_{\mathcal{G}}\left(L_{h-2},\ L_h\right)\right],$$

and $\Sigma = \left(\sigma_{ij}\right)$, where $\sigma_{ij} = \theta_{i-1}\theta_{j-1}E\left[\mathrm{Cov}_{\mathcal{G}}(L_{i-1},\ L_{j-1})\right]$. By (12), we have for $i \leq j$,

$$\sigma_{ij} = \begin{cases} \theta_{i-1}^2\mu_{i-1}, & i = j \geq 1, \\ \theta_{i-1}\theta_i\rho_i, & i \geq 1,\ j = i+1, \\ \theta_{i-1}\theta_{i+1}\nu_{i+1}, & i \geq 1,\ j = i+2, \\ 0, & j \geq i+3, \end{cases}$$

where $\theta_0 = 1$ and $\theta_1 = 2$. Thus, by (11)

$$\mathrm{Var}\left(S_L - \int_0^1 \sigma_s^2 \mathrm{d}s\right) = \mathbf{1}'\Sigma\mathbf{1}, \tag{13}$$

where $\mathbf{1} = (1, \ldots, 1)'_{(\ell+1)\times 1}$. In the following lemma, for simplicity of derivation, we derive the expressions for $\{\mu_i\}_{i=0}^{\ell}$, $\{\rho_i\}_{i=1}^{\ell}$ and $\{\nu_i\}_{i=2}^{\ell}$ when the drift term $u_t = 0$.

Lemma 2 *Let* $\lambda = E\left(\eta_t^4\right)/\left(E\left(\eta_t^2\right)\right)^2$ *denote the kurtosis of* η_t*. Then,*

$$\mu_0 = \frac{2}{n}E\left(\int_0^1 \sigma_s^4 \mathrm{d}s\right) + 4\sigma_\varepsilon^2 E\left(\int_0^1 \sigma_s^2 \mathrm{d}s\right) + (\lambda n - 1)\sigma_\varepsilon^4 + o(n^{-1}),$$
$$\mu_1 = \frac{1}{n}E\left(\int_0^1 \sigma_s^4 \mathrm{d}s\right) + 2\sigma_\varepsilon^2 E\left(\int_0^1 \sigma_s^2 \mathrm{d}s\right) + 2A_1 + \frac{(\lambda+4)n-6}{4}\sigma_\varepsilon^4 + o(n^{-1}),$$
$$\rho_1 = -2\sigma_\varepsilon^2 E\left(\int_0^1 \sigma_s^2 \mathrm{d}s\right) - 2A_1 - \frac{(\lambda+1)n-2}{2}\sigma_\varepsilon^4,$$
$$\mu_h = \frac{1}{n}E\left(\int_0^1 \sigma_s^4 \mathrm{d}s\right) + 2\sigma_\varepsilon^2 E\left(\int_0^1 \sigma_s^2 \mathrm{d}s\right) + A_h + B_h + \frac{3n-3h}{2}\sigma_\varepsilon^4 + o(n^{-1}),$$
$$\rho_h = -\sigma_\varepsilon^2 E\left(\int_0^1 \sigma_s^2 \mathrm{d}s\right) - \frac{1}{2}A_h - \frac{1}{2}B_h - \frac{2n-2h+1}{2}\sigma_\varepsilon^4,\ 2 \leqslant h \leqslant \ell$$
$$\nu_2 = \frac{n-1}{2}\sigma_\varepsilon^4,\ \nu_h = \frac{n-h+1}{4}\sigma_\varepsilon^4,\ 3 \leqslant h \leqslant \ell$$

where $A_h = \sigma_\varepsilon^2 E\left(\int_{1-h/n}^1 \sigma_s^2 \mathrm{d}s\right)$ *and* $B_h = \sigma_\varepsilon^2 E\left(\int_0^{h/n} \sigma_s^2 \mathrm{d}s\right),\ 1 \leqslant h \leqslant \ell.$

Remark 2 When the drift term $u_t \neq 0$, we can show that the expressions of $\{\mu_i\}_{i=0}^\ell$, $\{\rho_i\}_{i=1}^\ell$ and $\{\nu_i\}_{i=2}^\ell$ remain unchanged up to the order n^{-1}, yet additional terms of order n^{-1} will be included.

Theorem 1 *If* $\{\mu_i\}_{i=0}^\ell$, $\{\rho_i\}_{i=1}^\ell$ *and* $\{\nu_i\}_{i=2}^\ell$ *are known, then the proposed estimator of the integrated volatility is* $S_L(\boldsymbol{\theta}^*) = L_0 + 2L_1 + \sum_{h=2}^{\ell} \theta_h^* L_h$*, where* $\boldsymbol{\theta}^* = (\theta_2^*, \ldots, \theta_\ell^*)$ *are the solutions to the normal equations*

$$\frac{\partial \mathrm{Var}\left(S_L(\boldsymbol{\theta}) - \int_0^1 \sigma_s^2 \mathrm{d}s\right)}{\partial \theta_i} = 0, \tag{14}$$

$i = 2, 3, \ldots, \ell$*, see also* (16)*. Then, we have*

$$\left(S_L(\boldsymbol{\theta}^*) - \int_0^1 \sigma_s^2 \mathrm{d}s\right) \Big/ \sqrt{V_{S_L}} \xrightarrow{d} \mathcal{MN}(0, 1),$$

where $\mathcal{MN}$ *is a mixed normal distribution and*

$$V_{S_L} = \mu_0 + 4(\mu_1 + \rho_1) + \theta_2^*(\nu_2 + 2\rho_2) + 2\theta_3^*\nu_3, \tag{15}$$

is the minimum variance of S_L.

Proof Since $\widetilde{r}_i = r_i + \varepsilon_i$,

$$L_h = \sum_{i=1}^{n+1-h} r_i r_{i+h} + \sum_{i=1}^{n+1-h} (r_i \varepsilon_{i+h} + r_{i+h}\varepsilon_i) + \sum_{i=1}^{n+1-h} \varepsilon_i \varepsilon_{i+h}.$$

If we denote

$$R_1 = \left(\sum_{i=1}^{n+1} r_i^2, \sum_{i=1}^{n} r_i r_{i+1}, \ldots, \sum_{i=1}^{n+1-\ell} r_i r_{i+\ell} \right)',$$
$$R_2 = \left(2\sum_{i=1}^{n+1} r_i \varepsilon_i, \sum_{i=1}^{n} (r_i \varepsilon_{i+1} + r_{i+1}\varepsilon_i), \ldots, \sum_{i=1}^{n+1-\ell} (r_i \varepsilon_{i+\ell} + r_{i+\ell}\varepsilon_i) \right)',$$
$$R_3 = \left(\sum_{i=1}^{n+1} \varepsilon_i^2, \sum_{i=1}^{n} \varepsilon_i \varepsilon_{i+1}, \ldots, \sum_{i=1}^{n+1-\ell} \varepsilon_i \varepsilon_{i+\ell} \right)',$$

then $(L_0, \ldots, L_\ell)' = R_1 + R_2 + R_3$. By Lemma 1, we have

$$E(R_1 \big| \mathcal{G}) = \left(\int_0^1 \sigma_s^2 \mathrm{d}s, 0, \ldots, 0 \right)', \quad E(R_3) = \left(n\sigma_\varepsilon^2, -\frac{n}{2}\sigma_\varepsilon^2, 0, \ldots, 0 \right)'.$$

Also, by Lemma 2, we have

$$\mathrm{Cov}_{\mathcal{G}}(R_1) = M_1 \int_0^1 \sigma_s^4 \mathrm{d}s,\ \mathrm{Cov}_{\mathcal{G}}(R_2) = M_2 \int_0^1 \sigma_s^2 \mathrm{d}s \text{ and } \mathrm{Cov}(R_3) = M_3,$$

where $M_1 = \mathrm{diag}\left(\frac{2}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right)$, $M_2 = (m_{ij}^{(2)})\sigma_\varepsilon^2$ is a $(\ell+1) \times (\ell+1)$ symmetric matrix with upper triangular entries $(i \le j)$, $m_{11}^{(2)} = 4$, $m_{12}^{(2)} = -2$ and

$$m_{ij}^{(2)} = \begin{cases} 2, & i = j \ge 2, \\ -1, & i \ge 2,\ j = i+1, \\ 0, & j \ge i+2, \end{cases}$$

and $M_3 = (m_{ij}^{(3)})\sigma_\varepsilon^4$ is a $(\ell+1)\times(\ell+1)$ symmetric matrix with upper triangular entries, $m_{11}^{(3)} = (\lambda n - 1)$, $m_{12}^{(3)} = -(\lambda n + n - 2)/2$, $m_{13}^{(3)} = (n-1)/2$, $m_{22}^{(3)} = (\lambda n + 4n - 6)/4$ and

$$m_{ij}^{(3)} = \begin{cases} (3n - 3i)/2, & i = j \ge 3, \\ -(2n - 2i - 1)/2, & i \ge 2,\ j = i+1, \\ (n - i + 1)/4, & i \ge 2,\ j = i+2, \\ 0, & j \ge i+3. \end{cases}$$

Moreover, it is easy to show that $E_{\mathcal{G}}(R_i R_j') = \mathbf{0}$, $\forall i \ne j$ and $i, j = 1, 2, 3$. Recall that r_is are independent normal variables given the sigma-field $\mathcal{G}$ and $\{\varepsilon_i\}$ is a MA(1) process with finite eighth moment. Hence, conditional on $\mathcal{G}$, we have the following results by CLT as $n \to \infty$:

$$R_1 \xrightarrow{d} E\left(R_1 \big| \mathcal{G}\right) + \left(\int_0^1 \sigma_s^4 \mathrm{d}s\right)^{1/2} M_1^{1/2} Z_1,$$
$$R_2 \xrightarrow{d} \left(\int_0^1 \sigma_s^2 \mathrm{d}s\right)^{1/2} M_2^{1/2} Z_2, \quad R_3 \xrightarrow{d} E(R_3) + M_3^{1/2} Z_3,$$

where Z_1, Z_2 and Z_3 are independent standard normal variables. Since the proposed estimator is a linear combination of $R_1 + R_2 + R_3$, i.e.,

$$S_L(\boldsymbol{\theta}^*) = \left(\boldsymbol{\theta}^*\right)' (L_0, L_1, \cdots, L_\ell)' = \left(\boldsymbol{\theta}^*\right)' (R_1 + R_2 + R_3),$$

we conclude that $S_L(\boldsymbol{\theta}^*)$ follows a mixed normal distribution. The variance $V_{S_L} = \mathrm{Var}\left(S_L(\boldsymbol{\theta}^*) - \int_0^1 \sigma_s^2 \mathrm{d}s\right)$ is derived below.

By (13), the normal equations of (14) reduce to the following equations

$$\begin{cases} \mu_2\theta_2 + \rho_3\theta_3 + \nu_4\theta_4 & = -\nu_2 - 2\rho_2 \\ \rho_3\theta_2 + \mu_3\theta_3 + \rho_4\theta_4 + \nu_5\theta_5 & = -2\nu_3 \\ \nu_{h+2}\theta_h + \rho_{h+2}\theta_{h+1} + \mu_{h+2}\theta_{h+2} + \rho_{h+3}\theta_{h+3} + \nu_{h+4}\theta_{h+4} & = 0 \end{cases} \tag{16}$$

where $2 \leqslant h \leqslant \ell - 2$ and $\nu_{\ell+1} = \nu_{\ell+2} = \rho_{\ell+1} \equiv 0$. Hence, the optimal $\boldsymbol{\theta}^*$ satisfies the equations of (16), which implies the ith row sum of $\Sigma(\boldsymbol{\theta}^*)$,

$$\sum_{j=1}^{\ell+1} \sigma_{ij}(\boldsymbol{\theta}^*) = \theta_{i-1}^*\left(\nu_{i-1}\theta_{i-3}^* + \rho_{i-1}\theta_{i-2}^* + \mu_{i-1}\theta_{i-1}^* + \rho_i\theta_i^* + \nu_{i+1}\theta_{i+1}^*\right) = 0,$$

for $i \geq 3$. Consequently, together with (13), $V_{S_L} = \mathbf{1}'\Sigma(\boldsymbol{\theta}^*)\mathbf{1} = \sum_{i=1}^{2}\sum_{j=1}^{i+2} \sigma_{ij}(\boldsymbol{\theta}^*)$ as claimed in (15). □

It is difficult to obtain the optimal $\boldsymbol{\theta}^*$ directly from Eq. (16) since (μ_h, ρ_h, ν_h)s are unknown. In the next section, we obtain a simplified version of Eq. (16) by ignoring small order terms in Lemma 2 to resolve this problem. The new system of equations depends only on the signal-to-noise ratio S_{nr} and the ratio

$$q = \frac{Q}{\left(E \int_0^1 \sigma_s^2 \mathrm{d}s\right)^2},$$

where $Q = E\left(\int_0^1 \sigma_s^4 \mathrm{d}s\right)$ is called integrated quarticity.

3 A recursive algorithm for solving $S_L(\theta^*)$

If we divide both sides of (16) by $n\sigma_\varepsilon^4$, plug in μ_h, ρ_h and ν_h of Lemma 2 and ignore the $O(h/n^2)$ terms, then we have the following equations which only depend on S_{nr} and q,

$$\begin{cases} \dot\mu_2\theta_2 + \dot\rho_3\theta_3 + \dot\nu_4\theta_4 + \dot b_1 = 0 \\ \dot\rho_3\theta_2 + \dot\mu_3\theta_3 + \dot\rho_4\theta_4 + \dot\nu_5\theta_5 + \dot b_2 = 0 \\ \dot\nu_{h+2}\theta_h + \dot\rho_{h+2}\theta_{h+1} + \dot\mu_{h+2}\theta_{h+2} + \dot\rho_{h+3}\theta_{h+3} + \dot\nu_{h+4}\theta_{h+4} = 0 \end{cases} \tag{17}$$

where $2 \le h \le \ell - 2$ and

$$\dot b_1 = \dot\nu_2 + 2\dot\rho_2, \quad \dot b_2 = 2\dot\nu_3, \quad \dot\mu_h = q\, S_{nr}^2 + 2S_{nr} + \frac{3n-3h}{2n},$$
$$\dot\rho_h = -S_{nr} - \frac{2n-2h+1}{2n}, \quad \dot\nu_2 = \frac{n-1}{2n}, \quad \dot\nu_h = \frac{n-h+1}{4n},$$

for $2 \le h \le \ell$. To obtain the optimal $\boldsymbol\theta$, we replace $E\left(\int_0^1 \sigma_s^2 \mathrm{d}s\right)$ by the $S_L(\boldsymbol\theta)$, and set $S_{nr} = S_L(\boldsymbol\theta)/(n\sigma_\varepsilon^2)$ which is a function of $\boldsymbol\theta$ by (9). Integrating (9) and (17), we have the following system of equations for $\boldsymbol\theta$ and S_{nr},

$$F(\boldsymbol\theta, S_{nr}) \equiv \begin{pmatrix} & \dot{\mathbf{A}}(S_{nr}) & & \mathbf{0} \\ \frac{L_2}{n\sigma_\varepsilon^2} & \cdots & \frac{L_\ell}{n\sigma_\varepsilon^2} & -1 \end{pmatrix} \begin{pmatrix} \boldsymbol\theta \\ S_{nr} \end{pmatrix} + \begin{pmatrix} \mathbf{b}(S_{nr}) \\ \frac{L_0+2L_1}{n\sigma_\varepsilon^2} \end{pmatrix} = 0 \tag{18}$$

where $\dot{\mathbf{A}}(S_{nr})$ is the symmetric matrix with

$$\dot{\mathbf{A}}(S_{nr}) = \begin{pmatrix} \dot\mu_2 & \dot\rho_3 & \dot\nu_4 & 0 & \cdots & 0 \\ & \dot\mu_3 & \dot\rho_4 & \dot\nu_5 & \ddots & \vdots \\ & & \ddots & \ddots & \ddots & 0 \\ & & & \ddots & \ddots & \dot\nu_\ell \\ & \bullet & & & \dot\mu_{\ell-1} & \dot\rho_\ell \\ & & & & & \dot\mu_\ell \end{pmatrix}; \ \boldsymbol\theta = \begin{pmatrix} \theta_2 \\ \theta_3 \\ \vdots \\ \theta_\ell \end{pmatrix}; \ \mathbf{b}(S_{nr}) = \begin{pmatrix} \dot b_1 \\ \dot b_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

It is difficult to solve the system of nonlinear Eq. (18) directly. We employ alternately the Newton–Raphson method and the Gauss–Seidel method (the method of successive displacement) to get a numerical solution for (18). The solution at the $(k+1)$-th iterate, $\left(\boldsymbol\theta^{(k+1)}, S_{nr}^{(k+1)}\right)$ of the Newton–Raphson method satisfies the following iterative equation

$$J_k \begin{pmatrix} \boldsymbol\theta^{(k+1)} \\ S_{nr}^{(k+1)} \end{pmatrix} = J_k \begin{pmatrix} \boldsymbol\theta^{(k)} \\ S_{nr}^{(k)} \end{pmatrix} - F_k, \tag{19}$$

where $J_k \equiv J(\boldsymbol{\theta}^{(k)}, S_{nr}^{(k)})$ is the Jacobian matrix of $F(\boldsymbol{\theta}, S_{nr})$ with respect to $\theta_2, \ldots, \theta_\ell$ and S_{nr} evaluated at $(\boldsymbol{\theta}^{(k)}, S_{nr}^{(k)})$ where

$$J(\boldsymbol{\theta}, S_{nr}) = \begin{pmatrix} & \dot{\mathbf{A}}(S_{nr}) & & \overrightarrow{\mu}(\boldsymbol{\theta}) \\ \frac{L_2}{n\sigma_\varepsilon^2} & \cdots & \frac{L_\ell}{n\sigma_\varepsilon^2} & -1 \end{pmatrix},$$

$$\overrightarrow{\mu}(\boldsymbol{\theta}) = \left(\dot{\mu}_2'\theta_2 + \dot{\rho}_3'\theta_3 + \dot{\nu}_4'\theta_4 + \dot{b}_1',\ \dot{\rho}_3'\theta_2 + \dot{\mu}_3'\theta_3 + \dot{\rho}_4'\theta_4 + \dot{\nu}_5'\theta_5 + \dot{b}_2',\ \cdots,\ \dot{\mu}_\ell'\right)^T$$

and $F_k \equiv F(\boldsymbol{\theta}^{(k)}, S_{nr}^{(k)})$.

- The Gauss–Seidel method:
  We displace $S_{nr}^{(k+1)}$ on the LHS of (19) by $S_{nr}^{(k)}$ and change (19) to the equation LHS=RHS with

$$\text{LHS} = J_k \begin{pmatrix} \boldsymbol{\theta}^{(k+1)} \\ S_{nr}^{(k)} \end{pmatrix} = \begin{pmatrix} & \dot{\mathbf{A}}(S_{nr}^{(k)}) & & \overrightarrow{\mu}(\boldsymbol{\theta}^{(k)}) \\ \frac{L_2}{n\sigma_\varepsilon^2} & \cdots & \frac{L_\ell}{n\sigma_\varepsilon^2} & -1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}^{(k+1)} \\ S_{nr}^{(k)} \end{pmatrix}$$

$$\text{RHS} = J_k \begin{pmatrix} \boldsymbol{\theta}^{(k)} \\ S_{nr}^{(k)} \end{pmatrix} - F_k = \begin{pmatrix} \overrightarrow{\mu}(\boldsymbol{\theta}^{(k)}) S_{nr}^{(k)} - \mathbf{b}(S_{nr}^{(k)}) \\ -\frac{L_0 + 2L_1}{n\sigma_\varepsilon^2} \end{pmatrix}.$$

- The Newton–Raphson method:
  The parameter $\boldsymbol{\theta}^{(k+1)}$ is solved by the first $(\ell - 1)$ equations of the above LHS and RHS.

The remainder question is how to estimate the integrated quarticity Q and the microstructure noise variance σ_ε^2. For the estimation of Q, we used the following estimator of Jacod et al. (2009) (Remark 4 on p. 2256),

$$\hat{Q} = \frac{1}{3c^2\psi_2^2} \sum_{i=0}^{n-k_n+1} (\bar{r}_i^n)^4 - \frac{\psi_1}{nc^4\psi_2^2} \sum_{i=0}^{n-2k_n+1} (\bar{r}_i^n)^2 \sum_{j=i+k_n}^{i+2k_n-1} \tilde{r}_j^2 + \frac{\psi_1^2}{4nc^4\psi_2^2} \sum_{i=1}^{n-2} \tilde{r}_i^2 \tilde{r}_{i+2}^2,$$

see (8) for definitions of the notations. There are various methods proposed to estimate the microstructure noise variance σ_ε^2. For example, the sample mean of squared returns based on the highest frequency data, $\hat{\sigma}_1^2 = L_0/n$ has a biased term $n^{-1}E\int_0^1 \sigma_s^2 ds$; the negative twice lag-1 sample autocovariance divided by n, $\hat{\sigma}_2^2 = -2L_1/n$ which is unbiased but may take negative values. Barndorff-Nielsen et al. (2008) also propose a bias-corrected nonnegative estimator $\hat{\sigma}_3^2 = \exp\{\log(L_0/n) - T_k/L_0\}$. In this study, we consider the estimator $\hat{\sigma}_4^2 = L_0/(n + nS_{nr})$, which by Lemma 1 (i) is an unbiased nonnegative estimator of σ_ε^2. One advantage of $\hat{\sigma}_4^2$ is that it can be recursively updated with the S_{nr}, details are given in the following algorithm. We will compare the performance of the four estimators $\hat{\sigma}_i^2$, $i = 1, \ldots, 4$, in the simulation study of Sect. 4.1.

We summarize the proposed estimation procedure in the following. Set the initial estimators of $S_{nr}^{(0)}$ and $\sigma_{\varepsilon,0}^2$ as

$$S_{nr}^{(0)} = \frac{T_2^{(K)}}{n\sigma_{\varepsilon,0}^2}, \ \frac{T_m}{n\sigma_{\varepsilon,0}^2}, \ \frac{T_k}{n\sigma_{\varepsilon,0}^2} \ \text{or} \ \frac{T_p}{n\sigma_{\varepsilon,0}^2} \ \text{ with } \hat{\sigma}_{\varepsilon,0}^2 = \frac{L_0}{n},$$

where $T_2^{(K)}$, T_m, T_k and T_p are the two-scale, multi-scale, realized kernel and pre-averaging estimators defined in Sect. 1. The algorithm consists of the following three steps starting from $i = 1$.

(step 1) Solve $\boldsymbol{\theta}^{(i)}$ from the following equation

$$\dot{\mathbf{A}}(S_{nr}^{(i-1)})\boldsymbol{\theta}^{(i)} = -\mathbf{b}(S_{nr}^{(i-1)}).$$

(step 2) Obtain the estimator $S_L(\boldsymbol{\theta}^{(i)}) = \sum_{j=0}^{\ell} \theta_j^{(i)} L_j$.

(step 3) Update $S_{nr}^{(i)} = \dfrac{S_L(\boldsymbol{\theta}^{(i)})}{n\sigma_{\varepsilon,i-1}^2}$, $\hat{\sigma}_{\varepsilon,i}^2 = L_0/(n + nS_{nr}^{(i)})$ and $q = \dfrac{\hat{Q}}{S_L^2\left(\boldsymbol{\theta}^{(i)}\right)}$. If $\dfrac{|S_L(\boldsymbol{\theta}^{(i)}) - S_L(\boldsymbol{\theta}^{(i-1)})|}{S_L(\boldsymbol{\theta}^{(i-1)})} < 10^{-8}$, then stop and $\boldsymbol{\theta}^* = \boldsymbol{\theta}^{(i)}$ and $S_L(\boldsymbol{\theta}^*) = S_L(\boldsymbol{\theta}^{(i)})$; otherwise $i = i + 1$ and go to (step 1).

For the special constant volatility case, i.e., $\sigma_t \equiv \sigma$, Aït-Sahalia et al. (2005) proved that the maximum likelihood estimator (MLE) of σ is $n^{1/4}$-consistent and has an asymptotically normal distribution with variance

$$\widehat{\sigma}_{\text{MLE}}^2 = 4\sqrt{\frac{2\sigma^6\sigma_\varepsilon^2}{n} + \frac{\sigma^8}{n^2}} + \frac{2\sigma^4}{n}.$$

In the following proposition, we derive an asymptotic expansion for $\text{Var}(S_L(\boldsymbol{\theta}^*))$. The result shows that the leading order term of $\text{Var}(S_L(\boldsymbol{\theta}^*))$ is the same as $\widehat{\sigma}_{\text{MLE}}^2$ which indicates that $S_L(\boldsymbol{\theta}^*)$ is asymptotically efficient as the MLE for the constant volatility model.

Proposition 1 *If $\sigma_t \equiv \sigma$ $\forall t$ and $u_t = 0$, then we have*

$$\text{Var}(S_L(\boldsymbol{\theta}^*)) = 4\sqrt{\frac{2\sigma^6\sigma_\varepsilon^2}{n} + \frac{\sigma^8}{n^2}} + \frac{2\sigma^4 + 6\sigma^2\sigma_\varepsilon^2}{n} + O\left(n^{-3/2}\right).$$

Proof For the constant volatility model, the signal-to-noise ratio $S_{nr} = \sigma^2/(n\sigma_\varepsilon^2)$. By ignoring the $O(h/n)$ term, we can further simplify (17) to the following linear homogeneous difference equation

$$\theta_h - (4S_{nr} + 4)\,\theta_{h+1} + \left(4S_{nr}^2 + 8S_{nr} + 6\right)\theta_{h+2} - (4S_{nr} + 4)\,\theta_{h+3} + \theta_{h+4} = 0, \tag{20}$$

for $h = 2, \ldots, \ell$. The closed-form solution of (20) is

$$\theta_h^a = \left(1 + S_{nr} - \sqrt{2S_{nr} + S_{nr}^2}\right)^h (c_1 + hc_2), \ \forall\, 2 \leq h \leq \ell,$$

see Kelley and Peterson (2000). The constants c_1 and c_2 are obtained by plugging θ_h^a into the first two equations of (17), which gives

$$c_1 = 2 + O\left(n^{-2}\right), \quad c_2 = 2\sqrt{2S_{nr} + S_{nr}^2} + 2S_{nr} + O\left(n^{-2}\right).$$

In particular,

$$\begin{aligned}
\theta_2^a &= 2 - 4S_{nr} + 4S_{nr}\sqrt{S_{nr}(2 + S_{nr})} - 4S_{nr}^2 + O\left(n^{-2}\right)\\
\theta_3^a &= 2 - 12S_{nr} + 20S_{nr}\sqrt{S_{nr}(2 + S_{nr})} - 36S_{nr}^2 + 16S_{nr}^2\sqrt{S_{nr}(2 + S_{nr})} + O\left(n^{-2}\right).
\end{aligned}$$

When the volatility σ_t is a constant, the moments given in Lemma 2 have the following expression,

$$\begin{aligned}
\mu_0 &= \frac{2}{n}\sigma^4 + 4\sigma_\varepsilon^2\sigma^2 + (\lambda n - 1)\sigma_\varepsilon^4,\\
\mu_1 &= \frac{n-1}{n^2}\sigma^4 + \left(2 - \frac{2}{n}\right)\sigma_\varepsilon^2\sigma^2 + \frac{\lambda n + 4n - 6}{4}\sigma_\varepsilon^4,\\
\rho_1 &= -\left(2 - \frac{1}{n}\right)\sigma_\varepsilon^2\sigma^2 - \frac{\lambda n + n - 2}{2}\sigma_\varepsilon^4,\\
\rho_2 &= -\left(1 - \frac{2}{n}\right)\sigma_\varepsilon^2\sigma^2 - \frac{2n - 3}{2}\sigma_\varepsilon^4,
\end{aligned}$$

ν_2 and ν_3 are the same as given in Lemma 2. Finally, by (15), we have

$$\begin{aligned}
\mathrm{Var}(S_L(\boldsymbol{\theta}^*)) &= \mu_0 + 4(\mu_1 + \rho_1) + \theta_2^a(\nu_2 + 2\rho_2) + 2\theta_3^a\nu_3 + O\left(n^{-3/2}\right)\\
&= 4\sqrt{\frac{2\sigma^6\sigma_\varepsilon^2}{n} + \frac{\sigma^8}{n^2}} + \frac{2\sigma^4 + 6\sigma^2\sigma_\varepsilon^2}{n} + O\left(n^{-3/2}\right),
\end{aligned}$$

which completes the proof. □

Remark 3 When $u_t \neq 0$, the leading term of $\mathrm{Var}(S_L(\boldsymbol{\theta}^*))$ in Proposition 1, $4\sqrt{\frac{2\sigma^6\sigma_\varepsilon^2}{n} + \frac{\sigma^8}{n^2}}$, remains unchanged.

Although the weights θ_h^as are derived for the constant volatility model, they also provide good approximation to the optimal weights θ_h^* of SVM obtained from **(step 1)**–**(step 3)**. We perform a simulation study to compare θ_h^a with θ_h^*. Consider the Heston SVM defined by (21) with parameters $\kappa = 10$, $\omega = \sqrt{\kappa V}$ and $(V, \sigma_\varepsilon^2)$ satisfying the six setting given in Table 1. The maximum of $\max_{h\le 35}|\theta_h^* - \theta_h^a|/\theta_h^*$ for the six Heston models is only 0.12 % when the sample size $n = 500$ and the relative maximum errors decrease as the sample size increases. In Fig. 1, we plot $\ln\theta_h^*$ and $\ln\theta_h^a$ versus h for the Heston model defined by the first column of Table 1. The result indicates, as h increases, both θ_h^* and θ_h^a decay exponentially fast to zero, and their differences are negligible when h is small.

The proposed estimator $S_L(\boldsymbol{\theta}^*)$ and Theorem 1 are also applicable to the case when the volatility σ_t is a deterministic function of the time $t \in [0, 1]$. However, in such case, the quadratic type estimators based on global tuning parameters such as $S_L(\boldsymbol{\theta}^*)$, $T_2^{(K)}$, T_m, T_k and T_p have larger asymptotic variances than the estimator of Reiss (2011) (p.17 and Theorem 8.1) based on local tuning parameters.

4 Simulation and empirical studies

In Sect. 4.1, we perform a simulation to compare the root mean squared errors of the five estimators: the proposed estimator $S_L(\boldsymbol{\theta}^*)$, the two-scale estimator $T_2^{(K)}$, the multi-scale estimator T_m, the realized kernel estimator T_k and the pre-averaging estimator T_p. In Sect. 4.2, the proposed method is applied to estimate the intradaily integrated volatility of fifteen stocks listed on the NYSE.

4.1 Comparison of the five estimators

The Heston model Heston (1993) is a popular SVM for high frequency transaction data, which assumes that the efficient log price process $\{p_t\}$ satisfies

$$\begin{cases} \mathrm{d}p_t &= \sigma_t \mathrm{d}W_t \\ \mathrm{d}\sigma_t^2 &= \kappa(V - \sigma_t^2)\mathrm{d}t + \omega\sqrt{\sigma_t^2}\mathrm{d}B_t \end{cases}, \mathrm{Corr}(\mathrm{d}W_t, \mathrm{d}B_t) = \varphi \tag{21}$$

where the instantaneous volatility σ_t is modeled as a mean-reverting square-root diffusion process, also named the CIR process Cox et al. (1985). The parameter V corresponds to the expected long-term volatility, κ determines the convergence speed of the adjustment, ω is the volatility of σ_t, φ is the leverage parameter and $\{W_t,\ B_t : t \ge 0\}$ are scalar Brownian motions. In the simulation study, we consider a generalization of (21) in which the volatility process satisfies the following constant elasticity of variance (CEV) model Cox (1975); Chen et al. (2008),

$$\mathrm{d}\sigma_t^2 = \kappa(V - \sigma_t^2)\mathrm{d}t + \omega\sigma_t^{2\alpha}\mathrm{d}B_t, \quad 0 \le \alpha < 1. \tag{22}$$

Model (22) is called the Vasicek model when $\alpha = 0$ and reduces to the CIR process when $\alpha = 0.5$. The parameter settings are $\kappa = 1,\ \omega = \sqrt{\kappa V}/4$ when $\alpha = 0, 0.2, 0.4$ and $\kappa = 10,\ \omega = \sqrt{\kappa V}$ when $\alpha = 0.5, 0.6, 0.8$ and the values of $(V, \sigma_\varepsilon^2)$ are given in Table 1. Since the long run mean of the volatility $E(\sigma_t^2) = V$, $nS_{nr} = V/\sigma_\varepsilon^2$ which are also given in Table 1. The cases for $\varphi = 0$ (without leverage effect) and $\varphi = -0.5$ (the leverage effect model) are both considered. The sample sizes

$n = 500, 2000, 5000, 8000, 12000, 24000$ are considered. All the results are based on 1000 replications.

The tuning parameters of each estimators are set as follows. For the two-scaled estimator $T_2^{(K)}$, we set $K = c^* n^{2/3}$ with bias adjustment [see Eq. (64) in Zhang et al. 2005]. For the multi-scale estimator T_m, we set $K_i = i$, for $i = 1, \ldots, m$ with m $(5 \leq m \leq 10)$ chosen to minimize the MSE. For the realized kernel estimator T_k, we choose the optimal kernel function $k(x) = (1 + |x|)e^{-|x|}$ (see Proposition 1 of page 1498 in Barndorff-Nielsen et al. 2008) and $H = \xi n^{1/2}$ where ξ^2 are chosen among 0.1, 0.01 and $\hat{\sigma}_3^2\ \hat{Q}^{-1/2}$ which minimizes the MSE. For the pre-averaging estimator T_p, we set $c = 1/3$. For the proposed estimator S_L, we choose $\ell = 15$ for $n = 500$, $\ell = 20$ for $n = 2000, 5000$ and $\ell = 30$ for $n = 8000, 12000, 24000$.

For the estimation of the microstructure noise variance, the relative errors (the root MSE divided by the true value) of the four estimators $\hat{\sigma}_i^2$, $i = 1, \ldots, 4$ are reported in Table 2 for $nS_{nr}(\times 10^3) = 0.4, 1.2, 1.6, 3.2, 8, 12$ and $n = 500, 5000, 24000$. The results show that the proposed estimator $\hat{\sigma}_4^2$ has the smallest relative root mean squared errors for all cases.

For the estimation of the integrated volatility, the RMSE of an estimator $\hat{T}$ is defined as

$$\text{RMSE}(\hat{T}) = \sqrt{\frac{1}{r}\sum_{j=1}^{r}\left(\hat{T} - \sum_{t=1}^{n}\sigma_{t,j}^2\right)^2},$$

where $\{\sigma_{t,j}^2\}_{t=1}^n$ denotes the volatility path of the j-th replication, $j = 1, \ldots, r$. Herein, we use the sum of the volatilities $\sum_{t=1}^n \sigma_{t,j}^2$ to approximate the integrated volatility of the jth simulated path. And the relative error (RE) of $\hat{T}$ is defined as $\text{RE}(\hat{T}) = \text{RMSE}(\hat{T}) / \left(r^{-1}\sum_{j=1}^r \sum_{t=1}^n \sigma_{t,j}^2\right)$.

In Fig. 2, we plot the RE of the five estimators versus the six nS_{nr} for the Heston model with $\kappa = 10$, $\omega = \sqrt{\kappa V}$, $\varphi = 0$ and $(V, \sigma_\varepsilon^2)$ given in Table 1. For each nS_{nr}, the REs are shown sequentially from left to right for the sample size $n =$ 500, 2000, 5000, 8000, 12000, 24000. Similarly, in Fig. 3, we plot the RE for the Heston model with leverage effect ($\varphi = -0.5$). Both plots show that the proposed estimator $S_L(\boldsymbol{\theta}^*)$ attains the smallest RE for all cases. The RE of the other CEV models with $\alpha = 0, 0.2, 0.4, 0.6$ and 0.8 is similar to the results of the Heston model ($\alpha = 0.5$).

We also experiment on the case of deterministic volatility, which assumes

$$\sigma_t = m\left(0.000035 + 0.01(t - 0.5)^4\right). \tag{23}$$

The settings of m and σ_ε^2 are given in Table 3, which are chosen to produce nS_{nr} the same as in Table 1. Figure 4 plots the RE of the five estimators versus the six nS_{nr}. The results show that the proposed estimator $S_L(\boldsymbol{\theta}^*)$ attains the smallest RE for all cases.

We summarize the relative efficiencies of the four estimators with respect to S_L by the boxplots of the following RMSE ratios

$$\frac{\text{RMSE}(S_L)}{\text{RMSE}(T_2^{(K)})}, \quad \frac{\text{RMSE}(S_L)}{\text{RMSE}(T_m)}, \quad \frac{\text{RMSE}(S_L)}{\text{RMSE}(T_k)}, \quad \text{and} \quad \frac{\text{RMSE}(S_L)}{\text{RMSE}(T_p)}$$

in Fig. 5. Each boxplot consists of 468 relative efficiencies corresponding to (i) six α, two φ, six S_{nr} and six n for Model (22) and (ii) six S_{nr} and six n for Model (23). Since the ranges of the boxplots are all less than or equal to one, the proposed estimator S_L attains a better efficiency than the other four estimators.

4.2 Empirical study

For the empirical application, we consider the ultra-high-frequency tick-by-tick data of fifteen stocks listed on the NYSE (New York Stock Exchange): ABT, AMD, BAC, C, GE, JNJ, JPM, KO, MCD, MER, MRK, NOK, PEP, T, XOM. The normal trading hours of the NYSE is 6.5 hours (23400 s) from 9:30 to 16:00.

Since the intradaily trading times are non-regular, we employ the previous-tick interpolation scheme, see Dacorogna et al. (2001), to obtain equispaced data. Let $\{t_j, j = 1, 2, \ldots, n\}$ denote the observed transaction times, where n stands for the total number of transactions in a trading day. Define $\tau(0) = 0$ and $\tau(is) = \max\{t_j : t_j \le is,\ j = 1, 2, \ldots, n\}$, $i = 1, \ldots, [23400/s]$, the closest transaction time before and including time t, and set the log return at time t to be $\widetilde{r}_i = \widetilde{p}_{\tau(is)} - \widetilde{p}_{\tau((i-1)s)}$, where $s = 10,\ 5,\ 1$ sec.

The integrated volatilities are estimated based on non-zero log return data. We use RV (realized volatility), $T_2^{(K)},\ T_m,\ T_k,\ T_p$ and S_L, to estimate the daily integrated volatilities for the fifteen stocks based on the intraday high-frequency transaction data from 2002/01/02 $\sim$ 2002/01/31. The tuning parameters of T_m and T_k are set to be $M = 10$ and $\xi^2 = \hat{\sigma}_3^2/\sqrt{\hat{Q}}$, respectively. The monthly average of the six estimators for each sampling frequencies ($s = 10, 5, 1$) is given in Table 4. As expected, the realized volatility (RV) increases as the sampling frequency increases. Nevertheless, the other five estimators of the integrated volatility remain steady when the sampling frequency changes.

For each stock, per day we obtain the estimates $T_2^{(K)},\ T_m,\ T_k,\ T_p$ and S_L for each sampling frequency. And $(T_2^{(K)} - S_L)/S_L$, $(T_m - S_L)/S_L$, $(T_k - S_L)/S_L$ and $(T_p - S_L)/S_L$ denote the daily relative differences of $T_2^{(K)},\ T_m,\ T_k,\ T_p$ with respect to S_L. Figure 6 shows the boxplots of the four daily relative differences. Each boxplot contains approximately 900–945 daily relative differences of an estimator obtained from fifteen stocks and three sampling frequencies ($s = 10,\ 5,\ 1$ s) during the period 2002/01/02 to 2002/01/31. As shown, the two-scaled estimator $T_2^{(K)}$ and pre-averaging estimator T_p are close to S_L, while the multi-scale estimator T_m and the kernel estimator T_k tend to be larger than S_L. We also found that T_m and T_k can be closer to S_L if we choose different tuning parameters (other than the preset values

$M = 10$ and $\xi^2 = \hat{\sigma}_3^2/\sqrt{\hat{Q}}$) every day for each company. Nevertheless, the advantages and disadvantages of adjusting the tuning parameters are not clear which need further investigation.

5 Conclusion

Stochastic volatility models (SVM) with microstructure noise are prevailing for ultra-high-frequency data modeling. In this study, we proposed an optimal restricted quadratic estimators of integrated volatility for SVM with microstructure noise. The proposed estimator has an asymptotic mixed normal distribution and has the same efficiency as the MLE for the constant volatility model. A practical recursive algorithm is proposed to obtain the estimate. Both theoretical and simulation results strongly support the efficiency advantage of the proposed method compared with state-of-the-art methods including the two-scale estimator, the multi-scale estimator, the realized kernel estimator and the pre-averaging estimator. In future study, it is worthwhile to investigate the effects of using different integrated volatility estimates on statistical inference such as hypothesis testing, parameter estimation and portfolio selection.

Appendix A: Proofs

Proof of Lemma 1

First, we derive some relevant expectations needed in the proof.

$$\sum_{j=1}^{n+1} E_{\mathcal{G}}\left(r_j^2\right) = \sum_{j=1}^{n+1} E_{\mathcal{G}}\left(\int_{(j-1)/n}^{j/n} \sigma_s \mathrm{d}W_s\right)^2 = \sum_{j=1}^{n} E_{\mathcal{G}}\left(\int_{(j-1)/n}^{j/n} \sigma_s^2 \mathrm{d}s\right)$$
$$= \int_0^1 \sigma_s^2 \mathrm{d}s, \tag{24}$$

where the second equation is by the Itô isometry (Theorem 4.3.1 of Shreve 2004). Recall $\varepsilon_j = \eta_j - \eta_{j-1}$, we have

$$\sum_{j=1}^{n+1} E(\varepsilon_j^2) = E(\varepsilon_1^2) + \sum_{j=2}^{n} E(\varepsilon_j^2) + E(\varepsilon_{n+1}^2) = \frac{1}{2}\sigma_\varepsilon^2 + (n-1)\sigma_\varepsilon^2 + \frac{1}{2}\sigma_\varepsilon^2 = n\sigma_\varepsilon^2,$$
$$E\left(\varepsilon_j \varepsilon_{j+1}\right) = E(\eta_j - \eta_{j-1})(\eta_{j+1} - \eta_j) = -E\left(\eta_j^2\right) = -\frac{1}{2}\sigma_\varepsilon^2,\ j \geq 1,$$
$$E\left(\varepsilon_j \varepsilon_{j+h}\right) = E(\eta_j - \eta_{j-1})(\eta_{j+h} - \eta_{j+h-1}) = 0,\ h \geq 2,\ j \geq 1.$$

Thus, the expectations of L_hs are

$$E_{\mathcal{G}}(L_0) = E_{\mathcal{G}}\left[\sum_{j=1}^{n+1}(r_j+\varepsilon_j)^2\right] = \sum_{j=1}^{n+1}E_{\mathcal{G}}(r_j^2) + \sum_{j=1}^{n+1}E(\varepsilon_j^2) = \int_0^1 \sigma_s^2 \mathrm{d}s + n\sigma_\varepsilon^2,$$

$$E_{\mathcal{G}}(L_1) = \sum_{j=1}^{n}E_{\mathcal{G}}(r_j+\varepsilon_j)(r_{j+1}+\varepsilon_{j+1}) = \sum_{j=1}^{n}E\left(\varepsilon_j\varepsilon_{j+1}\right) = -\frac{n}{2}\sigma_\varepsilon^2$$

$$E_{\mathcal{G}}(L_h) = \sum_{j=1}^{n-h+1}E_{\mathcal{G}}(r_j+\varepsilon_j)(r_{j+h}+\varepsilon_{j+h}) = 0,\ h \geq 2.$$

Consequently, by setting $\theta_0 = 1$ and $\theta_1 = 2$, we have

$$E_{\mathcal{G}}(S_L) = \theta_0\left(\int_0^1 \sigma_s^2 \mathrm{d}s + n\sigma_\varepsilon^2\right) + \theta_1\left(-\frac{n}{2}\sigma_\varepsilon^2\right) = \int_0^1 \sigma_s^2 \mathrm{d}s,$$

which implies the unbiasness of S_L. □

Proof of Lemma 2

First, for the variance of microstructure noise, we have $\mathrm{Var}\left(\eta_t^2\right) = (\lambda-1)\sigma_\eta^4$ since $E(\eta_t^4) = \lambda\sigma_\eta^4$ and then

$$\mathrm{Var}\left(\varepsilon_j^2\right) = \begin{cases} \frac{\lambda+1}{2}\sigma_\varepsilon^4 & j = 2,\ldots,n, \\ \frac{\lambda-1}{4}\sigma_\varepsilon^4 & j = 1 \text{ or } n+1. \end{cases}$$

Next, let $t_i = i/n$, $i = 1, 2, \ldots, n$ be an equispaced partition of the unit interval [0, 1], and let $b_i = n^{1/2}\sigma_{t_{i-1}}(W_{t_i} - W_{t_{i-1}})$ and $\Delta_{ni} = n^{1/2}\int_{t_{i-1}}^{t_i}(\sigma_s - \sigma_{t_{i-1}})\mathrm{d}W_s$, then $n^{1/2}r_i = n^{1/2}(p_{t_i} - p_{t_{i-1}}) = n^{1/2}\int_{t_{i-1}}^{t_i}\sigma_s \mathrm{d}W_s = b_i + \Delta_{ni}$. Due to (4), we have

$$E|\Delta_{ni}|^{2q} = O(n^{-q}) \ \text{ for } q = 1, 2. \tag{25}$$

For instance, see Lee (2010). Recall from (25), the fourth moment of the nominal return r_j is

$$\begin{aligned} E_{\mathcal{G}}\left(\sum_{j=1}^{n+1} r_j^4\right) &= n^{-2}\sum_{j=1}^{n+1}E_{\mathcal{G}}\left(n^{1/2}r_j\right)^4 = n^{-2}\sum_{j=1}^{n+1}E_{\mathcal{G}}\left(b_j^4 + 6b_j^2\Delta_{nj}^2 + \Delta_{nj}^4\right) \\ &= n^{-2}\sum_{j=1}^{n+1}3\sigma_{t_{j-1}}^4 + 6n^{-2}\sum_{j=1}^{n+1}\sigma_{t_{j-1}}^2E(\Delta_{nj}^2) + n^{-2}\sum_{j=1}^{n+1}E(\Delta_{nj}^4) \\ &= \frac{3}{n}\int_0^1 \sigma_s^4 \mathrm{d}s + O(n^{-2}) \end{aligned} \tag{26}$$

where the last equality is by Riemann sum approximation to integral and (25). By (24) and (26), we have

$$\begin{aligned}
\mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n+1} r_j^2\right) &= E_{\mathcal{G}}\left(\sum_{j=1}^{n+1} r_j^2\right)^2 - \left(\sum_{j=1}^{n+1} E_{\mathcal{G}} r_j^2\right)^2 \\
&= \sum_{j=1}^{n+1} E_{\mathcal{G}}\left(r_j^4\right) + \sum_{i\neq j}^{n+1} E_{\mathcal{G}}\left(r_i^2\right) E_{\mathcal{G}}\left(r_j^2\right) - \left(\int_0^1 \sigma_s^2 \mathrm{d}s\right)^2 \\
&= \frac{3}{n}\int_0^1 \sigma_s^4 \mathrm{d}s + \left(\sum_{j=1}^{n+1} E_{\mathcal{G}} r_j^2\right)^2 - \sum_{j=1}^{n+1}\left(E r_j^2\right)^2 \\
&\quad - \left(\int_0^1 \sigma_s^2 \mathrm{d}s\right)^2 + O(n^{-2}) \\
&= \frac{3}{n}\int_0^1 \sigma_s^4 \mathrm{d}s - \frac{1}{n^2}\sum_{j=1}^{n+1} \sigma_{t_{j-1}}^4 + O(n^{-2}) \\
&= \frac{2}{n}\int_0^1 \sigma_s^4 \mathrm{d}s + o(n^{-1}). \qquad (27)
\end{aligned}$$

Also, for $h \geq 1$,

$$\begin{aligned}
\mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n+1} r_j r_{j+h}\right) &= E_{\mathcal{G}}\left(\sum_{j=1}^{n+1} r_j r_{j+h}\right)^2 = \sum_{j=1}^{n+1} E_{\mathcal{G}}\left(r_j^2 r_{j+h}^2\right) \\
&= \frac{1}{n^2}\sum_{j=1}^{n+1} \sigma_{t_{j-1}}^4 + O(hn^{-3/2}) = \frac{1}{n}\int_0^1 \sigma_s^4 \mathrm{d}s + o(n^{-1}), \qquad (28)
\end{aligned}$$

where $E_{\mathcal{G}}(r_{j+h}^2) = n^{-1}\sigma_{t_{j+h-1}}^2 + O(n^{-2}) = n^{-1}\sigma_{t_{j-1}}^2 + O(hn^{-3/2})$ using (4).

Next, we consider the variance of $L_h,\ h \geq 0$. By (27),

$$\begin{aligned}
\mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n+1} \widetilde{r}_j^2\right) &= \mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n+1} (r_j^2 + 2r_j\varepsilon_j + \varepsilon_j^2)\right) \\
&= \mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n+1} r_j^2\right) + 4\mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n+1} r_j\varepsilon_j\right) + \mathrm{Var}\left(\sum_{j=1}^{n+1} \varepsilon_j^2\right) \\
&= \frac{2}{n}\int_0^1 \sigma_s^4 \mathrm{d}s + o(n^{-1}) + 4\sigma_\varepsilon^2 \int_0^1 \sigma_s^2 \mathrm{d}s + (\lambda n - 1)\sigma_\varepsilon^4,
\end{aligned}$$

where

$$\begin{aligned}\mathrm{Var}\left(\sum_{j=1}^{n+1}\varepsilon_j^2\right) &= \sum_{j=1}^{n+1}\mathrm{Var}\left(\varepsilon_j^2\right) + 2\sum_{j=1}^{n} Cov(\varepsilon_j^2, \varepsilon_{j+1}^2)\\ &= 2\frac{\lambda-1}{4}\sigma_\varepsilon^4 + (n-1)\frac{\lambda+1}{2}\sigma_\varepsilon^4 + 2n\frac{\lambda-1}{4}\sigma_\varepsilon^4 = (\lambda n - 1)\sigma_\varepsilon^4\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}(\varepsilon_j^2, \varepsilon_{j+1}^2) &= \mathrm{Cov}(\eta_j^2 - 2\eta_j\eta_{j-1} + \eta_{j-1}^2, \eta_{j+1}^2 - 2\eta_{j+1}\eta_j + \eta_j^2) = \mathrm{Var}(\eta_j^2)\\ &= \frac{\lambda-1}{4}\sigma_\varepsilon^4.\end{aligned}$$

Thus, μ_0 is as claimed. Next, by (28),

$$\begin{aligned}\mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n}\widetilde{r}_j\widetilde{r}_{j+1}\right) &= \mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n}(r_jr_{j+1} + r_j\varepsilon_{j+1} + r_{j+1}\varepsilon_j + \varepsilon_j\varepsilon_{j+1})\right)\\ &= \mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n}r_jr_{j+1}\right) + 2\mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n}r_j\varepsilon_{j+1}\right)\\ &\quad +\mathrm{Var}\left(\sum_{j=1}^{n}\varepsilon_j\varepsilon_{j+1}\right)\\ &= \frac{1}{n}\int_0^1\sigma_s^4\mathrm{d}s + o(n^{-1}) + 2\sigma_\varepsilon^2\int_0^1\sigma_s^2\mathrm{d}s + A_1 + \frac{(\lambda+4)n-6}{4}\sigma_\varepsilon^4,\end{aligned}$$

where $A_1 = 2\sigma_\varepsilon^2\int_{1-1/n}^1\sigma_s^2\mathrm{d}s$,

$$\begin{aligned}\mathrm{Var}\left(\sum_{j=1}^{n}\varepsilon_j\varepsilon_{j+1}\right) &= \sum_{j=1}^{n}\mathrm{Var}\left(\varepsilon_j\varepsilon_{j+1}\right) + 2\sum_{j=1}^{n-1}\mathrm{Cov}(\varepsilon_j\varepsilon_{j+1}, \varepsilon_{j+1}\varepsilon_{j+2})\\ &= 2\frac{\lambda}{4}\sigma_\varepsilon^4 + \frac{(n-2)(\lambda+2)}{4}\sigma_\varepsilon^4 + 2n\frac{\sigma_\varepsilon^4}{4} = \frac{(\lambda+4)n-6}{4}\sigma_\varepsilon^4,\end{aligned}$$

$$\mathrm{Var}(\varepsilon_j\varepsilon_{j+1}) = \begin{cases}3\sigma_\eta^4 + \mathrm{Var}(\eta^2) = \frac{\lambda+2}{4}\sigma_\varepsilon^4 & j = 2, \ldots, n-1,\\ \sigma_\eta^4 + \mathrm{Var}(\eta^2) = \frac{\lambda}{4}\sigma_\varepsilon^4 & j = 1 \text{ or } n\end{cases} \tag{29}$$

and

$$\mathrm{Cov}\left(\varepsilon_j\varepsilon_{j+1}, \varepsilon_{j+1}\varepsilon_{j+2}\right) = E\left(\varepsilon_j\varepsilon_{j+1}^2\varepsilon_{j+2}\right) - \left[E\left(\varepsilon_j\varepsilon_{j+1}\right)\right]^2 = \frac{1}{4}\sigma_\varepsilon^4. \tag{30}$$

For $h \geq 2$,

$$\mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n-h+1} \widetilde{r}_j \widetilde{r}_{j+h}\right) = \mathrm{Var}_{\mathcal{G}}\left(\sum_{j=1}^{n-h+1} (r_j r_{j+h} + r_j \varepsilon_{j+h} + r_{j+h}\varepsilon_j + \varepsilon_j \varepsilon_{j+h})\right)$$
$$= \frac{1}{n}\int_0^1 \sigma_s^4 \mathrm{d}s + o(n^{-1}) + 2\sigma_\varepsilon^2 \int_0^1 \sigma_s^2 \mathrm{d}s + A_h + B_h + \frac{3n-3h}{2}\sigma_\varepsilon^4,$$

where $A_h = 2\sigma_\varepsilon^2 \int_{1-h/n}^1 \sigma_s^2 \mathrm{d}s$, $B_h = 2\sigma_\varepsilon^2 \int_0^{h/n} \sigma_s^2 \mathrm{d}s$ and by similar derivation as in (29) and (30), we have

$$\mathrm{Var}\left(\sum_{j=1}^{n-h+1} \varepsilon_j \varepsilon_{j+h}\right) = \sum_{j=1}^{n-h+1} \mathrm{Var}\left(\varepsilon_j \varepsilon_{j+h}\right) + 2\sum_{j=1}^{n-h} \mathrm{Cov}(\varepsilon_j \varepsilon_{j+h}, \varepsilon_{j+1}\varepsilon_{j+h+1})$$
$$= 2\frac{\lambda}{2}\sigma_\varepsilon^4 + (n-h-1)\sigma_\varepsilon^4 + 2(n-h)\frac{\sigma_\varepsilon^4}{4} = \frac{3n-3h}{2}\sigma_\varepsilon^4.$$

Then, μ_1 and μ_h, $h \geq 2$ are as claimed.

In the following, we consider the function $\rho_h = E\left[\mathrm{Cov}_{\mathcal{G}}(L_{h-1}, L_h)\right]$. Note that

$$\mathrm{Cov}_{\mathcal{G}}\left(\sum_{j=1}^{n+1} \widetilde{r}_j^2, \sum_{k=1}^{n} \widetilde{r}_k \widetilde{r}_{k+1}\right) = \sum_{j=1}^{n} \mathrm{Cov}_{\mathcal{G}}(\widetilde{r}_j^2, \widetilde{r}_j \widetilde{r}_{j+1}) + \sum_{j=1}^{n-1} \mathrm{Cov}_{\mathcal{G}}(\widetilde{r}_j^2, \widetilde{r}_{j+1}\widetilde{r}_{j+2})$$
$$+ \sum_{j=1}^{n} \mathrm{Cov}_{\mathcal{G}}(\widetilde{r}_{j+1}^2, \widetilde{r}_j \widetilde{r}_{j+1}) + \sum_{j=1}^{n-1} \mathrm{Cov}_{\mathcal{G}}(\widetilde{r}_{j+2}^2, \widetilde{r}_j \widetilde{r}_{j+1}),$$

where $\mathrm{Cov}_{\mathcal{G}}(\widetilde{r}_{j+2}^2, \widetilde{r}_j \widetilde{r}_{j+1}) = 0$ and

$$\mathrm{Cov}_{\mathcal{G}}(\widetilde{r}_j^2, \widetilde{r}_{j+1}\widetilde{r}_{j+2})$$
$$= \mathrm{Cov}_{\mathcal{G}}\left(r_j^2 + 2r_j\varepsilon_j + \varepsilon_j^2,\ r_{j+1}r_{j+2} + \varepsilon_{j+1}r_{j+2} + r_{j+1}\varepsilon_{j+2} + \varepsilon_{j+1}\varepsilon_{j+2}\right)$$
$$= \mathrm{Cov}\left(\eta_j^2 - 2\eta_j\eta_{j-1} + \eta_{j-1}^2, \eta_{j+1}\eta_{j+2} - \eta_{j+1}^2 - \eta_j\eta_{j+2} + \eta_j\eta_{j+1}\right) = 0.$$

Similarly,

$$\mathrm{Cov}_{\mathcal{G}}(\widetilde{r}_j^2, \widetilde{r}_j \widetilde{r}_{j+1}) = \mathrm{Cov}_{\mathcal{G}}\left(r_j^2 - 2r_j\varepsilon_j + \varepsilon_j^2,\ r_j r_{j+1} - \varepsilon_j r_{j+1} - r_j \varepsilon_{j+1} + \varepsilon_j \varepsilon_{j+1}\right)$$
$$= 2\mathrm{Cov}_{\mathcal{G}}\left(r_j\varepsilon_j, r_j\varepsilon_{j+1}\right) + \mathrm{Cov}\left(\varepsilon_j^2, \varepsilon_j\varepsilon_{j+1}\right)$$
$$= \begin{cases} -E_{\mathcal{G}}(r_j^2)\sigma_\varepsilon^2 - \frac{(\lambda+1)}{4}\sigma_\varepsilon^4, & j = 2, \ldots, n, \\ -E_{\mathcal{G}}(r_j^2)\sigma_\varepsilon^2 - \frac{(\lambda-1)}{4}\sigma_\varepsilon^4, & j = 1, \end{cases}$$

where

$$\begin{aligned}\mathrm{Cov}\left(\varepsilon_j^2, \varepsilon_j\varepsilon_{j+1}\right) &= \mathrm{Cov}\left(\eta_j^2 - 2\eta_j\eta_{j-1}+\eta_{j-1}^2, \eta_j\eta_{j+1}-\eta_j^2-\eta_{j-1}\eta_{j+1}+\eta_j\eta_{j-1}\right)\\ &= -\mathrm{Var}(\eta_j^2) - 2\mathrm{Var}(\eta_j\eta_{j-1}) = \begin{cases}-(\lambda+1)/4\sigma_\varepsilon^4, & j=2,\ldots,n,\\ -(\lambda-1)/4\sigma_\varepsilon^4, & j=1.\end{cases}\end{aligned}$$

The result of $\mathrm{Cov}_{\mathcal{G}}(\widetilde{r}_{j+1}^2, \widetilde{r}_j\widetilde{r}_{j+1})$ can be obtained similarly as above. Thus,

$$\begin{aligned}\rho_1 &= -2E\left[\sum_{j=1}^{n} E_{\mathcal{G}}\left(r_j^2\right)\sigma_\varepsilon^2 + \frac{(\lambda-1)+(n-1)(\lambda+1)}{4}\sigma_\varepsilon^4\right]\\ &= -2\sigma_\varepsilon^2 E\left(\int_0^1 \sigma_s^2 \mathrm{d}s\right) - 2A_1 - \frac{(\lambda+1)n-2}{2}\sigma_\varepsilon^4.\end{aligned}$$

For $h \geq 2$,

$$\begin{aligned}&\mathrm{Cov}_{\mathcal{G}}\left(\sum_{j=1}^{n-h+2}\widetilde{r}_j\widetilde{r}_{j+h-1}, \sum_{k=1}^{n-h+1}\widetilde{r}_k\widetilde{r}_{k+h}\right)\\ &= E_{\mathcal{G}}\left[\left(\sum_{j=1}^{n-h+2}\widetilde{r}_j\widetilde{r}_{j+h-1}\right)\left(\sum_{k=1}^{n-h+1}\widetilde{r}_k\widetilde{r}_{k+h}\right)\right]\\ &= \sum_{j=1}^{n-h+1} E_{\mathcal{G}}(\widetilde{r}_j^2)E(\varepsilon_{j+h-1}\varepsilon_{j+h})\\ &\quad + \sum_{j=1}^{n-h+1} E(\varepsilon_j\varepsilon_{j+1})E_{\mathcal{G}}(\widetilde{r}_{j+h}^2)\\ &= \sum_{j=1}^{n-h+1}\left(\int_{(j-1)/n}^{j/n}\sigma_s^2\mathrm{d}s + E(\varepsilon_j^2)\right)\left(-\frac{1}{2}\sigma_\varepsilon^2\right)\\ &\quad + \sum_{j=1}^{n-h+1}\left(-\frac{1}{2}\sigma_\varepsilon^2\right)\left(\int_{(j+h-1)/n}^{(j+h)/n}\sigma_s^2\mathrm{d}s + E(\varepsilon_{j+h}^2)\right)\\ &= -\sigma_\varepsilon^2\int_0^1\sigma_s^2\mathrm{d}s - \frac{1}{2}\sigma_\varepsilon^2\int_{1-h/n}^1\sigma_s^2\mathrm{d}s\\ &\quad -\frac{1}{2}\sigma_\varepsilon^2\int_0^{h/n}\sigma_s^2\mathrm{d}s - \left(n-h+\frac{1}{2}\right)\sigma_\varepsilon^4.\end{aligned}$$

Thus,

$$\rho_h = -\sigma_\varepsilon^2 E\left(\int_0^1\sigma_s^2\mathrm{d}s\right) - \frac{1}{2}A_h - \frac{1}{2}B_h - \frac{2n-2h+1}{2}\sigma_\varepsilon^4,$$

for $h \geq 2$. Finally, we consider the function $\nu_h = E\left[\mathrm{Cov}_{\mathcal{G}}(L_{h-2}, L_h)\right]$.

$$\nu_2 = E\left[\mathrm{Cov}_{\mathcal{G}}\left(\sum_{j=1}^{n+1}\widetilde{r}_j^2, \sum_{j=1}^{n-1}\widetilde{r}_j\widetilde{r}_{j+2}\right)\right] = E\left[E_{\mathcal{G}}\left(\sum_{j=1}^{n+1}\widetilde{r}_j^2\right)\left(\sum_{j=1}^{n-1}\widetilde{r}_j\widetilde{r}_{j+2}\right)\right]$$
$$= \sum_{j=1}^{n-1} E(\varepsilon_j\varepsilon_{j+1}^2\varepsilon_{j+2}) = \sum_{j=1}^{n-1} E(2\eta_j^2\eta_{j+1}^2) = 2(n-1)\sigma_\eta^4 = \frac{n-1}{2}\sigma_\varepsilon^4.$$

For $h \geq 3$,

$$\begin{aligned}\nu_h &= E\left[\mathrm{Cov}_{\mathcal{G}}\left(\sum_{j=1}^{n-h+3}\widetilde{r}_j\widetilde{r}_{j+h-2}, \sum_{k=1}^{n-h+1}\widetilde{r}_k\widetilde{r}_{k+h}\right)\right] \\ &= E\left[E_{\mathcal{G}}\left(\sum_{j=1}^{n-h+3}\widetilde{r}_j\widetilde{r}_{j+h-2}\right)\left(\sum_{k=1}^{n-h+1}\widetilde{r}_k\widetilde{r}_{k+h}\right)\right] \\ &= E\left(\sum_{j=1}^{n-h+1}\varepsilon_{j+1}\varepsilon_{j+h-1}\varepsilon_j\varepsilon_{j+h}\right) = \sum_{j=1}^{n-h+1} E(\varepsilon_j\varepsilon_{j+1})E(\varepsilon_{j+h-1}\varepsilon_{j+h}) \\ &= \frac{n-h+1}{4}\sigma_\varepsilon^4.\end{aligned}$$

□

Appendix B: Figures

See Figs. 1, 2, 3, 4, 5 and 6.

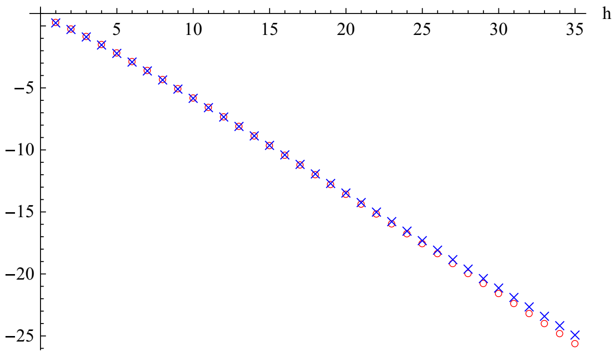

Fig. 1 $\ln\theta_h^*$ (*red circle*) and $\ln\theta_h^a$ (*blue cross*) v.s. h

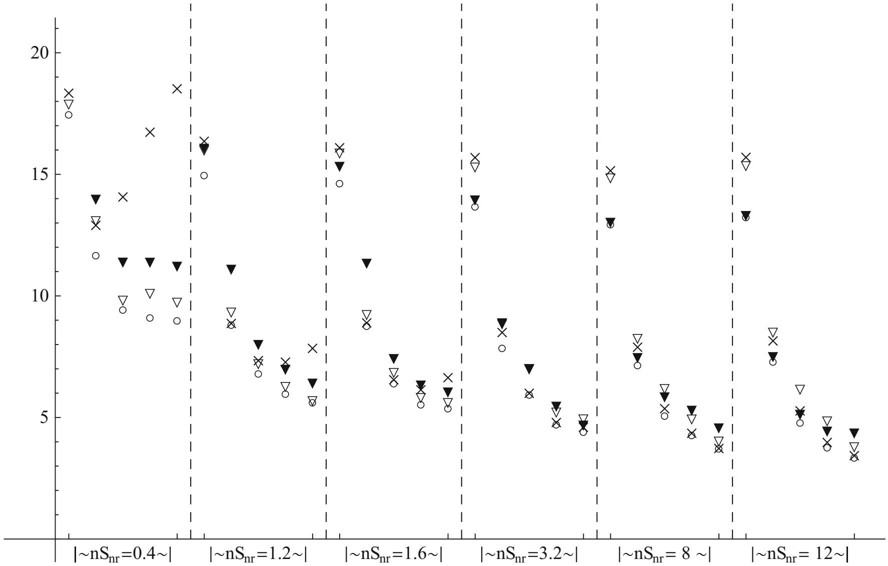

Fig. 2 The relative root mean squared errors (RE) of the five estimators (in percentage) for the Heston model: two-scaled (*inverted triangle*); multi-scale (*square*); kernel estimator (*cross*); pre-averaging estimator (*diamond*); S_L (*circle*). For each nS_{nr}, the RE are shown sequentially from left to right for the sample size $n = 500,\ 2000,\ 5000,\ 8000,\ 12000,\ 24000$

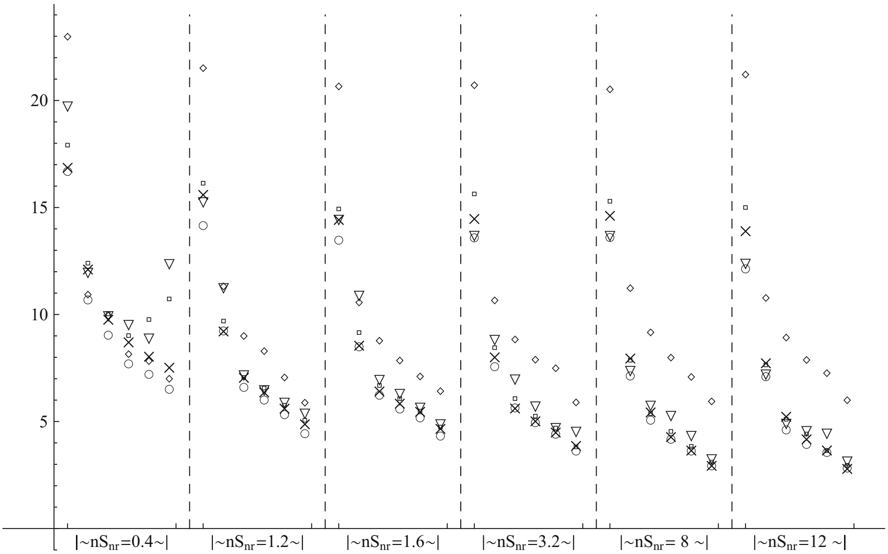

Fig. 3 The RE of the five estimators (in percentage) for the Heston model with leverage effect ($\varphi = -0.5$): two-scaled (*inverted triangle*); multi-scale (*square*); kernel estimator (*cross*); pre-averaging estimator (*diamond*); S_L (*circle*). For each nS_{nr}, the RE are shown sequentially from left to right for the sample size $n = 500,\ 2000,\ 5000,\ 8000,\ 12000,\ 24000$

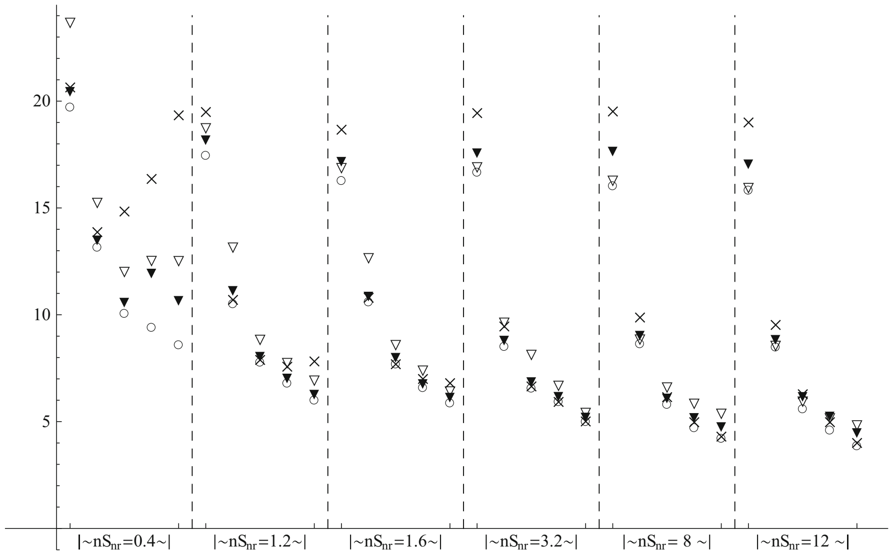

Fig. 4 The RE of the five estimators (in percentage) for the deterministic volatility: two-scaled (*inverted triangle*); multi-scale (*square*); kernel estimator (*cross*); pre-averaging estimator (*diamond*); S_L (*circle*). For each nS_{nr}, the RE are shown sequentially from left to right for the sample size $n = 500,\ 2000,\ 5000,\ 8000,\ 12000,\ 24000$

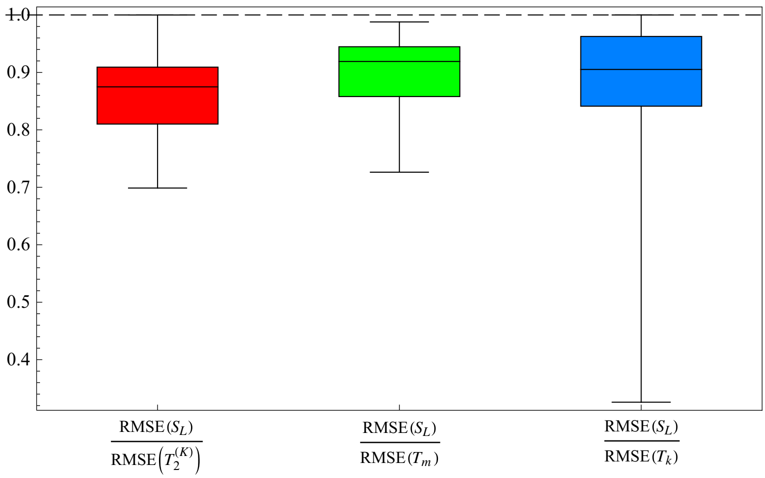

Fig. 5 The boxplots of the relative efficiencies of $T_2^{(K)}$, T_m, T_k and T_p with respect to S_L based on 468 cases

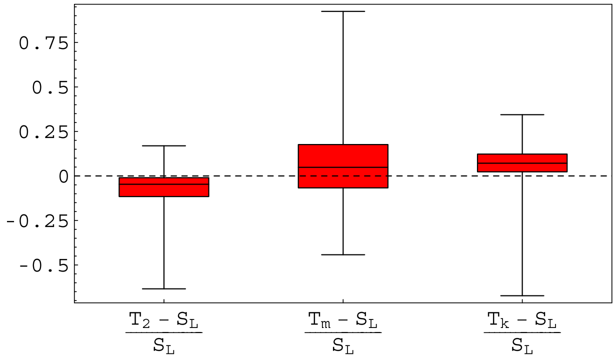

Fig. 6 The boxplots of the daily relative differences of $T_2^{(K)}$, T_m, T_k and T_p with respect to S_L based on fifteen companies and three sampling frequencies

Appendix C: Table

See Tables 1, 2, 3 and 4.

Table 1 The setting of $(V, \sigma_\varepsilon^2)$ and nS_{nr}

V $(\times 10^{-4})$	1.6	4.8	1.6	3.2	3.2	4.8
σ_ε^2 $(\times 10^{-7})$	4	4	1	1	0.4	0.4
nS_{nr} $(\times 10^3)$	0.4	1.2	1.6	3.2	8	12

Table 2 The RE of $\hat{\sigma}_i^2, i = 1, \ldots, 4$

nS_{nr} $(\times 10^3)$	0.4	1.2	1.6	3.2	8	12
$n = 500$						
$\hat{\sigma}_1^2$	0.4356	1.2409	1.6833	3.3540	8.2776	12.6152
$\hat{\sigma}_2^2$	0.1483	0.2254	0.2609	0.4437	0.9782	1.4006
$\hat{\sigma}_3^2$	0.1263	0.3340	0.4751	1.0692	2.8833	4.5186
$\hat{\sigma}_4^2$	0.1199	0.2099	0.2499	0.4233	0.8161	1.0886
$n = 5000$						
$\hat{\sigma}_1^2$	0.0474	0.1274	0.1661	0.3328	0.8208	1.2607
$\hat{\sigma}_2^2$	0.0386	0.0401	0.0395	0.0442	0.0576	0.0704
$\hat{\sigma}_3^2$	0.0273	0.0290	0.0297	0.0517	0.1705	0.3074
$\hat{\sigma}_4^2$	0.0258	0.0281	0.0282	0.0342	0.0504	0.0652
$n = 24000$						
$\hat{\sigma}_1^2$	0.0141	0.0287	0.0363	0.0697	0.1701	0.2595
$\hat{\sigma}_2^2$	0.0175	0.0179	0.0168	0.0183	0.0191	0.0198
$\hat{\sigma}_3^2$	0.0124	0.0125	0.0121	0.0134	0.0187	0.0301
$\hat{\sigma}_4^2$	0.0115	0.0117	0.0111	0.0126	0.0138	0.0155

Table 3 The setting of $(m, \sigma_\varepsilon^2)$

m	1	3	1	2	2	3
σ_ε^2 $(\times 10^{-7})$	4	4	1	1	0.4	0.4
nS_{nr} $(\times 10^3)$	0.4	1.2	1.6	3.2	8	12

Table 4 The monthly averages $(\times 10^{-4})$ of RV and the other five estimators based on fifteen stocks and three sampling frequencies (s) from 2002/01/02 ∼ 2002/01/31

	ABT			AMD			BAC		
	$s=1$	$s=5$	$s=10$	$s=1$	$s=5$	$s=10$	$s=1$	$s=5$	$s=10$
RV	2.638	2.520	2.397	12.47	11.61	10.94	3.279	3.094	2.825
T_2	1.629	1.640	1.688	9.423	9.679	9.621	2.016	2.023	2.066
T_m	1.651	1.638	1.601	10.51	10.54	10.54	2.182	2.191	2.192
T_k	1.681	1.696	1.719	9.982	10.22	10.16	2.108	2.123	2.166
T_p	1.564	1.547	1.553	9.716	9.774	9.394	2.092	2.085	2.072
S_L	1.630	1.641	1.678	9.174	9.471	9.427	2.033	2.010	2.041
	C			GE			JNJ		
	$s=1$	$s=5$	$s=10$	$s=1$	$s=5$	$s=10$	$s=1$	$s=5$	$s=10$
RV	4.378	4.022	3.657	7.441	5.745	5.140	3.309	2.985	2.579
T_2	3.365	3.411	3.503	2.118	2.712	2.840	1.283	1.324	1.405
T_m	3.529	3.566	3.513	3.175	3.261	3.404	1.458	1.463	1.461
T_k	3.499	3.546	3.602	2.692	2.941	3.110	1.386	1.419	1.476
T_p	3.384	3.405	3.376	3.130	3.102	3.107	1.391	1.388	1.401
S_L	3.292	3.337	3.448	2.658	2.704	2.798	1.318	1.333	1.390
	JPM			KO			MCD		
	$s=1$	$s=5$	$s=10$	$s=1$	$s=5$	$s=10$	$s=1$	$s=5$	$s=10$
RV	8.605	7.936	7.192	2.388	2.216	2.014	6.149	5.685	5.144
T_2	4.581	4.629	4.750	1.306	1.349	1.391	2.418	2.511	2.531
T_m	4.780	4.832	4.840	1.432	1.421	1.428	2.722	2.716	2.718
T_k	4.821	4.871	4.951	1.416	1.436	1.469	2.621	2.667	2.700
T_p	4.487	4.431	4.422	1.333	1.283	1.329	2.558	2.547	2.525
S_L	4.648	4.695	4.793	1.329	1.355	1.381	2.504	2.554	2.559

Table 4 continued

	MER			MRK			NOK		
	$s = 1$	$s = 5$	$s = 10$	$s = 1$	$s = 5$	$s = 10$	$s = 1$	$s = 5$	$s = 10$
RV	6.016	5.693	5.282	2.368	2.235	2.084	9.011	8.342	7.453
T_2	4.353	4.454	4.501	1.568	1.603	1.654	4.337	4.486	4.646
T_m	4.862	4.918	4.898	1.693	1.683	1.676	4.949	5.072	5.087
T_k	4.604	4.707	4.739	1.652	1.673	1.705	4.637	4.772	4.923
T_p	4.535	4.574	4.426	1.593	1.586	1.570	4.632	4.709	4.638
S_L	4.312	4.380	4.369	1.586	1.614	1.651	4.349	4.460	4.598
	PEP			T			XOM		
	$s = 1$	$s = 5$	$s = 10$	$s = 1$	$s = 5$	$s = 10$	$s = 1$	$s = 5$	$s = 10$
RV	2.206	2.110	1.984	5.488	5.075	4.640	4.725	4.247	3.694
T_2	1.357	1.359	1.396	3.044	3.122	3.213	2.132	2.185	2.270
T_m	1.417	1.409	1.407	3.516	3.514	3.468	2.268	2.247	2.203
T_k	1.435	1.440	1.477	3.468	3.519	3.506	2.258	2.292	2.365
T_p	1.343	1.321	1.327	3.149	3.168	3.093	2.181	2.156	2.157
S_L	1.354	1.355	1.387	3.001	3.049	3.089	2.144	2.174	2.242

Acknowledgments We are grateful to Kerby Shedden for suggesting the title of our paper. This research was supported in part by the Grants NSC 98-2118-M-110-001-MY2, NSC 100-2118-M-110-003 and NSC 102-2118-M-110-002-MY2 from the National Science Council of Taiwan and MOST 103-2811-M-110-003 from Ministry of Science and Technology.

References

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2000). Great realizations. *Risk*, *13*, 105–108.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*, 43–76.

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

Bandi, F. M., Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies*, *75*, 339–369.

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

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

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N. (2009). Realized kernels in practice: trades and quotes. *Econometrics Journal*, *12*, C1–C32.

Bibinger, M., Mykland, P.A. (2013). *Inference for multi-dimensional high-frequency data: equivalence of methods, central limit theorems, and an application to conditional independence testing*. Technical Report: University of Chicago.

Chen, S. X., Gao, J., Tang, C. (2008). A test for model specification of diffusion processes. *Annals of Statistics*, *36*, 167–198.

Cox, J. (1975). *Notes on option pricing I: constant elasticity of diffusions*. Discussion paper: Stanford University.

Cox, J. C., Ingersoll, J. E., Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, *53*, 385–408.

Dacorogna, M. M., Gencay, R., Müller, U. A., Olsen, R. B., Pictet, O. V. (2001). *An introduction to high-frequency finance*. San Diego: Academic Press.

Fan, J., Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, *102*, 1349–1362.

Hansen, P. R., Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, *24*, 127–218.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, *6*, 327–343.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and their Applications*, *119*, 2249–2276.

Kalnina, I., Linton, O. (2008). Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *Journal of Econometrics*, *147*(1), 47–59.

Kelley, W. G., Peterson, A. C. (2000). *Difference equations: an introduction with applications(2nd ed.)*. New York: Academic Press.

Koike, Y. (2014). *Time endogeneity and an optimal weight function in pre-averaging covariance estimation* (preprint). http://arxiv.org/abs/1403.7889

Lee, S. (2010). On the goodness of fit test for discretely observed sample from diffusion processes: divergence measure approach. *Journal of the Korean Mathematical Society*, *47*, 1137–1146.

Lee, S., Guo, M. (2012). Test for dispersion constancy in SDE models. *Applied Stochastic Models in Business and Industry*, *28*, 342–353.

Lin, L. C., Lee, S., Guo, M. (2013). Goodness-of-fit test for stochastic volatility models. *Journal of Multivariate Analysis*, *116*, 473–498.

Reiss, M. (2011). Asymptotic equivalence for inference in the volatility from noisy observations. *Annals of Statistics*, *39*(2), 772–802.

Shreve, S.E. (2004). *Stochastic calculus for finance II: continuous-time models(2nd ed.)*. New York: Springer.

Sun, Y. (2006). *Best quadratic unbiased estimators of integrated variance in the presence of market microstructure noise*. Working paper, University of California at San Diego.

Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics*, *159*, 235–250.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli*, *12*, 1019–1043.

Zhang, L., Mykland, P. A., Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, *100*, 1394–1411.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business and Economic Statistics*, *14*, 45–52.