City Research Online

City St George's, University of London

Citation: Zu, Y. & Peter Boswijk, H. (2014). Estimating spot volatility with high-frequency financial data. Journal of Econometrics, 181(2), pp. 117-135. doi: 10.1016/j.jeconom.2014.04.001

This is the unspecified version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/3849/

Link to published version: https://doi.org/10.1016/j.jeconom.2014.04.001

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Estimating spot volatility with high-frequency financial data

Yang Zu*
City University London

H. Peter Boswijk†
University of Amsterdam

February 7, 2014

Abstract

We construct a spot volatility estimator for high-frequency financial data which contain market microstructure noise. We prove consistency and derive the asymptotic distribution of the estimator. A data-driven method is proposed to select the scale parameter and the bandwidth parameter in the estimator. In Monte Carlo simulations, we compare the finite sample performance of our estimator with some existing estimators. Empirical examples are given to illustrate the potential applications of the estimator.

JEL Classification: C58, C13, C14.

Keywords: spot volatility, market microstructure noise, subsampling, scale selection, bandwidth selection.

*Email: Yang.Zu@city.ac.uk. Department of Economics, City University London, Northampton Square, EC1V 0HB, London, United Kingdom. Telephone: +44 (0)20 7040 8619. Fax: +44 (0)20 7040 8580.

†Email: H.P.Boswijk@uva.nl. Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE Amsterdam, The Netherlands. Telephone: +31 (0)20 525 4316.

1 Introduction

Spot volatility, also known as *instantaneous volatility*, measures the strength of return variations at a certain time point, expressed per unit of time (Andersen et al. (2010)). Spot volatility has important applications in studying the intraday patterns of the volatility process, testing price jumps (Lee and Mykland (2007), Veraart (2010)), and estimating parametric stochastic volatility models (Bandi and Reno (2009), Kanaya and Kristensen (2010)). In this paper, we are interested in the nonparametric estimation of *spot volatility* with high-frequency financial data.

Spot volatility estimation in the literature dates back to Merton (1980), who considered a constant volatility model. Later on, researchers tended to estimate volatility in the context of the ARCH model (Engle (1982)), the GARCH model (Bollerslev (1986)), and their numerous variations. Nonparametric estimation of spot volatility in the context of diffusion models was firstly considered by Foster and Nelson (1996). Andreou and Ghysels (2002) conducted simulation studies using Foster and Nelson's estimator and some related estimators. Recent contributions include Mykland and Zhang (2008), Fan and Wang (2008) and Kristensen (2010).

High-frequency financial data have become more accessible for academic research in recent years. In contrast to low frequency (daily, weekly or longer sampling frequency) financial datasets, high-frequency datasets are characterized by the large number of observations they contain and the existence of so-called market microstructure noise. O'Hara (1998) made theoretical studies of market microstructure noise; Andersen et al. (2000) and Hansen and Lunde (2006) analyzed the empirical characteristics of the noise.

Existing research on volatility measurement for high-frequency data focuses mainly on the *ex post* nonparametric measurement of the *integrated volatility* of the underlying efficient price process. Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002) made important early contributions to the use of *realized variance* to estimate the integrated volatility. However, they did not consider the effects of market microstructure noise, and the realized variance estimator can only be applied to sparsely sampled data, where the effects of noise are small. The problem of estimating integrated volatility under noise was first studied by Zhou (1996), who gave an unbiased but inconsistent estimator for integrated volatility. Aït-Sahalia et al. (2005) considered a constant variance model and gave a Maximum Likelihood Estimator for the constant variance. Later, four types of estimators were proposed for estimating integrated volatility in the presence of noise. These are the subsampling-based Two Scale Realized Variance (TSRV) estimator by Zhang et al. (2005) and the Multiscale Realized Variance (MSRV) estimator by Zhang (2006); the Realized Kernel (RK) estimator by Barndorff-Nielsen et al. (2008), which is based on Zhou (1996)'s first order moving average correction; the pre-averaging method by Podolskij and Vetter (2009), and Jacod et al. (2009); and the Quasi-Maximum Likeli-

hood Estimator (QMLE) by Xiu (2010), which is based on the estimator in Aït-Sahalia et al. (2005).

In this paper, we study the problem of estimating spot volatility with high-frequency data and we explicitly consider the effects of market microstructure noise. Our approach is closely related to the literature on integrated volatility measurement with noise. We construct our estimator based on the Two Scale Realized Variance estimator by Zhang et al. (2005) — our estimator calculates the increment of the Two Scale Realized Variance estimator over a small interval and applies an appropriate normalization. Under appropriate conditions, we prove consistency and derive the asymptotic distribution of our estimator and propose a data-driven procedure to select tuning parameters. In practically meaningful Monte Carlo simulations, we compare our estimator with existing methods in terms of several error measures and we demonstrate the improved accuracy in using our estimator.

Some recent research is closely related to this paper. Mykland and Zhang (2008) independently proposed the same estimator as in our paper, but did not provide a complete asymptotic theory. Bandi and Reno (2009), Ogawa and Sanfelici (2011), Bos et al. (2012), among others, have considered spot volatility estimators based on the Realized Kernel estimator and the Pre-Averaging estimator. In a concurrent paper, Mancini et al. (2012) (Section 3.1) have proposed a two-scale estimator for spot volatility weighted by the so-called delta sequence and have provided theoretical analysis; our estimator is a special case of their estimator with equal weights. We provide more comprehensive asymptotic and finite sample studies for our estimator; we also study the problem of bandwidth and scale parameters selection, which is important for practical implementation. In the presence of jumps but without noise, spot volatility has been studied by Aït-Sahalia and Jacod (2009), Ngo and Ogawa (2009) and Andersen et al. (2009). Munk and Schmidt-Hieber (2010b) and Munk and Schmidt-Hieber (2010a) studied the best possible convergence rate of any spot volatility estimator in a volatility model observed with noise, where the volatility process is assumed to be a deterministic function. Hoffmann et al. (2010) derived a minimax bound for the same problem in a genuine stochastic volatility model observed with noise, they showed that this bound is "nearly optimal" in their definition and they proposed a wavelet estimator that achieves this rate. Their rate is $n^{-1/8}$ if translated to the present context, up to some logarithmic corrections. Our estimator does not have the best rate of convergence in their sense, we discuss possible extensions to improve the convergence rate in Section 6.

The structure of this paper is as follows. Section 2 introduces the setup of the problem. Section 3 defines the estimator, studies its asymptotic properties and the problem of bandwidth and scale selection. Section 4 conducts Monte Carlo studies on the finite-sample properties of the estimator and Section 5 contains two empirical applications to Euro FX futures data. Section 6 discusses possible extensions to our model. Section

7 concludes the paper. Proofs are collected in Appendix A and technical lemmas are collected in Appendix B.

Throughout the paper, $\langle X, Y\rangle$ denotes the quadratic covariation of two processes X and Y; $\xrightarrow{d}$ denotes converge in distribution; $\xrightarrow{\text{st}}$ denotes stable convergence in distribution; $\xrightarrow{p}$ denotes converge in probability; for a real number x, $\lfloor x \rfloor$ denotes its integer part. We call σ_t^2 the *spot variance* at time t, and we call σ_t the *spot volatility* at time t. However, as in the financial econometrics literature, when we use the term *spot volatility* in general discussions, it could refer to either σ_t^2 or σ_t, depending on the context.

2 The model

Let $\{X_t\}$ be a univariate log price process, assumed to be a Brownian semimartingale, satisfying

$$\mathrm{d}X_t \;=\; \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad t \in [0,1],$$

where $\{W_t\}$ is a standard Brownian motion; $\{\mu_t\}$ is the spot drift process, and $\{\sigma_t\}$ is the spot volatility process; both are predictable. We further assume:

A1 the processes $\{\mu_t\}$ and $\{\sigma_t\}$ have continuous sample paths.

A2 the process $\{\sigma_t\}$ is positive.

Since X is a Brownian semimartingale, it has continuous sample paths, and its quadratic variation process satisfies

$$\langle X, X\rangle_t = \int_0^t \sigma_s^2 \mathrm{d}s, \quad t \in [0,1],$$

such that the spot volatility satisfies

$$\sigma_t^2 = \frac{\mathrm{d}\,\langle X, X\rangle_t}{\mathrm{d}t}. \tag{1}$$

Taking into account the market microstructure noise existing in high-frequency financial data, we further assume

A3 X is not observable, but

$$Y_t = X_t + \varepsilon_t$$

is observed over the interval $[0,1]$ in discrete time over a grid $t_i = i/n$ for $i = 0, 1, \ldots, n$ with equal distance $\Delta_n = 1/n$.

A4 $\{\varepsilon_{t_i}\}_{i=1}^n$ are independent and identically distributed (i.i.d.) with mean 0, variance ω^2, and with finite fourth moment. Furthermore, $\{\varepsilon_{t_i}\}_{i=1}^n$ are independent of the $\{X_t\}$ process.

The continuity assumption on the volatility sample paths accommodates a large class of spot variance processes such as diffusion processes, long memory, deterministic patterns as well as nonstationarity. The model allows for possible dependence between $\{W_t\}$ and $\{\sigma_t\}$, so leverage effects are allowed in this model.

The model specification and Assumption A1 exclude the possibility of jumps in both the price process and the volatility process. Assumption A4 excludes the possibility that the noise is dependent over time (so called dependent noise) and that the noise is dependent of the efficient price process (so-called endogenous noise). We discuss possible extensions to these cases in Section 6.

3 The estimator and its properties

3.1 The estimator

We are interested in estimating the realization of the spot variance process $\{\sigma_t^2\}$ at any time $t \in (0,1)$. Our estimator is based on the Two Scale Realized Variance estimator (TSRV) by Zhang et al. (2005).

The TSRV estimator uses a subsampled and averaged Realized Variance (RV) estimator over a scale K, together with a usual Realized Variance estimator to correct the effects of noise. It is defined as

$$\text{TSRV} = [Y,Y]^K - \frac{\bar{n}}{n}[Y,Y],$$

where

$$\begin{aligned} [Y,Y]^K &:= \frac{1}{K}\sum_{i=K}^{n}(Y_{t_i} - Y_{t_{i-K}})^2, \\ [Y,Y] &:= \sum_{i=1}^{n}(Y_{t_i} - Y_{t_{i-1}})^2, \\ \bar{n} &:= \frac{n-K+1}{K}. \end{aligned}$$

The definition of $[Y,Y]^K$ appears to be different from the original formulation in Zhang et al. (2005), but is an equivalent reformulation of their definition, which was also used by Zhang (2006).

Our estimator is constructed using equation (1) – a theoretical time derivative is approximated by a numerical derivative over a small interval. When the interval is

$[t-h,t]$, a filtering version of the estimator for the spot variance σ_t^2 is defined as:

$$\widehat{\sigma_t^2} = \frac{\text{TSRV}_t - \text{TSRV}_{t-h}}{h}. \tag{2}$$

We call it a filtering version of the estimator because it only uses data up to time t. Similarly, a smoothing version of the estimator which uses both lead and lagged data at time t could be constructed as follows:

$$\widehat{\sigma_t^2} = \frac{\text{TSRV}_{t+h/2} - \text{TSRV}_{t-h/2}}{h}.$$

Here we split the bandwidth evenly between leads and lags. In the following, all the theoretical results will be stated for the filtering version of the estimator. Analogous results apply to the smoothing version. In the Monte Carlo experiment, we study both versions of the estimator.

We call our estimator the Two Scale Realized Spot Variance (TSRSV) estimator. To have a direct expression for the estimator, we first introduce some notation. For a sequence $\{Z_{t_i}, i = 0, 1, \ldots, n\}$, define

$$[Z,Z]_t^{\prime K,h} := \frac{1}{h}\sum_{t-h\le t_i\le t} \frac{(\Delta^K Z_{t_i})^2}{K} = \frac{1}{h}\sum_{t-h\le t_i\le t} \frac{(Z_{t_i} - Z_{t_{i-K}})^2}{K},$$

where we use the notation Δ^K for the Kth difference operator; and we use a $'$ to signify this is a numerical time derivative; when $K = 1$, we write $[Z,Z]_t^{\prime h} := [Z,Z]_t^{\prime 1,h}$. Denote $\hat{V}_t^{K,h}$ as the filtering version of the TSRSV estimator at time t as in (2), it is defined as:

$$\hat{V}_t^{K,h} = [Y,Y]_t^{\prime K,h} - \frac{\bar{n}}{n}[Y,Y]_t^{\prime h},$$

where

$$\bar{n} = \frac{nh - K + 1}{Kh}.$$

The estimator depends on a subsampling size parameter K, which we call the scale parameter, and an interval length parameter h, which we call the bandwidth parameter.

3.2 Decomposition

To study the statistical properties of $\hat{V}_t^{K,h}$ as an estimator of σ_t^2, we first make a "bias-variance"-like decomposition for the difference of the two:

$$\begin{aligned}\hat{V}_t^{K,h} - \sigma_t^2 &= \left(\hat{V}_t^{K,h} - \frac{1}{h}\int_{t-h}^t \sigma_s^2 \mathrm{d}s\right) + \left(\frac{1}{h}\int_{t-h}^t \sigma_s^2 \mathrm{d}s - \sigma_t^2\right) \\ &=: P_0 + P_1,\end{aligned}$$

where P_0 and P_1 are defined implicitly and can be viewed as the variance and the bias part, respectively. However, note that this is not the bias-variance decomposition defined for usual nonparametric estimators, because under our assumption of possible existence of leverage effects, $E[\hat{V}_t^{K,h}] \neq \int_{t-h}^{t} \sigma_s^2 \mathrm{d}s/h$. The variance part P_0 is closely related to a similar quantity studied for the TSRV estimator, and can be analogously decomposed into two parts as in Zhang et al. (2005):

$$\begin{aligned} & \hat{V}_t^{K,h} - \frac{1}{h}\int_{t-h}^{t} \sigma_s^2 \mathrm{d}s \\ = \; & \left([Y,Y]_t^{\prime K,h} - [X,X]_t^{\prime K,h} - \frac{\bar{n}}{n}[Y,Y]_t^{\prime h}\right) + \left([X,X]_t^{\prime K,h} - \frac{1}{h}\int_{t-h}^{t} \sigma_s^2 \mathrm{d}s\right) \\ := \; & P_2 + P_3, \end{aligned}$$

where P_2 and P_3 are defined implicitly. Using the terminologies in the above mentioned paper, these are the "error due to noise" part and "error due to discretization" part, respectively.

We study the "bias" part P_1 first, for which we need make more specific assumptions on the path properties of the volatility process. We assume:

A5 The spot variance process $\{\sigma_t^2\}$ is an Itô process, satisfying

$$\mathrm{d}\sigma_t^2 = \Gamma_t \mathrm{d}t + \Lambda_t \mathrm{d}B_t,$$

where $\{\Gamma_t\}$ and $\{\Lambda_t\}$ are stochastic processes with continuous sample paths. $\{B_t\}$ is a standard Brownian Motion, possibly correlated with $\{W_t\}$.

This was the assumption used in Mykland and Zhang (2008) about the volatility process. Under this assumption, the limiting distribution of P_1 can be derived:

Proposition 1 *Under Assumptions A1, A2 and A5, as $n \to \infty$, $h \to 0$, $nh \to \infty$,*

$$h^{-1/2} P_1 \xrightarrow{\text{st}} \sqrt{\frac{1}{3}\Lambda_t^2} Z^{P_1},$$

where Z^{P_1} is a standard normal variable independent of the σ-algebra generated by the X process. The stable convergence is with respect to the σ-algebra generated by the X process.

Here P_1 converges to a mixed normal distribution stably, where Λ_t is the diffusion parameter of the spot variance process, which is random. The rate of convergence $h^{-1/2}$ is implied by the continuous semimartingale assumption we make for the spot variance process.

The diffusion assumption on the spot variance process is rather restrictive. To accommodate more classes of models for the spot variance process, we could relax Assumption A5 to a Hölder's type continuity assumption:

A5-1 For almost all paths of the spot variance process,

$$\left|\left(\sigma_t^2\right)^{(m)} - \left(\sigma_s^2\right)^{(m)}\right| \leqslant C\left|t-s\right|^{\alpha}, \quad \text{for all } t,s \in [0,1],$$

where $m \geqslant 0$ is an integer, the superscript $^{(m)}$ denotes the mth derivative with respect to time, and $0 < \alpha < 1$.

This assumption was used in Kristensen (2010). It is very general and accommodates diffusion process, long memory stochastic differential equations (as in Comte and Renault (1998)), deterministic patterns as well as nonstationarity in the spot variance process. In particular, continuous semimartingales are allowed for in the assumption with $m = 0$ and $\alpha < 1/2$ (see for example, Ch. V, Exercise 1.20 of Revuz and Yor (1998)). The Lévy's modulus of continuity type assumption $\sup_{|t-s|\leqslant h}|\sigma_t^2 - \sigma_s^2| = O_p\left(h^{1/2}\left|\log h\right|^{1/2}\right)$ (e.g. in Fan and Wang (2008)) is also included in this assumption, as well as a deterministic spot variance function. Under A5-1, a central limit theorem for P_1 is not available, but the following upper bound can be easily derived:

Proposition 2 *Under Assumptions A1, A2 and A5-1,* $P_1 = O_p\left(h^{m+\alpha}\right)$.

We then study the "due to noise" part P_2 and the "due to discretization" part P_3. We show that P_2 converges to a normal distribution conditional on the X process, while P_3 converges to a mixture of normal distributions stably. The results for these quantities will be extensions of Zhang et al. (2005)'s corresponding results for a fixed interval to our context of shrinking intervals.

Proposition 3 *Under Assumptions A1-A4, as* $n \to \infty$, $h \to 0$, $nh \to \infty$, $K \to \infty$ *and* $K/(nh) \to 0$,

$$\begin{aligned}\sqrt{\frac{K^2h}{n}} \times P_2 &= \sqrt{\frac{K^2h}{n}} \times \left([Y,Y]_t^{\prime K,h} - [X,X]_t^{\prime K,h} - \frac{\bar{n}}{n}[Y,Y]_t^{\prime h}\right) \\ &\xrightarrow{d} N\left(0, 8\omega^4\right),\end{aligned}$$

where the convergence is conditional on the X *process.*

Proposition 4 *Under Assumptions A1-A2, as* $n \to \infty$, $h \to 0$, $nh \to \infty$, $K \to \infty$ *and*

$K/(nh) \to 0$,

$$\begin{aligned}\sqrt{\frac{nh}{K}} \times P_3 &= \sqrt{\frac{nh}{K}} \times \left([X,X]_t^{\prime K,h} - \frac{1}{h}\int_{t-h}^{t} \sigma_s^2 \mathrm{d}s \right) \\ &\xrightarrow{\text{st}} \sqrt{\frac{4}{3}\sigma_t^4} Z^{P_3},\end{aligned}$$

where Z^{P_3} is a standard normal variable independent of the σ-algebra generated by the X process. The stable convergence above is with respect to the σ-algebra generated by the X process.

3.3 Asymptotic properties

The consistency of the estimator follows from the results of the previous propositions.

Theorem 1 *Under Assumptions A1 to A4, with either A5 or A5-1 satisfied, when $n \to \infty$, $h \to 0$, $nh \to \infty$, $K \to \infty$, $K/(nh) \to 0$ and $n/(K^2 h) \to 0$,*

$$\hat{V}_t^{K,h} \xrightarrow{p} \sigma_t^2.$$

The asymptotic distribution of the estimator under Assumption A5 is given in the following theorem:

Theorem 2 *Under Assumptions A1 to A5, when $K = c_K n^{2/3}$ and $h = c_h n^{-1/6}$, as $n \to \infty$,*

$$n^{1/12}\left(\hat{V}_t^{K,h} - \sigma_t^2\right) \xrightarrow{\text{st}} \left(8\omega^4 \frac{1}{c_K^2 c_h} + \frac{c_K}{c_h}\eta_t^2 + \frac{c_h}{3}\Lambda_t^2\right)^{\frac{1}{2}} N(0,1),$$

where $\eta_t^2 = 4\sigma_t^4/3$ and c_K and c_h are constants.

This limiting theorem balances the order of the "bias" term P_1 and the variance term $P_2 + P_3$ to achieve a convergence rate $n^{-1/12}$.

Alternatively, under the Hölder type continuity Assumption A5-1, the following limiting theorem can be derived:

Theorem 3 *Under Assumptions A1–A4 and A5-1, when $K = c_K n^{2/3}$ and $n^{1/6} h^{q+1/2} \to 0$ as $n \to \infty$,*

$$n^{1/6} h^{1/2}\left(\hat{V}_t^{K,h} - \sigma_t^2\right) \xrightarrow{\text{st}} \left(\frac{8\omega^4}{c_K^2} + c_K \eta_t^2\right)^{\frac{1}{2}} N(0,1),$$

where $\eta_t^2 = 4\sigma_t^4/3$, $q = m + \alpha$ is the smoothness parameter, and c_K is a constant.

Because the exact distribution of the "bias" part P_1 is not known, the theorem characterizes an undersmoothed version of the estimator, such that P_1 is dominated by P_2 and P_3. The convergence rate of the estimator is $n^{-1/6} h^{-1/2}$. As compared to the $n^{-1/6}$

rate of the TSRV estimator for Integrated Variance, the TSRSV estimator for spot variance loses the rate $h^{-1/2}$. For the bias part P_1 to be dominated, it is necessary that $n^{1/6}h^{q+1/2} \to 0$, this implies $h = o\left(n^{-1/[3(2q+1)]}\right)$, so that the loss in the convergence rate is $h^{-1/2} = n^{1/[6(2q+1)]}$. For continuous semimartingales, where q is any number that satisfies $0 < q < 1/2$, the loss in the rate is $n^{1/12}$ at least, so that the convergence rate with the current undersmoothing bandwidth can be infinitely close to, but is always lower than $n^{-1/12}$. As compared to the $n^{-1/12}$ rate in Theorem 2, the loss in the rate is because of the undersmoothing bandwidth used in this theorem.

The above estimator and the limiting distributions are for the spot variance. By continuous mapping theorem, it is easy to show that $\left(\hat{V}_t^{K,h}\right)^{1/2}$ is a consistent estimator for the spot volatility σ_t. Applying the delta method, the following limiting distribution for the spot volatility can be obtained by applying the transformation $f(x) = x^{1/2}$. Here we only give the limiting distribution for the spot volatility estimator under Assumption A5, the result under Assumption A5-1 can be obtained analogously and we omit it here.

Theorem 4 *Under Assumptions A1 to A5, when* $K = c_K n^{2/3}$ *and* $h = c_h n^{-1/6}$*, as* $n \to \infty$*,*

$$n^{1/12}\left(\left(\hat{V}_t^{K,h}\right)^{1/2} - \sigma_t\right) \xrightarrow{\text{st}} \frac{1}{2\sigma_t}\left(8\omega^4 \frac{1}{c_K^2 c_h} + \frac{c_K}{c_h}\eta_t^2 + \frac{c_h}{3}\Lambda_t^2\right)^{\frac{1}{2}} N(0,1),$$

where $\eta_t^2 = 4\sigma_t^4/3$ *and* c_K *and* c_h *are constants.*

Remark 1 The asymptotic distributions can be used to construct confidence interval, provided that consistent estimators are available for the unknown quantities in the expressions of the asymptotic variances. The noise variance ω^2 can be estimated consistently using the estimator in Zhang et al. (2005):

$$\widehat{\omega^2} = \frac{1}{2n}\sum_{i=1}^{n}\left(Y_{t_i} - Y_{t_{i-1}}\right)^2. \tag{3}$$

A consistent estimator for σ_t^4 can be simply obtained by taking squares on the estimator for the spot variance. However, a consistent estimator for Λ_t^2 is not known (to us). Heuristically, one could use the squared increment of the estimated spot variance process to approximate it.

3.4 Scale and bandwidth selection

To use the estimators in practice, the tuning parameters K and h need be chosen beforehand. In this section, we propose a method to choose these parameters based on the conditional Mean Integrated Squared Error (MISE) under Assumption A5. The MISE

is defined conditional on the volatility path, because practically what we are concerned with is estimating the *realized* but latent spot variance path.

Define

$$\text{MISE}^{\text{c}}(K,h) := E^{\sigma}\left[\int_0^1 \left(\hat{V}_t^{K,h} - \sigma_t^2\right)^2 \mathrm{d}t\right],$$

as the conditional MISE, where E^{σ} denotes the conditional expectation with respect to the sigma-algebra generated by $\{\sigma_t\}_{t\in[0,1]}$. This is a global measure for errors made on a generic interval $[0,1]$.

To evaluate $\text{MISE}^{c}(K,h)$, we make the following decomposition:

$$\begin{aligned} & E^{\sigma}\left[\int_0^1 \left(\hat{V}_t^{K,h} - \sigma_t^2\right)^2 \mathrm{d}t\right] \\ = & \int_0^1 E^{\sigma}\left[\left(\hat{V}_t^{K,h} - \sigma_t^2\right)^2\right] \mathrm{d}t \\ = & \int_0^1 \left\{E^{\sigma}\left[\left(\hat{V}_t^{K,h} - \sigma_t^2\right)\right]^2 + \overset{\sigma}{\text{Var}}\left[\left(\hat{V}_t^{K,h} - \sigma_t^2\right)\right]\right\} \mathrm{d}t, \end{aligned}$$

where Var^{σ} denotes the conditional variance with respect to the sigma-algebra generated by $\{\sigma_t\}_{t\in[0,1]}$. With this decomposition, and using the results of Theorem 2, we know that when $K = c_K n^{2/3}$ and $h = c_h n^{-1/6}$, MISE can be approximated by the following Asymptotic MISE (AMISE):

$$\text{AMISE}^{\text{c}}(K,h) := 8\frac{n}{K^2 h}\omega^4 + \frac{K}{nh}\frac{4}{3}\int_0^1 \sigma_t^4 \mathrm{d}t + \frac{1}{3}h\int_0^1 \Lambda_t^2 \mathrm{d}t. \tag{4}$$

From minimizing $\text{AMISE}^{c}(K,h)$, the parameters K and h can be solved explicitly. Denote the minimizers of AMISE as K^0 and h^0. From the first order conditions we obtain:

$$\begin{aligned} K^0 &= K^* \times n^{2/3}, \\ h^0 &= h^* \times n^{-1/6}, \end{aligned}$$

with constants

$$K^* = \left(\frac{12\omega^4}{\int_0^1 \sigma_t^4 \mathrm{d}t}\right)^{1/3}, \tag{5}$$

and

$$h^* = \left(\frac{\frac{8}{(K^*)^2}\omega^4 + \frac{4}{3}K^*\int_0^1 \sigma_t^4 \mathrm{d}t}{\frac{1}{3}\int_0^1 \Lambda_t^2 \mathrm{d}t}\right)^{1/2}. \tag{6}$$

K^* and h^* are still theoretical quantities and need to be estimated. For the constant K^*, ω^2 is the noise variance and can be estimated consistently using the estimator in (3); the integrated quarticity $\int_0^1 \sigma_t^4 \mathrm{d}t$ can be estimated consistently using the pre-averaging estimator in Podolskij and Vetter (2009). Alternatively the integrated quarticity can

be estimated using the Realized Quarticity estimator in Barndorff-Nielsen and Shephard (2002) for sparsely sampled returns. Simply plugging these two consistent estimators into (5) we get a consistent estimator for K^*.

For the constant h^*, the noise variance ω^2 and the integrated quarticity $\int_0^1 \sigma_t^4 \mathrm{d}t$ can be estimated in the same way as we discussed in the previous paragraph. For the quadratic variation of the spot variance process $\langle \sigma^2, \sigma^2 \rangle_1 = \int_0^1 \Lambda_t^2 \mathrm{d}t$, there is no established method to estimate it. We propose a heuristic estimator for it in the next paragraph. Plugging the estimates for all unknown quantities into (6) we get a heuristic plug-in type estimator for h^*, for which the theoretical properties are not known; we study its behaviour using Monte Carlo experiments in the next section.

The heuristic estimator for $\langle \sigma^2, \sigma^2 \rangle_1 = \int_0^1 \Lambda_t^2 \mathrm{d}t$ is constructed using its definition: we calculate the quadratic variation of a preliminarily estimated spot variance path. The preliminary estimate of the spot variance path can be obtained using our TSRSV estimator with heuristically (but reasonably) chosen scale and bandwidth parameters; or it can be obtained by applying the estimator in Mykland and Zhang (2008) or Kristensen (2010) to sparsely sampled returns (say 5-minute returns).

Remark 2 Under Assumption A5-1, only an upper bound for the "bias" part is known; the optimal rate of the bandwidth parameter can still be determined by optimizing the corresponding upper bound for the MISE. Since $P_2 + P_3 = O_p\left(n^{-1/6}h^{-1/2}\right)$, while the bias part $P_1 = O_p\left(h^q\right)$, solving for h to maximize the sum of the two upper bounds gives $h = O\left(n^{-1/[3(2q+1)]}\right)$. For a stochastic volatility model, when $q < 1/2$, this leads to a bandwidth $h = O\left(n^{-1/6}\right)$. The exact optimizing bandwidth is difficult to estimate as this is related to Hölder's constant in Assumption 5-1, which is hard to estimate from data. The selection of the scale parameter will be the same as in (5).

4 Simulation study

In this section, we study the finite sample properties of the estimator. A potential concern on the practical applicability of our estimator is the convergence rate $n^{-1/12}$ — one may be afraid that this rate will be too slow to be meaningful in practical sample sizes. We aim to address this concern in the Monte Carlo experiment. We compare the finite-sample properties of our estimator with two types of estimators: the Realized Spot Variance (RSV) estimator in Mykland and Zhang (2008) (see also Kristensen (2010) and Fan and Wang (2008)) and the TSRV estimator applied to intraday intervals. These estimators are developed in different contexts but can be used to estimate spot volatility.

4.1 Models, parameters and error measures

We use two models as our Data Generating Processes. The models and parameter values are the same as those used in Huang and Tauchen (2005), see also Barndorff-Nielsen et al. (2008).

The first considered model is the One Factor Stochastic Volatility (SV1F) model:

$$\begin{aligned} \mathrm{d}Y_t &= \mu \mathrm{d}t + \sigma_t \mathrm{d}W_t, \\ \sigma_t &= \exp\left(\beta_0 + \beta_1 \tau_t\right), \\ \mathrm{d}\tau_t &= \alpha \tau_t \mathrm{d}t + \mathrm{d}B_t, \end{aligned}$$

where $\mathrm{corr}(\mathrm{d}W_t, \mathrm{d}B_t) = \varphi$, with parameter values $\mu = 0.03$, $\beta_1 = 0.125$, $\alpha = -0.025$, and $\varphi = -0.3$. It is imposed that $\beta_0 = \beta_1/(2\alpha)$ to ensure the restriction $E\sigma_t^2 = 1$, and the τ process is assumed to initiate from its stationary distribution $N(0, -1/(2\alpha))$.

The second considered model is the Two Factor Stochastic Volatility (SV2F) model,

$$\begin{aligned} \mathrm{d}Y_t &= \mu \mathrm{d}t + \sigma_t \mathrm{d}W_t, \\ \sigma_t &= \mathrm{sexp}\left(\beta_0 + \beta_1 \tau_{1t} + \beta_2 \tau_{2t}\right), \\ \mathrm{d}\tau_{1t} &= \alpha_1 \tau_{1t} \mathrm{d}t + \mathrm{d}B_{1t}, \\ \mathrm{d}\tau_{2t} &= \alpha_2 \tau_{2t} \mathrm{d}t + (1 + \phi \tau_{2t}) \mathrm{d}B_{2t}, \end{aligned}$$

where $\mathrm{corr}(\mathrm{d}W_t, \mathrm{d}B_{1t}) = \varphi_1$ and $\mathrm{corr}(\mathrm{d}W_t, \mathrm{d}B_{2t}) = \varphi_2$ with parameter values $\mu = 0.03$, $\beta_0 = -1.2$, $\beta_1 = 0.04$, $\beta_2 = 1.5$, $\alpha_1 = -0.0037$, $\alpha_2 = -1.386$, $\phi = 0.25$, $\varphi_1 = -0.3$ and $\varphi_2 = -0.3$, and the function sexp(.) is defined as

$$\mathrm{sexp}(x) = \begin{cases} \exp(x), & \text{if } x \leqslant \log(1.5); \\ 1.5\sqrt{1 - \log(1.5) + x^2/\log(1.5)}, & \text{otherwise}. \end{cases}$$

It is a usual exponential function with a polynomial function splined in at high values to satisfy the growth condition of the differential equation system, to ensure its solution to exist and the Euler scheme to work. The first factor process is assumed to initiate from its stationary distribution $N(0, -1/(2\alpha_1))$, and the second factor process is assumed to initiate from 0.

In the simulations, time t is measured in days, and it is assumed that there are 6.5 hours in a day (stock market trading time), so in order to simulate second data, we take step size $\Delta = 1/(6.5 \times 60 \times 60) = 1/23400$. These models and parameter values will generate σ_t as percentage volatility. The SV2F model will have larger variation in volatility than the SV1F model.

For the market microstructure noise, we assume it is i.i.d. $N(0, \omega^2)$, where ω^2 is specified by three different magnitudes 0.0001, 0.001, 0.01. These noise magnitudes are

combined with the two models to make 6 scenarios, we call them SV1F1, SV1F2, SV1F3 and SV2F1, SV2F2, SV2F3, respectively.

We use four measurements of error in our comparisons, which are defined as follows:

1. Mean integrated squared error (MISE)

$$E\left[\int_0^1 \left(\hat{\sigma}_t^2 - \sigma_t^2\right)^2 \mathrm{d}t\right]^{1/2}$$

2. Mean integrated squared relative error (MISRE)

$$E\left[\int_0^1 \left(\frac{\hat{\sigma}_t^2 - \sigma_t^2}{\sigma_t^2}\right)^2 \mathrm{d}t\right]^{1/2}$$

3. Mean integrated absolute error (MIAE)

$$E\left[\int_0^1 \left|\hat{\sigma}_t^2 - \sigma_t^2\right| \mathrm{d}t\right]$$

4. Mean integrated absolute relative error (MIARE)

$$E\left[\int_0^1 \left|\frac{\hat{\sigma}_t^2 - \sigma_t^2}{\sigma_t^2}\right| \mathrm{d}t\right]$$

We estimate the spot variance for every second of a trading day (6.5 hours), so the integrals above are approximated by the discretization size of $1/23400$. The expectation is evaluated by the mean of 10000 Monte Carlo replications. The computations are conducted with OxMetrics 6.20, using the default random number generator (the George Marsaglia multiply-with-carry algorithm) with seed 20120505.

4.2 Comparison with existing methods

We compare our estimator with two existing methods of estimating spot volatility. The first method is the nonparametric RSV estimator as in Mykland and Zhang (2008) (see also Fan and Wang (2008) and Kristensen (2010)). This method does not assume the existence of market microstructure noise, and it has to be applied to sparsely sampled data. The second method is the TSRV estimator applied to intraday intervals.

4.2.1 TSRSV estimators

We implement both the filtering version and the smoothing version of the TSRSV. The bandwidth and scale parameters are computed using the method described in Section 3.4.

For the smoothing version of the TSRSV, the bandwidth is equally divided between lead and lagged observations; the noise variance is estimated as in (3); the integrated quarticity is estimated by the Realized Quarticity estimator applied to 5-minute returns — we did not use the preaveraging estimator to estimate the integrated quarticity, to reduce the computational burden; for the quadratic variation of the spot variance process, the preliminary spot variance path is estimated using the Realized Spot Variance estimator applied to 5-minute returns, with the bandwidth selected by the cross-validation method proposed in Kristensen (2010).

4.2.2 RSV estimator

Sparse sampling is usually used in practice to reduce the effects of microstructure noise when dealing with high-frequency financial data. Taking 5-minute returns and treating them as clean data is common practice when measuring integrated volatility. With sparsely sampled data, one can apply the nonparametric RSV estimator as in Mykland and Zhang (2008), which is a flat-weighted version of the kernel estimator of Fan and Wang (2008) and Kristensen (2010), to estimate the spot variance. We implement both the filtering version and the smoothing version for half-minute, 1-minute and 5-minute returns. To select the bandwidth, we use the cross-validation method proposed in Kristensen (2010).

4.2.3 TSRV over fixed intraday intervals

Current volatility measurement theories are well developed for the integrated variance estimation. Besides using these estimators for one day, we can divide the one day interval into pieces and use integrated variance estimators over these subintervals as well. If one could be satisfied with a spot variance estimator which is constant over those subintervals, these give estimators for the spot variance. We call these piecewise-constant spot variance estimators. The usual integrated variance estimator for a day is a special case — one just reports one volatility measurement for a day and uses it as a (constant) estimator for the spot variance. In this section, we use the Piecewise-Constant TSRV (PCTSRV) estimator for comparison purpose. The difference between the PCTSRV and the TSRSV is, the piecewise constant estimator is using non-overlapping subsamples, while our TSRSV estimators are using rolling subsamples.

We use equal-length subintervals for the PCTSRV estimator. That is, we divide the one day interval into equal-length intervals from the beginning of the day and we leave out the last interval when it is shorter than the subinterval length. For this estimator, one needs to choose the length of the subintervals h, as well as the scale parameter K. Since there is no established theory which can be used for these parameters, we use the bandwidth and scale parameter computed for the TSRSV estimator for the subinterval

length and the scale parameter, respectively.

4.2.4 Monte Carlo results

In Table 1, we give a list of the notations we used in the following tables. The Monte Carlo evaluated error measures, together with the corresponding Monte Carlo standard deviations are reported in Tables 2, 3, 4, 5, 6, and 7, each table corresponds to one of the 6 Monte Carlo scenarios.

We observe that all the RSV estimators are strictly dominated by TSRSV estimators in terms of all error measures. Of course we just choose 3 arbitrary sampling frequencies for the RSV estimators in our experiment, and there may exist a sampling frequency for these estimators to compete with our estimator. However, there is no established theory to select the sparse sampling frequency and investigating this problem is beyond the scope of this paper.

We also observe that the performance of the PCTSRV estimator is comparable to our estimators. Looking further into the results, we find that in almost all scenarios and for all error measures, the smoothing version of TSRSV gives the smallest error, with one exception for the MISRE in SV2F3, where the filtering version of TSRSV is the best. The relative performance of the filtering version of TSRSV and the PCTSRV estimator is not clear: for the SV1F1 scenario, the piecewise-constant estimator seems better; while for the SV1F3 scenario, the filtering version of TSRSV is better; the evidence for the SV1F2 scenario varies for different error measures; for the 3 scenarios in the SV2F model, the PCTSRV estimator is always better. Overall, the smoothing version of TSRSV is better than the PCTSRV estimator, which in turn seems to be better than the filtering version of TSRSV in more scenarios. But we remark that this comparison seems to be not fair for the filtering version of the TSRSV estimator, as it only uses lagged observations.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

We report the cross-validation selected bandwidths for the sparse-sampling based method and the scale/bandwidth parameters selected by our plug-in method described in Section 3.4. There is one set of such parameters being selected in each Monte Carlo sample, we report the means and variances of the 10000 realizations of these parameters in Tables 8, 9, 10, 11, 12 and 13 — each table corresponds to one of the six simulation scenarios. We observe that as the variance of the market microstructure noise increases, the selected scale parameter and the bandwidth parameters increase; and the SV2F model in general induces higher scale and bandwidth parameters than the SV1F model. For the three scenarios in the SV1F model, the selected scale parameters have mean values 8, 26 and 56 respectively; while for the three scenarios in the SV2F model the selected scale parameters have mean values 28, 39 and 57, respectively. For the three scenarios in the SV1F model, the selected bandwidth parameters have mean values corresponding to 27 minutes, 49 minutes and 71 minutes, respectively; while for the three scenarios in the SV2F model, the selected bandwidth parameters have mean values corresponding to 39 minutes, 60 minutes and 75 minutes, respectively.

[Table 8 about here.]

[Table 9 about here.]

[Table 10 about here.]

[Table 11 about here.]

[Table 12 about here.]

[Table 13 about here.]

5 Empirical applications

In this section, we give two empirical examples for possible applications of the TSRSV estimator. In the first example, we use the estimator to study the intraday variation of the volatility process. In the second example, the estimator is applied to calculating high-frequency Value-at-Risk.

The data we use are Euro FX Futures 1-second data obtained from tickdata.com. Euro FX Futures are currency futures contracts traded online at the Chicago Mercantile Exchange (CME) group. The contracts are traded 23 hours a day, a usual trading day starts from 17:00 of a day and ends at 16:00 of the next day. A trading week consists of 5 trading days, and a usual trading week starts from 17:00 on Sunday and ends at 16:00 on Friday. There are some special days, such as public holidays, when trading ends earlier.

5.1 Intraday variation of volatility

The spot volatility estimator can be used to study the intraday variation of volatility. Intraday variation of volatility was previously studied mostly with parametric approaches (e.g. Andersen and Bollerslev (1998), Taylor and Xu (1997), Boudt et al. (2011) and Bos et al. (2012)), where it is usually assumed that there exist a periodic part in the volatility process and one needs to take averages over many days (say, a month) to extract the periodic part. Although the existence of a periodic term in the volatility process is evident most of the time, we remark that this assumption is questionable when the volatility process undergoes a regime shift. Our approach is free of parametric assumptions and we need not assume the existence of the periodic part in the volatility process, nor do we need any specific form for it. As compared to the parametric approach of estimating the intraday variation of volatility, our estimator performs local averages of squared high-frequency returns around the time of interest, which we believe will better reflect the local variation of the volatility process. Our assumption on the volatility process allows for nonstationarity, structural change, periodicity and persistence in the volatility process.

To illustrate the method, we choose 4 weeks of 1-second data from 11 October, 2009 to 6 November, 2009. By preliminary analysis through the volatility signature plots for the first four days (Figure 1), and the autocorrelation function (ACF) plot (Figure 2) of high frequency returns, we confirm the existence of market microstructure noise (from significant first-order autocorrelation) and that the noise is dependent across time for 1-second high-frequency returns (because of non-zero higher-order autocorrelation). The estimated noise standard deviation is around 2×10^{-5}, which is small in magnitude as compared to the signal. From the ACF plots, we can identify that the market microstructure noise may be considered serially independent at the 1-minute frequency, and we can assume there is no noise at the 3-minute frequency.

[Figure 1 about here.]

[Figure 2 about here.]

We then apply the TSRSV and the PCTSRV estimator to the 1-minute returns, for which we can assume i.i.d. noise; and we apply the RSV estimator to the 3-minute returns, for which we can assume the absence of noise. We still use the cross-validation bandwidth for the RSV estimator and we use the plug-in method to select bandwidth/scale parameters for the TSRSV estimator and the PCTSRV estimator. Both the RSV estimator and the TSRSV estimator used are the smoothing version. The selected scale parameter is two, which roughly corresponds to two-minute intervals in the current context; and the selected bandwidth is around 0.11, which roughly corresponds to 2.5 hours for one-minute returns. The cross-validation method selects a 2 hours bandwidth for 3-minute returns. The piecewise constant estimator uses the same bandwidth and scale parameter

as TSRSV. To save space, we give the plots of the estimation results for the first 4 days in Figure 3. We also give a TSRV estimate for the daily integrated variance, which is viewed as a constant spot variance estimate for the whole day.

[Figure 3 about here.]

From the first 4 days, we can see some diurnal patterns in the spot variance. Usually there are 4 peaks of high volatility during one day, which may correspond to the morning and afternoon active trading hours in Asia, Europe and the US. We also note that the differences between the Realized Spot Variance estimator and the TSRSV estimator, as well as the piecewise constant estimator, are small. This may be due to the fact that the magnitude of noise is small for this example. However, we still recall from the Monte Carlo evidence in the previous section that the TSRSV is more accurate than the RSV and the PCTSRV estimator in terms of the MISE performance in most of the realistic scenarios. We also observe that the TSRV estimator is roughly the average of the spot variance estimators.

Finally, we give a plot of the estimated volatility path of a day with the 95% (pointwise) confidence interval in Figure 4, the confidence interval is constructed by the method described in Remark 1. Although the convergence rate of the estimator is slow, the confidence intervals seem not very wide in the current example.

[Figure 4 about here.]

5.2 Intraday Value-at-Risk based on spot volatility measures

In this section, we perform a simple exercise of predicting the Value-at-Risk (VaR) of 5-minute returns to illustrate the potential use of the spot volatility estimators in financial risk management. The dataset used here is the same as in Section 5.1.

At 5-minute level, the effect of market microstructure noise is negligible, so the observed returns can be assumed to follow the model

$$r_t = \sigma_t z_t, \tag{7}$$

where z_t is a standard normal variable, and σ_t is the spot volatility. The α-VaR in the next period is $-\sigma_t \Phi^{-1}(\alpha)$ for long positions and $\sigma_t \Phi^{-1}(1-\alpha)$ for short positions, where $\Phi^{-1}(.)$ is the quantile function of the standard normal distribution. The spot volatility σ_t can be estimated with different methods discussed earlier in this paper, for example the TSRSV, the RSV or the TSRV estimator. The TSRV estimator is used as a constant estimator for the spot volatility path of a whole day. Since we are performing a prediction exercise here, we include the filtering version the TSRSV estimator and the RSV estimator; for the TSRV estimator, we use the estimate from the previous day. In

the following, we give results by weeks to simplify the presentation. We have 20 trading days under investigation, so we have in total 4 weeks in the dataset.

Before performing the VaR prediction, we first give density estimates of the standardized returns $r_t/\widehat{\sigma}_t$. We compare the estimated density to a standard normal density; the closer to the standard normal, the better the model in (7) will provide an approximation. We give the results in Figure 5, which contains a 4 by 3 plot. Each row is a week and has 3 subplots which are the density estimates of 5-minute returns standardized by the TSRSV estimator, the RSV estimator and the TSRV estimator, respectively. We observe that the returns standardized by the TSRV is always fat-tailed, which just reflects that the 5-minute returns are fat-tailed. The standardized returns by the TSRSV and the RSV estimator are closer to the standard normal density, this is because time-varying volatility is able to model the fat-tailedness in data to a certain degree, if not all.

[Figure 5 about here.]

Now we perform the VaR prediction, and compare the 3 methods by the prediction performance. We predict the 1%, 5%, 95% and 99% VaR of 5-minute returns. The performance measure we use is the so called conditional coverage probability, which is the empirical frequency that the failure of VaR prediction happens: for long positions, this means that the realized return is smaller than the predicted VaR; for short positions, this means the realized return is larger than the predicted VaR. Theoretically, the conditional coverage probability needs be the same as the theoretical level α of VaR in the left tail, and $(1-\alpha)$ in the right tail. In practice the frequencies will differ from the theoretical value, and the closer to the theoretical value, the more accurate is the VaR.

We give the conditional coverage probability for the left tail and the right tail of the return distribution in Table 14 and Table 15, respectively. From the result we see that using an intraday measure, either the TSRSV estimator or the RSV estimator, will bring great improvement for the prediction performance as compared to the TSRV estimator. This again justifies the necessity of spot variance estimation. The RSV seems to be always a little better than the TSRSV estimator in terms of conditional coverage probability. However, we comment that this does not in particular mean the RSV is better than the TSRSV: the fact that the conditional coverage probabilities are larger than the theoretical values for both estimators means that the extreme returns in the high-frequency returns are not well described by either method. That the RSV gives a smaller conditional coverage probability is probably due to the fact that the RSV estimator provides a more conservative volatility estimate than the TSRSV estimator — actually this is exactly the case: the definition of the TSRSV estimator involves averaging the RSV over subsamples and making a further downward correction caused by the noise. We also note that the differences between TSRSV and RSV are small, which is actually consistent with the similarity in the path estimation result shown earlier in this section.

[Table 14 about here.]

[Table 15 about here.]

We remark that the result here does not contradict with the Monte Carlo evidence that TSRSV is more accurate than the RSV in estimating the realized volatility path, because there are also model uncertainties here in the VaR prediction.

6 Extensions and discussions

6.1 Multi-scale realized spot volatility estimator

In the discussion of the convergence rate after Theorem 3, we noted that the convergence rate of the TSRV estimator is playing a role in determining the convergence rate of the TSRSV estimator. Since the Multi-Scale Realized Variance (MSRV) estimator by Zhang (2006) converges to the Integrated Variance at an optimal rate $n^{-1/4}$, we conjecture that constructing a spot variance estimator based on the MSRV estimator will improve the convergence rate. In this section, we discuss this issue.

Analogous to the construction of the TSRSV estimator, a filtering version of the estimator can be constructed as:

$$\text{MSRSV}_{\text{t}} = \frac{\text{MSRV}_t - \text{MSRV}_{t-h}}{h}, \tag{8}$$

where

$$\text{MSRV}_{\text{t}} = \sum_{i=1}^{M} a_i \left[Y,Y\right]^{K_i} + 2\widehat{\omega^2},$$

where $\widehat{\omega^2}$ is given as in (3) and $a_i, i = 1, \ldots, M$ are the weights satisfying

$$a_i \;=\; \frac{1}{M^2} h\left(\frac{i}{M}\right) - \frac{1}{2M^2}\frac{i}{M} h'\left(\frac{i}{M}\right),$$

with $h\left(.\right)$ a continuously differentiable real-valued function satisfying

$$\int_0^1 h\left(x\right) \mathrm{d}x = 0, \quad \int_0^1 x h\left(x\right) \mathrm{d}x = 1.$$

The MSRV estimator is defined as a weighted sum of the subsampled and averaged RV estimators over M different scales $K_1, K_2, \ldots, K_M$.

Using the notation defined in Section 3, the MSRSV_t estimator can be written explicitly as

$$\text{MSRSV}_{\text{t}} = \sum_{i=1}^{M} a_i \left[Y,Y\right]^{\prime K_i,h}.$$

The MSRSV estimator can be decomposed as follows:

$$\begin{aligned}
\mathrm{MSRSV_t} - \sigma_t^2 &= \sum_{i=1}^{M} a_i \left[Y,Y\right]^{'K_i,h} - \sigma_t^2 \\
&= \left(\sum_{i=1}^{M} a_i \left[Y,Y\right]^{'K_i,h} - \frac{1}{h}\int_{t-h}^{t} \sigma_s^2 \mathrm{d}s\right) + \left(\frac{1}{h}\int_{t-h}^{t} \sigma_s^2 \mathrm{d}s - \sigma_t\right) + 2\widehat{\omega^2} \\
&= \left(\sum_{i=1}^{M} a_i \left[X,X\right]^{'K_i,h} - \frac{1}{h}\int_{t-h}^{t} \sigma_s^2 \mathrm{d}s\right)\left(\sum_{i=1}^{M} a_i \left[\varepsilon,\varepsilon\right]^{'K_i,h} + 2\omega^2\right) \\
&\quad + \left(\sum_{i=1}^{M} a_i \left[X,\varepsilon\right]^{'K_i,h}\right) + \left(\frac{1}{h}\int_{t-h}^{t} \sigma_s^2 \mathrm{d}s - \sigma_t\right) + O_p\left(n^{-1/2}\right) \\
&=: A + B + C + D + O_p\left(n^{-1/2}\right),
\end{aligned}$$

where A, B, C, D are defined implicitly.

Define the end term effects of the noise as

$$B_2 = -\frac{1}{h}\sum_{i=1}^{M} \frac{\alpha_i}{i}\left(\sum_{j=0}^{i-1} \varepsilon_{t_j}^2 + \sum_{j=nh-i+1}^{nh} \varepsilon_{t_j}^2\right) + 2\omega^2,$$

and $B_1 := B - B_2$, then the B term can be further decomposed into $B_1 + B_2$. It can be calculated that

$$\begin{aligned}
A &= O_p\left(\sqrt{\frac{M}{nh}}\right), \qquad B_1 = O_p\left(\sqrt{\frac{n}{M^3 h}}\right), \qquad B_2 = O_p\left(\sqrt{\frac{1}{Mh^2}}\right), \\
C &= O_p\left(\sqrt{\frac{1}{Mh}}\right), \qquad D = O_p\left(h^{1/2}\right),
\end{aligned}$$

where in B, the second term comes from the end effects.

Balancing these stochastic terms we find that by taking $M = O\left(n^{3/5}\right)$ and $h = O\left(n^{-1/5}\right)$, the convergence rate of the estimator can be $n^{-1/10}$, which is an improvement over the TSRSV rate $n^{-1/12}$. We also find that taking $M = O\left(n^{1/2}\right)$ can not improve the convergence rate, as in this case, A, C and B_1 are always dominated by B_2. Equating B_2 with D leads to $h = n^{-1/6}$ and the convergence rate of the estimator will remain at $n^{-1/12}$.

When $M = O\left(n^{3/5}\right)$ and $h = O\left(n^{-1/5}\right)$, it can be shown that

$$n^{1/10}\left(\mathrm{MSRSV_t} - \sigma_t^2\right) \xrightarrow{\mathrm{st}} \eta Z,$$

where

$$\eta^2 = \frac{4}{3}\sigma_t^4 \int_0^1 \mathrm{d}x \int_0^x h(y)\, h(x)\, y^2 (3x - y)\, \mathrm{d}y + 4 \operatorname{var}\left(\varepsilon^2\right) \int_0^1 \int_0^y x h(x)\, h(y)\, \mathrm{d}x \mathrm{d}y + \frac{1}{3}\Lambda_t^2,$$

and Z is a standard normal variable. The three terms in the asymptotic variance come from A, B_2 and D respectively, while $B_1 = O_p\left(n^{-3/10}\right)$ and $C = O_p\left(n^{-1/5}\right)$ are smaller order terms.

We see that by constructing an MSRSV estimator and using the same weighting scheme as in Zhang (2006), we do get improvement in the convergence rate of the estimator to $n^{-1/10}$. However, with this simple adaptation, a best convergence rate $n^{-1/8}$ is not obtainable. This is due to the end term effects B_2 blowing up too fast when localizing the MSRV estimator and dominating several other terms.

Constructing a multi-scale estimator which can reach the $n^{-1/8}$ rate is an interesting future research topic. Dealing with end effects will be crucial in the construction of such an estimator.

6.2 Dependent noise

We assume i.i.d. market microstructure noise, which is not realistic from an empirical point of view, where it is believed that the market microstructure noise is dependent over time (see Hansen and Lunde (2006)).

When the noise is weakly dependent across time, Aït-Sahalia et al. (2011) show that the classical TSRV estimator is no longer consistent for the integrated variance, they proposed an adjusted TSRV estimator, which is robust to the dependent noise and is consistent for the integrated variance. Intuitively, in the two-scale scheme correction for noise, the fast scale 1 is replaced by a slower scale J, by letting J, K both converge to ∞, a consistent estimator for IV is obtained.

Based on their TSRV estimator, the following TSRSV estimator can be defined:

$$\hat{V}_t^{K,J,h} = [Y,Y]_t^{\prime K,h} - \frac{\bar{n}_K}{\bar{n}_J}[Y,Y]_t^{\prime J,h},$$

where $\bar{n}_i := (nh - i + 1)/(ih)$, where $i = K, J$. We conjecture that this estimator is robust to dependent noise.

On the other hand, since the MSRV estimator is robust to dependent noise (Aït-Sahalia et al. (2011)), we conjecture that the MSRSV estimator we defined in the previous section should also display this robustness.

6.3 Endogenous noise

The market microstructure noise is assumed to be exogenous and homoscedastic in this paper. From Hansen and Lunde (2006) however, the empirical market microstructure noise seems to be endogenous and heteroscedastic. In the existing literature for estimating integrated variance, there is no treatment for general form of endogenous and heteroscedastic noise. Usually estimators are tuned to be robust to a certain specific type of endogeneity and heteroscedasticity. For example, Kalnina and Linton (2008)'s jitterred TSRV is robust to the following model for the market microstructure noise:

$$\begin{aligned} u_{t_i} &= v_{t_i} + \varepsilon_{t_i}, \\ v_{t_i} &= \delta\gamma_n \left(W_{t_i} - W_{t_{i-1}}\right), \\ \varepsilon_{t_i} &= m\left(t_i\right) + n^{-\alpha/2}\omega\left(t_i\right)\epsilon_{t_i}, \end{aligned}$$

where the noise $\{u_{t_i}\}$ is both endogenous and heteroscedastic. They impose the restriction that $\alpha \in [0, 1/2]$ and $\gamma_n^2 = n^{1-\alpha}$, such that the variance of the noise $\{u_{t_i}\}$ converges to 0 as $n \to \infty$, which they call a small noise assumption. Barndorff-Nielsen et al. (2008)'s realized kernel estimator is robust to the following linear model of endogeneity

$$u_{t_i} = \sum_{h=0}^{\bar{H}} \beta_h \left(Y_{t_i} - Y_{t_{i-1}}\right) + \bar{U}_{t_i},$$

where β_h's are constants, $\bar{U}$ is independent of the Y process, and $\bar{U}$ is a white noise process. It can be expected that by defining a spot volatility estimator based on the jittered TSRV, the endogeneity and heteroscedasticity of the specific form can be allowed.

In the context of spot volatility estimation, Ogawa and Sanfelici (2011) proposed a two-step regularized spot volatility estimator. Their estimator converges in the first absolute moment under a general form of endogeneity, however it is not clear if the robustness will be preserved if the convergence in distribution results are needed as in this paper.

6.4 Jumps

If there is a finite activity jump component in the volatility process $\{\sigma_t^2\}$, the theories developed in Section 3 are only valid at the points of continuity. For discontinuous points, we introduce the following two estimators:

$$\begin{aligned} \mathrm{TSRSV}_{\mathrm{t},+} &= \frac{\mathrm{TSRV}_t - \mathrm{TSRV}_{t+h}}{h}, \\ \mathrm{TSRSV}_{\mathrm{t},-} &= \frac{\mathrm{TSRV}_t - \mathrm{TSRV}_{t-h}}{h}. \end{aligned}$$

The notations are self-explanatory as the two estimators use data after the time t and before the time t, respectively. The two estimators are both consistent at continuity points. Now, let τ to be a jump point with jump size $J\tau$, it should hold that

$$\begin{aligned} \text{TSRSV}_{\tau,+} &\xrightarrow{p} \sigma_\tau^2 + J_\tau. \\ \text{TSRSV}_{\tau,-} &\xrightarrow{p} \sigma_\tau^2. \end{aligned}$$

By comparing the spot volatility estimators for a certain time point using data before and after the point, a jump in the volatility process is identified if the difference is big, and the jump size can be estimated by the difference of the two estimators. Gijbels et al. (2007) and Kristensen (2010) proposed a strategy to deal with jump points. Their estimator is defined based on a so-called Residual Sum of Squares (RSS) of a local linear estimator, however, the RSS is not directly available for the model in this paper, as the objective function $\{\sigma_t^2\}$ to be estimated is not observable. Adapting their strategy to our model will be an interesting extension.

Jumps in the log price process is difficult to deal with using the subsampling estimator. However, for practical purposes, by sampling sparsely, say at the 5-minute interval, one could use the estimator as defined in Equation (21) in Kristensen (2010), which is robust to jumps in the price process.

7 Conclusion

Motivated by the Two Scale Realized Variance estimator and the definition of spot variance, this paper develops a nonparametric estimator for spot variance with high-frequency financial data, which are contaminated by the market microstructure noise. Consistency of the estimator is shown, and different asymptotic distributions are derived. In particular, this paper proposes a data-driven plug-in method to select the scale and bandwidth parameters. In financially realistic scenarios, we conduct Monte Carlo experiments to study the finite sample properties of our estimator. Our estimator is applied to two empirical examples. Possible extensions of the model and the estimator are discussed.

Acknowledgment

The authors wish to thank the co-editor Yacine Aït-Sahalia and two anonymous referees for helpful comments, which greatly improved the paper. They also wish to thank Peter Hansen, Shin Kanaya, Bas Klein, Dennis Kristensen, Per Mykland, Roel Oomen, Neil Shephard, Kevin Sheppard, Georg Strasser, Aad van der Vaart, Bert van Es, Almut Veraart, Lan Zhang, and seminar and conference participants at the Tinbergen Institute, the Humboldt-Copenhagen 2009 Conference on Recent Developments in Financial Econo-

metrics, the 64th European Meeting of the Econometric Society, the 20th EC2 conference on Real Time Econometrics, and the Oxford-Man Institute of Quantitative Finance for helpful comments, discussions and suggestions. All remaining errors are our own. Part of the research was conducted while Yang Zu was visiting the Oxford-Man Institute of Quantitative Finance supported by a C. Willems Stichting travel grant; the institute's hospitality and the C. Willems Stichting's financial support are gratefully acknowledged.

References

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

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

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (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. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*, 885–905.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2010). Parametric and nonparametric volatility measurement. In Y. Aït-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*, Volume 1, pp. 67–137. Amsterdam: North Holland.

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

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2009). Duration-based volatility estimation. Global COE Hi-Stat Discussion Paper Series gd08-034, Institute of Economic Research, Hitotsubashi University.

Andreou, E. and E. Ghysels (2002). Rolling-sample volatility estimators: Some new theoretical, simulation, and empirical results. *Journal of Business & Economic Statistics 20*, 363–76.

Bandi, F. M. and R. Reno (2009). Nonparametric stochastic volatility. Global COE Hi-Stat Discussion Paper Series gd08-035, Institute of Economic Research, Hitotsubashi University.

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

Barndorff-Nielsen, O. E. and N. Shephard (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.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Bos, C. S., P. Janus, and S. J. Koopman (2012). Spot variance path estimation and its application to high-frequency jump testing. *Journal of Financial Econometrics 10*, 354–389.

Boudt, K., C. Croux, and S. Laurent (2011). Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance 18*, 353–367.

Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance 8*, 291–323.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Fan, J. and Y. Wang (2008). Spot volatility estimation for high-frequency data. *Statistics and Its Interface 1*, 279–288.

Foster, D. P. and D. B. Nelson (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica 64*, 139–74.

Gijbels, I., A. Lambert, and P. Qiu (2007). Jump-preserving regression and smoothing using local linear fitting: a compromise. *Annals of the Institute of Statistical Mathematics 59*, 235–272.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application*. New York: Academic Press.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics 24*, 127–161.

Hoffmann, M., A. Munk, and J. Schmidt-Hieber (2010). Nonparametric estimation of the volatility under microstructure noise: wavelet adaptation. *Preprint. Available at arXiv 1007*.

Huang, X. and G. Tauchen (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics 3*, 456.

Jacod, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law. In J. Azéma, M. Yor, and M. Émery (Eds.), *Séminaire de Probabilités XXXI*, pp. 232–246. Berlin: Springer.

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

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

Kanaya, S. and D. Kristensen (2010). Estimation of stochastic volatility models by nonparametric filtering. CREATES Research Papers 2010-67, School of Economics and Management, University of Aarhus.

Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory 26*, 60–93.

Lee, S. S. and P. A. Mykland (2007). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies 21*, 2535–2563.

Mancini, C., V. Mattiussi, and R. Renò (2012). Spot volatility estimation using delta

sequences. DiMaD Working Papers 2012-10, Dipartimento di Matematica per le Decisioni, Università degli Studi di Firenze.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics 8*, 323–361.

Munk, A. and J. Schmidt-Hieber (2010a). Lower bounds for volatility estimation in microstructure noise models. *Borrowing Strength: Theory Powering Applications-A Festschrift for Lawrence D. Brown, IMS Collection 6*, 43–55.

Munk, A. and J. Schmidt-Hieber (2010b). Nonparametric estimation of the volatility function in a high-frequency model corrupted by noise. *Electronic Journal of Statistics 4*, 781–821.

Mykland, P. A. and L. Zhang (2008). Inference for volatility-type objects and implications for hedging. *Statistics and Its interface 1*, 255–278.

Ngo, H.-L. and S. Ogawa (2009). A central limit theorem for the functional estimation of the spot volatility. *Monte Carlo Methods and Applications 15*, 353–380.

Ogawa, S. and S. Sanfelici (2011). An improved two-step regularization scheme for spot volatility estimation. *Economic Notes 40*, 107–134.

O'Hara, M. (1998). *Market Microstructure Theory*. New York: Wiley.

Podolskij, M. and M. Vetter (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli 15*, 634–658.

Podolskij, M. and M. Vetter (2010). Understanding limit theorems for semimartingales: a short survey. *Statistica Neerlandica 64*, 329–351.

Revuz, D. and M. Yor (1998). *Continuous Martingales and Brownian Motion* (3rd ed.). Berlin: Springer.

Taylor, S. J. and X. Xu (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance 4*, 317–340.

Veraart, A. (2010). Inference for the jump part of quadratic variation of Itô semimartingales. *Econometric Theory 26*, 331–368.

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., P. A. Mykland, and Y. Aït-Sahalia (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 & Economic Statistics 14*, 45–52.

Appendix A: Proofs

Proof (of Proposition 1) This is essentially part (c) of Theorem 1 of Mykland and Zhang (2008).

$$\begin{aligned}
P_1 &= \frac{1}{h}\int_{t-h}^{t}\sigma_s^2\mathrm{d}s-\sigma_t^2 \\
&= \frac{1}{h}\int_{t-h}^{t}\left(\sigma_s^2-\sigma_t^2\right)\mathrm{d}s \\
&= \frac{1}{h}\left(\sigma_s^2-\sigma_t^2\right)|_{t-h}^{t}-\frac{1}{h}\int_{t-h}^{t}s\mathrm{d}\sigma_s^2 \\
&= \frac{1}{h}\int_{t-h}^{t}\left[(t-h)-s\right]\mathrm{d}\sigma_s^2 \\
&= \frac{1}{h}\int_{t-h}^{t}\left[(t-h)-s\right]\Gamma_s\mathrm{d}s+\frac{1}{h}\int_{t-h}^{t}\left[(t-h)-s\right]\Lambda_s\mathrm{d}B_s.
\end{aligned}$$

where in the third equality integration by parts is used, and in the last equality Assumption A5 is used.

For the first term,

$$\begin{aligned}
\frac{1}{h}\int_{t-h}^{t}\left[(t-h)-s\right]\Gamma_s\mathrm{d}s &= \frac{1}{h}\int_{t-h}^{t}\left[(t-h)-s\right]\mathrm{d}s\left(\Gamma_t+O_p\left(h\right)\right) \\
&= -\frac{1}{2}h\left(\Gamma_t+O_p\left(h\right)\right) \\
&= O_p\left(h\right).
\end{aligned}$$

For the second term, note that $\frac{1}{h}\int_{t-h}^{t}\left[(t-h)-s\right]\Lambda_s\mathrm{d}B_s$ is an increment of a martingale $\frac{1}{h}\int_{0}^{t}\left[(t-h)-s\right]\Lambda_s\mathrm{d}B_s$, and we will use Theorem A.2 of Mykland and Zhang (2008) to derive its asymptotic properties. To calculate the increment of the quadratic variation process, we first calculate the quadratic variation process of $\frac{1}{h}\int_{0}^{t}\left[(t-h)-s\right]\Lambda_s\mathrm{d}B_s$ itself, first denote $M_t=\frac{1}{h}\int_{0}^{t}\left[(t-h)-s\right]\Lambda_s\mathrm{d}B_s$, so that

$$\begin{aligned}
&\langle M,M\rangle_{(s-(t-h))h+(t-h)}-\langle M,M\rangle_{t-h} \\
= &\frac{1}{h^2}\int_{(t-h)}^{t-h+(s-(t-h))h}\left[(t-h)-u\right]^2\Lambda_u^2\mathrm{d}u \\
= &\frac{1}{h^2}\int_{(t-h)}^{t-h+(s-(t-h))h}\left[(t-h)-u\right]^2\mathrm{d}u\left(\Lambda_{t-h}^2+O_p\left(h\right)\right) \\
= &\frac{1}{h^2}\times\frac{1}{3}\left[u-(t-h)\right]^3|_{t-h}^{t-h+(s-(t-h))h}\left(\Lambda_{t-h}^2+O_p\left(h\right)\right) \\
= &\frac{1}{h^2}\times\frac{1}{3}\left(s-(t-h)\right)^3h^3\left(\Lambda_{t-h}^2+O_p\left(h\right)\right) \\
= &\frac{1}{3}h\times\left(s-t+h\right)^3\left(\Lambda_{t-h}^2+O_p\left(h\right)\right)
\end{aligned}$$

such that

$$h^{-1}\left(\langle M,M\rangle_{(s-(t-h))h+(t-h)}-\langle M,M\rangle_{t-h}\right)\xrightarrow{p}\frac{1}{3}\left(s-t\right)^{3}\Lambda_{t}^{2},\forall s\geqslant t,$$

so in the notation of Theorem A.2, $\alpha=1$, $f_t(s)=s^3$ and $\eta_t^2=\Lambda_t^2/3$, and it follows from Theorem A.2 that

$$\left(X_{u,0\leqslant u\leqslant t},h^{-1/2}\times\frac{1}{h}\int_{t-h}^{t}\left[(t-h)-s\right]\Lambda_{s}\mathrm{d}B_{s}\right)\xrightarrow{d}\left(X_{u,0\leqslant u\leqslant t},\frac{1}{\sqrt{3}}\Lambda_{t}\times Z\right),$$

where Z is a standard normal variable independent of the X process. □

Proof (of Proposition 2) This is a simple consequence of expanding the spot variance process up to the mth derivative and Assumption A5-1. □

Proof (of Proposition 3) Note that because of the additive noise specification, we have

$$[Y,Y]_{t}^{\prime K,h}=[X,X]_{t}^{\prime K,h}+[\varepsilon,\varepsilon]_{t}^{\prime K,h}+2[X,\varepsilon]_{t}^{\prime K,h}.$$

Note that this decomposition also holds for the case $K=1$. First,

$$\begin{aligned}
&[Y,Y]_{t}^{\prime K,h}-[X,X]_{t}^{\prime K,h}-\frac{\bar{n}}{n}[Y,Y]_{t}^{\prime h}\\
=&\left([Y,Y]_{t}^{\prime K,h}-[X,X]_{t}^{\prime K,h}-2\bar{n}\omega^{2}\right)-2\bar{n}\left(\frac{1}{2n}[Y,Y]_{t}^{\prime h}-\omega^{2}\right)\\
=&\left([\varepsilon,\varepsilon]_{t}^{\prime K,h}-2\bar{n}\omega^{2}+2[X,\varepsilon]_{t}^{\prime K,h}\right)\\
&-\frac{\bar{n}}{n}\left([\varepsilon,\varepsilon]_{t}^{\prime h}-2n\omega^{2}+[X,X]_{t}^{\prime h}+2[X,\varepsilon]_{t}^{\prime h}\right)\\
=&\left([\varepsilon,\varepsilon]_{t}^{\prime K,h}-2\bar{n}\omega^{2}\right)-\frac{\bar{n}}{n}\left([\varepsilon,\varepsilon]_{t}^{\prime h}-2n\omega^{2}\right)\\
&+\left(2[X,\varepsilon]_{t}^{\prime K,h}-\frac{\bar{n}}{n}[X,X]_{t}^{\prime h}-2\frac{\bar{n}}{n}[X,\varepsilon]_{t}^{\prime h}\right).
\end{aligned}$$

From Lemma 1 below,

$$\sqrt{\frac{h}{n}}K\left(\begin{array}{c}[\varepsilon,\varepsilon]_{t}^{\prime K,h}-2\bar{n}\omega^{2}\\ \frac{\bar{n}}{n}\left([\varepsilon,\varepsilon]_{t}^{\prime h}-2n\omega^{2}\right)\end{array}\right)\xrightarrow{p}N\left(\left(\begin{array}{c}0\\0\end{array}\right),\left(\begin{array}{cc}4E\varepsilon^{4} & 4\,\mathrm{Var}(\varepsilon^{2})\\ 4\,\mathrm{Var}(\varepsilon^{2}) & 4E\varepsilon^{4}\end{array}\right)\right).$$

Because ε is independent of X, this convergence in distribution still holds conditionally on the X process.

For the quantities in the last bracket, using Lemma 2 below we know that conditional on the X process, $[X,\varepsilon]_{t}^{\prime K,h}=O_{p}\left(1/\sqrt{Kh}\right)$ and $[X,\varepsilon]_{t}^{\prime h}=O_{p}\left(1/\sqrt{h}\right)$, and note that $[X,X]_{t}^{\prime h}=O_{p}\left(1\right)$ because this is essentially a spot volatility estimator with clean data, such as that in Mykland and Zhang (2008) and Kristensen (2010), so that the last bracket is of order $o_{p}\left(\sqrt{n/(hK^{2})}\right)$ conditional on the X process. So the claimed conditional

convergence in distribution holds. □

Proof (of Proposition 4) First note that a similar quantity in Zhang et al. (2005), $[X,X]_t^{K,h}$ has the following decomposition

$$\begin{aligned}
[X,X]_t^{\prime K,h} &= [X,X]_t^{\prime h} + 2\frac{1}{h}\sum_{t-h\leqslant t_i\leqslant t}(\Delta X_{t_i})\sum_{j=1}^{K\wedge i}\left(1-\frac{j}{K}\right)(\Delta X_{t_{i-j}}) + O_p\left(\frac{K}{nh}\right) \\
&= \frac{1}{h}\int_{t-h}^{t}\sigma_s^2\mathrm{d}s + O_p\left(\frac{1}{\sqrt{nh}}\right) + 2\frac{1}{h}\sum_{t-h\leqslant t_i\leqslant t}(\Delta X_{t_i})\sum_{j=1}^{K\wedge i}\left(1-\frac{j}{K}\right)(\Delta X_{t_{i-j}}) \\
&\quad + O_p\left(\frac{K}{nh}\right),
\end{aligned}$$

where we use $[X,X]_t^{\prime h} = \int_{t-h}^{t}\sigma_s^2\mathrm{d}s/h + O_p\left(1/\sqrt{nh}\right)$, which can be found in for example, Theorem 1 in Mykland and Zhang (2008) or Theorem 3 in Kristensen (2010). So if we denote

$$D_t^{K,h} = \frac{2}{h}\sum_{t-h\leqslant t_i\leqslant t}(\Delta X_{t_i})\sum_{j=1}^{K\wedge i}\left(1-\frac{j}{K}\right)(\Delta X_{t_{i-j}}),$$

we can write

$$\begin{aligned}
P_3 &= [X,X]_t^{\prime K,h} - \frac{1}{h}\int_{t-h}^{t}\sigma_s^2\mathrm{d}s \\
&= D_t^{K,h} + O_p\left(\frac{1}{\sqrt{nh}}\right) + O_p\left(\frac{K}{nh}\right).
\end{aligned}$$

In the following, we will show that

$$\sqrt{\frac{nh}{K}}D_t^{K,h} \xrightarrow{\text{st}} N\left(0, \frac{4}{3}\times\sigma_t^4\right), \tag{9}$$

as $nh/K \to \infty$, for which we need to use Theorem 2.2 in Mykland and Zhang (2008). Observe that the claim of the proposition holds easily with (9). Note that Mykland and Zhang (2008)'s theorem is in particular needed for the present purpose, because the usual stable convergence theorems for stochastic processes, such as Theorem 2.6 of Podolskij and Vetter (2010), see also Jacod (1997), only work for fixed intervals and are not suited

for the shrinking interval situation here. Note that

$$\begin{aligned} D_t^{K,h} &= \frac{2}{h}\sum_{t-h\leqslant t_i\leqslant t}(\Delta X_{t_i})\sum_{j=1}^{K\wedge i}\left(1-\frac{j}{K}\right)(\Delta X_{t_{i-j}}) \\ &= \frac{2}{h}\sum_{0\leqslant t_i\leqslant t}(\Delta X_{t_i})\sum_{j=1}^{K\wedge i}\left(1-\frac{j}{K}\right)(\Delta X_{t_{i-j}}) \\ &\quad -\frac{2}{h}\sum_{0\leqslant t_i\leqslant t-h}(\Delta X_{t_i})\sum_{j=1}^{K\wedge i}\left(1-\frac{j}{K}\right)(\Delta X_{t_{i-j}}), \end{aligned}$$

which is the increment of the continuous martingale

$$M_t^n := \sum_{0\leqslant t_i\leqslant t}(\Delta X_{t_i})\sum_{j=1}^{K\wedge i}\left(1-\frac{j}{K}\right)(\Delta X_{t_{i-j}}). \tag{10}$$

To prove the limit distribution of $D_t^{K,h}$, we calculate the increment of the quadratic variation of M_t^n with step size $\lfloor s-(t-h)\rfloor h$. By Lemma 3 below,

$$\begin{aligned} & \frac{nh}{K}\left(\langle M^n,M^n\rangle_{\lfloor s-(t-h)\rfloor h+(t-h)}-\langle M^n,M^n\rangle_{t-h}\right) \\ = & \frac{h}{([s-(t-h)]h)^2}\sum_{\lfloor s-(t-h)\rfloor h+(t-h)\leqslant t_i\leqslant t-h}\sigma_{t_i}^4 g_i\Delta t(1+o_p(1)) \\ \to & \frac{1}{s-t}\times\frac{4}{3}\sigma_t^4, \end{aligned}$$

as $h\to 0$.

Because of Assumption A5, the filtration we are working with is generated by a finite number of Brownian motions, so applying Theorem A.3 by Mykland and Zhang (2008), we get

$$\sqrt{\frac{nh}{K}}D_t^{K,h}\xrightarrow{d}\sqrt{\frac{4}{3}\sigma_t^4}Z,$$

jointly with the process $(X_{u,0\leqslant u\leqslant 1})$, and where Z is a standard normal variable independent of the X process, which is also called mixing-global convergence. As discussed in Remark 1 in Section 2.2 of the same paper, mixing convergence is the same as stable convergence with the σ-algebra generated by the X process in the present context, so the proposition is proved. □

Proof (of Theorem 1) From the results of Propositions 1–4, we know that P_1, P_2 and P_3 are $o_p(1)$ when either A5 or A5-1 holds and $n/(K^2h)\to 0$; this implies the consistency. □

Proof (of Theorem 2) By Proposition 1 and Proposition 4, P_1 and P_3 converge to mixed normal distributions stably, respectively. The joint convergence in distribution of P_1 and

P_3 follows easily; while for the covariance terms, we can choose Z^{P_1} and Z^{P_3} such that they are independent of each other, and at the same time independent of the σ-algebra generated by the X process, so $P_1 + P_3$ converges stably to the sum of two mixed normal distributions. For the joint convergence between $P_1 + P_3$ and P_2, the covariance can be shown to be asymptotically zero using a conditional argument, and note the fact that each item is conditionally of mean zero. Then the stable convergence of $P_1 + P_2 + P_3$ follows. The properties of stable convergence ensure that this convergence in distribution is stable with respect to the σ-algebra generated by the X process. □

Proof (of Theorem 3) This is essentially the same as Theorem 2, other than that the bias part P_1 vanishes under the choice of the order of the bandwidth. □

Proof (of Theorem 4) This is a simple application of the delta method. □

Appendix B: Technical lemmas

The following are lemmas that are used in the proofs of the propositions and theorems.

Lemma 1 *Let ε_{t_i} be independent and identically distributed with mean zero and variance ω^2, assume further that $E|\varepsilon|^{4+\delta} < \infty$ for some $\delta > 0$. Let K a sequence of positive integers satisfying $K/(nh) \to 0$ as $nh \to \infty$, then we have*

$$\begin{pmatrix} \sqrt{\frac{h}{n}}\left([\varepsilon,\varepsilon]_t'^{h} - 2n\omega^2\right) \\ \sqrt{\frac{h}{n}}K\left([\varepsilon,\varepsilon]_t'^{K,h} - 2\bar{n}\omega^2\right) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4E\varepsilon^4 & 4\,\mathrm{Var}(\varepsilon^2) \\ 4\,\mathrm{Var}(\varepsilon^2) & 4E\varepsilon^4 \end{pmatrix}\right).$$

Proof First note that

$$\begin{aligned} & \sqrt{\frac{h}{n}}K\left([\varepsilon,\varepsilon]_t'^{K,h} - 2\bar{n}\omega^2\right) \\ = & \frac{1}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}(\varepsilon_{t_i}^2-\omega^2) + \frac{1}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+1}^{\lfloor tn\rfloor-K}(\varepsilon_{t_i}^2-\omega^2) - \frac{2}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}\varepsilon_{t_i}\varepsilon_{t_{i-K}} \\ = & \frac{2}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}(\varepsilon_{t_i}^2-\omega^2) - \frac{2}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}\varepsilon_{t_i}\varepsilon_{t_{i-K}} + R \\ = & \frac{2}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}(\varepsilon_{t_i}^2-\omega^2) - \frac{2}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}\varepsilon_{t_i}\varepsilon_{t_{i-K}} + O_p\left(\sqrt{\frac{K}{nh}}\right). \end{aligned}$$

The remainder term is given by

$$R := \frac{1}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+1}^{\lfloor (t-h)n\rfloor+K}(\varepsilon_{t_i}^2-\omega^2) - \frac{1}{\sqrt{nh}}\sum_{i=\lfloor tn\rfloor-K+1}^{\lfloor tn\rfloor}(\varepsilon_{t_i}^2-\omega^2),$$

and it is easy to show that $R = O_p\left(\sqrt{K/(nh)}\right)$ because $ER^2 = \mathrm{Var}(R) \leq 2\frac{K}{nh}\mathrm{Var}(\varepsilon^2)$.

Similarly, we can show that

$$\begin{aligned}
&\sqrt{\frac{h}{n}}\left([\varepsilon,\varepsilon]_t'^h - 2n\omega^2\right) \\
= \;& \frac{2}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+2}^{\lfloor tn\rfloor}(\varepsilon_{t_i}^2-\omega^2) - \frac{2}{\sqrt{nh}}\sum_{t_i=\lfloor (t-h)n\rfloor+2}^{\lfloor tn\rfloor}\varepsilon_{t_i}\varepsilon_{t_{i-1}} + O_p\left(\sqrt{\frac{1}{nh}}\right).
\end{aligned}$$

To save notations, define

$$\begin{aligned}
M_1 &= \frac{1}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+2}^{\lfloor tn\rfloor}(\varepsilon_{t_i}^2-\omega^2), \\
M_2 &= \frac{1}{\sqrt{nh}}\sum_{t_i=\lfloor (t-h)n\rfloor+2}^{\lfloor tn\rfloor}\varepsilon_{t_i}\varepsilon_{t_{i-1}}, \\
M_3 &= \frac{1}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}\varepsilon_{t_i}\varepsilon_{t_{i-K}}.
\end{aligned}$$

Note that

$$\frac{1}{\sqrt{nh}}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}(\varepsilon_{t_i}^2-\omega^2) = M_1 + o_p(1),$$

in view of $K/(nh)\to 0$.

Therefore we have

$$\begin{pmatrix} \sqrt{\frac{h}{n}}\left([\varepsilon,\varepsilon]_t'^h - 2n\omega^2\right) \\ \sqrt{\frac{h}{n}}K\left([\varepsilon,\varepsilon]_t'^{K,h} - 2\bar{n}\omega^2\right) \end{pmatrix} = \begin{pmatrix} 2(M_1-M_3) \\ 2(M_1-M_2) \end{pmatrix} + o_p(1).$$

In the following, we show that jointly

$$\begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix} \xrightarrow{d} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathrm{Var}(\varepsilon^2) & 0 & 0 \\ 0 & \omega^4 & 0 \\ 0 & 0 & \omega^4 \end{pmatrix}\right), \tag{11}$$

and it follows that

$$\begin{pmatrix} 2(M_1-M_3) \\ 2(M_1-M_2) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4E\varepsilon^4 & 4\,\mathrm{Var}(\varepsilon^2) \\ 4\,\mathrm{Var}(\varepsilon^2) & 4E\varepsilon^4 \end{pmatrix}\right),$$

and the result of the lemma follows easily.

Now we show (11). Note that M_1, M_2 and M_3 are all sums of martingale difference sequences, we can apply the (multivariate version of the) central limit theorem of Hall

and Heyde (1980), Corollary 3.1. By the assumption $E|\varepsilon|^{4+\delta} < \infty$ for some $\delta > 0$, the conditional Lindeberg's condition is easily satisfied. For the conditional variance,

$$\begin{aligned}
\frac{1}{nh}\sum_{i=\lfloor (t-h)n\rfloor+2}^{\lfloor tn\rfloor} \mathrm{Var}(\varepsilon_i^2 - E\varepsilon^2|\mathcal{F}_{i-1}) &= \mathrm{Var}(\varepsilon^2) + o_p(1) \\
\frac{1}{nh}\sum_{i=\lfloor (t-h)n\rfloor+2}^{\lfloor tn\rfloor} \mathrm{Var}(\varepsilon_i\varepsilon_{i-1}|\mathcal{F}_{i-1}) &= \frac{\omega^2}{nh}\sum_{i=\lfloor (t-h)n\rfloor+2}^{\lfloor tn\rfloor} \varepsilon_{i-1}^2 \\
&= \omega^2(\omega^2 + o_p(1)), \\
&= \omega^4 + o_p(1) \\
\frac{1}{nh}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor} \mathrm{Var}(\varepsilon_i\varepsilon_{i-K}|\mathcal{F}_{i-1}) &= \frac{\omega^2}{nh}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor} \varepsilon_{i-K}^2 \\
&= \omega^4 + o_p(1).
\end{aligned}$$

For the conditional covariance,

$$\begin{aligned}
\frac{1}{nh}\sum_{i=\lfloor (t-h)n\rfloor+2}^{\lfloor tn\rfloor} \mathrm{cov}(\varepsilon_i^2 - E\varepsilon^2, \varepsilon_i\varepsilon_{i-1}|\mathcal{F}_{i-1}) &= \frac{E\varepsilon^3}{nh}\sum_{i=\lfloor (t-h)n\rfloor+2}^{\lfloor tn\rfloor} \varepsilon_{i-1} = o_p(1), \\
\frac{1}{nh}\sum_{i=\lfloor (t-h)n\rfloor+2}^{\lfloor tn\rfloor} \mathrm{cov}(\varepsilon_i^2 - E\varepsilon^2, \varepsilon_i\varepsilon_{i-K}|\mathcal{F}_{i-1}) &= \frac{E\varepsilon^3}{nh}\sum_{i=\lfloor (t-h)n\rfloor+2}^{\lfloor tn\rfloor} \varepsilon_{i-K} = o_p(1), \\
\frac{1}{nh}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor} \mathrm{cov}(\varepsilon_i\varepsilon_{i-1}, \varepsilon_i\varepsilon_{i-K}|\mathcal{F}_{i-1}) &= \frac{E\varepsilon^2}{nh}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor} \varepsilon_{i-1}\varepsilon_{i-K} = o_p(1).
\end{aligned}$$

So the convergence in distribution in (11) follows easily. □

Lemma 2 *Let K be fixed positive integer such that $K < n$, then conditionally on the X process, we have $[X,\varepsilon]_t^{'K,h} = O_p\left(\frac{1}{\sqrt{h}}\right)$. Furthermore, if $K \to \infty, K = o(n)$, we have $[X,\varepsilon]_t^{'K,h} = O_p\left(\frac{1}{\sqrt{hK}}\right)$.*

Proof By definition

$$\begin{aligned}
[X,\varepsilon]_t'^{K,h} &= \frac{1}{hK}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}(\Delta^K X_{t_i})(\Delta^K \varepsilon_{t_i}),\\
&= \frac{1}{hK}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}\Delta^K X_{t_i}\varepsilon_{t_i} - \frac{1}{hK}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}\Delta^K X_{t_i}\varepsilon_{t_{i-K}}\\
&= \frac{1}{hK}\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor-K}\left(\Delta^K X_{t_i}-\Delta^K X_{t_{i+K}}\right)\varepsilon_{t_i} - \frac{1}{hK}\sum_{i=\lfloor (t-h)n\rfloor+1}^{\lfloor (t-h)n\rfloor+K}\Delta^K X_{t_{i+K}}\varepsilon_{t_i}\\
&\quad +\frac{1}{hK}\sum_{i=\lfloor tn\rfloor-K+1}^{\lfloor tn\rfloor}\Delta^K X_{t_i}\varepsilon_{t_i},
\end{aligned}$$

which is organized in such a way that the summed terms are conditionally independent, so that we can calculate the conditional variance easily:

$$\begin{aligned}
&\mathrm{Var}([X,\varepsilon]_t'^{K,h}|X)\\
=&\left(\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor-K}\left(\Delta^K X_{t_i}-\Delta^K X_{t_{i+K}}\right)^2+\sum_{i=\lfloor (t-h)n\rfloor+1}^{\lfloor (t-h)n\rfloor+K}\left(\Delta^K X_{t_{i+K}}\right)^2\right.\\
&\left.+\sum_{i=\lfloor tn\rfloor-K+1}^{\lfloor tn\rfloor}\left(\Delta^K X_{t_i}\right)^2\right)\times\frac{1}{h^2K^2}\omega^2\\
=&\left(2\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}\left(\Delta^K X_{t_i}\right)^2-2\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor-K}\Delta^K X_{t_i}\Delta^K X_{t_{i+K}}\right)\frac{1}{h^2K^2}\omega^2\\
\le&\left(4\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}\left(\Delta^K X_{t_i}\right)^2+\sum_{i=\lfloor tn\rfloor-K+1}^{\lfloor tn\rfloor}\left(\Delta^K X_{t_i}\right)^2\omega^2\right.\\
&\left.-\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor (t-h)n\rfloor+2K}\left(\Delta^K X_{t_i}\right)^2\right)\frac{1}{h^2K^2}\omega^2\\
=&\left(O_p\left(hK\right)+O_p\left(\frac{K^2}{n}\right)\right)\frac{1}{h^2K^2}\omega^2\\
=&\left(O_p\left(\frac{1}{hK}\right)+o_p\left(\frac{1}{h}\right)\right)\omega^2, \qquad (12)
\end{aligned}$$

as $n\to\infty, h\to 0, nh\to\infty$. In the inequality, we use Cauchy-Schwarz for the cross product term. The order of $\sum_{i=\lfloor (t-h)n\rfloor+K+1}^{\lfloor tn\rfloor}\left(\Delta^K X_{t_i}\right)^2$, which is just $[X,X]_h^{avg}K$ as in Zhang et al. (2005), is $O_p(hK)$ by the proof of Lemma 1. Because X is a diffusion process, $\Delta^K X_{t_i}=O_p(K/n)$, there are $2K$ of such terms, so the order of the remaining terms are $O_p\left(K^2/n\right)$.

So in (12), when K is a fixed integer, this is $O_p(1/h)$, while when $K \to \infty$, this is $O_p(1/(hK))$, and the claims of the lemma follow easily. □

Lemma 3 *The increment over step size δ of the martingale defined in (10) satisfies*

$$\langle M^n, M^n \rangle_t - \langle M^n, M^n \rangle_{t-\delta} = \frac{K}{n\delta} \eta_{n,\delta}^2 + o_p\left(\frac{K}{n\delta}\right),$$

where

$$\eta_{n,\delta}^2 = \frac{1}{\delta} \sum_{t-\delta \leq t_i \leq t} \sigma_{t_i}^2 g_i \Delta t,$$

and

$$g_i = 4 \sum_{j=1}^{K \wedge i} \left(1 - \frac{j}{K}\right) \frac{n}{K}.$$

Proof

$$\begin{aligned}
& \langle M^n, M^n \rangle_t - \langle M^n, M^n \rangle_{t-\delta} \\
= & \frac{4}{\delta^2} \sum_{t-\delta \leqslant t_i \leqslant t} \Delta \langle X, X \rangle_{t_i} \left(\sum_{j=1}^{K \wedge i} \left(1 - \frac{j}{K}\right) (\Delta X_{t_{i-j}}) \right)^2 \\
= & \frac{4}{\delta^2} \sum_{t-\delta \leqslant t_i \leqslant t} \Delta \langle X, X \rangle_{t_i} \sum_{j=1}^{K \wedge i} \left(\left(1 - \frac{j}{K}\right) (\Delta X_{t_{i-j}}) \right)^2 \\
& + 8 \frac{1}{\delta^2} \sum_{t-\delta \leqslant t_i \leqslant t} \Delta \langle X, X \rangle_{t_i} \zeta_i \\
= & (I) + (II),
\end{aligned}$$

where (I) and (II) are defined implicitly, and

$$\zeta_i := \sum_{l>r>0}^{i-1} (\Delta X_{t_l}) (\Delta X_{t_r}) \left(1 - \frac{i-l}{K}\right)^+ \left(1 - \frac{i-r}{K}\right)^+,$$

is just short-hand notation.

We first calculate (I):

$$
\begin{aligned}
& (I) \\
= & \frac{4}{\delta^2} \sum_{t-\delta \leqslant t_i \leqslant t} \Delta \langle X, X \rangle_{t_i} \sum_{j=1}^{K \wedge i} \left(\left(1 - \frac{j}{K} \right) \left(\Delta X_{t_{i-j}} \right) \right)^2 \\
= & \frac{4}{\delta^2} \sum_{t-\delta \leqslant t_i \leqslant t} \left(\sigma_{t_i}^2 \Delta t + (\Delta t)^{3/2} \right) \sum_{j=1}^{K \wedge i} \left(1 - \frac{j}{K} \right)^2 \left(\sigma_{t_{i-j}}^2 \Delta t + (\Delta t)^{3/2} \right) \\
= & \frac{4}{\delta^2} \sum_{t-\delta \leqslant t_i \leqslant t} \left(\sigma_{t_i}^2 \Delta t \right) \sum_{j=1}^{K \wedge i} \left(1 - \frac{j}{K} \right)^2 \left(\sigma_{t_{i-j}}^2 \Delta t \right) \left(1 + o_p \left(1 \right) \right) \\
= & \frac{4}{\delta^2} \sum_{t-\delta \leqslant t_i \leqslant t} \left(\sigma_{t_i}^2 \Delta t \right) \sum_{j=1}^{K \wedge i} \left(1 - \frac{j}{K} \right)^2 \left(\sigma_{t_i}^2 + O_p \left((\Delta t)^{1/2} \left| \log \left(\Delta t \right) \right|^{1/2} \right) \Delta t \right) \left(1 + o_p \left(1 \right) \right) \\
= & \frac{4}{\delta^2} \sum_{t-\delta \leqslant t_i \leqslant t} \sigma_{t_i}^4 \Delta t \sum_{j=1}^{K \wedge i} \left(1 - \frac{j}{K} \right)^2 \Delta t \left(1 + o_p \left(1 \right) \right) \\
= & \frac{K}{n\delta^2} \sum_{t-\delta \leqslant t_i \leqslant t} \sigma_{t_i}^4 g_i \Delta t \left(1 + o_p \left(1 \right) \right) \\
= & \frac{K}{n\delta} \eta_{n,\delta}^2 + o_p \left(\frac{K}{n\delta} \right),
\end{aligned}
$$

where we use the definition of g_i and the fact that $\Delta t = 1/n$ in the second to last equality, and the definition of $\eta_{n,\delta}^2$ in the last equality.

We now calculate (II) following the strategy as in Zhang et al. (2005). First we define

$$
(II)' = 8 \frac{1}{\delta^2} \sum_{t-\delta \leqslant t_i \leqslant t} \sigma_{t_i}^2 \left(\Delta t \right) \zeta_i.
$$

We then prove

$$
(II) \quad = \quad (II)' + o_p \left(\frac{K}{n\delta} \right), \tag{13}
$$

and

$$
(II)' = o_p \left(\frac{K}{n\delta} \right). \tag{14}
$$

Then we prove

$$
(II) = o_p \left(\frac{K}{n\delta} \right),
$$

and obtain

$$
\langle M^n, M^n \rangle_t - \langle M^n, M^n \rangle_{t-\delta} = \frac{K}{n\delta} \eta_{n,\delta}^2 + o_p \left(\frac{K}{n\delta} \right).
$$

So we are now left with proving (13) and (14). To prove (13), note that using Hölder's inequality and the smoothness of volatility path, we have

$$\begin{aligned} E\left|(II)-(II)'\right| &\leqslant 8\frac{1}{\delta^2}\sum_{t-\delta\leqslant t_i\leqslant t} E\left|\Delta\left\langle X,X\right\rangle_{t_i}-\sigma_{t_i}^2\left(\Delta t\right)\right|\left|\zeta_i\right| \\ &\leqslant 8\frac{1}{\delta^2}\sum_{t-\delta\leqslant t_i\leqslant t}\left|\Delta\left\langle X,X\right\rangle_{t_i}-\sigma_{t_i}^2\left(\Delta t\right)\right|_2\left|\zeta_i\right|_2 \\ &\leqslant 8\frac{1}{\delta^2}\sum_{t-\delta\leqslant t_i\leqslant t}\sup_i\left|\Delta\left\langle X,X\right\rangle_{t_i}-\sigma_{t_i}^2\left(\Delta t\right)\right|_2\sup_i\left|\zeta_i\right|_2 \\ &\leqslant \frac{8}{\delta^2}n\delta\times O\left(\left(\Delta t\right)^{3/2}\left|\log\left(\Delta t\right)\right|^{1/2}\right)\times O\left(\left(\Delta t\right)K\right) \\ &= o_p\left(\frac{K}{n\delta}\right), \end{aligned}$$

where we also use the result that

$$\sup_i\left|\zeta_i\right|_2 = O\left(\left(\Delta t\right)K\right),$$

which is a result from equation (A.29) in Zhang et al. (2005); the definition for ζ_i is exactly the same as there.

Next we prove (14). Note that in all cases, we are summing over the interval $t-h\leqslant t_i\leqslant t$; in the following, to save notation, we re-index the t_i in the above interval as $t_0,t_1,\ldots.$, and there are $nh-1$ items to sum.

$$(II)' = 8\frac{1}{\delta^2}\sum_{i=0}^{n\delta-1}\sigma_{t_i}^2\left(\Delta t\right)\sum_{l>r>0}^{i-1}\left(\Delta X_{t_l}\right)\left(\Delta X_{t_r}\right)\left(1-\frac{i-l}{K}\right)^+\left(1-\frac{i-r}{K}\right)^+.$$

Changing the order of summation, we can rewrite it as

$$(II)' = 8\frac{1}{\delta^2}\sum_{l=0}\left(\Delta X_{t_l}\right)\sum_{r=0}^{l-1}\left(\Delta X_{t_r}\right)\sum_{i=l+1}^{n\delta-1}\sigma_{t_i}^2\left(\Delta t\right)\left(1-\frac{i-l}{K}\right)^+\left(1-\frac{i-r}{K}\right)^+.$$

Note that

$$\left(\Delta X_{t_l}\right)\sum_{r=0}^{l-1}\left(\Delta X_{t_r}\right)\sum_{i=l+1}^{n\delta-1}\sigma_{t_i}^2\left(\Delta t\right)\left(1-\frac{i-l}{K}\right)^+\left(1-\frac{i-r}{K}\right)^+$$

is a martingale difference sequence indexed by l, with respect to the natural filtration generated by X_{t_i} sequence, thus $(II)'$ is a martingale. Therefore we can use the Burkholder-Davis-Gundy's inequality to evaluate the second moment by calculating its quadratic variation. Then to know its stochastic order, also using the boundedness of the σ process

by $\bar{\sigma}$, we have

$$\begin{aligned}
& E\left((II)'\right)^2 \\
\leqslant\ & \frac{64}{\delta^4} E \sum_{l=0}^{n\delta-1} \Delta \left\langle X, X \right\rangle_{t_l} \left(\sum_{r=0}^{l-1} (\Delta X_{t_r}) \sum_{i=l+1}^{n\delta-1} \sigma_{t_i}^2 (\Delta t) \left(1 - \frac{i-l}{K}\right)^+ \left(1 - \frac{i-r}{K}\right)^+ \right)^2 \\
\leqslant\ & \frac{64}{\delta^4} (\bar{\sigma})^2 (\Delta t) \sum_{l=0}^{n\delta-1} E \left(\sum_{r=0}^{l-1} (\Delta X_{t_r}) \sum_{i=l+1}^{n\delta-1} \sigma_{t_i}^2 (\Delta t) \left(1 - \frac{i-l}{K}\right)^+ \left(1 - \frac{i-r}{K}\right)^+ \right)^2 .
\end{aligned}$$

The brackets contain the summation of a martingale difference indexed by r, so applying the Burkholder-Davis-Gundy's inequality again, we have

$$\begin{aligned}
& E\left((II)'\right)^2 \\
\leqslant\ & \frac{64}{\delta^4} (\bar{\sigma})^4 (\Delta t)^2 \sum_{l=0}^{n\delta-1} \sum_{r=0}^{l-1} E \left(\sum_{i=l+1}^{n\delta-1} \sigma_{t_i}^2 (\Delta t) \left(1 - \frac{i-l}{K}\right)^+ \left(1 - \frac{i-r}{K}\right)^+ \right)^2 \\
\leqslant\ & \frac{64}{\delta^4} (\bar{\sigma})^4 (\Delta t)^2 \times (\bar{\sigma})^4 (\Delta t)^2 \times \sum_{l=0}^{n\delta-1} \sum_{r=0}^{l-1} E \left(\sum_{i=l+1}^{n\delta-1} \left(1 - \frac{i-l}{K}\right)^+ \left(1 - \frac{i-r}{K}\right)^+ \right)^2 \\
\leqslant\ & \frac{64}{\delta^4} (\bar{\sigma})^8 (\Delta t)^4 (n\delta \times K) \left(K^2\right) \\
=\ & O\left(\left(\frac{K}{n\delta} \right)^3 \right) = o\left(\frac{K}{n\delta} \right).
\end{aligned}$$

□

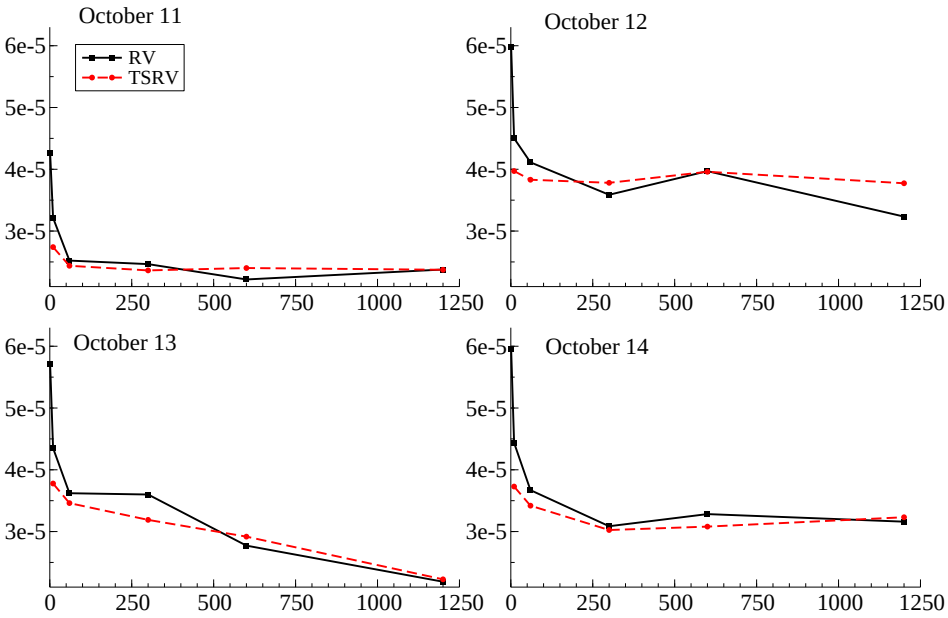

Figure 1: Volatility signature plots, October 11–14, 2009
Note: The horizontal axis represents sampling frequency in seconds

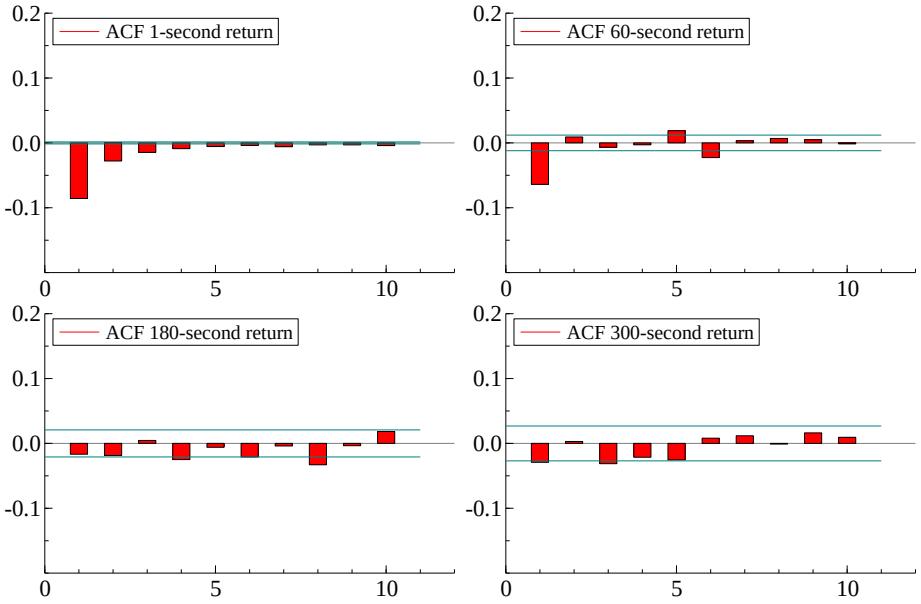

Figure 2: ACF plots for different sampling frequencies

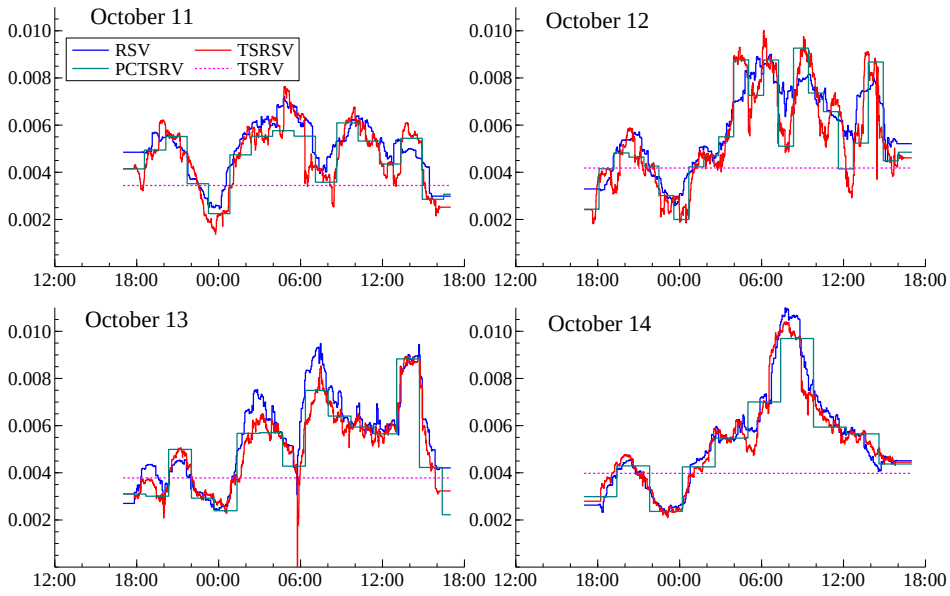

Figure 3: Spot variance estimates, October 11–14, 2009

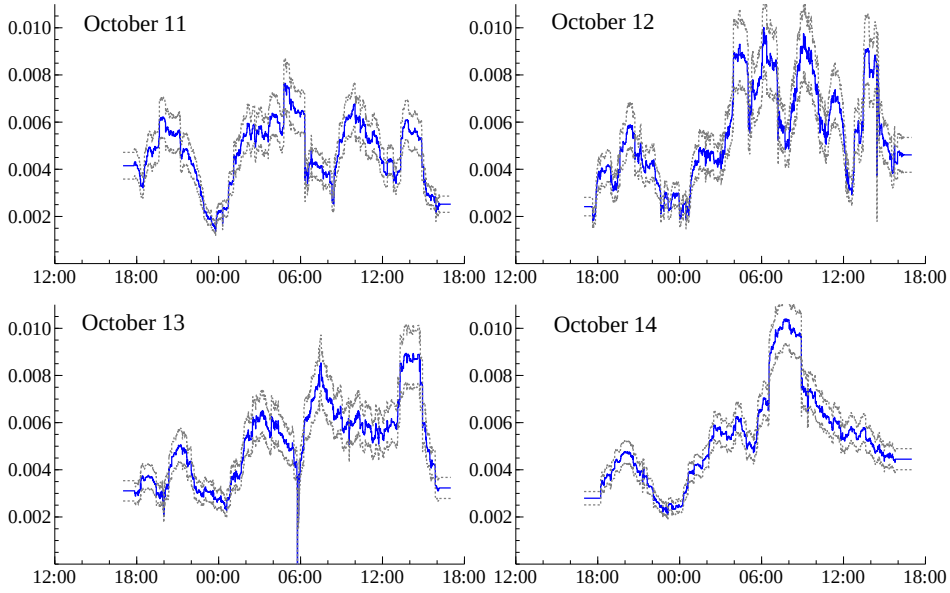

Figure 4: TSRSV with confidence interval, October 11–14, 2009

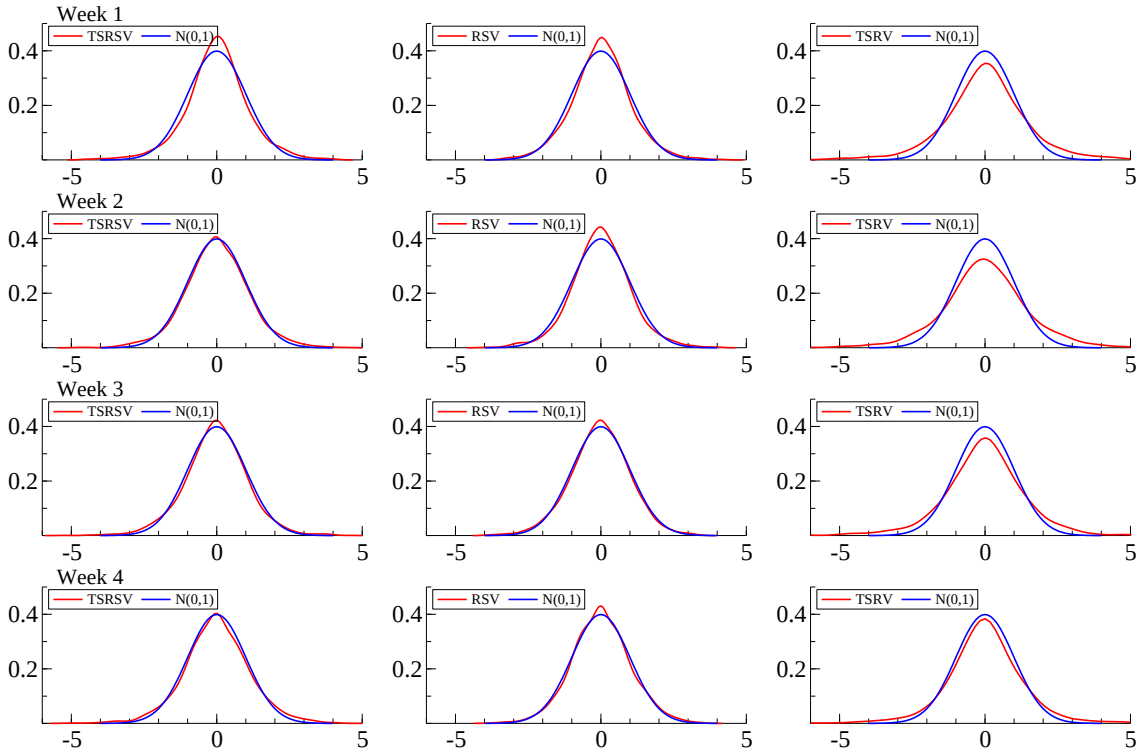

Figure 5: Densities of 3-minute returns standardized by TSRSV, RSV and TSRV, Week 1–4

E1	RSV estimator with 0.5-minute returns (filtering)
E2	RSV estimator with 0.5-minute returns (smoothing)
E3	RSV estimator with 1-minute returns (filtering)
E4	RSV estimator with 1-minute returns (smoothing)
E5	RSV estimator with 5-minute returns (filtering)
E6	RSV estimator with 5-minute returns (smoothing)
E7	TSRSV estimator (filtering)
E8	TSRSV estimator (smoothing)
E9	PCTSRV estimator
MC sd	Monte Carlo standard deviation

Table 1: Notations used in later tables

	MISE	MC sd	MISRE	MC sd	MIAE	MC sd	MIARE	MC sd
E1	0.268	0.00310	0.601	0.00900	0.331	0.00394	0.690	0.00944
E2	0.271	0.00320	0.607	0.00903	0.336	0.00406	0.699	0.00955
E3	0.254	0.00412	0.385	0.00425	0.320	0.00524	0.471	0.00462
E4	0.246	0.00391	0.383	0.00427	0.309	0.00488	0.465	0.00460
E5	0.311	0.00510	0.317	0.00127	0.379	0.00626	0.389	0.00161
E6	0.250	0.00405	0.261	0.00117	0.299	0.00488	0.313	0.00137
E7	0.097	0.00138	0.106	0.00026	0.121	0.00172	0.132	0.00031
E8	**0.094**	0.00133	**0.103**	0.00026	**0.117**	0.00166	**0.128**	0.00031
E9	0.096	0.00135	0.105	0.00028	0.119	0.00170	0.131	0.00034

Table 2: Monte Carlo error measures in SV1F1

	MISE	MC sd	MISRE	MC sd	MIAE	MC sd	MIARE	MC sd
E1	1.571	0.00251	5.481	0.09184	1.699	0.00339	5.792	0.09632
E2	1.585	0.00254	5.513	0.09223	1.726	0.00343	5.870	0.09720
E3	0.833	0.00353	2.740	0.04542	0.956	0.00473	2.986	0.04801
E4	0.838	0.00339	2.766	0.04561	0.956	0.00441	3.002	0.04819
E5	0.376	0.00501	0.672	0.00877	0.463	0.00614	0.808	0.00985
E6	0.324	0.00399	0.646	0.00899	0.385	0.00483	0.737	0.00960
E7	0.122	0.00160	0.154	0.00079	0.151	0.00198	0.189	0.00100
E8	**0.118**	0.00152	**0.151**	0.00082	**0.146**	0.00190	**0.185**	0.00103
E9	0.122	0.00157	0.155	0.00084	0.151	0.00196	0.191	0.00108

Table 3: Monte Carlo error measures in SV1F2

	MISE	MC sd	MISRE	MC sd	MIAE	MC sd	MIARE	MC sd
E1	15.53	0.01024	54.80	0.91817	16.22	0.01179	57.40	0.96356
E2	15.62	0.01017	55.09	0.92219	16.44	0.01154	58.16	0.97239
E3	7.731	0.00783	27.27	0.45471	8.246	0.00947	29.01	0.48089
E4	7.804	0.00785	27.50	0.45707	8.288	0.00898	29.17	0.48270
E5	1.585	0.00567	5.395	0.09134	1.807	0.00698	5.970	0.10051
E6	1.605	0.00518	5.524	0.09215	1.740	0.00593	5.863	0.09742
E7	0.227	0.00210	0.449	0.00517	0.276	0.00258	0.549	0.00740
E8	**0.223**	0.00196	**0.446**	0.00511	**0.272**	0.00242	**0.546**	0.00731
E9	0.230	0.00206	0.458	0.00534	0.281	0.00255	0.575	0.00984

Table 4: Monte Carlo error measures in SV1F3

	MISE	MC sd	MISRE	MC sd	MIAE	MC sd	MIARE	MC sd
E1	0.397	0.00710	19.92	1.2040	0.592	0.01150	28.61	1.7447
E2	0.336	0.00553	19.94	1.2200	0.483	0.00906	28.85	1.7914
E3	0.388	0.00859	10.08	0.6034	0.614	0.01356	14.67	0.8722
E4	0.312	0.00668	10.03	0.6118	0.482	0.01069	14.45	0.8765
E5	0.509	0.01313	2.738	0.1143	0.794	0.01976	4.176	0.1668
E6	0.441	0.01120	2.744	0.1139	0.668	0.01608	4.104	0.1626
E7	0.244	0.00509	0.701	0.0070	0.400	0.00799	1.082	0.0151
E8	**0.156**	0.00330	**0.497**	0.0071	**0.257**	0.00524	**0.758**	0.0149
E9	0.188	0.00398	0.569	0.0069	0.312	0.00632	0.882	0.0143

Table 5: Monte Carlo error measures in SV2F1

	MISE	MC sd	MISRE	MC sd	MIAE	MC sd	MIARE	MC sd
E1	1.678	0.00545	197.6	12.041	1.871	0.00927	282.9	17.444
E2	1.654	0.00435	198.5	12.212	1.822	0.00731	287.4	17.930
E3	0.987	0.00757	98.26	6.0369	1.207	0.01225	142.1	8.7261
E4	0.940	0.00597	98.54	6.1194	1.105	0.00967	141.7	8.7664
E5	0.596	0.01292	19.92	1.1503	0.875	0.01952	29.26	1.6650
E6	0.540	0.01107	20.09	1.1343	0.753	0.01588	28.61	1.6112
E7	0.312	0.00622	1.566	0.0605	0.494	0.00954	2.678	0.1238
E8	**0.207**	0.00410	**1.351**	0.0724	**0.331**	0.00627	**2.302**	0.1415
E9	0.246	0.00490	1.443	0.0698	0.391	0.00751	2.520	0.1373

Table 6: Monte Carlo error measures in SV2F2

	MISE	MC sd	MISRE	MC sd	MIAE	MC sd	MIARE	MC sd
E1	15.51	0.01020	1974	120.4	16.20	0.01201	2825	174.4
E2	15.61	0.01013	1984	122.1	16.43	0.01165	2873	179.3
E3	7.773	0.00799	981.0	60.63	8.311	0.01073	1419	88.57
E4	7.846	0.00797	984.2	61.20	8.337	0.00987	1416	87.65
E5	1.804	0.01160	192.8	11.49	2.119	0.01750	281.3	16.62
E6	1.811	0.01024	194.7	11.43	2.016	0.01418	274.9	15.99
E7	0.440	0.00820	**8.229**	0.563	0.660	0.01233	15.65	1.069
E8	**0.319**	0.00536	8.255	0.655	**0.479**	0.00811	**15.60**	1.213
E9	0.364	0.00646	8.376	0.616	0.548	0.00972	16.25	1.134

Table 7: Monte Carlo error measures in SV2F3

	Mean	Standard Deviation
cross-validation bandwidth E1	1268.9	450.99
cross-validation bandwidth E2	632.10	244.20
cross-validation bandwidth E3	1820.0	701.98
cross-validation bandwidth E4	969.72	360.32
cross-validation bandwidth E5	4582.4	1697.6
cross-validation bandwidth E6	3385.1	800.51
plug-in scale E7, E8, E9	8.2311	5.8312
plug-in bandwidth E7, E8, E9	1612.1	751.91

Table 8: Tuning parameters in SV1F1

	Mean	Standard Deviation
cross-validation bandwidth E1	1263.8	476.26
cross-validation bandwidth E2	573.09	244.38
cross-validation bandwidth E3	1809.6	728.48
cross-validation bandwidth E4	925.53	348.86
cross-validation bandwidth E5	4573.2	1710.8
cross-validation bandwidth E6	3389.7	807.64
plug-in scale E7, E8, E9	26.511	13.311
plug-in bandwidth E7, E8, E9	2957.6	1301.8

Table 9: Tuning parameters in SV1F2

	Mean	Standard Deviation
cross-validation bandwidth E1	1247.3	502.17
cross-validation bandwidth E2	518.79	231.64
cross-validation bandwidth E3	1782.1	779.58
cross-validation bandwidth E4	880.50	339.59
cross-validation bandwidth E5	4606.7	1760.0
cross-validation bandwidth E6	3408.8	828.19
plug-in scale E7, E8, E9	52.563	13.090
plug-in bandwidth E7, E8, E9	4300.9	1608.1

Table 10: Tuning parameters in SV1F3

	Mean	Standard Deviation
cross-validation bandwidth E1	979.86	478.93
cross-validation bandwidth E2	566.07	275.16
cross-validation bandwidth E3	1410.1	741.30
cross-validation bandwidth E4	921.54	408.56
cross-validation bandwidth E5	4457.7	1973.1
cross-validation bandwidth E6	3554.6	1132.7
plug-in scale E7, E8, E9	18.762	16.989
plug-in bandwidth E7, E8, E9	2343.7	1670.0

Table 11: Tuning parameters in SV2F1

	Mean	Standard Deviation
cross-validation bandwidth E1	1193.1	508.21
cross-validation bandwidth E2	547.62	265.68
cross-validation bandwidth E3	1674.7	779.33
cross-validation bandwidth E4	910.98	388.16
cross-validation bandwidth E5	4532.1	1875.5
cross-validation bandwidth E6	3504.0	1013.9
plug-in scale E7, E8, E9	39.516	21.056
plug-in bandwidth E7, E8, E9	3630.6	2036.3

Table 12: Tuning parameters in SV2F2

	Mean	Standard Deviation
cross-validation bandwidth E1	1266.8	513.08
cross-validation bandwidth E2	521.07	239.33
cross-validation bandwidth E3	1792.3	783.84
cross-validation bandwidth E4	885.03	364.87
cross-validation bandwidth E5	4624.2	1823.9
cross-validation bandwidth E6	3454.1	909.05
plug-in scale E7, E8, E9	57.457	15.854
plug-in bandwidth E7, E8, E9	4526.7	1905.3

Table 13: Tuning parameters in SV2F3

	TSRSV	RSV	TSRV	TSRSV	RSV	TSRV
	1% VaR	1% VaR	1% VaR	5% VaR	5% VaR	5% VaR
week 1	0.0239	0.0209	0.0439	0.0561	0.0548	0.0957
week 2	0.0257	0.0200	0.0613	0.0579	0.0509	0.0940
week 3	0.0274	0.0226	0.0535	0.0670	0.0618	0.1131
week 4	0.0331	0.0217	0.0331	0.0735	0.0548	0.0661
All	0.0275	0.0213	0.0480	0.0636	0.0556	0.0922

Table 14: Left tail

	TSRSV	RSV	TSRV	TSRSV	RSV	TSRV
	95% VaR	95% VaR	95% VaR	99% VaR	99% VaR	99% VaR
week 1	0.0600	0.0600	0.0979	0.0235	0.0235	0.0500
week 2	0.0631	0.0539	0.1018	0.0252	0.0213	0.0657
week 3	0.0548	0.0496	0.1061	0.0200	0.0191	0.0461
week 4	0.0739	0.0648	0.0696	0.0352	0.0252	0.0365
All	0.0630	0.0571	0.0938	0.0260	0.0223	0.0496

Table 15: Right tail