SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2010-67

Estimation of Stochastic Volatility Models by Nonparametric Filtering

Shin Kanaya and Dennis Kristensen

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Estimation of Stochastic Volatility Models by Nonparametric Filtering*

Shin Kanaya†
University of Oxford

Dennis Kristensen‡
Columbia University and CREATES§

October 2010

Abstract

A two-step estimation method of stochastic volatility models is proposed: In the first step, we estimate the (unobserved) instantaneous volatility process using the estimator of Kristensen (2010, *Econometric Theory* 26). In the second step, standard estimation methods for fully observed diffusion processes are employed, but with the filtered volatility process replacing the latent process. Our estimation strategy is applicable to both parametric and nonparametric stochastic volatility models, and we give theoretical results for both. The resulting estimators of the drift and diffusion terms of the volatility model will carry additional biases and variances due to the first-step estimation, but under regularity conditions these vanish asymptotically and our estimators inherit the asymptotic properties of the infeasible estimators based on observations of the volatility process. A simulation study examines the finite-sample properties of the proposed estimators.

Keywords: Realized spot volatility; stochastic volatility; kernel estimation; nonparametric; semiparametric.

JEL codes: C14; C32; C58.

---

*The authors wish to thank Bruce Hansen, Bent Nielsen, Neil Shephard, Yoshinori Kawasaki, Kotaro Hitomi, Yoshihiko Nishiyama, Russell Davidson, Howell Tong, Pascale Valery and Karim Abadir for fruitful discussions and helpful suggestions. They also wish to thank participants at seminars at University of Wisconsin-Madison, the CREATES opening Conference, the 2007 EC^2 Meeting, the 2009 Japanese Statistical Society Spring Meeting, the 4th London and Oxbridge Time Series Workshop, the 43rd Annual Conference of the Canadian Economics, the 2009 North American Summer Meeting of the Econometric Society, and the 2009 Far East and South Asia Meeting of the Econometric Society for comments and suggestions. Kristensen gratefully acknowledges financial support from the Danish National Research Foundation (through a grant to CREATES) and the National Science Foundation (grant no. SES-0961596). Part of this research was conducted while Kanaya was visiting the Institute of Statistical Methematics and Kristensen was visiting Princeton University and University of Copenhagen, whose hospitality is gratefully acknowledged.

†Department of Economics, Oxford-Man Institute and Nuffield College. Address: The Oxford-Man Institute, Eagle House, Walton Well Road, Oxford, OX2 6ED, UK. E-mail: `shin.kanaya@economics.ox.ac.uk`.

‡Department of Economics, Columbia University, 1018 International Affairs Building, 420 West 118th Street, New York, NY 10027. E-mail: `dk2313@columbia.edu`.

§Center for Research in Econometric Analysis of Time Series, University of Aarhus.

1 Introduction

One of the key components in asset pricing models is the conditional second moment or volatility of the processes. It has long been recognized that volatility varies over time, and considerable efforts have been put into modelling and forecasting this variable. Within a continuous-time framework, stochastic volatility (SV) diffusion models, such as the Heston (1993) and Hull and White (1987) models, represent a popular class (see Shephard, 2005 for an overview). Unfortunately, the estimation and forecasting of such models are hampered by the fact that the volatility is a latent and unobserved variable, and most proposed estimation methods are not easily implemented and require considerable computational power.

In recent years, nonparametric estimators of the integrated volatility have emerged. These are model-free estimators of the integral of the (realised) volatility process over a given time interval, see Andersen et al. (2010) for an overview. Furthermore, they are very simple to compute and as such do not suffer from the aforementioned computational problems. However, realised volatility estimators are only able to give a model-free measure of the volatility in-sample, and are not informative about future, or out-of-sample, volatility.

In this paper, we propose a general estimation strategy for SV diffusion models that combines a simple, model-free realised volatility estimator with the additional structure imposed by the Markov diffusion model. The resulting estimators are simple to implement and require little, if any, numerical optimization. The estimation strategy allows for both nonparametric and fully parametric specifications of the SV model, and as such is very flexible.

The estimation method proceeds in two steps: In the first step, a nonparametric estimator of the spot (or instantaneous) volatility as proposed in Kristensen (2010a) is computed. This estimator is simply a kernel-weighted version of standard integrated volatility estimators and supplies us with an estimate of the spot volatility process over the sample interval. The main idea is then to combine the spot volatility estimator with existing estimation methods for fully observed diffusion models: If the volatility process indeed had been observed, we could use any of many existing estimation methods to estimate the SV model. We have not observed the volatility process, but what we do have is its consistent estimator which we can use in its place. Thus, in the second step, the spot volatility estimator is plugged into a given existing estimation method for fully observed diffusion models.

We derive the asymptotics of the resulting estimators for two leading estimation methods: For nonparametric SV models, we consider the nonparametric kernel estimator à la Bandi and Phillips (2003). For fully parametric models, we consider estimators akin to the ones proposed in Prakasa Rao (1988) or Bandi and Phillips (2007). For each of the two estimators, we give conditions under which they are consistent and asymptotically normally distributed. In the nonparametric case, our estimation problem is similar to the one considered in Rilstone (1996) where the kernel regression with generated regressors is considered; see also Newey, Powell and Vella (1999) and Xiao, Linton, Carroll and Mammen (2003) for similar nonparametric two-step estimators. The parametric estimators can be seen as a two-step semiparametric estimation procedure, where a parametric estimator relies on a preliminary nonparametric estimator; see e.g. Kanaya (2010a) and Kristensen (2010b) for related estimators in a diffusion setting.

Our estimators rely on certain nuisance parameters that need to be chosen in the implementation. In particular, a bandwidth has to be chosen in the estimation of the spot volatility. Our theoretical results offer some guidance regarding how this and other parameters should be chosen. Based on these, we discuss in some detail how the estimators can be implemented in practice. We also investigate the finite-sample performance of our estimator through a simulation study with particular emphasis on its sensitivity towards the choice of nuisance parameters. We find that the estimators are quite robust and fairly precise for reasonable sample sizes.

Within the class of Markov diffusion models, a number of studies have proposed ways to identify and estimate the parameters of an underlying parametric SV model. If only low-frequency data is available, the estimation problem is harder since the amount of information available to the econometrician is more limited. In a few specific examples, one can derive analytical expressions of certain moment functions and use these in the estimation (Chacko and Viceira, 2003), but in general numerical methods need to be used to deal with latent variables (see e.g. Altissimo and Mele, 2009, Andersen and Lund, 1997; Chib et al., 2006; Gallant et al.,1997; Brownlees, Kristensen and Shin, 2010). These are all situated within a parametric framework, and require substantial computer power to implement.

In the case where high-frequency data is available, the use of integrated volatility estimators has greatly facilitated the estimation of models of the volatility. A number of studies have proposed to estimate parametric diffusion models of the volatility by matching certain conditional moments of the integrated volatility with their estimated ones using GMM-type methods. Examples of this approach are Barndorff-Nielsen and Shephard (2001), Bollerslev and Zhou (2002), Corradi and Distaso (2006) and Todorov (2009). These methods appear to work well when the volatility process is assumed to solve a simple parametric model where the conditional moments can be expressed analytically. However, for more general parametric models, the closed form of the conditional moments can be difficult to derive, and as a result the extension of this type of estimation strategy to a more general setting will require the use of simulation-based or other computationally burdensome methods. We also note that how to obtain nonparametric estimators of the SV model from integrated volatility is not clear.

In related studies, Comte, Genon-Catalot and Rozenholc (2009), Renò (2006, 2008) and Bandi and Renò (2009) propose estimators similar to ours, but they only consider nonparametric volatility models. Furthermore, Comte et al (2009) assume that the integrated volatility is observable (if their setting is read in the context of the volatility estimation). Renò (2006) and Bandi and Renò (2009) do not develop a complete asymptotic theory of their estimators. A key result needed in deriving the theoretical results of the proposed estimators here and in these studies is the uniform consistency (and its rate) of the preliminary spot volatility estimator over a growing time interval. We are able to give primitive conditions for the uniform consistency result for our specific estimator, where these primitive conditions allow for most models found in the literature. The proof of this uniform consistency result proves to be technically very demanding due to two properties of the object of interest, namely the sample path of the volatility process: First, it is not smooth, and second it is unbounded as time diverges. This is in contrast to standard nonparametric estimation problems (e.g. density and regression estimation), and we have to use some novel theoretical techniques in order to establish our result for an expanding time interval, in particular a new result on the global modulus of continuity of stochastic processes. In comparison, Renò (2008) only establishes consistency of the preliminary estimator over a fixed time interval. This in turn means that he can only show results for the estimation of the diffusion coefficient of the volatility model. Furthermore, his consistency result relies on some strong assumptions on the model, including that the domain of the volatility has to be a compact interval. This rules out standard models found in the literature. Bandi and Renò (2009) avoid this issue by simply imposing the high-level assumption of the modulus of continuity of the diffusion processes over the expanding time interval, as well as imposing some bounded conditions on the volatility process and its transformation function f. On the other hand, their framework is more general than ours in that they allow for the presence of jumps in both observable and latent processes, and as such their and our studies complement each other.

The remains of the paper is organised as follows: In the next section, we outline our proposed estimation method for the nonparametric and fully parametric case. In Section 3, the asymptotic properties of our estimators are derived under regularity conditions, while the practical implementation of the estimator is discussed in Section 4. The results of a simulation study investigating the finite-sample properties of our estimator are presented in Section 5. Section 6 concludes. All proofs and lemmas are found in Appendix A, while tables and figures can be found in Appendices B and C.

We use the following notation throughout: The symbols $\stackrel{P}{\to}$ and $\stackrel{d}{\to}$ denote convergence in probability and distribution, respectively. The abbreviation a.s. is for "almost surely." The transpose of a vector or matrix A is denoted $A^{\bigstar}$. For a vector or matrix $B = [b_{ij}]$, $||B||$ denotes $\sum_{i,j} |b_{ij}|$. For definitional equations, we use the notations: $C := D$ and $E =: F$, where the former means that C is defined by D, and the latter means that E is defined by F.

2 A General Estimation Method for SV Models

Let $\{X_t\} = \{X_t : t \geq 0\}$ be a Brownian semimartingale solving

$$\begin{cases} dX_t = \mu_t dt + \sigma_t dW_t; \\ d\sigma_t^2 = \alpha\left(\sigma_t^2\right) dt + \beta\left(\sigma_t^2\right) dZ_t, \end{cases} \tag{1}$$

where $\{W\}_{t\geq 0}$ and $\{Z_t\}_{t\geq 0}$ are two (possibly correlated) standard Brownian motions, while $\{\mu_t\}_{t\geq 0}$ and $\{\sigma_t\}_{t\geq 0}$ are adapted stochastic processes. The process $\left\{\sigma_t^2\right\}$ is usually denoted the (spot) volatility process of $\{X_t\}$, while $\{\mu_t\}$ is the drift process. The second part of the model in eq. (1), stating the dynamics of the volatility process, is referred to as a stochastic volatility (SV) model.

Suppose we have observed $X_{t_0}, X_{t_1}, ..., X_{t_n}$ at $n+1$ discrete time points $0 = t_0 < t_1 < ... < t_n = T$. Given these observations, we wish to draw inference about the underlying drift and diffusion terms, $\alpha(\cdot)$ and $\beta^2(\cdot)$. Since we have not observed the process $\left\{\sigma_t^2\right\}$, the estimation of these two terms involve a latent variable which in general has to be integrated out. This could for example be done using particle filtering which is computationally demanding; see, for example, Brownlees et al (2010). We here propose simple-to-compute estimators relying on a first-step kernel estimator of the volatility.

To motivate our estimators, consider for the moment the counter-factual situation where the volatility process has been observed at discrete, equidistant time points, $0 \leq \tau_0 < \tau_1 < ... < \tau_N \leq T$ with $\delta = |\tau_j - \tau_{j-1}|$ being the time-distance between consecutive observations. Then we can estimate the drift $\alpha(\cdot)$ and the diffusion $\beta^2(\cdot)$ through standard estimation methods for discretely sampled diffusion processes. Two specific estimators are considered subsequently, nonparametric and fully parametric ones. For semiparametric models, where either the drift or the diffusion term is left unspecified, the estimation can proceed by combining the proposed non- and fully parametric estimators. Alternatively, the semiparametric MLE approach in Kanaya (2010a) may be employed.

For fully nonparametric estimation of the drift and diffusion function, kernel estimators have been considered in Bandi and Phillips (2003), Florens-Zmirou (1993) and Jiang and Knight (1997) amongst others. These are given by:

$$\tilde{\alpha}(x) = \frac{\sum_{j=1}^{N-1} \mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)\left[\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right]}{\delta \sum_{j=1}^{N} \mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)}, \quad \tilde{\beta}^2(x) = \frac{\sum_{j=1}^{N-1} \mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)\left[\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right]^2}{\delta \sum_{j=1}^{N} \mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)}, \quad (2)$$

where $\mathcal{K}_b(z) = \mathcal{K}(z/b)/b$, $\mathcal{K}$ is a kernel, and $b > 0$ a bandwidth; see Silverman (1986) for an introduction to kernel estimation.

If parametric forms for either or both of the two functions are specified, a number of estimators offer themselves, see for example Florens-Zmirou (1989), Dacunha-Castelle and Florens-Zmirou (1986), Jacod (2006), Sørensen (2009) and Yoshida (1992). We here follow Bandi and Phillips (2007) and consider least-square estimators of the parameters. Suppose that the drift and/or diffusion functions belong to known parametric families, $\alpha(\cdot) = \alpha(\cdot; \theta_1^*)$ and/or $\beta^2(\cdot) = \beta^2(\cdot; \theta_2^*)$ for two parameters $\theta_1^* \in \Theta_1 \subseteq \mathbb{R}^{d_1}$ and $\theta_2^* \in \Theta_2 \subseteq \mathbb{R}^{d_2}$. We then specify our estimators as slightly modified versions of the ones in Bandi and Phillips (2007): Define the two objective functions, $\tilde{Q}_1(\theta_1)$ and $\tilde{Q}_2(\theta_2)$, by

$$\begin{aligned} \tilde{Q}_1(\theta_1) &= \sum\nolimits_{j=1}^{N-1} \left[(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}) - \alpha(\sigma^2_{\tau_j}; \theta_1)\delta\right]^2; \\ \tilde{Q}_2(\theta_2) &= \sum\nolimits_{j=1}^{N-1} \left[(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j})^2 - \beta^2(\sigma^2_{\tau_j}; \theta_2)\delta\right]^2. \end{aligned} \quad (3)$$

We then choose our estimators as $\tilde{\theta}_k = \arg\min_{\theta_k \in \Theta_k} \tilde{Q}_k(\theta_k)$ for $k = 1, 2$.

Now, let us return to the actual situation where the volatility is unobserved. Thus, the above estimators are infeasible. Instead, we here suggest to substitute nonparametric estimates of the spot volatilities for the unobserved ones. A number of estimators have been proposed in the literature. We here focus on the kernel estimator of Kristensen (2010a):

$$\hat{\sigma}^2_\tau = \sum_{i=1}^{n} K_h(t_{i-1} - \tau)\left[X_{t_i} - X_{t_{i-1}}\right]^2 \quad (4)$$

where $K_h(z) = K(z/h)/h$, K is another kernel, and $h > 0$ another bandwidth. As $\Delta := \max_{i=1,...,n}|t_i - t_{i-1}| \to 0$ and $h \to 0$ at a suitable rate, Kristensen (2010a) demonstrates that $\hat{\sigma}_\tau^2 \xrightarrow{P} \sigma_\tau^2$.

With this estimator, we can compute $\hat{\sigma}_\tau^2$ at any given set of discrete time points τ_j, $j = 0, ..., N$. These time points are under the control of the econometrician and may potentially differ from the actual observation time points, $t_0, ..., t_n$. We therefore refer to $\{\tau_j\}$ as pseudo-sampling times. When deriving the asymptotics of our estimators, we will impose certain restrictions on these. Now, simply replace the spot volatilities with its estimates in eq. (2) to obtain feasible nonparametric estimators:

$$\hat{\alpha}(x) = \frac{\sum_{j=1}^{N-1} \mathcal{K}_b\left(\hat{\sigma}_{\tau_j}^2 - x\right)\left[\hat{\sigma}_{\tau_{j+1}}^2 - \hat{\sigma}_{\tau_j}^2\right]}{\delta \sum_{j=1}^{N} \mathcal{K}_b\left(\hat{\sigma}_{\tau_{j-1}}^2 - x\right)}; \tag{5}$$

$$\hat{\beta}^2(x) = \frac{\sum_{j=1}^{N-1} \mathcal{K}_b\left(\hat{\sigma}_{\tau_j}^2 - x\right)\left[\hat{\sigma}_{\tau_{j+1}}^2 - \hat{\sigma}_{\tau_j}^2\right]^2}{\delta \sum_{j=1}^{N} \mathcal{K}_b\left(\hat{\sigma}_{\tau_j}^2 - x\right)}. \tag{6}$$

Similarly, for the parametric estimators, we simply replace $\sigma_{\tau_j}^2$ by $\hat{\sigma}_{\tau_j}^2$ in eq. (3),

$$\hat{Q}_1(\theta_1) = \sum_{j=1}^{N-1}\left[\left(\hat{\sigma}_{\tau_{j+1}}^2 - \hat{\sigma}_{\tau_j}^2\right) - \alpha\left(\hat{\sigma}_{\tau_j}^2; \theta_1\right)\delta\right]^2; \tag{7}$$

$$\hat{Q}_2(\theta_2) = \sum_{j=1}^{N-1}\left[\left(\hat{\sigma}_{\tau_{j+1}}^2 - \hat{\sigma}_{\tau_j}^2\right)^2 - \beta^2\left(\hat{\sigma}_{\tau_j}^2; \theta_2\right)\delta\right]^2. \tag{8}$$

and the feasible estimators are then defined as:

$$\hat{\theta}_k = \arg\min_{\theta_k \in \Theta_k} \hat{Q}_k(\theta_k) \quad \text{for} \quad k = 1, 2. \tag{9}$$

We here have proposed specific estimators in nonparametric and fully parametric settings. It should be clear though that the filtered spot volatility can be combined with any other existing estimation methods for fully observed diffusion models as cited above to obtain estimators for SV models.

3 Asymptotics of the Estimators

In this section, we derive the asymptotics of the non- and parametric estimators proposed in the previous section. We first obtain a uniform convergence result of the preliminary kernel filtered estimator of the volatility process, $\{\hat{\sigma}_t^2\}$. This result is in turn used to demonstrate that the proposed two-step estimators are asymptotically equivalent to the infeasible estimators based on actual observations of $\{\sigma_t^2\}$.

3.1 The Spot Volatility Estimator

We here derive a uniform rate of convergence of the spot volatility estimator. Since the uniform convergence (over time) of $\{\hat{\sigma}_t^2\}$ may be of independent interests in other applications, we do not restrict the true volatility process to be a Markov diffusion (as imposed in eq. (1)). Instead, we only require that the drift and volatility processes, μ_t and σ_t^2, satisfy certain moment conditions, and that the volatility process is sufficiently smooth. It could, for example, be long memory type model (as found in Comte and Renault, 1996) or general Brownian semimartingales and as such be used as an input in the estimation of more general models.

We first introduce a class of smooth kernels:

$\mathbb{K}(m,r)$ A kernel $k:\mathbb{R}\to\mathbb{R}$ is said to belong to $\mathbb{K}(m,r)$, $m,r\geq 0$, if:

(i) $\int_{-\infty}^{\infty} k(x)\,dx = 1$, $\int_{-\infty}^{\infty} xk(x)\,dx = 0$ and $\int_{-\infty}^{\infty} x^2 k(x)\,dx = 1$.

(ii) There exists some positive constants $\bar{K}<\infty$ and $C<\infty$ such that the i-th derivative $k^{(i)}$ satisfies

$$\sup_{x\in\mathbb{R}} |k^{(i)}(x)| \leq \bar{K}, \quad \int_{-\infty}^{\infty} |x|^m |k^{(i)}(x)|\,dx \leq \bar{K}; \quad \text{and}$$

$|k^{(i)}(x)|$ is not decreasing on $(-\infty,-C]$ and not increasing on $[C,\infty)$, for each $i = 0,1,...,r$.

It can be shown that most standard kernels, including the Gaussian one, belong to this class. The last condition in (ii), the monotonicity of the absolute derivative for large $|x|$, may be unfamiliar but is actually satisfied by many kernels (e.g., the Gaussian kernel, and kernels which have a compact support). This condition is useful to obtain sharp convergence rates. We will require that the two kernels, K and $\mathcal{K}$, belong to this class for suitable values of (m,r). For the kernel K which is used for the spot volatility estimation in eq. (4), the condition $\int_{-\infty}^{\infty} xK(x)\,dx = 0$ is not required and thus one-sided kernels, e.g., one suggested in Zhang and Karunamini (1998) may be used as suggested in Kristensen (2010a).

Next, we impose conditions on the drift and volatility process of $\{X_t\}$:

A.1 There exist constants $p>0$ and $l_1,l_2\geq 0$: $\sup_{t\leq T} E\left[|\mu_t|^{2+p}\right] = O\left(T^{l_1}\right)$ and $\sup_{s\leq T} E\left[|\mu_t|^4\right] = O\left(T^{l_2}\right)$ as $T\to\infty$.

A.2 (i) There exist constants $m_1,m_2\geq 0$: $\sup_{t\leq T} E\left[\sigma_t^2\right] = O\left(T^{m_1}\right)$ and $\sup_{t\leq T} E\left[\sigma_t^4\right] = O\left(T^{m_2}\right)$ as $T\to\infty$. (ii) There exist constants $\lambda>0$, $\rho>0$ and $C>0$ such that $E\left[\left|\sigma_t^2-\sigma_s^2\right|^{\lambda}\right] \leq C\left|t-s\right|^{1+\rho}$.

The uniform moment conditions imposed in (A.1) and (A.2.i) are used to strengthen the convergence results of Kristensen (2010a) from uniform convergence over the interval $[0,T]$ with $T<\infty$ fixed to the case where $T\to\infty$. If we only wanted to show convergence for fixed $T<\infty$, these moments could be disposed of. However, we need $T\to\infty$ in order to estimate the drift function

$\alpha(\cdot)$, since it is not identified from data observed within a fixed interval, c.f. Merton (1980) and Kristensen (2010a, Theorem 5).

If the drift is zero, $\mu_t = 0$ for all t, we can choose $l_1 = l_2 = -\infty$ in (A.1). We also note that if $\{\mu_t\}$ is stationary with finite fourth moment, we can choose $p = 2$ and $l_1 = l_2 = 0$ in (A.1). The condition is however also satisfied for non-stationary sequences; an instructive example of this is a standard Brownian motion, say $\{B_t\}$: if $\mu_t = B_t$, we can choose $p = 2$ and $l_1 = l_2 = 2$.

In the case where the latent volatility process $\{\sigma_s^2\}$ is recurrent, we can easily find some examples that satisfy the condition (A.2.i): most parametric diffusion models found in the literature, including Ornstein-Uhlenbech (OU) and CIR/Feller's square-root models, are positive recurrent and stationary with the second moment finite, where we can set $m_1 = m_2 = 0$. Even when $\{\sigma_s^2\}$ is null recurrent, many processes satisfy $m_2 \leq 2$ and hence $m_1 \leq 1$ (note that $m_1 \leq m_2/2$ always holds). In fact, it is known that any diffusion process (on the whole real line) whose drift function has compact support and a (uniformly) bounded diffusion function is null recurrent (see e.g. Has'minskiĭ, 1980, Chapter IV), and m_2 is commonly equal to or less than 2 for such a diffusion process.

The condition (A.2.ii) is a smoothness condition of $\{\sigma_t^2\}$ in the L_λ-norm. A useful implication of (A.2.ii) is that it delivers bounds on the modulus of continuity of the volatility process which is defined as

$$\omega_{[0,T]}(\Delta) = \max_{s,t\in[0,T]:|t-s|\leq\Delta} \left|\sigma_t^2 - \sigma_s^2\right|.$$

The properties of the modulus of continuity of a diffusion process are well-known for the case where the time horizon is finite ($T = \bar{T} < \infty$), c.f. Revuz and Yor (1994, Theorems 1.8 and 2.1, pp. 18, 25). However, we have not been able to find any results in the literature for the long span case where $T \to \infty$, and this is exactly what is needed in our case. In Appendix A.6, we therefore establish a new result that the standard rate for the modulus of continuity can be extended to hold over an infinite time interval $[0,\infty)$. In particular, we show that $\omega_{[0,\infty)}(\Delta) = O_{a.s.}(\Delta^\gamma)$ for any $\gamma \in [0, \rho/\lambda)$ as $T \to \infty$. This result is often needed when one considers nonparametric estimators for continuous-time processes under the long span asymptotics, and should have an independent interest (see also Kanaya, 2010b for related results). The condition (A.2.ii) is automatically satisfied with $\rho = \lambda/2 - 1$ if $\{\sigma_t^2\}$ is a stationary diffusion process whose drift and diffusion functions satisfy $E[|\alpha(\sigma_t^2)|^\lambda] < \infty$ and $E[|\beta(\sigma_t^2)|^\lambda] < \infty$ for some $\lambda > 2$. These conditions are in turn satisfied for any $\lambda > 0$ if, for example, $\{\sigma_t\}$ is an OU or CIR process.

Finally, we restrict the set of feasible bandwidth sequences that can be used to estimate the trajectory of $\{\sigma_t^2\}$:

B.1 For some value $\gamma \in (0, \rho/\lambda)$, the bandwidth $h \to 0$ is chosen such that, as $T \to \infty$ and $\Delta \to 0$:

$$\frac{\Delta^{5/2} T^{3+3l_2/2}}{h^{4+3\gamma}} [\log T + \log(1/h)] = O(1), \quad \frac{\Delta T^{3+3m_2/2}}{h^{4+3\gamma}} [\log T + \log(1/h)] = O(1).$$

These two conditions are required to ensure uniform consistency: The first condition is used to control the bias due to the presence of the drift $\{\mu_t\}$; it implies that the bias incurred from this term has negligible impact in the estimation uniformly as $T \to \infty$. In particular, if the drift is not present then the first condition is automatically satisfied. The second condition of (B.1) is a strengthening of the classical condition of "rapidly increasing experimental design" normally used in the estimation of diffusion models, $\Delta T \left(= \Delta^2 n\right) \to 0$. This type of condition was originally introduced in Prakasa Rao (1988) for the parametric estimation of diffusion models, and is widely used to establish properties of diffusion estimators under infill asymptotics, $\Delta \to 0$. In our case, since we are using local estimators, we need to impose faster shrinking of the time interval Δ.

The two conditions in (B.1) are stronger than those imposed in Kristensen (2010a) where uniform convergence over a compact time interval is shown. Similarly, the conditions are also stronger than those normally found for uniform consistency of kernel estimators over an unbounded interval (of the state space, not the time). This is due to the fact that our estimation problem is much harder than standard kernel estimation problems: A standard assumption in the kernel estimation literature is that the object of interest is uniformly bounded; for example, in Kristensen (2010a) this is obtained by only giving uniform convergence over the interval $[0, T]$ for $T (< \infty)$ fixed. A similar idea is used in Renò (2008) where only results for $T < \infty$ fixed are given. In our setting, we need $T \to \infty$ since our "target" is the full trajectory $\left\{\sigma_\tau^2 : 0 < \tau < \infty\right\}$, which is required for the identification of the drift component $\alpha(\cdot)$. Therefore, if we followed a standard proof strategy, we would need to assume that $\left\{\sigma_\tau^2 : 0 < \tau < \infty\right\}$ was uniformly bounded almost surely. However, this in turn would rule out all standard volatility models since these have unbounded trajectories as $T \to \infty$. Instead, we here utilize the moment restrictions imposed in (A.2) to control the behaviour of the trajectories of $\{\mu_t\}$ and $\left\{\sigma_t^2\right\}$. This also explains why the parameters l_2 and m_2 appear in (B.1). When both $\{\mu_t\}$ and $\left\{\sigma_t^2\right\}$ are stationary, the conditions in (B.1) can be simplified (see B.1' in the next section).

Theorem 1 *Assume that (A.1)-(A.2) hold, and that $K \in \mathbb{K}(1,1)$. If h satisfies (B.1), then, for any $\gamma \in (0, \rho/\lambda)$:*

$$\sup_{\tau \in [0,T]} \left|\hat{\sigma}_\tau^2 - \sigma_\tau^2\right| = O_P\left(h^\gamma\right) + O_P\left(\Delta^{1/2} T^{\{2+p(1+m_1)/2+l_1+m_1\}/(2+p)} h^{-(2+p/2)/(2+p)}\right) \tag{10}$$

as $T \to \infty$ with $\Delta \to 0$.

Proof. See Appendix A.1. ■

Given this rate of convergence of $\hat{\sigma}_\tau^2$, we can now derive the asymptotics of the estimators in the second step.

3.2 Nonparametric Estimation of the SV Model

We here derive the asymptotic properties of the nonparametric two-step estimators of the SV model. The estimation problem in the second step can be seen as kernel estimation with errors-in-variables. The implications of this for kernel regression was analyzed in Mammen, Rothe and Schienle (2010), Rilstone (1996) and Sperlich (2009) in a cross-sectional framework while kernel density estimation of stochastic processes with errors-in-variables was considered in Blanke and Pumo (2003). We follow a similar strategy to Mammen, Rothe and Schienle (2010), Rilstone (1996) and Sperlich (2009) when analyzing the impact of the first-step estimation of $\{\sigma_t^2\}$ on the nonparametric estimators of the SV model in the second step: We split up the total estimation error into two components: One component due to the estimation of $\{\sigma_t^2\}$ in the first step, and a second component due to the sampling error of the estimator based on the actual process. For example, for the nonparametric drift estimator $\hat{\alpha}(x)$ proposed in eq. (5), we write the total estimation error as

$$\hat{\alpha}(x) - \alpha(x) = [\hat{\alpha}(x) - \tilde{\alpha}(x)] + [\tilde{\alpha}(x) - \alpha(x)], \tag{11}$$

where $\tilde{\alpha}(x)$ is the infeasible drift estimator given in eq. (2) based on observations of $\{\sigma_t^2\}$. The asymptotics of the second term follow from arguments as in Bandi and Phillips (2003) under the regularity conditions stated below. Theorem 1 is used to demonstrate that the first term converges to zero in probability at a sufficiently fast rate when the number of grid points $N \to \infty$ is chosen appropriately. The rate can be chosen so that the first term is asymptotically negligible, implying that the feasible estimator shares the asymptotic properties of the infeasible one. We impose the following conditions to ensure that this heuristic argument holds theoretically:

A.2' The process $\{\sigma_t^2\}$ has range $I = (0, \bar{\sigma})$, where $\bar{\sigma} \leq \infty$, and satisfies:

- **(i)** $\alpha(x)$ and $\beta^2(x)$ are at least twice continuously differentiable.
- **(ii)** $\beta^2(\cdot) > 0$ on I.
- **(iii)** The scale measure $S(x) = \int_c^x s(y)\,dy$, where

$$s(y) := \exp\left\{-2\int_c^y \alpha(u)\,\beta^{-2}(u)\,du\right\},$$

  for some generic constant $c \in I$, satisfies

$$S(x) \to -\infty \quad \text{as} \quad x \to 0; \quad S(x) \to \infty \quad \text{as} \quad x \to \bar{\sigma}.$$

  Also, $\int_0^r \beta^{-2}(x)\,s(x)\,dx < \infty$.
- **(iv)** $E\left[\sigma_t^4\right] < \infty$, $E\left[\left|\alpha\left(\sigma_t^2\right)\right|^\lambda\right] < \infty$ and $E\left[\left|\beta\left(\sigma_t^2\right)\right|^\lambda\right] < \infty$ for some $\lambda > 2$.

Condition (A.2') is a strengthening of (A.2) which holds under (A.2'). It is a fairly standard regularity condition that is often imposed when deriving asymptotics of diffusion estimators. (A.2'.i) and (A.2'.ii) are sufficient for the existence of a unique strong solution up to an explosion time (Karatzas and Shreve, 1991, Theorem 5.5.15 and Corollary 5.3.23). In conjunction with (A.2'.i)-(A.2'.ii), (A.2'.iii) is sufficient for the process to be nonexplosive and positive recurrent and for its invariant density to exist (see Proposition 5.5.22 of Karatzas and Shreve, 1991 and Chapter 15 of Karlin and Taylor, 1981). We will in the following let $\pi(x)$ denote the invariant density of $\{\sigma_t^2\}$, and assume that the process has been initialized at this distribution. Given these, the process is stationary and we can set $m_1 = m_2 = 0$ in (A.2). The positive recurrence condition is not strictly necessary to derive asymptotic results for our estimators. We can extend our results to null recurrent volatility processes by using arguments similar to those in Bandi and Phillips (2003). However, under the null recurrence, the convergence rates of bandwidths and time intervals become stochastic since they depend on the local time, and the required conditions and proofs become much more complicated. We therefore maintain the stationarity assumption for simplicity. The final condition, (A.2'.iv), imposes two moment conditions on the volatility process. The condition is satisfied by many volatility models, including CIR and GARCH-diffusion models. If one is only interested in estimating the drift and not the diffusion, (A.2'.iv) can be weakened to $E\left[\left|\alpha\left(\sigma_t^2\right)\right|\right] < \infty$ and $E\left[\beta\left(\sigma_t^2\right)\right] < \infty$.

Given (A.2'), we can simplify the conditions in (B.1) (by "$m_1 = m_2$"). For further simplicity, we set $p = 2$ and thus $l_1 = l_2$. Now, the conditions in (B.1) simplify to the following:

B.1' Condition (A.1) holds with $p = 2$. For some $\gamma \in (0, \rho/\lambda)$, the bandwidth $h \to 0$ is chosen such that, as $T \to \infty$ and $\Delta \to 0$: $\Delta T^{l_2} = O(1)$ and $\dfrac{\Delta T^3}{h^{4+3\gamma}}\left[\log T + \log(1/h)\right] = O(1)$.

We will in the following work with Condition (B.1') instead of (B.1) since this facilitates our subsequent analyses: In particular, under (B.1'), the second term in the right-hand side of eq. (10) is $o_P(h^\gamma)$, and the rate in Theorem 1 simplifies to

$$\sup_{\tau\in[0,T]} \left|\hat{\sigma}_\tau^2 - \sigma_\tau^2\right| = O_P(h^\gamma) \quad (\text{as } T \to \infty,\ \Delta \to 0). \tag{12}$$

Suppose now that l_2 is relatively small. Then, for any given γ, the best possible rate of $\hat{\sigma}_\tau^2$ satisfying (B.1') (in terms of T and Δ) is

$$\sup_{\tau\in[0,T]} \left|\hat{\sigma}_\tau^2 - \sigma_\tau^2\right| = O_P\left(\left(\Delta T^3 \log T\right)^{\frac{\gamma}{4+3\gamma}}\right) \quad \text{if } \ Th \to \infty.$$

If $Th = O(1)$, then the best rate can be written as $O_P(\left[\mathfrak{W}\left(1/\Delta T^3\right)\right]^{\frac{-\gamma}{4+3\gamma}})$ where $\mathfrak{W}(\cdot)$ is the Lambert W function, which is defined as the inverse function of $f(w) = we^w$. Note that $\left(\Delta T^3\right)^{\frac{\gamma}{4+3\gamma}+\varepsilon} < \left[\mathfrak{W}\left(1/\Delta T^3\right)\right]^{\frac{-\gamma}{4+3\gamma}} < \left(\Delta T^3\right)^{\frac{\gamma}{4+3\gamma}}$ for any $\varepsilon > 0$. Given the aforementioned technical difficulties, it is not surprising that the rate obtained here is slower than possible rates for fixed T (see Section 3 of Kristensen, 2010a).

In addition to (B.1'), we constrain the set of feasible bandwidths and pseudo-sampling points. The first set, (B-NDR), is used to derive the asymptotic properties of the drift estimator, while the second, (B-NDI), is employed when analyzing the ones of the diffusion estimator.

B-NDR As $T \to \infty$, $\Delta \to 0$, $N \to \infty$, $\delta \to 0$ and $b \to 0$:

- **(i)** $h^\gamma\left(b^{-1}\delta^{-1/2} + \delta^{-1}\right) \to 0$, $\delta^\gamma/b \to 0$ and $Tb \to \infty$.
- **(ii)** $h^{2\gamma}N\left(b^{-1} + b\delta^{-1}\right) \to 0$, $b^5T = O(1)$ and $Tb\delta^{2\gamma} \to 0$.

B-NDI As $T \to \infty$, $\Delta \to 0$, $N \to \infty$, $\delta \to 0$ and $b \to 0$:

- **(i)** $h^\gamma\left(b^{-1} + \delta^{-1+\gamma}\right) \to 0$, $\delta^\gamma/b \to 0$ and $Nb \to \infty$.

(ii) $h^{2\gamma}N\left(b^{-1}+b\delta^{-2+2\gamma}\right)\to 0$, $b^5N=O(1)$ and $Nb\delta^{2\gamma}\to 0$.

The conditions in (B-NDR) and (B-NDI) that do not involve h are similar to the ones imposed in Bandi and Phillips (2003) for the case of stationary diffusion processes. The additional assumptions involving h are introduced to ensure that the error due to the preliminary estimation of $\left\{\sigma_t^2\right\}$ does not affect the asymptotic properties. Roughly speaking, we need to set the first-step bandwidth h smaller than the second-step one b. Similar conditions are employed in Newey et al (1999) and Xiao et al (2003) to establish theoretical results of their two-step nonparametric estimators.

We are now ready to state the asymptotic distributions of our two-step drift and diffusion estimators:

Theorem 2 *Assume that (A.1), (A.2') and (B.1') are satisfied, $K\in\mathbb{K}(1,1)$ and $\mathcal{K}\in\mathbb{K}(2,2)$. If (B-NDR.i) holds, then $\hat{\alpha}(x)\xrightarrow{P}\alpha(x)$. Further, if (B-NDR.ii) holds, then*

$$\sqrt{Tb}\left[\hat{\alpha}(x)-\alpha(x)-b^2\times bias_{\alpha}(x)\right]\xrightarrow{d}N\left(0,\frac{\beta^2(x)}{\pi(x)}\int\mathcal{K}^2(z)\,dz\right)$$

where

$$bias_{\alpha}(x):=\frac{\partial\alpha(x)}{\partial x}\frac{\partial\pi(x)}{\partial x}/\pi(x)+\frac{1}{2}\frac{\partial^2\alpha(x)}{\partial x^2}.$$

Proof. See Appendix A.2. ■

Theorem 3 *Assume that (A.1), (A.2') and (B.1') are satisfied, $K\in\mathbb{K}(1,1)$ and $\mathcal{K}\in\mathbb{K}(2,2)$. If (B-NDI.i) holds, then $\hat{\beta}(x)\xrightarrow{P}\beta(x)$. Further, if additionally (B-NDI.ii) holds, then*

$$\sqrt{T\delta^{-1}b}\left[\hat{\beta}^2(x)-\beta^2(x)-b^2\times bias_{\beta^2}(x)\right]\xrightarrow{d}N\left(0,\frac{2\beta^4(x)}{\pi(x)}\int\mathcal{K}^2(z)\,dz\right)$$

where

$$bias_{\beta^2}(x):=\frac{\partial\beta^2(x)}{\partial x}\frac{\partial\pi(x)}{\partial x}/\pi(x)+\frac{1}{2}\frac{\partial^2\beta^2(x)}{\partial x^2}.$$

Proof. See Appendix A.3. ■

If the condition $b^5T=O(1)$ in (B-NDR.ii) is strengthened to $b^5T\to 0$, the bias component in Theorem 2 vanishes fast enough to have no impact on the asymptotic distribution. Similarly in Theorem 3, if $b^5N=o(1)$, the bias term can be ignored.

The above results show that the feasible estimators behave asymptotically in the same way as the infeasible ones based on actual observations of $\left\{\sigma_t^2\right\}$ at the pseudo-sampling points under the regularity conditions imposed. In particular, our asymptotic results do not include additional bias and variance components due to the first step in our estimation procedure, and the feasible and infeasible estimators are asymptotically equivalent. This is due to (B-NDR) and (B-NDI) respectively that ensure that the first-step estimation errors are asymptotically negligible. In finite sample, the first step will obviously have an effect on the final estimators and it would be desirable to be able to quantify these. However, we have not been able to derive explicit expression of the uniform bias and variance of $\hat{\sigma}_t^2$, and its impact on the second step. This is not special to this paper.

For example, in the literature on semiparametric two-step estimators involving kernel estimation in the first step, all theoretical results are usually stated such that the first-step bias and variance vanishes asymptotically. Similarly, the theoretical results for the two-step nonparametric estimators developed in Newey et al. (1999) and Xiao et al. (2003) do not include first-step estimation errors.

Furthermore, note that the estimation errors from the first step will be smaller than those in the second step if we set the pseudo sampling interval δ of $\{\hat{\sigma}_t^2\}$ (in estimating $\alpha(x)$ and $\beta^2(x)$) relatively larger than the observation interval Δ of $\{X_t\}$. A realistic scenario would be that intra-daily observations of $\{X_t\}$ are available. Then by choosing δ corresponding to, for example, daily frequencies, we expect that the first-step estimation error will be negligible. This is supported by Fan (2006a,b), Jiang and Knight (1999) and Phillips and Yu (2005, 2006) where it is demonstrated that the Nadaraya-Watson type estimators for (observable) diffusion processes exhibit good performance even for relatively large choices of δ, i.e., the discretization bias is less serious than other biases. In total, by choosing δ (relatively) larger than Δ, the above asymptotic distribution should be a reasonable approximation even though it neglects the first step estimation error. We will discuss the specific choice of δ in further detail in Section 4.

3.3 (Semi-) Parametric Estimation of the SV Model

We here give results for the parametric estimators of the SV model. The estimation problem in this section can be seen as a two-step semiparametric estimators where in the first step a nonparametric estimator is obtained which in turn in the second step is used to obtain a parametric estimator; for similar problems within the framework of observable diffusion processes, we refer to e.g. Bandi and Phillips (2007), Kanaya (2010a) and Kristensen (2010b).

The proof strategy is the same as in the previous section: We split up the total estimation error into two components, where the first part, due to pre-estimation of $\left\{\sigma_t^2\right\}$, is shown to be negligible asymptotically under suitable conditions on the bandwidth and pseudo-sampling points.

First, we present a set of conditions required for our estimators of the parameters in the drift term:

A-SDR **(i)** θ_1^* is an interior point of some compact subset Θ_1 of $\mathbb{R}^{d_1}$; and

$$\int_I [\alpha(x;\theta_1) - \alpha(x)]^2 \pi(x)\, dx = 0$$

if and only if $\theta_1 = \theta_1^*$.

(ii) $\alpha(x;\theta_1)$ is twice continuously differentiable in θ_1 and there exists some functions $A_k(\cdot)$, $k = 1, 2$, such that uniformly over $\theta_1 \in \Theta_1$:

$$\left|\alpha(x;\theta_1) - \alpha\left(x;\theta_1'\right)\right| \leq A_1(x)\left|\left|\theta_1 - \theta_1'\right|\right|, \quad \left|\left|\frac{\partial \alpha(x;\theta_1)}{\partial \theta_1}\right|\right| + \left|\left|\frac{\partial^2 \alpha(x;\theta_1)}{\partial \theta_1 \partial \theta_1^{\bigstar}}\right|\right| \leq A_2(x),$$

where $E\left[A_k^2\left(\sigma_s^2\right)\right] < \infty$, $k = 1, 2$.

(iii) $\alpha(x;\theta_1)$, $\partial_{\theta_1}\alpha(x;\theta_1)$ and $\partial_{\theta_1\theta_1^\star}\alpha(x;\theta_1)$ are differentiable in x for each $\theta_1 \in \Theta_1$. There exist some constants $C > 0$ and $v_1 > 0$ such that uniformly over $\theta_1 \in \Theta_1$:

$$\left\|\frac{\partial\alpha(x;\theta_1)}{\partial x}\right\| + \left\|\frac{\partial^2\alpha(x;\theta_1)}{\partial x\partial\theta_1}\right\| + \left\|\frac{\partial^3\alpha(x;\theta_1)}{\partial x\partial\theta_1\partial\theta_1^\star}\right\| \le C\left[1 + |x|^{v_1}\right],$$

and $E[|\sigma_t^2|^{2v_1}] < \infty$.

The first two conditions of (A-SDR) are standard for the parametric diffusion estimation, and are similar to those imposed in, for example, Jacod (2006), Kessler (1997) and Yoshida (1992): (A-SDR.i) ensures identification of θ_1 through the drift function. The smoothness assumptions in (A-SDR.ii) implies that the objective function and its limit are twice differentiable functions of θ_1, which in turn enable us to use a standard Taylor expansion argument for deriving the asymptotic distribution. The moment conditions are needed to ensure that the variance of the estimator is well-defined. The final condition, which may be atypical, is used to demonstrate that the error from replacing σ_t by $\hat{\sigma}_t$ in the estimation is asymptotically negligible. All the conditions are satisfied by standard volatility models such as CIR and GARCH diffusion models.

The above conditions are used to show both consistency and asymptotic normality. For consistency only, the conditions could be weakened considerably, but for simplicity we maintain the conditions for both properties.

Finally, we restrict the pseudo-sampling points and the bandwidth:

B-SDR As $T \to \infty$, $\Delta \to 0$, $N \to \infty$ and $\delta \to 0$: (i) $h^\gamma/\delta \to 0$; (ii) $\sqrt{T}\left[\delta + h^\gamma/\delta\right] \to 0$.

The conditions on the shrinking rates of the bandwidth h and the sampling time δ in (B-SDR) are simpler than the ones in (B-NDR) for the nonparametric estimation, since no smoothing parameter has to be chosen in the second step. Without the first-step estimation, the condition would simplify to $\sqrt{T}\delta \to 0$, under which the discretization error of the infeasible estimator is negligible. Given these conditions, we have the following theorem:

Theorem 4 *Suppose that (A.1), (A.2'), (B.1') and (A-SDR) are satisfied; and $K \in \mathbb{K}(1,1)$. If (B-SDR.i) holds, then $\hat{\theta}_1 \xrightarrow{P} \theta_1^*$. If additionally (B-SDR.ii) holds, then*

$$\sqrt{T}(\hat{\theta}_1 - \theta_1^*) \xrightarrow{d} N\left(0, H_1^{*-1}\Omega_1^* H_1^{*-1}\right), \quad \textit{where}$$

$$\Omega_1^* := 4E\left[\partial_{\theta_1}\alpha\left(\sigma_t^2;\theta_1^*\right)\partial_{\theta_1}\alpha\left(\sigma_t^2;\theta_1^*\right)^\star \beta^2\left(\sigma_t^2\right)\right], \quad \textit{and} \quad H_1^* := 2E\left[\partial_{\theta_1}\alpha\left(\sigma_t^2;\theta_1^*\right)\partial_{\theta_1}\alpha\left(\sigma_t^2;\theta_1^*\right)^\star\right].$$

Proof. See Appendix A.4. ■

Similarly to the nonparametric case, this theorem gives conditions under which $\hat{\theta}_1$ is first-order equivalent to the infeasible estimator, $\tilde{\theta}_1$. The shared asymptotic distribution is completely standard for estimation of ergodic diffusion models, see e.g. Sørensen (2009) or Yoshida (1992). The asymptotic variance component of the estimator can easily be estimated by replacing population moments and true values with sample versions and true values and estimated ones, respectively.

To derive properties of the estimator of the diffusion parameters, the following conditions are imposed on the SV model:

A-SDI **(i)** θ_2^* is an interior point of some compact subset Θ_2 of $\mathbb{R}^{d_2}$; and

$$\int_I \left[\beta^2(x;\theta_2) - \beta^2(x)\right]^2 \pi(x)\, dx = 0$$

if and only if $\theta_2 = \theta_2^*$.

(ii) $\beta^2(x;\theta_2)$ is twice continuously differentiable in θ_2 and there exist functions $B_k(\cdot)$, $k = 1, 2$, such that uniformly over $\theta_2 \in \Theta_2$:

$$\left|\beta^2(x;\theta_2) - \beta^2(x;\theta_2')\right| \leq B_1(x)\left\|\theta_2 - \theta_2'\right\|, \quad \left\|\frac{\partial \beta^2(x;\theta_2)}{\partial \theta_2}\right\| + \left\|\frac{\partial^2 \beta^2(x;\theta_2)}{\partial \theta_2 \partial \theta_2^\star}\right\| \leq B_2(y),$$

where $E\left[B_k^2\left(\sigma_t^2\right)\right] < \infty$, $k = 1, 2$.

(iii) $\beta^2(x;\theta_2)$, $\partial_{\theta_2}\beta^2(x;\theta_2)$ and $\partial_{\theta_2\theta_2^\star}\beta^2(x;\theta_2)$ are differentiable in x for each $\theta_2 \in \Theta_2$. There exist constants $C > 0$ and $v_2 > 0$ such that uniformly over $\theta_2 \in \Theta_2$:

$$\left\|\frac{\partial \beta^2(x;\theta_2)}{\partial x}\right\| + \left\|\frac{\partial^2 \beta^2(x;\theta_2)}{\partial x \partial \theta_2}\right\| + \left\|\frac{\partial^3 \beta^2(x;\theta_2)}{\partial x \partial \theta_2 \partial \theta_2^\star}\right\| \leq C\left[1 + |x|^{v_1}\right],$$

where $E\left[\left|\sigma_t^2\right|^{2v_2}\right] < \infty$.

The conditions imposed here on the diffusion function are analogous to the ones imposed on the drift, and we refer to the discussion following after condition (A-SDR). We impose the following conditions on the pseudo-sampling points and the bandwidth:

B-SDI $N \to \infty$, $\delta \to 0$ and $b \to 0$ satisfy: (i) $h^\gamma/\delta^{1-\gamma} \to 0$; (ii) $\sqrt{N}\left[\delta + h^\gamma/\delta^{1-\gamma}\right] \to 0$.

Again, these are similar to those for the drift estimation, except that now the rates for the estimator, bandwidth h, and the pseudo time distance are different due to the faster convergence of the diffusion estimator. We here impose the classical condition of "rapidly increasing experimental design," $\sqrt{N}\delta(= \sqrt{T\delta}) \to 0$, while for the drift estimator we only required $\sqrt{T}\delta \to 0$.

Theorem 5 *Suppose that (A.1), (A.2'), (B.1') and (A-SDI) are satisfied; and $K \in \mathbb{K}(1,1)$. If the condition (B-SDI.i) holds, then $\hat{\theta}_2 \xrightarrow{P} \theta_2^*$. If additionally (B-SDI.ii) holds, then*

$$\sqrt{N}(\hat{\theta}_2 - \theta_2^*) \xrightarrow{d} N\left(0, H_2^{*-1}\Omega_2^* H_2^{*-1}\right), \quad \textit{where}$$

$$\Omega_2^* := 8E\left[\partial_{\theta_2}\beta^2\left(\sigma_t^2;\theta_2^*\right)\partial_{\theta_2}\beta^2\left(\sigma_t^2;\theta_2^*\right)^\star \beta^4\left(\sigma_t^2\right)\right], \quad H_2^* := 2E\left[\partial_{\theta_2}\beta^2\left(\sigma_t^2;\theta_2^*\right)\partial_{\theta_2}\beta^2\left(\sigma_t^2;\theta_2^*\right)^\star\right].$$

Proof. See Appendix A.4. ■

Similarly to the parametric drift estimator, this theorem states that the two-step estimator is first-order asymptotically equivalent to the infeasible estimator $\tilde{\theta}_2$. We also note that, analogous to the nonparametric estimators, the convergence rate of the diffusion estimator is faster than that of the drift estimator. Again, the two matrices Ω_2^* and H_2^* can be estimated by standard moment estimators.

Note again that the asymptotic distributions obtained in Theorems 4 and 5 are not affected by the first step estimation errors as long as the stated conditions on δ in (B-SDR) and (B-SDI) are satisfied. While these conditions require large δ (relatively to Δ, or small N relatively to n), a natural choice might be setting $\delta = \Delta$ (or equivalently $N = n$), which is often used in semi-parametric two-step estimation or generated-regressor problems (see, e.g., Kanaya, 2010a, and Kristensen, 2010b). Small δ (and large pseudo sample size N) might enhance efficiency in the second step. We discuss this point in the following.

Consider first the drift estimation: As can be seen in Theorem 4, the normalization factor for the asymptoptic distribution is given as $\sqrt{T}$, which is independent of a specific choice of δ (and N). Thus, in view of the convergence rate of $\hat{\theta}_1$, there is no gain by using small δ (such as $\delta = N$). Again, the variance of the limit distribution in Theorem 4 is the same as the one obtained when the true volatility process σ_τ^2 is observable, which is guaranteed by using relatively large δ. Therefore, we can further say that as far as the first-order asymptotic approximation is concerned, there is no loss by using relatively large δ and estimated $\hat{\sigma}_\tau$ (instead of σ_τ^2) in estimating θ_1.

A situation is somewhat different in the diffusion estimation: The convergence rate of the estimator $\hat{\theta}_2$ is directly affected by that of δ (N). Thus, choosing large N (e.g., $N = n$) might improve the convergence rate of $\hat{\theta}_2$. However, in order to obtain the convergence and distributional result under such N, we need to quantify the impacts of the (uniform) bias and variance of $\hat{\sigma}_t^2$ on the second step. As in the nonparametric case, we have not been able to quantify them. We also note that it is generally uncertain if we could obtain the asymptotic normality under relatively large N such as $N = n$.

The above theoretical results are similar to the ones obtained in Todorov (2009), where estimators of the integrated volatility is used in the estimation of SV models: He gives conditions under which the first-step estimation error from using estimated integrated volatilities instead of the actual ones does not affect the asymptotic distribution of his parametric GMM estimators, but refrain from a higher-order analysis of the impact of the first-step estimation error.

4 Bandwidth Selection and Sampling

All of the estimators analyzed in Sections 3.2 and 3.3 involve nuisance parameters in the form of bandwidths and/or pseudo-sampling intervals. We here discuss how these should be selected in practice. The purpose here is to propose practical working rules. As such, we only provide an informal analysis since a full theoretical description would be quite involved and outside the scope of this study. Some of the proposed selection rules may give shrinking rates of bandwidths or sampling intervals which violate some of the conditions stated for Theorems 2-5 to hold. However, it seems difficult to obtain simple data-driven selection rules which are formally consistent with the theoretical conditions. This is often the case in the literature on non- and semiparametric two-step estimators.[1]

[1] For example, Xiao, Linton, Carroll and Mammen (2003) consider bandwidth selection rules in their simulation study that work well in practice, but do not satisfy conditions imposed in deriving their asymptotic results.

To compute the first-step volatility estimator, a natural, data-driven bandwidth selection method is cross-validation. Kristensen (2010a) argues that the following cross-validation criteria should lead to asymptotically optimal bandwidths: $h_{CV} = \arg\min_{h>0} CV_{\sigma^2}(h)$, where

$$CV_{\sigma^2}(h) = \sum_{i=1}^{n} \mathbb{I}\{T_l \leq t_i \leq T_u\} \left[\left(X_{t_{i+1}} - X_{t_i} \right)^2 / \Delta - \hat{\sigma}^2_{-i,t_i} \right]^2,$$

for some $0 \leq T_l < T_u \leq T$, and where $\hat{\sigma}^2_{-i,t_i}$ is the leave-one-out estimator. This criterion is tailored to minimize the integrated squared error of the volatility estimator, $\int_{T_u}^{T_l} \left[\sigma_t^2 - \hat{\sigma}_t^2 \right]^2 dt$. Since the end goal is to obtain precise estimates of the SV model, we should indeed choose h to optimize some criterion for the second step estimator (e.g., the mean squared error of $\hat{\alpha}(x)$, $\hat{\beta}^2(x)$, $\hat{\theta}_1$ or $\hat{\theta}_2$). In this respect, h_{CV} is not ideal. According to the theoretical results, undersmoothing appears to be required, so we recommend that one chooses an initial bandwidth by cross-validation which in turn is scaled down by an appropriate factor.

Once $\left\{ \hat{\sigma}_t^2 \right\}$ has been obtained, we have to choose an additional bandwidth b and a (pseudo-) sampling frequency N (or equivalently δ) for the computation of the drift and diffusion estimates. We here propose to choose $\delta > 0$ at a daily frequency such that we use daily (estimated) volatilities in the second step of our estimation procedure. The primary reason for this choice is that in practice the volatility is known to have intradaily seasonal patterns; by choosing daily frequencies in the second step, we can ignore these in the estimation. Secondly, by choosing δ to correspond to daily observations, we hope that the additional time series dependence in $\left\{ \hat{\sigma}_t^2 \right\}$ due to the first-step estimation is controlled so that the second-step estimation error dominates (as is the case in our theoretical results).

Given the choice of $\delta > 0$, we also propose to use cross-validation in the second step; the precise procedure is described in Kanaya and Kristensen (2010) who develop bandwidth selection procedures for diffusion processes. Their results assume uncontaminated observations of the diffusion process, but we expect that with δ chosen at a daily frequency, the estimation error in $\left\{ \hat{\sigma}^2_{i\delta} : i = 1, ..., N \right\}$ can be ignored.

The semiparametric estimators only require the choice of the first-step bandwidth, h, and the second-step sampling frequency, δ. Given that our estimation strategy corresponds to a two-step semiparametric estimation procedure, we expect in general that undersmoothing should be used in the first-step. Regarding the choice of δ, we now briefly analyze how this impacts on the MSE's of the parametric estimators. For this purpose, we also assume that the error in the first step estimation can be ignored and consider the MSE of the infeasible estimators based on the objective functions in eq. (3). First, the MSE for the estimator of the drift parameter, $\tilde{\theta}_k$, is given by

$$\mathrm{MSE}_{\theta_k} := E\left[\left(\tilde{\theta}_k - \theta_k^* \right)^{\star} \left(\tilde{\theta}_k - \theta_k^* \right) \right], \quad k = 1, 2.$$

Using the standard Taylor expansion,

$$\tilde{\theta}_k - \theta_k^* = - \left[H_k^* + o_P(1) \right]^{-1} \hat{S}_k \left(\theta_k^*, \sigma^2 \right), \tag{13}$$

for $k=1,2$, where $\hat{S}_k$ is the score function (see the Appendix A.4) and H_k^* is the limit of the Hessian function evaluated at the true value θ_k^* (given in Theorems 4 and 5). Since it is not easy to directly analyze MSE_{θ_k}, we consider its approximation $\mathrm{MSE}^*_{\theta_k}$, which we construct by following the same strategy as in analyzing the MSE of nonparametric estimators (see Sec. 3.3 of Pagan and Ullah, 1999 and Kanaya and Kristensen, 2010). Define

$$\begin{aligned}\mathrm{MSE}^*_{\theta_k} \quad : \quad &= E\left[\left(H_k^{*-1}\hat{S}_k\left(\theta_k^*,\sigma^2\right)\right)^{\bigstar} H_k^{*-1}\hat{S}_k\left(\theta_k^*,\sigma^2\right)\right] \\ &= \mathrm{tr}\left\{H_k^{*-1}E\left[\hat{S}_k\left(\theta_k^*,\sigma^2\right)\hat{S}_k\left(\theta_k^*,\sigma^2\right)^{\bigstar}\right]H_k^{*-1}\right\}, \end{aligned} \tag{14}$$

for $k=1,2$, where $\mathrm{tr}\{A\}$ is the sum of diagonal elements of the matrix A. By the same arguments as in the nonparametric case, this should be a good approximation to MSE_{θ_k}. Our semi-parametric $\tilde{\theta}_k$ has no smoothing bias unlike the nonparametric estimators (see Kanaya and Kristensen, 2010), and therefore we can decompose $\mathrm{MSE}^*_{\theta_k}$ into three terms:

$$\mathrm{MSE}^*_{\theta_k} = \mathrm{tr}\left\{\mathcal{B}_{\theta_k}\mathcal{B}_{\theta_k}^{\bigstar}\right\} + \mathrm{tr}\left\{\mathcal{V}_{\theta_k}\right\} + \mathrm{tr}\left\{\mathcal{C}_{\theta_k}\right\} \tag{15}$$

where $\mathcal{B}_{\theta_k,\mathrm{ds}}$ is the discretization bias and $\mathcal{V}_{\theta_k}$ and $\mathcal{C}_{\theta_k}$ are the variance and covariance components respectively; see Appendix A.5 for their finite-sample expressions. We derive asymptotic expressions of these terms under the following assumptions:

C-SDR The functions $||\partial_{\theta_1}\alpha\left(x;\theta_1^*\right)||$, $||\partial_{x\theta_1}\alpha\left(x;\theta_1^*\right)||$, $||\partial_{xx\theta_1}\alpha\left(x;\theta_1^*\right)||$, $|\alpha\left(x\right)|$, $|\alpha'\left(x\right)|$, $|\alpha''\left(x\right)|$ and $\beta^2\left(x\right)$ are all bounded by some function $\psi\left(x\right)$ satisfying $E\left[\left|\psi\left(\sigma_s^2\right)\right|^6\right]<\infty$.

C-SDI The functions $\left|\left|\partial_{\theta_2}\beta^2\left(x;\theta_2^*\right)\right|\right|$, $\left|\left|\partial_{x\theta_2}\beta^2\left(x;\theta_2^*\right)\right|\right|$, $\left|\left|\partial_{xx\theta_2}\beta^2\left(x;\theta_2^*\right)\right|\right|$, $\beta^2\left(x\right)$, $\left|\partial_x\beta^2\left(x\right)\right|$, $\left|\partial_{xx}\beta^2\left(x\right)\right|$ and $|\alpha\left(x\right)|$ are all bounded by some function $\psi\left(x\right)$ satisfying $E\left[\left|\psi\left(\sigma_s^2\right)\right|^6\right]<\infty$.

Theorem 6 *Suppose that (A.2') holds.*
(i) If (A-SDR.i), (A-SDR.ii) and (C-SDR) are satisfied, then it holds that

$$\mathcal{B}_{\theta_k} = \delta\bar{\mathcal{B}}_{\theta_k} + o\left(\delta\right), \quad \mathcal{C}_{\theta_k} = O\left(\delta^2\right), \tag{16}$$

with $k=1$ and

$$\bar{\mathcal{B}}_{\theta_1} := E\left[\partial_{\theta_1}\alpha\left(\sigma_t^2;\theta_1^*\right)\left[\alpha'\left(\sigma_t^2\right)\alpha\left(\sigma_t^2\right)+\alpha''\left(\sigma_t^2\right)\beta^2\left(\sigma_t^2\right)/2\right]\right],$$

and also

$$\mathcal{V}_{\theta_1} = \frac{1}{T}\mathrm{tr}\left\{H_1^{*-1}\Omega_1^*H_1^{*-1}\right\} + o\left(\frac{1}{T}\right). \tag{17}$$

(ii) If (A-SDI.i), (A-SDI.ii) and (C-SDI) are satisfied, then eqs. in (16) hold with $k=2$ and

$$\bar{\mathcal{B}}_{\theta_2} := E\left[\partial_{\theta_2}\beta^2\left(\sigma_t^2;\theta_2^*\right)\left[\partial_x\beta^2\left(\sigma_t^2\right)\alpha\left(\sigma_t^2\right)+\partial_{xx}\beta^2\left(\sigma_t^2\right)\beta^2\left(\sigma_t^2\right)/2\right]\right],$$

and also

$$\mathcal{V}_{\theta_2} = \frac{1}{n}\mathrm{tr}\left\{H_2^{*-1}\Omega_2^*H_2^{*-1}\right\} + o\left(\frac{1}{n}\right).$$

Proof. See Appendix A.5. ■

Thus, we have

$$\mathrm{MSE}^*_{\theta_1} = O\left(\delta^2\right) + O\left(1/T\right); \quad \text{and} \quad \mathrm{MSE}^*_{\theta_2} = O\left(\delta^2\right) + O\left(1/n\right) = O\left(\delta^2\right) + O\left(\delta/T\right).$$

These expressions are consistent to those derived in Tang and Chen (2009), who consider the estimation of diffusion processes with the linear drift function. From this theorem, we see that the optimal choice of δ (for any given T) is always to let it shrink to zero at the fastest possible rate. Again, in practice we will however use the daily frequency in the second step since it allows us to ignore intradaily patterns in the volatility.

5 A Simulation Study

We here examine the performance of our non- and semi-parametric estimators. We assume that the following stochastic volatility model is a data-generating process:

$$\begin{cases} dX_t = \sigma_t dW_t; \\ d\sigma_t^2 = \beta\left(\alpha - \sigma_t^2\right) dt + \kappa\sigma_t^2 dZ_t, \end{cases} \tag{18}$$

where W_t and Z_t are independent standard Brownian motions. This is the continuous-time limit version of the GARCH model as derived in Nelson (1990) and Drost and Werker (1996), and satisfies the conditions imposed in Section 3. We measure time in days and consider the following two sample frequencies: $\Delta^{-1} = 60 \times 24$ and 12×24 which correspond to sampling every 1 and 5 minutes respectively. We choose the parameter values as $\alpha = 0.476$, $\beta = 0.510$ and $\kappa^2 = 0.0518$, and the time span as $T = 3 \times 250$ days which roughly corresponds to 3 year with 250 business days per year. In order to simulate data from the model, we employ the Euler discretisation scheme (see Kloeden and Platten, 1999),

$$\begin{cases} \Delta X_{id} = \sigma_{(i-1)d}\sqrt{d}\varepsilon_{1,i}; \\ \Delta\sigma_{id}^2 = \beta(\alpha - \sigma_{(i-1)d}^2)d + \kappa\sigma_{(i-1)d}^2\sqrt{d}\varepsilon_{2,i}, \end{cases}$$

where $\{\varepsilon_{1,i}\}$ and $\{\varepsilon_{2,i}\}$ are i.i.d. $N(0,1)$ with $\{\varepsilon_{1,i}\}$ and $\{\varepsilon_{2,i}\}$ independent. Here, $d > 0$ is the length of the discretization step; it is chosen as $d = \underline{\Delta}/100$, where $\underline{\Delta}^{-1} = 60 \times 24$ corresponds to the highest sampling frequency used in the simulation study.

Throughout, we implement the first-step kernel estimator of σ_t^2 using a Gaussian kernel. The bandwidth h is chosen as $h = 0.10$ for $\Delta^{-1} = 60 \times 24$ and $h = 0.14$ for $\Delta^{-1} = 12 \times 24$. These two bandwidth choices were found by running the standard cross-validation procedure described in Section 4 for five trial Monte Carlo samples yielding h_i^*, $i = 1, ..., 5$. For all the subsequent Monte Carlo samples that our simulation study is based on, we then fixed the bandwidth at the average across these five cross-validated bandwidth choices divided by two, $h = \bar{h}^*/2$, and are in effect undersmoothing in the first step. The reason for not running the cross-validation procedure for each sample is that the procedure is rather time-consuming.

In the second step, we have to choose the pseudo-sampling frequency, δ, for both the non- and semiparametric estimator. We here experiment with three different choices: In the case where $\Delta^{-1} = 60 \times 24$, we chose $\delta = 1/8$, $\delta = 1/4$ and $\delta = 1/2$, and for $\Delta^{-1} = 12 \times 24$, we chose $\delta = 1/2$, $\delta = 1$ and $\delta = 2$. Here, $\delta = 1/2$, for example, corresponds to two pseudo-observations per day. For the nonparametric estimator, we also have to choose a second kernel, $\mathcal{K}$, and bandwidth, b. The kernel $\mathcal{K}$ was chosen as the Gaussian one. As with the bandwidth choice for b, we also here ran cross-validation procedure for five trial samples and then fixed the bandwidth b at the average over these cross-validated bandwidths. Again, this was done in order to speed up the simulation study.

As noted earlier, our two-step estimators suffer from double sampling error: One component is due to the sample variation in the unobserved process $\left\{\sigma_t^2\right\}$, and a second one due to only observing $(\Delta X_t)^2/\Delta$ which is a contaminated version of σ_t^2. In order to evaluate how much of the resulting sampling error is due to the contamination, we also computed the corresponding infeasible estimators using the actual values of $\sigma_\delta^2, \sigma_{2\delta}^2, \sigma_{3\delta}^2, \ldots$.

To evaluate the performance of the nonparametric estimators, we computed approximate integrated bias, variance and MSE for $x = [0.3, 0.8]$ (the volatility process spent 95% of the time within that interval in our samples). The integrated squared bias of the drift estimates was estimated by $\mathrm{BIAS}^2 = \int_{0.3}^{0.8} [\alpha(x) - \bar{\alpha}(x)]^2 dx$ where $\bar{\alpha}(x) = \frac{1}{S}\sum_{s=1}^{S} \hat{\alpha}_s(x)$ and $\hat{\alpha}_s(x)$ was the estimated drift in the s-th sample over $S(=400)$ Monte Carlo replications we generated. Similarly, the integrated variance and MSE were estimated by $\mathrm{VAR} = \frac{1}{S}\sum_{s=1}^{S}\int_{0.3}^{0.8} [\hat{\alpha}_s(x) - \bar{\alpha}(x)]^2 dx$ and $\mathrm{MSE} = \frac{1}{S}\sum_{s=1}^{S}\int_{0.3}^{0.8} [\hat{\alpha}_s(x) - \alpha(x)]^2 dx = \mathrm{BIAS}^2 + \mathrm{VAR}$.

In Table 1, we report integrated squared bias, variance and MSE of the drift and diffusion estimators for the first sampling scheme, $\Delta^{-1} = 60 \times 24$. In column 1 and 2, the performance of the infeasible and feasible nonparametric drift estimator is reported. As predicted by theory, the performance of the infeasible estimator deteriorates as the sampling frequency δ^{-1} decreases. As expected, this is not in general the case for the feasible two-step estimator however: Too small or too large choices of δ^{-1} yield poor estimates; here, $\delta = 1/4$ gives the best performance of the three different choices. A similar pattern is found in column 3 and 4 where the results of the diffusion estimators are reported: The MSE of the infeasible diffusion estimator increases with δ, while the feasible one performs best at the intermediate choice of $\delta = 1$.

In Figure 1-4, we have plotted the pointwise means of the infeasible and feasible estimators for $\delta = 1/4$ together with their 95% confidence intervals. The plots mirror the results of Table 1 with little difference between the 1-step and 2-step estimators which is rather encouraging.

In Table 2, we report the same results but now for the second sampling scheme, $\Delta^{-1} = 12 \times 24$. In general the performance of the feasible estimator is worse due to less precise estimates of $\left\{\sigma_t^2\right\}$ in the first step. To control the added estimation error in the second step, we here have chosen $\delta = 1/2$, $\delta = 1$ and $\delta = 2$ in the second step. The same picture appears as for the higher frequency. Again, the intermediate choice of $\delta = 1$ yields the most precise estimates with a too low or too high choices of δ reducing precision.

Two results of the simulation study that may seem surprising are: First, the 2-step estimators outperform the 1-step ones in some cases ($\Delta^{-1} = 60 \times 24$ and $\delta = 1/4$; $\Delta^{-1} = 12 \times 24$ and

$\delta = 1$). This seems to indicate that the pre-smoothing of data actually improves on the performance of the Nadraya-Watson estimators in some cases. Second, the MSE of the drift estimator with $\Delta^{-1} = 12 \times 24$ and $\delta = 1$ is lower than the one with $\Delta^{-1} = 60 \times 24$ and $\delta = 1/4$. This is most likely due to the fact that the bandwidths h and b in our simulation study have been chosen in a rather ad hoc manner. It further emphasizes the importance of developing good, data-driven bandwidth selection procedures for our estimators.

We next analyze the finite-sample performance of the parametric estimators. We maintain the SV model in eq. (18) as the DGP with the same parameter values. For this model, the parametric least-squares criteria of eqs. (7)-(8) are

$$\hat{Q}_1(\theta_1) = \sum_{j=1}^{N-1} \left[\Delta\hat{\sigma}^2_{\tau_{j+1}} - \beta\left(\alpha - \hat{\sigma}^2_{\tau_j}\right)\delta\right]^2, \quad \hat{Q}_2(\theta_2) = \sum_{j=1}^{N-1} \left[\left(\Delta\hat{\sigma}^2_{\tau_{j+1}}\right)^2 - \kappa^2\hat{\sigma}^4_{\tau_j}\delta\right]^2,$$

and the corresponding estimators can be written in the closed form:

$$\hat{\alpha} = -\frac{a_{\mathrm{OLS}}}{b_{\mathrm{OLS}}}, \quad \hat{\beta} = -b_{\mathrm{OLS}}, \quad \hat{\kappa}^2 = c_{\mathrm{OLS}},$$

where, with $X_{\tau_j} = (1, \hat{\sigma}^2_{\tau_j})^\star$,

$$\begin{pmatrix} a_{\mathrm{OLS}} \\ b_{\mathrm{OLS}} \end{pmatrix} = \frac{1}{\delta}\left(\sum_{j=1}^{N-1} X_{\tau_j}X'_{\tau_j}\right)^{-1}\left(\sum_{j=1}^{N-1} X_j\Delta\hat{\sigma}^2_{\tau_{j+1}}\right), \quad c_{\mathrm{OLS}} = \frac{1}{\delta}\left(\sum_{j=1}^{N-1}\hat{\sigma}^8_{\tau_j}\right)^{-1}\left(\sum_{j=1}^{N-1}\hat{\sigma}^4_{\tau_j}\left(\Delta\hat{\sigma}^2_{\tau_{j+1}}\right)^2\right).$$

Tables 3 and 4 report results for the cases $\Delta^{-1} = 60 \times 24$ and $\Delta^{-1} = 12 \times 24$ respectively. For both sampling frequencies, we chose, after some experimentation, three pseudo-sampling frequencies, $\delta = 1/12$, $1/6$ and $1/4$. We here note that we use smaller pseudo frequencies compared to the nonparametric case. It appears as if parametric estimators are less affected by the first-step error, such that we can choose a smaller δ.

In contrast to the nonparametric estimators, the infeasible estimators outperform our 2-step estimators in all cases. Otherwise, patterns similar to those for the nonparametric estimators appear: First, more data available in the first step ($\Delta = 1/(24 \times 60)$ versus $\Delta = 1/(24 \times 12)$) improves the quality of the spot volatility estimator which in turn leads to better performance of the final estimators. Second, a small level of δ is not necessarily optimal; for example, with $\Delta = 1/(24 \times 12)$, the estimation results based on $\delta = 1/6$ generally outperform the ones using $\delta = 1/12$. Otherwise, the performance of the parametric estimators are somewhat mixed across the different parameters. The long-run level, α, is estimated consistently well across all sampling schemes and is close to the infeasible estimator based on observing the volatility process. On the other hand, relatively large biases are incurred when implementing our estimator for the mean-reversion parameter, β: For example, in the case with $\Delta = 1/(24 \times 12)$ and $\delta = 1/6$, the smallest squared bias of our estimator is 14.4633×10^{-4} compared to 2.3990×10^{-4} for the infeasible estimator. Finally, the performance of our estimator of κ^2 falls somewhere in between these two cases.

6 Conclusion and Extensions

We have proposed a method for the estimation of SV models in the presence of high-frequency data. The asymptotic properties of the estimator were derived and their finite-sample precision examined in a simulation study. It is of interest to extend our estimation method in a number of directions; we discuss these below.

In the observation equation, it would be of interest to allow for market microstructure noise and for jumps. The presence of these would affect the performance of the kernel filter $\hat{\sigma}_t^2$. In Kristensen (2010a), methods to handle noise and jumps in the kernel filtering are proposed; these could without problems be used also in our context. However, the asymptotic and finite-sample impact of using these different first-step estimators should be analyzed.

In the state equation for σ_t^2, it could also be of interest to allow for jumps. Again, Kristensen (2010a) discusses how these can be handled in the pre-estimation of σ_t^2. This estimator could then be combined with the nonparametric estimation procedure for jump-diffusions proposed in Bandi and Nguyen (2003); see also Bandi and Renò (2008).

Finally, our theoretical results ignore the first-step sampling error. It would be useful to extend our asymptotic results to include both first- and second-step sampling errors. A first step in this direction has been made by Mammen, Rothe and Schienle (2010) in a cross-sectional setting.

References

Altissimo, F. & A. Mele (2009) Simulated Nonparametric Estimation of Dynamic Models. *Review of Economic Studies* 76, 413-450.

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 & F.X. Diebold (2010) Parametric and Nonparametric Measurement of Volatility. In *Handbook of Financial Econometrics*, vol. 1 (eds. Y. Aït-Sahalia & L.P. Hansen), pp. 67-137. Amsterdam: North Holland.

Andersen, T.G. & J. Lund (1997) Estimating Continuous-Time Stochastic Volatility Models of the Short Term Interest Rate. *Journal of Econometrics* 77, 343-377.

Bandi, F.M. & T. Nguyen (2003) On the functional estimation of jump-diffusion models. *Journal of Econometrics* 116, 293-328.

Bandi, F.M. & P.C.B. Phillips (2003) Fully Nonparametric Estimation of Scalar Diffusion Models. *Econometrica* 71, 241-283.

Bandi, F.M. & P.C.B. Phillips (2007) A simple approach to the parametric estimation of potentially nonstationary diffusions. *Journal of Econometrics*, 137, 354-395.

Bandi, F.M. & R. Renò (2009) Nonparametric Stochastic Volatility. Manuscript, Graduate School of Business, University of Chicago.

Barndorff-Nielsen, O.E. & N. Shephard (2001) Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society, Series B* 64, 253-280.

Blanke, D. & B. Pumo (2003) Optimal Sampling for Density Estimation in Continuous Time. *Journal of Time Series Analysis* 24, 1-23.

Bollerslev, T. & H. Zhou (2002) Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility. *Journal of Econometrics* 109, 33-65.

Brownlees, C., D. Kristensen & Y. Shin (2010) Smooth Filtering and Maximum Likelihood Estimation of Dynamic Latent Variables Models. Manuscript, Department of Economics, Columbia University.

Chacko, G. & L.M. Viceira (2003) Spectral GMM Estimation of Continuous-Time Processes. *Journal of Econometrics* 116, 259-292.

Chib, S., M.K. Pitt & N. Shephard (2006) Likelihood based inference for diffusion driven models. Manuscript, Department of Economics, University of Oxford.

Comte, F. ,V. Genon-Catalot & Y. Rozenholc (2009) Nonparametric Adaptive Estimation for Integrated Diffusions. *Stochastic Processes and their Applications* 119, 811-834.

Comte, F. & E. Renault (1996) Long memory continuous time models, *Journal of Econometrics* 73, 101-149.

Corradi, V. & W. Distaso (2006) Semiparametric Comparison of Stochastic Volatility Models via Realized Measures. *Review of Economic Studies* 73, 635-677.

Dacunha-Castelle, D. & D. Florens-Zmirou (1986) Estimation of the Coefficients of a Diffusion from Discrete Observations. *Stochastics* 19, 263-284.

Drost, F.C. & B.J.M. Werker (1996) Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics* 74, 31-57.

Dzhaparidze, K & J.H. van Zanten (2001) On Bernstein-type inequalities for martingales. *Stochastics Process and their Applications* 93, 109-117.

Fan, J. (2006a) A Selective Overview of Nonparametric Methods in Financial Econometrics, *Statistical Science*, 20, 317-337.

Fan, J. (2006b) Rejoinder: A Selective Overview of Nonparametric Methods in Financial Econometrics, *Statistical Science*, 20, 351-357.

Florens-Zmirou, D. (1989) Approximate Discrete Time Schemes for Statistics of Diffusion Processes. *Statistics* 20, 547-557.

Florens-Zmirou, D. (1993) On Estimating the Diffusion Coefficient from Discrete Observations. *Journal of Applied Probability* 30, 790-804.

Gallant, A. R., D. Hsieh, & G. Tauchen (1997) Estimation of Stochastic Volatility Models with Diagnostics. *Journal of Econometrics* 81, 159–192.

Jiang, G.J. & J.L. Knigh (1999) Finite Sample Comparison of Alternative Estimators of Itô Diffusion Processes: A Monte Carlo Study. *Journal of Computational Finance* 2, 5–38.

Has'minskiĭ, R.Z. (1980) *Stochastic Stability of Differential Equations,* Stijthoff & Noordoff International Publishers B.V., Alphen aan denRijn.

Heston, S. (1993) A Closed-form Solutions for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies* 6, 327–343.

Hull, J.C. & A. White (1987) The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance* 42, 281-300.

Jacod, J. (2006) Parametric Inference for Discretely Observed Non-Ergodic Diffusions. *Bernoulli* 12, 383-401.

Jiang, G.J. & Knight, J.L. (1997) A Nonparametric Approach to the Estimation of Diffusion Processes, with An Application to A Short-term Interest Rate Model. *Econometric Theory* 13, 615-645.

Kanaya, S. (2010a) Semi-parametric Maximum-Likelihood Estimation for Diffusion Processes. Manuscript, Department of Economics, University of Oxford.

Kanaya, S. (2010b) Uniform Convergence Rates of Kernel-Based Nonparametric Estimators for Continuous Time Diffusion Processes: A Damping Function Approach. Manuscript, Department of Economics, University of Oxford.

Kanaya, S. & D. Kristensen (2010) Optimal Sampling and Bandwidth Selection for Nonparametric Kernel Estimators of Diffusion Processes. Manuscript, Department of Economics, University of Oxford.

Karlin, S. & H.M. Taylor (1981) *A Second Course in Stochastic Processes*, Academic Process, New York.

Karatzas, I. & S.E. Shreve (1991) *Brownian Motion and Stochastic Calculus*. Second edition, Springer.

Kessler, M. (1997) Estimation of an Ergodic Diffusion from Discrete Observations. *Scandinavian Journal of Statistics* 24, 211-229.

Kloeden, P.E. & E. Platen (1999) Numerical Solution of Stochastic Differential Equations, 2nd ed. Springer-Verlag.

Kristensen, D. (2010a) Nonparametric Filtering of the Realized Spot Volatility: A Kernel-based Approach. *Econometric Theory* 26, 60-93.

Kristensen, D. (2010b) Pseudo-Maximum Likelihood Estimation in Two Classes of Semiparametric Diffusion Models. *Journal of Econometrics* 156, 239-259.

Mammen, E., C. Rothe & M. Schienle (2010) Nonparametric Regression with Nonparametrically Generated Covariates. Working paper, Toulouse School of Economics.

Merton, R. (1980) On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics* 8, 323-361.

Nelson, D.B. (1990) ARCH models as diffusion approximations. *Journal of Econometrics* 45, 7-39.

Newey, W.K., J.L. Powell & F. Vella (1999) Nonparametric Estimation of Triangular Simultaneous Equations Models. *Econometrica* 67, 565-603.

Newey, W.K. & D. McFadden (1994) Large Sample Estimation and Hypothesis Testing. *Handbook of Econometrics* Vol. IV, 2111-2245 (R. Engle and D. McFadden, eds.), North Holland.

Pagan, A. & A. Ullah (1999) *Nonparametric Econometrics*, Cambridge University Press.

Phillips, P.C.B & Yu, J. (2005) Jackknifing Bond Option Prices. *The Review of Financial Studies*, 18, 707-742.

Phillips, P.C.B & Yu, J. (2006) Comment: A Selective Overview of Nonparametric Methods in Financial Econometrics, *Statistical Science*, 20, 338-343.

Prakasa Rao, B.L.S (1988) Statistical Inference from Sampled Data for Stochastic Processes. *Contemporary Mathematics*, 80 249-284. American Mathematical Society, Providence, RI.

Renò, R. (2006) Nonparametric Estimation of Stochastic Volatility Models. *Economics Letters* 90, 390-395.

Renò, R. (2008) Nonparametric Estimation of the Diffusion Coefficient of Stochastic Volatility Models. *Econometric Theory* 24, 1174-1206.

Revuz, D. & M. Yor (1999) *Continuous Martingales and Brownian Motion*. Third edition, Springer.

Rilstone, P. (1996) Nonparametric Estimation of Models with Generated Regressors. *International Economic Review* 37, 299-313.

Sperlich, S. (2009) A Note on Nonparametric Estimation with Predicted Variables. *Econometrics Journal* 12, 382-395.

Shephard, N. (2005) *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Sørensen, M. (2009) Efficient Estimation for Ergodic Diffusions Sampled at High Frequency. Manuscript, Department of Mathematical Sciences, University of Copenhagen.

Tang, C.Y. & S.X. Chen (2009) Parameter Estimation and Bias Correction for Diffusion Processes. *Journal of Econometrics* 149, 65-81.

Todorov, V. (2009) Estimation of Continuous-Time Stochastic Volatility Models with Jumps Using High-Frequency Data. *Journal of Econometrics* 148, 131-148.

Xiao, Z. O.B. Linton, R.J. Carroll & E. Mammen (2003) More Efficient Local Polynomial Estimation in Nonparametric Regression With Autocorrelated Errors. *Journal of the American Statistical Association* 98, 980-992.

Yoshida, N. (1992) Estimation for Diffusion Processes from Discrete Observations. *Journal of Multivariate Analysis* 41, 220-242.

Zhang, S & R.J. Karunamuni (1998) On Kernel Density Estimation near Endpoints. *Journal of Statistical Planning and Inference* 70, 301-316.

A Proofs

A.1 Proof of Theorem 1: Spot Volatility Estimator

As noted in the main text, our proof of the uniform convergence result will rely on almost sure Hölder continuity of the process $t \mapsto \sigma_t^2$ uniformly over the infinite time interval $[0, \infty)$. The following lemma shows that indeed the spot volatility process has this property under weak conditions. The lemma follows from a more general result on uniform Hölder continuity of stochastic processes stated in Appendix A.6.

Lemma 1 *Suppose that (A.3) holds. Then, for any $\gamma \in (0, \rho/\lambda)$, there exists some constant $D\,(>0)$ such that*

$$\Pr\left[\omega \in \Omega \left| \exists \bar{\Delta}(\omega) \;\; s.t. \sup_{|t-s| \in \left(0, \bar{\Delta}(\omega)\right);\; s,t \in [0,\infty)} \frac{\left|\sigma_t^2(\omega) - \sigma_s^2(\omega)\right|}{|t-s|^{\gamma}} \leq D \right.\right] = 1. \tag{19}$$

Proof. From (A.3) and Lemma 9, there exists a continuous modification $\{\tilde{\sigma}_t^2\}$ of $\left\{\sigma_t^2\right\}$ which is a.s. Hölder globally over $[0, \infty)$. Identifying $\left\{\sigma_t^2\right\}$ with $\{\tilde{\sigma}_t^2\}$, we have eq. (19). ■

Next, we expand the spot volatility estimator and analyze each of the terms in this expansion. In what follows, we extend the processes $\{\mu_t\}$ and $\left\{\sigma_t^2\right\}$ by letting $\mu_t = \sigma_t^2 = 0$ if $t < 0$. By Ito's lemma for continuous semimartingales,

$$\begin{aligned}
\left(\Delta X_{t_i}\right)^2 &= 2\int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \mu_u du + \int_{t_{i-1}}^{s} \sigma_u dW_u\right) \mu_s ds \\
&\quad +2\int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \mu_u du + \int_{t_{i-1}}^{s} \sigma_u dW_u\right) \sigma_s dW_s + \int_{t_{i-1}}^{t_i} \sigma_s^2 ds.
\end{aligned}$$

Thus, the left-hand side of (10) is bounded by

$$\sup_{\tau \in [0,T]} \left|\hat{\sigma}_\tau^2 - \sigma_\tau^2\right| \leq 2R_1 + 2R_2 + 2R_3 + 2R_4 + R_5,$$

where

$$\begin{aligned}
R_1 &:= \sup_{\tau \in [0,T]} \left|\sum_{i=2}^{n} K_h\left(t_{i-1} - \tau\right) \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \mu_u du\right) \mu_s ds\right|; \\
R_2 &:= \sup_{\tau \in [0,T]} \left|\sum_{i=2}^{n} K_h\left(t_{i-1} - \tau\right) \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \sigma_u dW_u\right) \mu_s ds\right|; \\
R_3 &:= \sup_{\tau \in [0,T]} \left|\sum_{i=2}^{n} K_h\left(t_{i-1} - \tau\right) \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \mu_u du\right) \sigma_s dW_s\right|; \\
R_4 &:= \sup_{\tau \in [0,T]} \left|\sum_{i=2}^{n} K_h\left(t_{i-1} - \tau\right) \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \sigma_u dW_u\right) \sigma_s dW_s\right|; \\
R_5 &:= \sup_{\tau \in [0,T]} \left|\sum_{i=2}^{n} K_h\left(t_{i-1} - \tau\right) \int_{t_{i-1}}^{t_i} \sigma_s^2 ds - \sigma_\tau^2\right|.
\end{aligned}$$

We show below that

$$
\begin{aligned}
R_1 &= O_P\left(\Delta T^{(2+2l_1)/(2+p)} h^{-2/(2+p)}\right); && (20)\\
R_2 &= O_P\left(\Delta^{1/2} T^{\{2+p(1+m_1)/2+l_1+m_1\}/(2+p)} h^{-(2+p/2)/(2+p)}\right); && (21)\\
R_3 &= O_P\left(h^{\gamma}\right); && (22)\\
R_4 &= O_P\left(h^{\gamma}\right); && (23)\\
R_5 &= O_P\left(\Delta T^{1+m_1}/h^2\right) + O_{a.s.}\left(h^{\gamma}\right). && (24)
\end{aligned}
$$

Together, eqs. (20)-(24) establish the desired result (10) (we can see that R_1 is of smaller order than R_2 by comparing eq. (28) with eq. (30)).

Proof of eq. (20). By Jensen's inequality and $\max_{1\leq i\leq n}|t_i - t_{i-1}| \leq \Delta$,

$$
\left|\int_{t_{i-1}}^{t_i}\left(\int_{t_{i-1}}^{s}\mu_u du\right)\mu_s ds\right| \leq \int_{t_{i-1}}^{t_i}\left(\int_{t_{i-1}}^{t_i}|\mu_u| du\right)|\mu_s| ds = \left(\int_{t_{i-1}}^{t_i}\mu_s ds\right)^2 \leq \Delta\int_{t_{i-1}}^{t_i}\mu_s^2 ds.
$$

Thus, for a sequence $\phi_T \to \infty$ as $T \to \infty$,

$$
\begin{aligned}
R_1 &\leq \Delta \sup_{\tau\in[0,T]}\sum_{i=2}^{n}|K_h(t_{i-1}-\tau)|\int_{t_{i-1}}^{t_i}|\mu_s|^2 ds\\
&= \Delta \sup_{\tau\in[0,T]}\sum_{i=2}^{n}|K_h(t_{i-1}-\tau)|\int_{t_{i-1}}^{t_i}|\mu_s|^2 1\{|\mu_s|\leq\phi_T\} ds\\
&+ \Delta \sup_{\tau\in[0,T]}\sum_{i=2}^{n}|K_h(t_{i-1}-\tau)|\int_{t_{i-1}}^{t_i}|\mu_s|^2 1\{|\mu_s|>\phi_T\} ds\\
&=: R_{11} + R_{12}. && (25)
\end{aligned}
$$

Here, R_{11} is the truncated version of R_1 and satisfies

$$
\begin{aligned}
R_{11} &= \Delta \sup_{\tau\in[0,T]}\sum_{i=2}^{n}\frac{1}{h}\int_{t_{i-1}}^{t_i}\left|K\left(\frac{s-\tau}{h}+\frac{t_{j-1}-s}{h}\right)\right||\mu_s|^2 1\{|\mu_s|\leq\phi_T\} ds\\
&= \Delta \sup_{\tau\in[0,T]}\frac{1}{h}\int_{0}^{T}\left|K\left(\frac{s-\tau}{h}+O(\Delta/h)\right)\right||\mu_s|^2 1\{|\mu_s|\leq\phi_T\} ds\\
&= \Delta \sup_{\tau\in[0,T]}\int_{-\tau/h}^{(T-\tau)/h}|K(u+O(\Delta/h))|\left|\mu_{uh+\tau}\right|^2 1\left\{\left|\mu_{uh+\tau}\right|\leq\phi_T\right\} du\\
&\leq \Delta(\phi_T)^2 \times \int_{-\infty}^{\infty}|K(u+O(\Delta/h))| du = O\left(\Delta(\phi_T)^2\right). && (26)
\end{aligned}
$$

Note that $\int_{-\infty}^{\infty}|K(u+O(\Delta/h))| du \to \int_{-\infty}^{\infty}|K(u)| du$ as $\Delta/h \to 0$ by the assumption that $K \in \mathbb{K}(1,1)$ and the bounded convergence theorem. As for R_{12},

$$
E[R_{12}] \leq E\left[\Delta\bar{K}\frac{1}{h}\int_0^T|\mu_s|^2 1\{|\mu_s|>\phi_T\} ds\right] \leq \frac{\bar{K}}{h(\phi_T)^p}\Delta T\sup_{s\leq T}E\left[|\mu_s|^{2+p}\right] = O\left(\frac{\Delta T^{1+l_1}}{h(\phi_T)^p}\right), \quad (27)
$$

where the last equality follows from assumption (A.1). Now, choose $\phi_T = T^{(1+l_1)/(p+2)} h^{-1/(p+2)}$. Then, eqs. (25)-(27) establish that

$$R_1 = O\left(\Delta\left(\phi_T\right)^2\right) + O_P\left(\frac{\Delta T^{1+l_1}}{h\left(\phi_T\right)^p}\right) = O_P\left(\Delta T^{(2+2l_1)/(2+p)} h^{-2/(2+p)}\right). \tag{28}$$

Proof of eq. (21). By an application of Hölder's inequality, we have

$$R_2 \leq \sup_{\tau\in[0,T]} \sum_{i=2}^{n} \left|K_h\left(t_{i-1}-\tau\right)\right| \int_{t_{i-1}}^{t_i} \left|\mu_s\right| ds \max_{s\in[t_{i-1},t_i]} \left|\int_{t_{i-1}}^{s} \sigma_u dW_u\right| \leq \sqrt{R_{21}\times R_{22}}, \tag{29}$$

where

$$\begin{aligned} R_{21} &:= \sup_{\tau\in[0,T]} \sum_{i=2}^{n} \left|K_h\left(t_{i-1}-\tau\right)\right| \left(\int_{t_{i-1}}^{t_i} \left|\mu_s\right| ds\right)^2; \text{ and} \\ R_{22} &:= \sup_{\tau\in[0,T]} \sum_{i=2}^{n} \left|K_h\left(t_{i-1}-\tau\right)\right| \left|\max_{s\in[t_{i-1},t_i]} \left(\int_{t_{i-1}}^{s} \sigma_u dW_u\right)\right|^2. \end{aligned}$$

With $\{\phi_T\}_{T\geq 0}$ being as before: $R_{21} = O\left(\Delta\left(\phi_T\right)^2\right) + O_P\left(\frac{\Delta T^{(1+l_1)}}{h(\phi_T)^p}\right)$ by the same arguments as for R_1, while

$$\begin{aligned} E\left[R_{22}\right] &= E\left[\sup_{\tau\in[0,T]} \sum_{i=2}^{n} \left|K_h\left(t_{i-1}-\tau\right)\right| \left|\max_{s\in[t_{i-1},t_i]} \left(\int_{t_{i-1}}^{s} \sigma_u dW_u\right)\right|^2\right] \\ &\leq \frac{\bar{K}}{h} \sum_{i=1}^{n} CE\left[\int_{t_{i-1}}^{t_i} \sigma_u^2 du\right] \leq \bar{K}C\frac{T}{h} \sup_{s\leq T} E\left[\sigma_s^2\right] = O\left(\frac{T^{1+m_1}}{h}\right), \end{aligned}$$

by the Burkholder-Davis-Gundy inequality and assumption (A.1). Given these with eq. (29) and the choice of $\phi_T = T^{(1+l_1)/(p+2)} h^{-1/(p+2)}$, we obtain

$$\begin{aligned} R_2 &= \left\{O\left(\Delta\left(\phi_T\right)^2\right) + O_P\left(\frac{\Delta T^{(1+l_1)}}{h\left(\phi_T\right)^p}\right)\right\}^{1/2} \left\{O_P\left(\frac{T^{1+m_1}}{h}\right)\right\}^{1/2} \\ &= O_P\left(\Delta^{1/2} T^{\{2+p(1+m_1)/2+l_1+m_1\}/(2+p)} h^{-(2+p/2)/(2+p)}\right). \end{aligned} \tag{30}$$

Proof of eq. (22). First, let $\mathcal{T}_k := \{\tau\in[0,T] : |\tau-\tau_k| \leq T/\nu_T\}$, $k = 1,\ldots,\nu_T$, be a covering of $[0,T]$, where ν_T is the number of intervals, and τ_k is the center of each interval $\mathcal{T}_k$. Then, R_3 is bounded by

$$\begin{aligned} R_3 &\leq \max_{k\in\{1,\ldots,\nu_T\}} \sup_{\tau\in\mathcal{T}_k} \left|\sum_{i=2}^{n} \left[K_h\left(t_{i-1}-\tau\right) - K_h\left(t_{i-1}-\tau_k\right)\right] \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \mu_u du\right) \sigma_s dW_s\right| \\ &\quad + \max_{k\in\{1,\ldots,\nu_T\}} \left|\sum_{i=2}^{n} K_h\left(t_{i-1}-\tau_k\right) \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \mu_u du\right) \sigma_s dW_s\right| \\ &=: R_{31} + R_{32}. \end{aligned}$$

The following bound of R_{31} then holds:

$$\begin{aligned}
R_{31} &\leq \max_{k\in\{1,\ldots,\nu_T\}} \sup_{\tau\in\mathcal{T}_k} \left\{ \sum_{i=2}^{n} \left[K_h\left(t_{i-1}-\tau\right) - K_h\left(t_{i-1}-\tau_k\right)\right]^2 \right\}^{1/2} \\
&\quad \times \left\{ \sum_{i=2}^{n} \left(\int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \mu_u du \right) \sigma_s dW_s \right)^2 \right\}^{1/2} \\
&\leq \frac{1}{h} \left\{ n\bar{K}^2 \left(\frac{T}{h\nu_T} \right)^2 \times O_P\left(\Delta^2 T^{1+\eta}\right) \right\}^{1/2} = O\left(\frac{\Delta^{1/2} T^{2+\eta/2}}{h^2 \nu_T} \right), \qquad (31)
\end{aligned}$$

where $\eta := \max\{l_2, m_2\}$. The first inequality holds by Holder's inequality; and the second by the global Lipschitz continuity of K and by the following result:

$$\begin{aligned}
&E\left[\sum_{i=2}^{n} \left(\int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \mu_u du \right) \sigma_s dW_s \right)^2 \right] = E\left[\sum_{i=2}^{n} \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \mu_u du \right)^2 \sigma_s^2 ds \right] \\
&\leq E\left[\Delta \sum_{i=2}^{n} \int_{t_{i-1}}^{t_i} \mu_u^2 du \int_{t_{i-1}}^{t_i} \sigma_s^2 ds \right] \qquad (32) \\
&\leq E\left[\frac{\Delta^2}{2} \sum_{i=2}^{n} \int_{t_{i-1}}^{t_i} \mu_u^4 du + \int_{t_{i-1}}^{t_i} \sigma_s^4 ds \right] = O\left(\Delta^2 T^{1+\eta}\right), \qquad (33)
\end{aligned}$$

where Jensen's inequality and the inequality $AB \leq \left(A^2 + B^2\right)/2$ have been used to derive eqs. (32) and (33), respectively. Next, consider the term R_{32}: Define a function $\varrho : [0,\infty) \to [0,\infty)$ as

$$\varrho\left(s\right) := t_{i-1} \ \text{ if } \ s \in [t_{i-1}, t_i). \qquad (34)$$

Using this function, we can re-write the inside of the absolute-value sign of R_{32} as

$$\sum_{i=2}^{n} K\left(\frac{t_{i-1}-\tau_k}{h} \right) \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{s} \mu_u du \right) \sigma_s dW_s = \int_0^T K\left(\frac{\varrho\left(s\right)-\tau_k}{h} \right) \left(\int_{\varrho(s)}^{s} \mu_u du \right) \sigma_s dW_s = M_T^k\left(1\right),$$

where $M_T^k\left(r\right) := \int_0^{rT} K\left(\frac{\varrho(s)-\tau_k}{h}\right)\left(\int_{\varrho(s)}^{s} \mu_u du\right) \sigma_s dW_s$. Note that $\left\{M_T^k\left(r\right)\right\}_{r\in[0,1]}$ is a continuous martingale (for each k and T) which vanishes at zero and has the quadratic variation process

$$\left\langle M_T^k \right\rangle_r = \int_0^{rT} K^2\left(\frac{\varrho\left(s\right)-\tau_k}{h} \right) \left(\int_{\varrho(s)}^{s} \mu_u du \right)^2 \sigma_s^2 ds. \qquad (35)$$

Then, for any non-stochastic $c\,(>0)$, it holds that

$$\Pr\left(\left| M_T^k\left(1\right) \right| \geq c \right) \leq \Pr\left(\left| M_T^k\left(1\right) \right| \geq c,\ \left\langle M_T^k \right\rangle_1 \leq \xi_T \right) + \Pr\left(\left\langle M_T^k \right\rangle_1 > \xi_T \right), \qquad (36)$$

where $\{\xi_T\}$ is a sequence that depends on T. We show that each term of the right-hand of eq. (36) tends to zero (for an appropriate choice of c and ξ_T). Applying the exponential inequality for continuous martingales (see, e.g., Exercise 3.15 in Ch. IV of Revuz and Yor, 1999 or Dzhaparidze and van Zanten, 2001) to the first term of the right-hand side of eq. (36), we have

$$\Pr\left(\left|M_T^k(1)\right| \geq c,\ \left\langle M_T^k \right\rangle_1 \leq \xi_T\right) \leq 2\exp\left\{-\frac{c^2}{2\xi_T}\right\}. \tag{37}$$

To find the bound for the second term of eq. (36), $\Pr\left(\langle M_T^k\rangle_1 > \xi_n\right)$, we derive an upper bound for $\langle M^k\rangle_T$:

$$\begin{aligned}
\left\langle M_T^k\right\rangle_1 &\leq \bar{K}^2\int_0^T\left(\int_{\varrho(s)}^s \mu_u du\right)^2\sigma_s^2 ds \leq \bar{K}^2\sum_{i=2}^n\int_{t_{i-1}}^{t_i}\left(\int_{t_{i-1}}^{t_i}\mu_u du\right)^2\sigma_s^2 ds \\
&\leq \Delta\bar{K}^2\sum_{i=2}^n\left(\int_{t_{i-1}}^{t_i}\mu_u^2 du\right)\int_{t_{i-1}}^{t_i}\sigma_s^2 ds \\
&\leq \Delta\left(\bar{K}^2/2\right)\sum_{i=2}^n\left[\left(\int_{t_{i-1}}^{t_i}\mu_u^2 du\right)^2+\left(\int_{t_{i-1}}^{t_i}\sigma_s^2 ds\right)^2\right] \\
&\leq \Delta^2\left(\bar{K}^2/2\right)\sum_{i=2}^n\left[\int_{t_{i-1}}^{t_i}\mu_u^4 du+\int_{t_{i-1}}^{t_i}\sigma_s^4 ds\right] = \Delta^2\left(\bar{K}^2/2\right)\left[\int_0^T\mu_u^4 du+\int_0^T\sigma_s^4 ds\right]
\end{aligned}$$

where Jensen's inequality and the inequality $AB \leq \left(A^2+B^2\right)/2$ have been used. This implies that

$$\begin{aligned}
\Pr\left(\left\langle M_T^k\right\rangle_1 > \xi_T\right) &\leq \Pr\left(\Delta^2\left(\bar{K}^2/2\right)\left[\int_0^T\mu_u^4 du+\int_0^T\sigma_s^4 ds\right] > \xi_T\right) \\
&\leq \left(\bar{K}^2/2\right)\frac{\Delta^2 T}{\xi_T}\left(\sup_{s\leq T}E\left[\mu_s^4\right]+\sup_{s\leq T}E\left[\sigma_s^4\right]\right) = O\left(\frac{\Delta^2 T^{1+\eta}}{\xi_T}\right)
\end{aligned} \tag{38}$$

by Markov's inequality and assumption (A.1)-(A.2), where $\eta := \max\{l_2, m_2\}$. By eqs. (36), (37) and (38),

$$\Pr\left(M_T^k(1) \geq c\right) \leq 2\exp\left\{-\frac{c^2}{2\xi_n}\right\}+O\left(\frac{\Delta^2 T^{1+\eta}}{\xi_T}\right). \tag{39}$$

Note that the right-hand side of eq. (39) does not depend on k. Thus, setting $c = Ja_T$ where J is any positive constant and $\{a_T\}$ is a sequences which tend to zero as $T\to\infty$,

$$\begin{aligned}
\Pr\left(R_{32} \geq Ja_T\right) &= \Pr\left(\frac{1}{h}\max_{k\in\{1,\ldots,\nu_T\}}\left|M_T^k(1)\right| \geq Ja_T\right) \\
&\leq \sum_{k=1}^{\nu_T}\Pr\left(\left|M_T^k(1)\right| \geq Ja_T h\right) = 2\nu_T\exp\left\{-\frac{J^2 a_T^2 h^2}{2\xi_T}\right\}+O\left(\frac{\nu_T\Delta^2 T^{1+\eta}}{\xi_T}\right).
\end{aligned} \tag{40}$$

If the right-hand side of eq. (40) shrinks to zero as $J\to\infty$, then $R_{32} = O_P\left(a_T\right)$.

Now, letting $a_T = h^\gamma$, $\nu_T = \Delta^{1/2}T^{2+\eta/2}/h^{2+\gamma}$ and $\xi_T = J\nu_T\Delta^2 T^{1+\eta}$, we have $R_{31} = O\left(h^\gamma\right)$ and

$$\Pr\left(R_{32} \geq Jh^\gamma\right) = 2\exp\left\{\log\left(\frac{\Delta^{1/2}T^{2+\eta/2}}{h^{2+\gamma}}\right)-J\frac{h^{4+3\gamma}}{\Delta^{5/2}T^{3+3\eta/2}}\right\}+O\left(J^{-1}\right).$$

Note that the first and second conditions in (B.1) implies that $\Delta^{5/2}T^{3+3\eta/2}/h^{4+3\gamma} \to 0$, and therefore

$$
\begin{aligned}
& \log\left(\frac{\Delta^{1/2}T^{2+\eta/2}}{h^{2+\gamma}}\right) - J\frac{h^{4+3\gamma}}{\Delta^{5/2}T^{3+3\eta/2}} \\
= & \underbrace{\frac{h^{4+3\gamma}}{\Delta^{5/2}T^{3+3\eta/2}}}_{\to\infty}\left\{\underbrace{\left(\log\Delta^{1/2}\right)\frac{\Delta^{5/2}T^{3+3\eta/2}}{h^{4+3\gamma}}}_{=o\left(\log\Delta^{1/2}\right)} - J\underbrace{\left[\log T^{(2+\eta/2)} + \log(1/h)^{(2+\gamma)}\right]\frac{\Delta^{5/2}T^{3+3\eta/2}}{h^{4+3\gamma}}}_{=O(1)}\right\} \\
\to & -\infty.
\end{aligned}
$$

This implies that $\Pr(R_{32} \geq Jh^{\gamma}) \to 0$ for any J, which establishes the desired result.

Proof of eq. (23). The convergence rate of R_4 can be derived in the similar manner as for R_3. Construct a covering of $[0, T]$, $\{\mathcal{U}_k\}_{k=1}^{v_T}$. Each $\mathcal{U}_k$ has the radius T/v_T from the center τ_k. Then,

$$
\begin{aligned}
R_4 \leq & \max_{k\in\{1,\ldots,v_T\}}\sup_{\tau\in\mathcal{U}_k}\left|\sum_{i=2}^{n}\left[K_h(t_{i-1}-\tau) - K_h(t_{i-1}-\tau_k)\right]\int_{t_{i-1}}^{t_i}\left(\int_{t_{i-1}}^{s}\sigma_u dW_u\right)\sigma_s dW_s\right| \\
& + \max_{k\in\{1,\ldots,v_T\}}\left|\sum_{i=1}^{n}K_h(t_{i-1}-\tau_k)\int_{t_{i-1}}^{t_i}\left(\int_{t_{i-1}}^{s}\sigma_u dW_u\right)\sigma_s dW_s\right| \\
& =: R_{41} + R_{42}.
\end{aligned} \tag{41}
$$

By arguments similar to those in deriving the bound of R_{31} in eq. (31),

$$
\begin{aligned}
R_{41} \leq & \left\{\max_{k\in\{1,\ldots,v_T\}}\sup_{\tau\in\mathcal{U}_k}\sum_{i=2}^{n}\left[K_h(t_{i-1}-\tau) - K_h(t_{i-1}-\tau_k)\right]^2\right\}^{1/2} \\
& \times\left\{\sum_{i=2}^{n}\left[\int_{t_{i-1}}^{t_i}\left(\int_{t_{i-1}}^{s}\sigma_u dW_u\right)\sigma_s dW_s\right]^2\right\}^{1/2} \\
\leq & \frac{1}{h}\left\{\bar{K}^2\frac{nT^2}{h^2 v_T^2}\times O_P\left(\Delta T^{(1+m_2)}\right)\right\}^{1/2} = O_P\left(\frac{T^{(2+m_2/2)}}{h^2 v_T}\right),
\end{aligned} \tag{42}
$$

To find the probability bound of R_{42}, define a continuous martingale $\left\{N_T^k(r)\right\}_{r\in[0,1]}$ for each k and T by

$$
N_T^k(r) := \int_0^{rT} K\left(\frac{\varrho(s)-\tau_k}{h}\right)\left(\int_{\varrho(s)}^{s}\sigma_u dW_u\right)\sigma_s dW_s,
$$

whose quadratic variation process $\left\{\langle N_T^k\rangle_r\right\}$ is computed analogously to $\left\{\langle M_T^k\rangle_r\right\}$. Then, by the same arguments as for R_{32},

$$
\Pr(|R_{42}| > Jb_T) \leq \sum_{k=1}^{v_T}\left[\Pr\left(\left|N_T^k(1)\right| \geq Jb_T h,\ \left\langle N_T^k\right\rangle_1 \leq \zeta_T\right) + \Pr\left(\left\langle N_T^k\right\rangle_1 > \zeta_T\right)\right].
$$

Since

$$
\Pr\left(\left|N_T^k(1)\right| \geq Jb_T h,\ \left\langle N_T^k\right\rangle_1 \leq \zeta_T\right) \leq 2v_T\exp\left\{-J^2 b_T^2 h^2/2\zeta_T\right\},
$$

and $\Pr\left(\langle N_T^k\rangle_1 > \zeta_T\right) = O\left(\Delta T^{(1+m_2)}/\zeta_T\right)$, which can be derived analogously to eqs. (37) and (38), it holds that

$$\Pr\left(|R_{42}| > Jb_T\right) \le 2v_T \exp\left\{-\frac{J^2 b_T^2 h^2}{2\zeta_T}\right\} + O\left(\frac{v_n \Delta T^{(1+m_2)}}{\zeta_T}\right). \tag{43}$$

With $b_T = h^\gamma$, $v_T = T^{(2+m_2/2)}/h^{2+\gamma}$ and $\zeta_T = Jv_T\Delta T^{(1+m_2)}$, we have $R_{41} = O_P\left(h^\gamma\right)$ and

$$\Pr\left(|R_{42}| > Jh^\gamma\right) \le 2\exp\left\{\log\left(\frac{T^{(2+m_2/2)}}{h^{2+\gamma}}\right) - \frac{Jh^{4+3\gamma}}{2\Delta T^{(3+m_2/2)}}\right\} + O\left(J^{-1}\right). \tag{44}$$

Given the second condition in (B.1),

$$\begin{aligned}
& \log\left(\frac{T^{(2+m_2/2)}}{h^{2+\gamma}}\right) - \frac{Jh^{4+3\gamma}}{2\Delta T^{(3+m_2/2)}} \\
= \; & \underbrace{\frac{h^{4+3\gamma}}{\Delta T^{3+3m_2/2}}}_{\to\infty} \left\{\underbrace{\frac{\Delta T^{(3+m_2/2)}}{h^{4+3\gamma}}\left[(2+m_2/2)\log T + (2+\gamma)\log\left(1/h\right)\right]}_{=O(1)} - \frac{J}{2}\right\} \to -\infty \text{ for large } J.
\end{aligned}$$

Therefore, the left-hand side of eq. (44) tends to zero as $J \to \infty$, and $R_{42} = O_P\left(h^\gamma\right)$.

Proof of eq. (24). Assumption (A.2) and Lemma 1 imply that there exists some $\bar{\Delta}\,(>0)$ such that for any $\Delta \le \bar{\Delta}$, $|\sigma_t - \sigma_s| \le D\,|t-s|^\gamma$ a.s. We have

$$\begin{aligned}
R_5 &= \sup_{\tau\in[0,T]}\left|\sum_{i=2}^n \frac{1}{h}\int_{t_{i-1}}^{t_i} K\left(\frac{s-\tau}{h} + \frac{t_{i-1}-s}{h}\right)\sigma_s^2 ds - \sigma_\tau^2\right| \\
&= \sup_{\tau\in[0,T]}\left|\frac{1}{h}\int_0^T K\left(\frac{s-\tau}{h} + k\left(s\right)\right)\sigma_s^2 ds - \sigma_\tau^2\right| \\
&\le \sup_{\tau\in[0,T]}\frac{1}{h}\int_{-\tau/h}^{(T-\tau)/h} \left|K\left(u+k\left(s\right)\right)\right|\left|\sigma_{uh+\tau}^2 - \sigma_\tau^2\right| du \\
&\le Dh^\gamma \int_{-\infty}^{\infty}\left|K\left(u+k\left(s\right)\right)\right|\left|u\right|^\gamma du,
\end{aligned} \tag{45}$$

where $k\left(s\right) = \left(\varrho\left(s\right) - s\right)/h$ and $\varrho\left(s\right)$ is defined in eq. (34). Noting that $k\left(s\right) = O\left(\Delta/h\right)$ uniformly over s, we can show $\int_{-\infty}^{\infty}\left|K\left(u+k\left(s\right)\right)\right|\left|u\right|^\gamma du = \int_{-\infty}^{\infty}\left|K\left(u+O\left(\Delta/h\right)\right)\right|\left|u\right|^\gamma du < \infty$ by using the dominated convergence theorem with $K^*\left(u\right)\left|u\right|^\gamma$ a dominant function, where K^* is constructed in the following manner:

$$K^*\left(u\right) := \begin{cases} \bar{K} & \text{if } |u| \le C + \bar{\varepsilon}, \\ \dfrac{|K\left(C\right)| - \bar{K}}{\bar{\varepsilon}}\left[u - C - \bar{\varepsilon}\right] + \bar{K} & \text{if } |u| \in (C+\bar{\varepsilon}, C+2\varepsilon], \\ |K\left(u-2\bar{\varepsilon}\right)| & \text{if } |u| > C + 2\bar{\varepsilon}. \end{cases} \tag{46}$$

where $\bar{\varepsilon}\,(>0)$ is some constant. Since K is monotone for $|x| \ge C$, it holds that $\left|K\left(u+k\left(s\right)\right)\right| \le K^*\left(u\right)$ for any s if $\sup_s \varrho\left(s\right) \le \bar{\varepsilon}$ (recall that $\varrho\left(s\right) = O\left(\Delta/h\right)$ uniformly over any s). We can also show that $\int K^*\left(u\right)\left|u\right|^\gamma du < \infty$ if $\int\left|K\left(u\right)\right|\left|u\right|^\gamma du < \infty$, which follows from $K \in \mathbb{K}\left(1,1\right)$ and $\gamma < 1/2$. Eq. (45) now yields the desired result.

A.2 Proof of Theorem 2: Nonparametric Drift Estimator

As in eq. (11), we split up $\hat{\alpha}(x) - \alpha(x)$ into two terms. The first term of the right-hand side of eq. (11) converges to zero by Lemma 2 under (B-NDR.i). Noting that $h^{\gamma}/b\delta^{1/2} \to 0$ and $\delta^{\gamma}/b \to 0$ respectively imply $h^{\gamma}/b < \delta^{1/2}$ and $\delta^{1/2} < b^{1/2\gamma}$, the condition of Lemma 2 is satisfied with $q = 1/2\gamma$. The convergence of the second term is investigated in Lemma 3. The condition (B-NDR.ii) ensures that the first term $\hat{\alpha}(x) - \tilde{\alpha}(x)$ has no effect such that the asymptotic distribution is completely determined by $\tilde{\alpha}(x) - \alpha(x)$.

Lemma 2 *Assume that (A.1), (A.2') and (B.1') are satisfied,* $K \in \mathbb{K}(1,1)$ *and* $\mathcal{K} \in \mathbb{K}(2,2)$*. If there exists some* $q\,(>0)$ *such that* $h^{\gamma}/b = O(b^{q})$*, then*

$$\hat{\alpha}(x) - \tilde{\alpha}(x) = O_P\left(h^{\gamma} \Big/ b\delta^{1/2}\right) + O_P\left(h^{\gamma}/\delta\right). \tag{47}$$

as $T, N \to \infty$ *and* $\delta \to 0$ *with* $b \to 0$*.*

Proof. We first split up the left-hand side of eq. (47) into three terms:

$$\hat{\alpha}(x) - \tilde{\alpha}(x) = A_1 + A_2 + A_3,$$

where

$$A_1 := \frac{(1/T)\sum_{j=1}^{N-1}\mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)\left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)}{(\delta/T)\sum_{j=1}^{N}\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j} - x\right)} - \frac{(1/T)\sum_{j=1}^{N-1}\mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)\left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)}{(\delta/T)\sum_{j=1}^{N}\mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)};$$

$$A_2 := \frac{(1/T)\sum_{j=1}^{N-1}\left[\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j} - x\right) - \mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)\right]\left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)}{(\delta/T)\sum_{j=1}^{N}\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j} - x\right)};$$

$$A_3 := \frac{(1/T)\sum_{j=1}^{N-1}\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_{j-1}} - x\right)\left[\left(\hat{\sigma}^2_{\tau_{j+1}} - \hat{\sigma}^2_{\tau_j}\right) - \left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)\right]}{(\delta/T)\sum_{j=1}^{N}\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j} - x\right)}.$$

We below show that $A_1 = O_P\left(h^{\gamma(d+1)}/b^{d+2}\right)$, $A_2 = O_P\left(h^{\gamma}/\left(b\delta^{1/2}\right)\right)$ and $A_3 = O_P\left(h^{\gamma}/\delta\right)$ which establish the result stated in eq. (47). Before showing the convergence result of each term, we present the following useful result which we use repeatedly: for any $d > 0$,

$$\frac{\delta}{T}\sum_{j=1}^{N}\left[\mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right) - \mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j} - x\right)\right] = O_P\left(\frac{h^{\gamma}}{b}\right). \tag{48}$$

Proof of eq. (48). To find the probability bound, look at

$$
\begin{aligned}
&\text{the left-hand side of eq. (48)}\\
&= \frac{1}{N}\sum_{j=1}^{N}\frac{1}{b}\mathcal{K}'\left(\left(\sigma^2_{\tau_j}-x\right)b^{-1}-w_j\left(\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right)b^{-1}\right)\times w_j\left(\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right)b^{-1}\\
&\leq \sup_{1\leq j\leq N} w_j\left|\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right|b^{-1}\\
&\times\left\{\frac{1}{N}\sum_{j=1}^{N}\frac{1}{b}\left|\mathcal{K}'\left(\left(\sigma^2_{\tau_j}-x\right)b^{-1}-w_j\left(\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right)b^{-1}\right)\right|\mathbf{1}\left\{w_j\left|\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right|b^{-1}\leq\bar{\epsilon}\right\}\right.\\
&\left.+\frac{1}{N}\sum_{j=1}^{N}\frac{1}{b}\left|\mathcal{K}'\left(\left(\sigma^2_{\tau_j}-x\right)b^{-1}-w_j\left(\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right)b^{-1}\right)\right|\mathbf{1}\left\{w_j\left|\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right|b^{-1}>\bar{\epsilon}\right\}\right\} \qquad (49)\\
&\leq O_P\left(h^\gamma/b\right)\\
&\times\left\{\frac{1}{N}\sum_{j=1}^{N}\frac{1}{b}\mathcal{K}^*\left(\left(\sigma^2_{\tau_j}-x\right)b^{-1}\right)+\frac{1}{N}\sum_{j=1}^{N}\frac{\bar{K}}{b}\times\left(\frac{\sup_{1\leq j\leq N}w_j|\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}|b^{-1}}{\bar{\epsilon}}\right)^{1/q}\right\}\\
&= O_P\left(h^\gamma/b\right)\times\left\{O_P(1)+\frac{\bar{K}}{b}\times O_P\left(\left(\frac{b^q}{\bar{\epsilon}}\right)^{1/q}\right)\right\}, \qquad (50)
\end{aligned}
$$

where the first equality holds by the mean-value theorem with $w_j\in[0,1]$, $j=1,\ldots,N$, being some random variables. The inequality (49) holds by eq. (12) and

$$
\left|\mathcal{K}'\left(\left(\sigma^2_{\tau_j}-x\right)b^{-1}+w_j\left(\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right)b^{-1}\right)\right|\mathbf{1}\left\{w_j\left|\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right|b^{-1}\leq\bar{\epsilon}\right\}\leq\mathcal{K}^*\left(\left(\sigma^2_{\tau_j}-x\right)b^{-1}\right) \qquad (51)
$$

where $\mathcal{K}^*$ is some function and $\bar{\epsilon}$ is some positive constant, such that

$$
\sup_{|\epsilon|\leq\bar{\epsilon}}\left|\mathcal{K}'\left(\left(\sigma^2_{\tau_j}-x\right)b^{-1}+\epsilon\right)\right|\leq\mathcal{K}^*\left(\left(\sigma^2_{\tau_j}-x\right)b^{-1}\right).
$$

Since $\mathcal{K}\in\mathbb{K}(0,1)$, such $\mathcal{K}^*$ and $\bar{\epsilon}$ exist by the same argument in showing the convergence of R_5 in the proof of Theorem 1, where we can let $\mathcal{K}^*(u)=K^*(u)$ as given in eq. (46). The equality in eq. (50) holds since $N^{-1}\sum_{j=1}^{N}\mathcal{K}^*_b\left(\sigma^2_{\tau_j}-x\right)=O_P(1)$ and $w_j\left(\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right)b^{-1}=O_P\left(h^\gamma/b\right)$ (uniformly over any j), where the former follows from the positive recurrence of the process $\left\{\sigma^2_t\right\}$ and standard arguments on the kernel estimation with the uniform boundedness of $\mathcal{K}^*$ and its tail decay condition.

Convergence of A_1. The term A_1 can be re-written as

$$
A_1=\frac{(1/T)\sum_{j=1}^{N-1}\mathcal{K}_b\left(\sigma^2_{\tau_j}-x\right)\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)}{(\delta/T)\sum_{j=1}^{N}\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j}-x\right)\times(\delta/T)\sum_{j=1}^{N}\mathcal{K}_b\left(\sigma^2_{\tau_j}-x\right)}\times\frac{1}{N}\sum_{j=1}^{N}\left[\mathcal{K}_b\left(\sigma^2_{\tau_j}-x\right)-\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j}-x\right)\right].
$$

By eq. (48), and

$$\frac{1}{T}\sum_{j=1}^{N-1}\mathcal{K}_b\left(\sigma^2_{\tau_j}-x\right)\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)=\alpha\left(x\right)\pi\left(x\right)+o_P\left(1\right); \tag{52}$$

$$\frac{\delta}{T}\sum_{j=1}^{N}\mathcal{K}_b\left(\sigma^2_{\tau_j}-x\right)=\pi\left(x\right)+o_P\left(1\right), \tag{53}$$

we have

$$A_1=\frac{\alpha\left(x\right)+o_P\left(1\right)}{\left(\pi_0\left(x\right)+o_P\left(1\right)\right)^2}\times O_P\left(\frac{h^\gamma}{b}\right)=O_P\left(\frac{h^\gamma}{b}\right).$$

Convergence of A_2. Write the numerator of the term A_2 as:

$$\frac{1}{T}\sum_{j=1}^{N-1}\left[\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j}-x\right)-\mathcal{K}_b\left(\sigma^2_{\tau_j}-x\right)\right]\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)=A_{21}^{nu}+A_{22}^{nu}. \tag{54}$$

where

$$A_{21}^{nu}:=\frac{1}{T}\sum_{j=1}^{N-1}\left[\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j}-x\right)-\mathcal{K}_b\left(\sigma^2_{\tau_j}-x\right)\right]\int_{\tau_j}^{\tau_{j+1}}\alpha\left(\sigma^2_s\right)ds,$$
$$A_{22}^{nu}:=\frac{1}{T}\sum_{j=1}^{N-1}\left[\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j}-x\right)-\mathcal{K}_b\left(\sigma^2_{\tau_j}-x\right)\right]\int_{\tau_j}^{\tau_{j+1}}\beta\left(\sigma^2_s\right)dZ_s.$$

By a similar argument to show (50), we can find the bound of A_{21}^{nu}:

$$\begin{aligned}
A_{21}^{nu} &\leq \sup_{1\leq j\leq N} w_j\left|\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right|b^{-1}\times\left\{\frac{1}{T}\sum_{j=1}^{N-1}\frac{1}{b}\int_{\tau_j}^{\tau_{j+1}}\mathcal{K}^*\left(\left(\sigma^2_s-x\right)b^{-1}\right)\left|\alpha\left(\sigma^2_s\right)\right|ds\right.\\
&\quad\left.+\frac{\bar{K}}{b}\frac{1}{T}\sum_{j=1}^{N-1}\int_{\tau_j}^{\tau_{j+1}}\left|\alpha\left(\sigma^2_s\right)\right|ds\times\left(\frac{\sup_{1\leq j\leq N}w_j|\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}|b^{-1}}{\bar{\epsilon}}\right)^{1/q}\right\}\\
&= O_P\left(h^\gamma/b\right)\times\left\{O_P\left(1\right)+\frac{\bar{K}}{b}\times O_p\left(1\right)\times O_P\left(\left(\frac{b}{\bar{\epsilon}}\right)^{1/q}\right)\right\}\\
&= O_P\left(h^\gamma/b\right),
\end{aligned} \tag{55}$$

where we have used

$$\frac{1}{T}\sum_{j=1}^{N-1}\frac{1}{b}\int_{\tau_j}^{\tau_{j+1}}\mathcal{K}^*\left(\left(\sigma^2_s-x\right)b^{-1}\right)\left|\alpha\left(\sigma^2_s\right)\right|ds=O_P\left(1\right),\quad \frac{1}{T}\sum_{j=1}^{N}\int_{\tau_{j-1}}^{\tau_j}\left|\alpha\left(\sigma^2_s\right)\right|ds=O_P\left(1\right).$$

Note that both the equalities holds by the condition (A.2'). Next, look at the term A_{22}^{nu}. We use a similar argument to show eq. (48):

$$\begin{aligned}
A_{22}^{nu} &= \frac{1}{T}\sum_{j=1}^{N-1}\left[\mathcal{K}_b\left(\hat{\sigma}_{\tau_j}^2 - x\right) - \mathcal{K}_b\left(\sigma_{\tau_j}^2 - x\right)\right]\int_{\tau_j}^{\tau_{j+1}}\beta\left(\sigma_s^2\right)dZ_s \\
&= \frac{1}{T}\sum_{j=1}^{N-1}\frac{1}{b}\mathcal{K}'\left(\left(\sigma_{\tau_j}^2 - x\right)b^{-1} + w_j\left(\sigma_{\tau_j}^2 - \hat{\sigma}_{\tau_j}^2\right)b^{-1}\right)w_j\left(\sigma_{\tau_j}^2 - \hat{\sigma}_{\tau_j}^2\right)b^{-1}\int_{\tau_j}^{\tau_{j+1}}\beta\left(\sigma_s^2\right)dZ_s \\
&\leq \sup_{1\leq j\leq N-1} w_j\left|\sigma_{\tau_j}^2 - \hat{\sigma}_{\tau_j}^2\right|b^{-1} \\
&\times\left\{\frac{1}{T}\sum_{j=1}^{N-1}\frac{1}{b}\left|\mathcal{K}'\left(\left(\sigma_{\tau_j}^2 - x\right)b^{-1} + w_j\left(\sigma_{\tau_j}^2 - \hat{\sigma}_{\tau_j}^2\right)b^{-1}\right)\right|\left|\int_{\tau_j}^{\tau_{j+1}}\beta\left(\sigma_s^2\right)dZ_s\right|\right. \\
&\times\mathbf{1}\left\{w_j\left|\sigma_{\tau_j}^2 - \hat{\sigma}_{\tau_j}^2\right|b^{-1}\leq\bar{\epsilon}\right\} \\
&+\frac{1}{T}\sum_{j=1}^{N-1}\frac{1}{b}\left|\mathcal{K}'\left(\left(\sigma_{\tau_j}^2 - x\right)b^{-1} + w_j\left(\sigma_{\tau_j}^2 - \hat{\sigma}_{\tau_j}^2\right)b^{-1}\right)\right|\left|\int_{\tau_j}^{\tau_{j+1}}\beta\left(\sigma_s^2\right)dZ_s\right| \\
&\quad\left.\times\mathbf{1}\left\{w_j\left|\sigma_{\tau_j}^2 - \hat{\sigma}_{\tau_j}^2\right|b^{-1}>\bar{\epsilon}\right\}\right\} \\
&\leq \sup_{1\leq j\leq N-1} w_j\left|\sigma_{\tau_j}^2 - \hat{\sigma}_{\tau_j}^2\right|b^{-1}\times\left\{\frac{1}{T}\sum_{j=1}^{N-1}\frac{1}{b}\mathcal{K}^*\left(\left(\sigma_{\tau_j}^2 - x\right)b^{-1}\right)\left|\int_{\tau_j}^{\tau_{j+1}}\beta\left(\sigma_s^2\right)dZ_s\right|\right. \\
&+\left.\frac{1}{T}\sum_{j=1}^{N-1}\frac{\bar{K}}{b}\left|\int_{\tau_j}^{\tau_{j+1}}\beta\left(\sigma_s^2\right)dZ_s\right|\times\left(\frac{\sup_{1\leq j\leq N-1}w_j|\sigma_{\tau_j}^2 - \hat{\sigma}_{\tau_j}^2|b^{-1}}{\bar{\epsilon}}\right)^d\right\} \\
&= O_P\left(h^\gamma/b\right)\times\left\{O_P\left(\delta^{-1/2}\right) + \frac{\bar{K}}{b}\times O_P\left(\delta^{-1/2}\right)\times O_P\left(\left(\frac{h^\gamma}{b}\right)^d\right)\right\},
\end{aligned} \tag{56}$$

where the last equality uses

$$\frac{1}{T}\sum_{j=1}^{N-1}\frac{1}{b}\mathcal{K}^*\left(\left(\sigma_{\tau_j}^2 - x\right)b^{-1}\right)\left|\int_{\tau_j}^{\tau_{j+1}}\beta\left(\sigma_s^2\right)dZ_s\right| = O_P\left(\delta^{-1/2}\right);\ \text{and} \tag{57}$$

$$\frac{1}{T}\sum_{j=1}^{N}\left|\int_{\tau_{j-1}}^{\tau_j}\beta\left(\sigma_s^2\right)dZ_s\right| = O_P\left(\delta^{-1/2}\right). \tag{58}$$

Eq. (58) follows by (B-NDR.i) and the Burkholder-Davis-Gundy inequality, while eq. (57) can be shown as follows:

$$
\begin{aligned}
& E\left[\frac{1}{T}\sum_{j=1}^{N-1}\frac{1}{b}\left|\int_{\tau_{j-1}}^{\tau_j}\mathcal{K}^*\left(\left(\sigma_{\tau_j}^2-x\right)b^{-1}\right)\beta\left(\sigma_s^2\right)dZ_s\right|\right] \\
\leq\ & \frac{1}{T}\sum_{j=1}^{N-1}\frac{1}{b}E\left[\left|\int_{\tau_{j-1}}^{\tau_j}\left[\mathcal{K}^*\left(\left(\sigma_{\tau_j}^2-x\right)b^{-1}\right)\beta\left(\sigma_s^2\right)\right]^2 ds\right|^{1/2}\right] \\
\leq\ & \frac{1}{T}\sum_{j=1}^{N-1}\frac{1}{b}\left\{E\left[\left\{\mathcal{K}^*\left(\left(\sigma_{\tau_j}^2-x\right)b^{-1}\right)\right\}^2\right]E\left[\int_{\tau_{j-1}}^{\tau_j}\beta^2\left(\sigma_s^2\right)ds\right]\right\}^{1/2} \\
\leq\ & \frac{1}{T}\sum_{j=1}^{N-1}E\left[\mathcal{K}^*\left(\left(\sigma_{\tau_j}^2-x\right)b^{-1}\right)\right]\left\{E\left[\int_{\tau_{j-1}}^{\tau_j}\beta^2\left(\sigma_s^2\right)ds\right]\right\}^{1/2}=O\left(\delta^{-1/2}\right),
\end{aligned}
$$

where Ito's isometry and Holder's and Schwartz's inequalities are used. The last equality holds since $b^{-1}E\left[\mathcal{K}^*\left((\sigma_{\tau_j}^2-x)b^{-1}\right)\right]=O(1)$ and $E\left[\int_{\tau_{j-1}}^{\tau_j}\beta^2\left(\sigma_s^2\right)ds\right]=O(\delta)$. Noting that we can show that the denominator of the term A_2 is $O_P(1)$ by using eqs. (48) and (53), we have

$$
A_2=\frac{A_{21}^{nu}+A_{22}^{nu}}{O_P(1)}=\frac{O_P\left(h^\gamma/b\right)+O_P\left(h^\gamma/b\delta^{1/2}\right)}{O_P(1)}=O_P\left(h^\gamma/b\delta^{1/2}\right).
$$

by eqs. (54), (55) and (56).

Convergence of A_3.

$$
A_3\leq\frac{(1/T)\sum_{j=1}^{N-1}\left|\mathcal{K}_b\left(\hat{\sigma}_{\tau_{j-1}}^2-x\right)\right|\times O\left(2\sup_{\tau\in[0,T]}\left|\hat{\sigma}_\tau^2-\sigma_\tau^2\right|\right)}{(\delta/T)\sum_{j=1}^{N}\mathcal{K}_b\left(\hat{\sigma}_{\tau_j}^2-x\right)}=O_P\left(h^\gamma/\delta\right).
$$

Now, the proof is completed. ■

Lemma 3 *Assume that (A.2') holds, and $\mathcal{K}\in\mathbb{K}(2,2)$. If $N,T\to\infty$ and $b\to 0$ with $\delta^\gamma/b\to 0$ and $Tb\to\infty$, then $\tilde{\alpha}(x)\xrightarrow{P}\alpha(x)$. If in addition $Tb\delta^{2\gamma}\to 0$ and $Tb^5=O(1)$, then*

$$
\sqrt{Tb}\left[\tilde{\alpha}(x)-\alpha(x)-b^2\times bias_\alpha(x)\right]\xrightarrow{d}N\left(0,\beta^4(x)\int\mathcal{K}^2(z)\,dz\right),
$$

where $bias_\alpha(x)$ is given in Theorem 2.

Proof. We follow similar steps as in Bandi and Phillips (2003) noting the followings: Their theorems are based on the assumptions that (i) the process is not necessarily stationary/ergodic, but only Harris recurrent; and (ii) the path of the diffusion process is a.s. uniformly continuous with the degree of $\sqrt{\delta\log(1/\delta)}$. Bandi and Phillips (2003) then require that $\sqrt{\delta\log(1/\delta)}\bar{L}(x,T)/b\to 0$ for both the consistency and the asymptotic normality, with $\bar{L}(x,T)$ denoting the chronological local time. However, when the ergodicity of the process is assumed, this condition can be weakened to $\sqrt{\delta\log(1/\delta)}/b$ and $\sqrt{Tb}\times\sqrt{\delta\log(1/\delta)}\to 0$ for the consistency and asymptotic normality, respectively.

Finally, note that the uniform continuity assumption in (ii) may not be always justified under the long-span scheme (see Kanaya, 2010b). Instead, we only assume that the degree of the uniform continuity of the diffusion process is δ^{γ} (for some $\gamma \in (0, 1/2)$) instead of $\sqrt{\delta \log(1/\delta)}$, which can be justified by the condition (A.2.ii) and arguments in Appendix A.6. ■

A.3 Proof of Theorem 3: Nonparametric Diffusion Estimator

We follow the same strategy as in the proof of Theorem 2: First, write

$$\hat{\beta}^2(x) - \beta^2(x) = \left[\hat{\beta}^2(x) - \tilde{\beta}^2(x)\right] + \left[\tilde{\beta}^2(x) - \beta^2(x)\right],$$

where the two terms in the right-hand side are analyzed in Lemma 4 and 5 respectively. Note that $h^{\gamma}/\delta^{1-\gamma} \to 0$ and $\delta^{\gamma}/b \to 0$ respectively imply $h^{\gamma} < \delta^{1-\gamma}$ and $\delta^{1-\gamma} < b^{\frac{1-\gamma}{\gamma}}$, and thus $h^{\gamma}/b < b^{\frac{1-\gamma}{\gamma}-1} = b^{\frac{1-2\gamma}{\gamma}}$. Therefore, the condition of Lemma 4 is satisfied with $q = (1-2\gamma)/\gamma$. The condition (B-NDI.ii) ensures that $\hat{\beta}^2(x) - \tilde{\beta}^2(x) = o_P(1/\sqrt{Nb})$, and thus the asymptotic distribution is determined by $\sqrt{Nb}\left[\tilde{\beta}^2(x) - \beta^2(x)\right]$.

Lemma 4 *Assume that (A.1), (A.2') and (B.1') are satisfied, $K \in \mathbb{K}(1,1)$ and $\mathcal{K} \in \mathbb{K}(2,2)$. If there exists some $q\,(>0)$ such that $h^{\gamma}/b = O(b^{q})$, then*

$$\hat{\beta}^2(x) - \tilde{\beta}^2(x) = O_P(h^{\gamma}/b) + O_P\left(h^{\gamma}/\delta^{1-\gamma}\right),$$

as $N, T \to \infty$ with $b, \delta \to 0$ and $h^{\gamma}/\delta^{1-\gamma} \to 0$.

Proof. Similarly to the proof of Lemma 2,

$$\hat{\beta}^2(x) - \tilde{\beta}^2(x) = B_1 + B_2 + B_3,$$

where

$$B_1 := \frac{(1/T)\sum_{j=1}^{N-1}\mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)\left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)^2}{(\delta/T)\sum_{j=1}^{N}\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j} - x\right)} - \frac{(1/T)\sum_{j=1}^{N-1}\mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)\left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)^2}{(\delta/T)\sum_{j=1}^{N}\mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)},$$

$$B_2 := \frac{(1/T)\sum_{j=1}^{N-1}\left[\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j} - x\right) - \mathcal{K}_b\left(\sigma^2_{\tau_j} - x\right)\right]\left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)^2}{(\delta/T)\sum_{j=1}^{N}\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j} - x\right)},$$

$$B_3 := \frac{(1/T)\sum_{j=1}^{N-1}\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_{j-1}} - x\right)\left[\left(\hat{\sigma}^2_{\tau_{j+1}} - \hat{\sigma}^2_{\tau_j}\right)^2 - \left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)^2\right]}{(\delta/T)\sum_{j=1}^{N}\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_j} - x\right)}.$$

We claim that $B_1 = O_P(h^{\gamma}/b)$, $B_2 = O_P(h^{\gamma}/b)$, and $B_3 = O_P\left(h^{\gamma}/\delta^{1-\gamma}\right)$. First, the convergence of B_1 follows from the same arguments as for A_1 in Lemma 1. Next, look at the numerator of the term B_2:

$$
\begin{aligned}
& \frac{1}{Tb}\sum_{j=1}^{N-1}\mathcal{K}'\left(\frac{\sigma^2_{\tau_j}-x}{b}+w_j\left(\hat{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}\right)b^{-1}\right)\left(\hat{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}\right)b^{-1}\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)^2 \\
= \;& \frac{1}{Tb}\sum_{j=1}^{N-1}\mathcal{K}'\left(\frac{\sigma^2_{\tau_j}-x}{b}+w_j\left(\hat{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}\right)b^{-1}\right)\left(\hat{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}\right)b^{-1} \\
& \times\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)^2\mathbf{1}\left\{w_j\left(\hat{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}\right)b^{-1}\leq\bar{\epsilon}\right\} \\
& +\frac{1}{Tb}\sum_{j=1}^{N-1}\mathcal{K}'\left(\frac{\sigma^2_{\tau_j}-x}{b}+w_j\left(\hat{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}\right)b^{-1}\right)\left(\hat{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}\right)b^{-1} \\
& \times\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)^2\mathbf{1}\left\{w_j\left(\hat{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}\right)b^{-1}>\bar{\epsilon}\right\} \\
\leq \;& \sup_{1\leq j\leq N-1}\left|\hat{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}\right|b^{-1}\times\left\{\frac{1}{Tb}\sum_{j=1}^{N-1}\mathcal{K}^*\left(\frac{\sigma^2_{\tau_j}-x}{b}\right)\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)^2\right. \\
& \left.+\frac{\bar{K}}{Tb}\sum_{j=1}^{N-1}\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)^2\times\left(\frac{\sup_{1\leq j\leq N-1}w_j|\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}|b^{-1}}{\bar{\epsilon}}\right)^{1/q}\right. \\
= \;& O_P\left(h^\gamma/b\right)\times\left\{O_P(1)+\frac{\bar{K}}{b}\times O_P(1)\times O_P\left(\left(\frac{b^q}{\bar{\epsilon}}\right)^{1/q}\right)\right\},
\end{aligned}
$$

where the first equality holds by the mean-value theorem with $w_j\in[0,1]$, $j=1,\ldots,N-1$, being some random variables. The last equality holds since we can show

$$
\frac{1}{Tb}\sum_{j=1}^{N-1}\mathcal{K}^*\left(\frac{\sigma^2_{\tau_j}-x}{b}\right)\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)^2=O_P(1),\quad \frac{1}{T}\sum_{j=1}^{N-1}\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)^2=O_P(1).
$$

Finally, to show the convergence of B_3, look at the following quantity:

$$
\begin{aligned}
& \sup_{1\leq j\leq N-1}\left|\left(\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right)^2-\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)^2\right| \\
& \leq 2\sup_{\tau\in[0,T]}\left|\hat{\sigma}^2_\tau-\sigma^2_\tau\right|\times\sup_{1\leq j\leq N-1}\left|\left(\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right)+\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)\right| \\
& =2\sup_{\tau\in[0,T]}\left|\hat{\sigma}^2_\tau-\sigma^2_\tau\right|\times\sup_{1\leq j\leq N-1}\left|\left(\hat{\sigma}^2_{\tau_{j+1}}-\sigma^2_{\tau_{j+1}}+\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)+\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}+\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right)\right| \\
& \leq 2\sup_{\tau\in[0,T]}\left|\hat{\sigma}^2_\tau-\sigma^2_\tau\right|\times\left[2\sup_{\tau\in[0,T]}\left|\hat{\sigma}^2_\tau-\sigma^2_\tau\right|+2\sup_{1\leq j\leq N-1}\left|\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right|\right] \\
& =O_P\left(h^{2\gamma}\right)+O_P\left(h^\gamma\right)\times O_{a.s.}\left(\delta^\gamma\right),
\end{aligned}
\tag{59}
$$

where the last equality by eq. (12) and Lemma 1. Note that $T^{-1}\sum_{j=1}^{N-1}\mathcal{K}_b\left(\hat{\sigma}^2_{\tau_{j-1}}-x\right)=O_P\left(\delta^{-1}\right)$. Note also that $\gamma\in(0,1/2)$ and $h^\gamma/\delta^{1-\gamma}\to0$ imply $h^\gamma/\delta^\gamma\to0$, and therefore $h^{2\gamma}\delta^{-1}=h^\gamma/\delta^{1-\gamma}\times h^\gamma/\delta^\gamma\leq h^\gamma/\delta^{1-\gamma}$, which, together with eq. (59), establishes the desired result $B_3=O_P\left(h^\gamma/\delta^{1-\gamma}\right)$. This completes the proof. ■

Lemma 5 *Assume that (A.2') holds, and $\mathcal{K} \in \mathbb{K}(2,2)$. If $N, T \to \infty$, $b \to 0$ with $\delta^{\gamma}/b \to 0$ and $Nb \to \infty$, then $\tilde{\beta}^2(x) \xrightarrow{P} \beta^2(x)$. If in addition $Nb\delta^{2\gamma} \to 0$ and $Nb^5 = (1)$, then*

$$\sqrt{Nb}\left[\tilde{\beta}^2(x) - \beta^2(x) - b^2 \times bias_{\beta^2}(x)\right] \xrightarrow{d} N\left(0, 4\beta^2(x)\int \mathcal{K}^2(z)\,dz\right),$$

where $bias_{\beta^2}(x)$ is given in Theorem 3.

Proof. This follows from the same arguments as in Bandi and Phillips (2003); see also the remarks made in the proof of Lemma 3. ∎

A.4 Proof of Theorems 4-5: Semiparametric Estimators

To derive the asymptotic results for the proposed estimator, we re-define the objective functions. Instead of $\hat{Q}_k(\theta_k)$ (for $k = 1,2$) in eq. (7), we consider the following objective functions:

$$\begin{aligned}
\hat{R}_1(\theta_1, \sigma^2) &:= \frac{1}{T}\sum_{i=1}^{N-1}\alpha\left(\sigma^2_{\tau_{j+1}};\theta_1\right)\left[\alpha\left(\sigma^2_{\tau_{j+1}};\theta_1\right)\delta - 2\left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)\right],\\
\hat{R}_2(\theta_2, \sigma^2) &:= \frac{1}{T}\sum_{i=1}^{N-1}\beta^2\left(\sigma^2_{\tau_{j+1}};\theta_2\right)\left[\beta^2\left(\sigma^2_{\tau_{j+1}};\theta_2\right)\delta - 2\left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)^2\right].
\end{aligned}$$

The maximizer of $R_k\left(\theta_k, \hat{\sigma}^2\right)$ is equal to the original estimator $\hat{\theta}_k$ defined in eq. (9) as the maximizer of $\hat{Q}_k(\theta_k)$. This re-definition facilitates our subsequent analyses (note that for any normalization factor a_T, $a_T^{-1}\tilde{Q}_1(\theta_1)$ does not converge to the mean of the squared difference $\int_I [\alpha(y;\theta_1) - \alpha(y)]^2 \pi(y)\,dy$; the same argument also applies to $a_T^{-1}\tilde{Q}_2(\theta_2)$). We also introduce their limits:

$$\begin{aligned}
R_1(\theta_1) &:= \int_I \alpha(y;\theta_1)\left[\alpha(y;\theta_1) - 2\alpha(y)\right]\pi(y)\,dy;\\
R_2(\theta_2) &:= \int_I \beta^2(y;\theta_2)\left[\beta^2(y;\theta_2) - 2\beta^2(y)\right]\pi(y)\,dy.
\end{aligned}$$

The first and second order derivatives of the above objective functions are

$$\begin{aligned}
\hat{S}_1(\theta_1, \sigma^2) &:= \frac{2}{T}\sum_{j=1}^{N-1}\partial_{\theta_1}\alpha\left(\sigma^2_{\tau_j};\theta_1\right)\left[\alpha\left(\sigma^2_{\tau_j};\theta_1\right)\delta - \left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)\right],\\
\hat{H}_1(\theta_1, \sigma^2) &:= \frac{2}{T}\sum_{j=1}^{N-1}\Big\{\partial_{\theta_1}\alpha\left(\sigma^2_{\tau_j};\theta_1\right)\partial_{\theta_1}\alpha\left(\sigma^2_{\tau_j};\theta_1\right)^{\star}\delta\\
&\quad + \partial_{\theta_1\theta_1^{\star}}\alpha\left(\sigma^2_{\tau_j};\theta_1\right)\left[\alpha\left(\sigma^2_{\tau_j};\theta_1^*\right)\delta - \left(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}\right)\right]\Big\};\\
H_1(\theta_1) &:= 2E\left[\partial_{\theta_1}\alpha\left(\sigma^2_t;\theta_1\right)\partial_{\theta_1}\alpha\left(\sigma^2_t;\theta_1\right)^{\star}\right];
\end{aligned}$$

and

$$\begin{aligned}
\hat{S}_2\left(\theta_2,\sigma^2\right) &: = \frac{2}{T}\sum_{j=1}^{N-1}\partial_{\theta_2}\beta^2\left(\sigma_{\tau_j}^2;\theta_2\right)\left[\beta^2\left(\sigma_{\tau_j}^2;\theta_2\right)\delta-\left(\sigma_{\tau_{j+1}}^2-\sigma_{\tau_j}^2\right)^2\right];\\
\hat{H}_2\left(\theta_2,\sigma^2\right) &: = \frac{2}{T}\sum_{j=1}^{N-1}\left\{\partial_{\theta_2}\beta^2\left(\sigma_{\tau_j}^2;\theta_2\right)\partial_{\theta_2}\beta^2\left(\sigma_{\tau_j}^2;\theta_2\right)^{\bigstar}\delta\right.\\
&\quad\left.+\partial_{\theta_2\theta_2^{\bigstar}}\beta^2\left(\sigma_{\tau_j}^2;\theta_2\right)\left[\beta^2\left(\sigma_{\tau_j}^2;\theta_2^*\right)\delta-\left(\sigma_{\tau_{j+1}}^2-\sigma_{\tau_j}^2\right)^2\right]\right\};\\
H_2\left(\theta_2\right) &: = 2E\left[\partial_{\theta_2}\beta^2\left(\sigma_t^2;\theta_2\right)\partial_{\theta_2}\beta^2\left(\sigma_t^2;\theta_2\right)^{\bigstar}\right].
\end{aligned}$$

We first state the asymptotic distribution of the infeasible estimators:

Lemma 6 *Suppose that the conditions in (A.2') are satisfied, and let $T\to\infty$ and $\delta\to 0$.*
(i) If (A-SDR) holds, then

$$\sup_{\theta_k\in\Theta_k}\left|\hat{R}_k\left(\theta_k,\sigma^2\right)-R_k\left(\theta_k\right)\right| = o_P(1); \quad \text{and} \quad \sup_{\theta_k\in\Theta_k}\left\|\hat{H}_k\left(\theta_k,\sigma^2\right)-H_k\left(\theta_k\right)\right\| = o_P(1), \tag{60}$$

with $k=1$ and thus $\tilde{\theta}_1\xrightarrow{P}\theta_1^*$. Moreover, if $T\delta^2\to 0$, then

$$\sqrt{T}\hat{S}_1\left(\theta_1^*,\sigma^2\right)\xrightarrow{d}N\left(0,\Omega_1^*\right); \quad \text{and} \quad \sqrt{T}\left[\tilde{\theta}_1-\theta_1^*\right]\xrightarrow{d}N\left(0,H_1^{*-1}\Omega_1^*H_1^{*-1}\right),$$

where Ω_1^* and H_1^* are given in Theorem 4.
(ii) If (A-SDI) holds, then the results (60) hold with $k=2$ and thus $\tilde{\theta}_2\xrightarrow{P}\theta_2^*$. Moreover, if $T\delta\to 0$, then

$$\sqrt{n}\hat{S}_n\left(\theta_2^*,\sigma^2\right)\xrightarrow{d}N\left(0,\Omega_2^*\right); \quad \text{and} \quad \sqrt{n}\left[\tilde{\theta}_2-\theta_2^*\right]\xrightarrow{d}N\left(0,H_2^{*-1}\Omega_2^*H_2^{*-1}\right),$$

where Ω_2^* and H_2^* are given in Theorem 5.

Proof. This follows along the same lines as in Sørensen (2009) or Yoshida (1992), and so we omit the proof. ■

Next, we derive the stochastic difference between the feasible and infeasible objective function and its derivatives

Lemma 7 *Assume that (A.1), (A.2') and (B.1') are satisfied, $K\in\mathbb{K}(1,1)$. If the condition (A-SDR.iii) holds, then*

$$\left|\hat{R}_1\left(\theta_1,\hat{\sigma}^2\right)-\hat{R}_1\left(\theta_1,\sigma^2\right)\right| = O_P\left(h^\gamma/\delta\right); \tag{61}$$

$$\sqrt{T}\left\|\hat{S}_1\left(\theta_1^*,\hat{\sigma}^2\right)-\hat{S}_1\left(\theta_1^*,\sigma^2\right)\right\| = O_P\left(T^{1/2}h^\gamma/\delta\right); \tag{62}$$

$$\left\|\hat{H}_1\left(\theta_1,\hat{\sigma}^2\right)-\hat{H}_1\left(\theta_1,\sigma^2\right)\right\| = O_P\left(h^\gamma/\delta\right), \tag{63}$$

where (61) and (63) hold uniformly over any $\theta_1\in\Theta_1$.

Proof. We only prove eq. (61) since the proofs of eqs. (62)-(63) are analogous. Write $\hat{R}_1\left(\theta_1,\hat{\sigma}^2\right)-\hat{R}_1\left(\theta_1,\sigma^2\right)=\mathbf{a}\left(\theta_1\right)-2\mathbf{b}\left(\theta_1\right)$, where

$$\begin{aligned}
\mathbf{a}\left(\theta_1\right) &:= \frac{1}{T}\sum_{j=1}^{N-1}\left[\alpha^2\left(\sigma^2_{\tau_j};\theta_1\right)-\alpha^2\left(\hat{\sigma}^2_{\tau_j};\theta_1\right)\right]\delta,\\
\mathbf{b}\left(\theta_1\right) &:= \frac{1}{T}\sum_{j=1}^{N-1}\left[\alpha\left(\sigma^2_{\tau_j};\theta_1\right)\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)-\alpha\left(\hat{\sigma}^2_{\tau_j};\theta_1\right)\left(\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right)\right].
\end{aligned}$$

We below derive the convergence rate of each term. By the mean-value theorem, for some $w_j\in[0,1]$,

$$\begin{aligned}
\mathbf{a}\left(\theta_1\right) &= \frac{1}{N}\sum_{j=1}^{N-1}\left[\partial_y\alpha\left(\sigma^2_{\tau_j}+w_j\left(\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right);\theta_1\right)\alpha\left(\sigma^2_{\tau_j}+w_j\left(\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right);\theta_1\right)\right]\left(\sigma^2_{\tau}-\hat{\sigma}^2_{\tau}\right)\\
&\leq \frac{1}{N}\sum_{j=1}^{N-1}C\left[1+\left|\sigma^2_{\tau_j}+w_j\left(\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right)\right|^{v_1}\right]^2\times\sup_{\tau\in[0,T]}\left|\sigma^2_\tau-\hat{\sigma}^2_\tau\right|\\
&\leq \frac{1}{N}\sum_{j=1}^{N-1}C\left[1+\left|\sigma^2_{\tau_j}\right|^{2v_1}+\left|\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right|^{2v_1}\right]\times O_P\left(h^\gamma\right)=O_P\left(h^\gamma\right),
\end{aligned}$$

where the second inequality follows from (A-SDR.iii); the third inequality from the inequality: $(A+B)^{v_1}\leq C_{v_1}(|A|^{2v_1}+|B|^{2v_1})$ for some constant $C_{v_1}\,(>0)$, and the last equality holds since $N^{-1}\sum_{j=1}^{N-1}|\sigma^2_{\tau_j}|^{2v_1}=O_p(1)$, which follows from $E[\left|\sigma^2_s\right|^{2v_1}]<\infty$, and $|\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}|=O_P\left(h^\gamma\right)$ uniformly over any j. Next, by the same arguments, it holds that

$$\begin{aligned}
\mathbf{b}\left(\theta_1\right) &= \frac{1}{T}\sum_{j=1}^{N-1}\alpha\left(\sigma^2_{\tau_j};\theta_1\right)\left[\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)-\left(\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right)\right]\\
&\quad+\frac{1}{T}\sum_{j=1}^{N-1}\left[\alpha\left(\sigma^2_{\tau_j};\theta_1\right)-\alpha\left(\hat{\sigma}^2_{\tau_j};\theta_1\right)\right]\left(\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right)\\
&\leq \frac{1}{N}\sum_{j=1}^{N-1}\left|\alpha\left(\sigma^2_{\tau_j};\theta_1\right)\right|\times 2\sup_{\tau\in[0,T]}\left|\sigma^2_\tau-\hat{\sigma}^2_\tau\right|/\delta\\
&\quad+\frac{1}{N}\sum_{j=1}^{N-1}\left|\partial_y\alpha\left(\sigma^2_{\tau_j}+w_j\left(\hat{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}\right);\theta_1\right)\right|\times\sup_{\tau\in[0,T]}\left|\sigma^2_\tau-\hat{\sigma}^2_\tau\right|\times\sup_{1\leq j\leq N-1}\left|\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right|/\delta\\
&= O_P\left(h^\gamma/\delta\right)+O_P\left(h^\gamma\left[h^\gamma+\delta^\gamma\right]/\delta\right),
\end{aligned}$$

where the last equality holds since $N^{-1}\sum_{j=1}^{N-1}\left|\partial_y\alpha\left(\sigma^2_{\tau_j}+w_j\left(\hat{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}\right);\theta_1\right)\right|=O_P(1)$ (uniformly over θ_1), and

$$\begin{aligned}
\sup_{1\leq j\leq N-1}\left|\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right| &= \sup_{1\leq j\leq N-1}\left|\hat{\sigma}^2_{\tau_{j+1}}-\sigma^2_{\tau_{j+1}}+\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}+\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right|\\
&\leq 2\sup_{\tau\in[0,T]}\left|\sigma^2_\tau-\hat{\sigma}^2_\tau\right|+\sup_{1\leq j\leq N-1}\left|\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right|\\
&= O_P\left(h^\gamma\right)+O_{a.s.}\left(\delta^\gamma\right)=O_P\left(h^\gamma+\delta^\gamma\right). \qquad (64)
\end{aligned}$$

Now, we have $\mathbf{b}\left(\theta_1\right)=O_P\left(h^\gamma/\delta^{1-\gamma}\right)$ uniformly over θ_1. ■

Lemma 8 *Assume that (A.1), (A.2') and (B.1') are satisfied, $K \in \mathbb{K}(1,1)$. If the condition (A-SDI.iii) holds, then*

$$\begin{aligned}
\left|\hat{R}_2\left(\theta_2,\hat{\sigma}^2\right)-\hat{R}_2\left(\theta_2,\sigma^2\right)\right| &= O_P\left(h^{\gamma}/\delta^{1-\gamma}\right); &(65)\\
\sqrt{N}\left\|\hat{S}_2\left(\theta_2^*,\hat{\sigma}^2\right)-\hat{S}_2\left(\theta_2^*,\sigma^2\right)\right\| &= O_P\left(\sqrt{N}h^{\gamma}/\delta^{1-\gamma}\right); &(66)\\
\left\|\hat{H}_2\left(\theta_2,\hat{\sigma}^2\right)-\hat{H}_2\left(\theta_2,\sigma^2\right)\right\| &= O_P\left(h^{\gamma}/\delta^{1-\gamma}\right), &(67)
\end{aligned}$$

where (65) and (67) hold uniformly over any $\theta_2 \in \Theta_2$.

Proof. We only prove eq. (65). Eqs. (66) and (67) can be shown analogously. Write $\hat{R}_2\left(\theta_2,\hat{\sigma}^2\right) - \hat{R}_2\left(\theta_2,\sigma^2\right) = \mathbf{c}(\theta_2) - 2\mathbf{d}(\theta_2)$, where

$$\begin{aligned}
\mathbf{c}(\theta_2) &: = \frac{1}{N}\sum_{i=1}^{N-1}\left[\beta^4\left(\hat{\sigma}^2_{\tau_{j+1}};\theta_2\right)-\beta^4\left(\sigma^2_{\tau_{j+1}};\theta_2\right)\right];\\
\mathbf{d}(\theta_2) &: = \frac{1}{T}\sum_{i=1}^{N-1}\left[\beta^2\left(\hat{\sigma}^2_{\tau_{j+1}};\theta_2\right)\left(\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right)^2-\beta^2\left(\sigma^2_{\tau_{j+1}};\theta_2\right)\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)^2\right].
\end{aligned}$$

We can show that $\mathbf{c}(\theta_2) = O_P(h^{\gamma})$ (uniformly over θ_2) by using the same arguments as for the convergence of $\mathbf{a}(\theta_1)$ in the proof of Lemma 7. As for $\mathbf{d}(\theta_2)$,

$$\begin{aligned}
\mathbf{d}(\theta_2) &= \frac{1}{T}\sum_{i=1}^{N-1}\left[\beta^2\left(\hat{\sigma}^2_{\tau_{j+1}};\theta_2\right)-\beta^2\left(\sigma^2_{\tau_{j+1}};\theta_2\right)\right]\left(\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right)^2\\
&\quad+\frac{1}{T}\sum_{i=1}^{N-1}\beta^2\left(\sigma^2_{\tau_{j+1}};\theta_2\right)\left[\left(\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right)^2-\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)^2\right]\\
&\leq \underbrace{\frac{1}{N}\sum_{i=1}^{N-1}\partial_y\beta^2\left(\sigma^2_{\tau_j}+w_j\left(\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}\right);\theta_2\right)}_{=O_P(1)}\times\sup_{\tau\in[0,T]}\left|\sigma^2_{\tau}-\hat{\sigma}^2_{\tau}\right|\times\sup_{1\leq j\leq N-1}\left|\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right|^2/\delta\\
&\quad+\underbrace{\frac{1}{N}\sum_{i=1}^{N-1}\beta^2\left(\sigma^2_{\tau_{j+1}};\theta_2\right)}\sup_{1\leq j\leq N-1}\left|\left(\hat{\sigma}^2_{\tau_{j+1}}-\hat{\sigma}^2_{\tau_j}\right)^2-\left(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}\right)^2\right|/\delta\\
&= O_P(h^{\gamma})\times O_P\left(h^{2\gamma}+\delta^{2\gamma}\right)\times(1/\delta)+\left[O_P\left(h^{2\gamma}\right)+O_P(h^{\gamma})\times O_{a.s.}(\delta^{\gamma})\right]\times(1/\delta),
\end{aligned}$$

where the last equality follows from eq. (64) and eq. (59). Noting that $\gamma \in (0,1/2)$, we have $\mathbf{d}(\theta_2) = O_P\left(h^{\gamma}/\delta^{1-\gamma}\right)$. ■

We are now ready to prove Theorems 4-5:

Proof of Theorem 4. To prove consistency, we verify the conditions in Newey and McFadden (1994, Theorem 2.1): (i) compactness of the parameter space; (ii) continuity of the objective function and its limit function; (iii) uniform convergence of the objective function; and (iv) identifiability. Conditions (A-SDR.i) and (A-SDR.ii) imply (i), (ii) and (iv), and we only need to show uniform convergence. Write

$$\hat{R}_1\left(\theta_1,\hat{\sigma}^2\right)-R_1(\theta_1)=\left[\hat{R}_1\left(\theta_1,\hat{\sigma}^2\right)-\hat{R}_1\left(\theta_1,\sigma^2\right)\right]+\left[\hat{R}_1\left(\theta_1,\sigma^2\right)-R_1(\theta_1)\right], \tag{68}$$

where the two terms in the right-hand side converge uniformly by Lemma 7 and 6 respectively.

For the asymptotic normality, we introduce the score and Hessian functions. By the Taylor expansion,

$$\sqrt{T}\left[\hat{\theta}_1 - \theta_1^*\right] = \hat{H}_1^{-1}\left(\bar{\theta}_1, \hat{\sigma}^2\right)\sqrt{T}\hat{S}_1\left(\theta_1^*, \hat{\sigma}^2\right),$$

where $\bar{\theta}_1$ is on the line segment connecting $\hat{\theta}_1$ to θ_1^*. By (62) in Lemma 8, $\sqrt{T}[\hat{S}_1\left(\theta_1^*, \hat{\sigma}^2\right) - \hat{S}_1\left(\theta_1^*, \sigma^2\right)] = o_P(1)$ with (B-SDR), while the Hessian satisfies

$$\hat{H}_1\left(\theta_1, \hat{\sigma}^2\right) - H_1(\theta_1) = \left[\hat{H}_1\left(\theta_1, \hat{\sigma}^2\right) - \hat{H}_1\left(\theta_1, \sigma^2\right)\right] + \left[\hat{H}_1\left(\theta_1, \sigma^2\right) - H_1(\theta_1)\right] \to^P 0$$

uniformly θ_1 by Lemmas 7 and 6. Thus, $\hat{\theta}_1$ has the same asymptotic distribution as the infeasible estimator $\tilde{\theta}_1$, which is given in Lemma 6. ■

Proof of Theorem 5. This follows along the same lines as the proof of Theorem 4. ■

A.5 Proof of Theorem 6: Expansions of approximated MSEs

Let

$$\begin{aligned} U_1(j) &:= 2\partial_{\theta_1}\alpha\left(\sigma_{\tau_j}^2;\theta_1^*\right)\left[\alpha(\sigma_{\tau_j}^2;\theta_1^*)\delta - (\sigma_{\tau_{j+1}}^2 - \sigma_{\tau_j}^2)\right]; \\ U_2(j) &:= 2\partial_{\theta_2}\beta^2\left(\sigma_{\tau_j}^2;\theta_2^*\right)[\beta^2(\sigma_{\tau_j}^2;\theta_2^*)\delta - (\sigma_{\tau_{j+1}}^2 - \sigma_{\tau_j}^2)^2]. \end{aligned}$$

Then, we can write $\hat{S}_k\left(\theta_k^*, \sigma^2\right) = T^{-1}\sum_{j=1}^{N-1} U_k(j)$, and

$$H_k^{*-1}E\left[\hat{S}_k\left(\theta_k^*, \sigma^2\right)\hat{S}_k\left(\theta_k^*, \sigma^2\right)^{\star}\right]H_k^{*-1} = \mathcal{B}_{\theta_k}\mathcal{B}_{\theta_k}^{\star} + \mathcal{V}_{\theta_k} + \mathcal{C}_{\theta_k},$$

where

$$\begin{aligned} \mathcal{B}_{\theta_k} &:= -H_k^{*-1}E\left[\hat{S}_k\left(\theta_k^*, \sigma^2\right)\right]; \\ \mathcal{V}_{\theta_k} &:= H_k^{*-1}\left(T^{-2}\sum_{j=1}^{N-1}E\left[\left(U_k(j) - E[U_k(j)]\right)\left(U_k(j) - E[U_k(j)]\right)^{\star}\right]\right)H_k^{*-1}; \\ \mathcal{C}_{\theta_k} &:= H_k^{*-1}\left(T^{-2}\sum_{1\le i\neq j\le N-1}E\left[\left(U_k(i) - E[U_k(i)]\right)\left(U_k(j) - E[U_k(j)]\right)^{\star}\right]\right)H_k^{*-1}. \end{aligned}$$

We below provide the proof for part (i) ($k = 1$) only. Part (ii) ($k = 2$) can be proved in the same way, and its proof is omitted. Let $\mathcal{L}$ be the differential operator defined by $\mathcal{L}f(x) = f'(x)\alpha(x) + f''(x)\beta^2(x)/2$ for any twice differentiable function f. We first consider the expression of $\mathcal{B}_{\theta_1}$:

$$\begin{aligned} E[U_k(j)] &= -2E\left[\partial_{\theta_1}\alpha\left(\sigma_{\tau_j}^2;\theta_1^*\right)\int_{\tau_j}^{\tau_{j+1}}\int_{\tau_j}^{s}\mathcal{L}\alpha\left(\sigma_u^2\right)duds\right] \\ &= -2E\left[\int_{\tau_j}^{\tau_{j+1}}\int_{\tau_j}^{s}\partial_{\theta_1}\alpha\left(\sigma_u^2;\theta_1^*\right)\mathcal{L}\alpha\left(\sigma_u^2\right)duds\right] \\ &\quad +2\int_{\tau_j}^{\tau_{j+1}}\int_{\tau_j}^{s}E\left[\left[\partial_{\theta_1}\alpha\left(\sigma_u^2;\theta_1^*\right) - \partial_{\theta_1}\alpha\left(\sigma_{\tau_j}^2;\theta_1^*\right)\right]\mathcal{L}\alpha\left(\sigma_u^2\right)\right]duds \\ &= -\delta^2E\left[\partial_{\theta_1}\alpha\left(\sigma_u^2;\theta_1^*\right)\mathcal{L}\alpha\left(\sigma_u^2\right)\right] + 2\int_{\tau_j}^{\tau_{j+1}}\int_{\tau_j}^{s}\int_{\tau_j}^{u}E\left[\mathcal{L}\partial_{\theta_1}\alpha\left(\sigma_v^2;\theta_1^*\right)\mathcal{L}\alpha\left(\sigma_u^2\right)\right]dvduds \\ &= -\delta^2E\left[\partial_{\theta_1}\alpha\left(\sigma_t^2;\theta_1^*\right)\mathcal{L}\alpha\left(\sigma_t^2\right)\right][1 + O(\delta)], \qquad (69) \end{aligned}$$

uniformly over j, where the first and third equalities use the martingale property of stochastic integrals and Ito's lemma, which is applied to $\alpha\left(\sigma_s^2\right)-\alpha(\sigma_{\tau_j}^2)$ and $\partial_{\theta_1}\alpha\left(\sigma_u^2;\theta_1^*\right)-\partial_{\theta_1}\alpha\left(\sigma_{\tau_j}^2;\theta_1^*\right)$; and the last equality holds since

$$E\left[\mathcal{L}\partial_{\theta_1}\alpha\left(\sigma_v^2;\theta_1^*\right)\mathcal{L}\alpha\left(\sigma_u^2\right)\right]\leq\left\{E\left[\left|\mathcal{L}\partial_{\theta_1}\alpha\left(\sigma_v^2;\theta_1^*\right)\right|^2\right]E\left[\left|\mathcal{L}\alpha\left(\sigma_u^2\right)\right|^2\right]\right\}^{1/2}\leq E\left[\left|\psi\left(\sigma_t^2\right)\right|^4\right]=O\left(1\right),$$

uniformly over any u and v, which follows from the moment conditions in (C-SDR). Now, the above definition of $\mathcal{B}_{\theta_k}$ and eq. (69) implies the expression (16).

To find the expression of $\mathcal{V}_{\theta_k}$, first note that

$$\begin{aligned}
&E\left[U_1\left(j\right)U_1\left(j\right)^{\bigstar}\right]=4E\Bigg[\partial_{\theta_1}\alpha\left(\sigma_t^2;\theta_1^*\right)\partial_{\theta_1}\alpha\left(\sigma_t^2;\theta_1^*\right)^{\bigstar}\\
&\times\Bigg\{\alpha^2\left(\sigma_{\tau_j}^2\right)\delta^2-2\alpha\left(\sigma_{\tau_j}^2\right)\delta\left(\int_{\tau_j}^{\tau_{j+1}}\alpha\left(\sigma_s^2\right)ds+\int_{\tau_j}^{\tau_{j+1}}\beta\left(\sigma_s^2\right)dZ_s\right)\\
&+2\int_{\tau_j}^{\tau_{j+1}}\left(\int_{\tau_j}^{s}\alpha\left(\sigma_u^2\right)du+\int_{\tau_j}^{s}\beta\left(\sigma_u^2\right)dZ_u\right)\alpha\left(\sigma_s^2\right)ds\\
&\quad+2\int_{\tau_j}^{\tau_{j+1}}\left(\int_{\tau_j}^{s}\alpha\left(\sigma_u^2\right)du+\int_{\tau_j}^{s}\beta\left(\sigma_u^2\right)dZ_u\right)\beta\left(\sigma_s^2\right)dZ_s+\int_{\tau_j}^{\tau_{j+1}}\beta^2\left(\sigma_s^2\right)ds\Bigg\}\Bigg]\\
=\;&\underbrace{\delta E\left[\partial_{\theta_1}\alpha\left(\sigma_t^2;\theta_1^*\right)\partial_{\theta_1}\alpha\left(\sigma_t^2;\theta_1^*\right)^{\bigstar}\beta^2\left(\sigma_t^2\right)\right]}_{=\Omega_1^*}\left[1+O\left(\delta\right)\right],\text{ uniformly over } j
\end{aligned}$$

where Ito's lemma is applied to $(\sigma_{\tau_{j+1}}^2-\sigma_{\tau_j}^2)^2$ in the first equality; and the second equality follows from arguments similar to those in deriving eq. (69). By the definition of $\mathcal{V}_{\theta_k}$ and the result that $E\left[U_k\left(j\right)\right]=O\left(\delta^2\right)$ uniformly over j,

$$\mathcal{V}_{\theta_k}=H_k^{*-1}\left(T^{-2}\sum\nolimits_{j=1}^{N-1}\left[\delta\Omega_1^*\left[1+o\left(1\right)\right]-O\left(\delta^4\right)\right]\right)H_k^{*-1}=T^{-1}H_k^{*-1}\Omega_1^*H_k^{*-1}\left[1+O\left(\delta\right)\right],$$

implying eq. (17). To find the expression of $\mathcal{C}_{\theta_k}$, we write

$$\partial_{\theta_1}\alpha\left(\sigma_{\tau_j}^2;\theta_1^*\right)\left[\alpha(\sigma_{\tau_j}^2;\theta_1^*)\delta-(\sigma_{\tau_{j+1}}^2-\sigma_{\tau_j}^2)\right]=:\Upsilon_1\left(j\right)+\Upsilon_2\left(j\right),$$

where $\Upsilon_1\left(j\right):=-\partial_{\theta_1}\alpha\left(\sigma_{\tau_j}^2;\theta_1^*\right)\int_{\tau_j}^{\tau_{j+1}}\int_{\tau_j}^{s}\mathcal{L}\alpha\left(\sigma_u^2\right)duds$ and

$$\Upsilon_2\left(j\right):=\partial_{\theta_1}\alpha\left(\sigma_{\tau_j}^2;\theta_1^*\right)\left\{\int_{\tau_j}^{\tau_{j+1}}\int_{\tau_j}^{s}\alpha'\left(\sigma_u^2\right)\beta\left(\sigma_u^2\right)dZ_uds-\int_{\tau_j}^{\tau_{j+1}}\beta\left(\sigma_s^2\right)dZ_s\right\}.$$

Then, by the martingale property of stochastic integrals, Fubini's theorem and the conditions in (C-SDR), $E\left[\Upsilon_1\left(i\right)\Upsilon_2\left(j\right)^{\bigstar}\right]=0$ and $E\left[\Upsilon_2\left(i\right)\Upsilon_2\left(j\right)^{\bigstar}\right]=0$ for $i\neq j$. Given the moment conditions in (C-SDR), we can show that $E\left[\Upsilon_1\left(i\right)\Upsilon_1\left(j\right)^{\bigstar}\right]=O\left(\delta^4\right)$ uniformly over any $i\neq j$, by using arguments analogous to those for $\mathcal{B}_{\theta_k}$ and $\mathcal{V}_{\theta_k}$. This, together with eq. (69),

$$E\left[\left(U_k\left(i\right)-E\left[U_k\left(i\right)\right]\right)\left(U_k\left(j\right)-E\left[U_k\left(j\right)\right]\right)^{\bigstar}\right]=E[\Upsilon_1\left(i\right)\Upsilon_1\left(j\right)^{\bigstar}]-E\left[U_k\left(i\right)\right]E\left[U_k\left(j\right)\right]^{\bigstar}=O\left(\delta^4\right),$$

which, together with the definition of $\mathcal{C}_{\theta_k}$, implies that $\mathcal{C}_{\theta_k}=O\left(\delta^2\right)$. Now, the proof is completed.

A.6 Path Continuity of Stochastic Processes

In this section, we prove a general result for the uniform Hölder continuity of a stochastic process over an infinite interval $[0, \infty)$. This plays a key role in deriving the uniform consistency rate of the spot volatility estimator.

Lemma 9 *Suppose that a stochastic process $\{\Gamma_t\}_{t \geq 0}$ on a probability space $(\Omega, \mathfrak{F}, P)$ satisfies the condition:*

$$E\left[\left|\Gamma_t - \Gamma_s\right|^a\right] \leq C \left|t - s\right|^{1+b}, \tag{70}$$

for some positive constants a, b and C each of which is independent of s and t. Then, there exists a continuous modification $\{\tilde{\Gamma}_t\}_{t \geq 0}$ of $\{\Gamma_t\}_{t \geq 0}$, which is a.s. Hölder continuous with exponent d for every $d \in (0, b/a)$ with a coefficient $\vartheta := \sum_{J=1}^{\infty} J^{2+d} \left(1/J!\right)^d$:

$$\Pr\left[\omega \in \Omega \left| \exists \bar{\Delta}(\omega) \ \ s.t. \sup_{|t-s| \in \left(0, \bar{\Delta}(\omega)\right);\ s,t \in [0,\infty)} \frac{\left|\tilde{\Gamma}_t(\omega) - \tilde{\Gamma}_s(\omega)\right|}{\left|t - s\right|^d} \leq \vartheta \right.\right] = 1. \tag{71}$$

Proof. The following arguments proceed along the lines of the proof of Theorem 2.2.8 in Karatzas and Shreve (1991) where s, t are supposed to take values in some finite interval $[0, T]$ ($T = \bar{T}$ fixed). We first prove the global Hölder property of the process on an enlarging interval, i.e., $[0, T]$ where $T \to \infty$, and next show that it actually holds over the infinite interval $[0, \infty)$.

For any $\varepsilon > 0$, we have

$$\Pr\left[\left|\Gamma_t - \Gamma_s\right| \geq \varepsilon\right] \leq \frac{E\left[\left|\Gamma_t - \Gamma_s\right|^a\right]}{\varepsilon^a} \leq C\varepsilon^{-a} \left|t - s\right|^{1+b} \tag{72}$$

by Čebyšev's inequality, and thus $\Gamma_t \xrightarrow{P} \Gamma_s$ as $s \to t$. Setting $t = km/m!$, $s = (k-1)\, m/m!$ and $\varepsilon = (m/m!)^d$ (where $d \in (0, b/a)$) in eq. (72), we obtain

$$\Pr\left[\left|\Gamma_{km/m!} - \Gamma_{(k-1)m/m!}\right| \geq (m/m!)^d\right] \leq C\,(m/m!)^{(1+b-ad)}$$

and consequently,

$$\Pr\left[\max_{1 \leq k \leq m!} \left|\Gamma_{km/m!} - \Gamma_{(k-1)m/m!}\right| \geq (m/m!)^d\right] \leq Cm^{(1+b-ad)} \Big/ (m!)^{b-ad}.$$

By the fact that $\sum_{m=1}^{\infty} m^{(1+b-ad)} \Big/ (m!)^{b-ad}$ exists and the Borel-Cantelli lemma, there exists a set $\Omega^* \in \mathfrak{F}$ with $\Pr(\Omega^*) = 1$ such that

$$\forall \omega \in \Omega^*,\ \exists m^*(\omega),\ \forall m \geq m^*(\omega), \quad \max_{1 \leq k \leq m!} \left|\Gamma_{km/m!} - \Gamma_{(k-1)m/m!}\right| < (m/m!)^d, \tag{73}$$

where m^* is a positive and integer-valued random variable.

For each integer $m\,(\geq 1)$ and any integer $l \geq m$, consider the following sets:

$$E_l^m := \{km/l! \mid k = 0, 1, \ldots, l!\}; \quad \text{and} \quad E^m := \textstyle\bigcup_{l=1}^{\infty} E_l^m.$$

The set E_l^m consists of $(l!+1)$ points in $[0,m]$, while E^m consists of infinitely many points in $[0,m]$. Note that E^m is dense in $[0,m]$ for any m. Now fix $\omega\,(\in \Omega^*)$ and $m\,(\geq m^*(\omega))$. We shall show that for every $l\,(> m)$,

$$\begin{gathered}\forall t,s \in E_l^m \text{ with } |t-s| \in (0, m/m!)\,,\\ |\Gamma_t(\omega)-\Gamma_s(\omega)| \leq 2\sum_{J=m+1}^{l} J^2 \left[\frac{J+1}{(J+1)!}\right]^d.\end{gathered} \tag{74}$$

To show this, we use the inductive method. First, we prove that eq. (74) is valid for $l=m+1$. For any $s,t \in E_{m+1}^m$ with $|t-s| \in (0,m/m!)$, there exist some $k_1,k_2 \in \{0,1,\ldots,(m+1)!\}$ with $0 \leq k_2-k_1 \leq m$ such that

$$\begin{aligned}|\Gamma_t(\omega)-\Gamma_s(\omega)| \leq{}& \left|\Gamma_{k_1(m+1)/(m+1)!}(\omega)-\Gamma_{(k_1+1)(m+1)/(m+1)!}(\omega)\right|\\ &+\left|\Gamma_{(k_1+1)(m+1)/(m+1)!}(\omega)-\Gamma_{(k_1+2)(m+1)/(m+1)!}(\omega)\right|\\ &+\cdots+\left|\Gamma_{(k_2-1)(m+1)/(m+1)!}(\omega)-\Gamma_{k_2(m+1)/(m+1)!}(\omega)\right|.\end{aligned}$$

Each term of the right-hand side is bounded by $m\left[(m+2)/(m+2)!\right]^d$, which is implied by the fact $E_{m+1}^m \subset E_{m+2}^{m+2}$ and the inequality eq. (73). Thus, by the triangle inequalities, we have

$$|\Gamma_t(\omega)-\Gamma_s(\omega)| \leq m^2\left[\frac{(m+2)}{(m+2)!}\right]^d.$$

Second, suppose that eq. (74) is valid for $l=m+1,\ldots,L-1$. For $s<t$, $(s,t \in E_L^m)$ with $|t-s| \in (0,m/m!)$, consider the numbers $s_1 := \min\left\{u \in E_{L-1}^m : u \geq s\right\}$ and $t_1 := \max\left\{u \in E_{L-1}^m : u \leq t\right\}$, and notice that $s_1,t_1 \in E_{L-1}^m \subset E_L^m \subset E_{L+1}^{L+1}$; $s,t \in E_L^m \subset E_{L+1}^{L+1}$; $s_1-s < m/(L-1)!$; and $t-t_1 < m/(L-1)!$. By the inequality eq. (73) with $m=L+1$,

$$\begin{aligned}|\Gamma_{s_1}(\omega)-\Gamma_s(\omega)| &\leq mL\left((L+1)/(L+1)!\right)^d;\\ |\Gamma_t(\omega)-\Gamma_{t_1}(\omega)| &\leq mL\left((L+1)/(L+1)!\right)^d.\end{aligned} \tag{75}$$

There are two possible relationships among s,t,s_1 and t_1: (i) if $|t-s| \geq m/(L-1)!$, it holds that $s \leq s_1 \leq t_1 \leq t$ (with at least one inequality strict); (ii) if $|t-s| < m/(L-1)!$, either of $|t-s| < |s_1-t_1| = m/(L-1)!$ or $|s_1-t_1| = 0$. Thus, we have $|t_1-s_1| \leq \max\{m/(L-1)!, |t-s|\} \leq m/m!$, and use the induction assumption (74) with $l=L-1$:

$$|\Gamma_{t_1}(\omega)-\Gamma_{s_1}(\omega)| \leq 2\sum_{J=m+1}^{L-1} J^2\left[\frac{J+1}{(J+1)!}\right]^d. \tag{76}$$

Therefore, by eqs. (75) and (76) together with the triangle inequality,

$$|\Gamma_t(\omega)-\Gamma_s(\omega)| \leq 2mL\left((L+1)/(L+1)!\right)^d + 2\sum_{J=m+1}^{L-1} J^2\left[\frac{J+1}{(J+1)!}\right]^d < 2\sum_{J=m+1}^{L} J^2\left[\frac{J+1}{(J+1)!}\right]^d.$$

We have shown eq. (74) for any $l\,(> m)$, as desired.

We can now show that $\{\Gamma_t(\omega) \mid t \in E^m\}$ is uniformly Hölder in t for $\forall\omega \in \Omega^*$ for any m. Consider any numbers $s,t \in E^m$ with $m \geq m^*\,(= m^*(\omega))$ and $|t-s| < \bar{\Delta}(\omega) \equiv m^*/m^*!$. Note that

$E^m \subseteq E^{m'}$ for $m \leq m'$. We can pick some $m' (\geq m)$ such that $s, t \in E^{m'}$ with $(m'+2) / (m'+2)! \leq t - s < (m'+1) / (m'+1)!$. Then, by eq. (74), we obtain

$$|\Gamma_t(\omega) - \Gamma_s(\omega)| \leq 2 \sum_{J=m'+1}^{\infty} J^2 \left[\frac{J+1}{(J+1)!} \right]^d \leq \left[(m'+2) / (m'+2)! \right]^d \times \sum_{J=1}^{\infty} J^{2+d} (1/J!)^d$$

and thus, $|\Gamma_t(\omega) - \Gamma_s(\omega)| / |t-s|^d \leq c_h$ where $\vartheta := \sum_{J=1}^{\infty} J^{2+d} (1/J!)^d$. Note that the existence of ϑ can be checked by d'Alembert's criterion for any $d (> 0)$.

We define $\{\tilde{\Gamma}_t\}_{t \geq 0}$ as follows. For $\omega \notin \Omega^*$, set $\tilde{\Gamma}_t(\omega) = 0$, $t \in [0, m]$. For $\omega \in \Omega^*$ and $t \in E^m$, set $\tilde{\Gamma}_t(\omega) = \Gamma_t(\omega)$. For $\omega \in \Omega^*$ and $t \in [0, m] \cap (E^m)^c$, choose a sequence $\{s_n\}_{n=1}^{\infty}$with $s_n (\in E^m) \to t$; by the uniform continuity and the fact that s_n is Cauchy, $\{\Gamma_{s_n}(\omega)\}_{n=1}^{\infty}$ is also Cauchy, whose limit depends of t but not on the particular sequence $\{s_n\}$; and thus let $\tilde{\Gamma}_t(\omega) = \lim_{s_n \to t} \Gamma_{s_n}(\omega)$. Thus, the resulting process $\{\tilde{\Gamma}_t\}_{t \in [0,m]}$ is continuous, and is also uniformly Hölder in $t \in [0, m]$. We will show $\{\tilde{\Gamma}_t\}$ is a modification of $\{\Gamma_t\}$: observe that for $t \in E^m$, $\tilde{\Gamma}_t = \Gamma_t$ a.s.; for $t \in [0, m] \cap (E^m)^c$ and $\{s_n\}$ with $s_n (\in E^m) \to t$, we have $\Gamma_{s_n} \to \Gamma_t$ in probability (by eq. (70)) and $\Gamma_{s_n} \to \tilde{\Gamma}_t$ a.s., which implies $\tilde{\Gamma}_t = \Gamma_t$ a.s.

Let $m = [T] + 1$ with $[T]$ denoting the largest integer less than or equal to T. Now, we have proved that for any $\omega \in \Omega^*$, there exist some $m^*(\omega)$ and $\bar{\Delta}(\omega) (\equiv m^*/m^*!)$ such that $\forall m \geq m^*(\omega)$

$$\sup_{\substack{|t-s| \in (0, \bar{\Delta}(\omega)) \\ t,s \in [0,T]}} \left| \tilde{\Gamma}_t(\omega) - \tilde{\Gamma}_s(\omega) \right| \Big/ |t-s|^d \leq \sup_{\substack{|t-s| \in (0, \bar{\Delta}(\omega)) \\ t,s \in [0,m]}} \left| \tilde{\Gamma}_t(\omega) - \tilde{\Gamma}_s(\omega) \right| \Big/ |t-s|^d \leq \vartheta,$$

which implies that $\Pr(\Omega_1) = 1$, where

$$\Omega_1 := \left\{ \exists \bar{\Delta}(\omega), \ \exists T^*, \ \forall T (\geq T^*), \quad \sup_{|t-s| \in (0, \bar{\Delta}(\omega)); \ s,t \in [0,T]} \frac{\left| \tilde{\Gamma}_t(\omega) - \tilde{\Gamma}_s(\omega) \right|}{|t-s|^d} \leq \vartheta \right\}. \tag{77}$$

Note that

$$\begin{aligned} \Omega_1 &\subset \left\{ \exists \bar{\Delta}(\omega), \ \exists T^*, \quad \sup_{|t-s| \in (0, \bar{\Delta}(\omega)); \ s,t \in [0,T^*]} \frac{|\tilde{\Gamma}_t(\omega) - \tilde{\Gamma}_s(\omega)|}{|t-s|^d} \leq \vartheta \right\} \\ &= \left\{ \exists \bar{\Delta}(\omega), \ \exists T^*, \ \forall T (\leq T_*), \quad \sup_{|t-s| \in (0, \bar{\Delta}(\omega)); \ s,t \in [0,T]} \frac{|\tilde{\Gamma}_t(\omega) - \tilde{\Gamma}_s(\omega)|}{|t-s|^d} \leq \vartheta \right\} =: \Omega_2. \end{aligned} \tag{78}$$

Since $\Pr(\Omega_1) = 1$, we then have $\Pr(\Omega_2) = 1$. For any events $E, F \in \mathfrak{F}$, we have the inequality: $\Pr(E \cap F) \geq \Pr(E) + \Pr(F) - 1$. With $E = \Omega_1$ and $F = \Omega_2$, we obtain $\Pr(\Omega_1 \cap \Omega_2) = 1$, which, together with

$$\Omega_1 \cap \Omega_2 = \left\{ \omega \in \Omega \middle| \exists \bar{\Delta}(\omega), \ \forall T, \quad \sup_{|t-s| \in (0, \bar{\Delta}(\omega)); \ s,t \in [0,T]} \frac{|\tilde{\Gamma}_t(\omega) - \tilde{\Gamma}_s(\omega)|}{|t-s|^d} \leq \vartheta \right\},$$

implies the desired result, eq. (71). ■

B Tables

	Drift		Diffusion	
	Infeasible 1-step	Feasible 2-step	Infeasible 1-step	Feasible 2-step
	0.6036	5.0377	0.0110	0.2038
$\delta = 1/8$	1.6332	0.6911	0.0147	0.0073
	2.2368	5.7288	0.0257	0.2111
	0.8503	0.6218	0.0258	0.0145
$\delta = 1/4$	1.5776	1.3627	0.0220	0.0201
	2.4279	1.9845	0.0478	0.0347
	1.3417	0.7511	0.0607	0.0304
$\delta = 1/2$	1.3348	1.3689	0.0289	0.0353
	2.6764	2.1200	0.0896	0.0657

Table 1: Performance of infeasible and feasible nonparametric drift and diffusion estimators, $\Delta = 1/(24 \times 60)$. In each cell, integrated squared bias ($\times 10^{-4}$), variance ($\times 10^{-4}$) and MSE ($\times 10^{-4}$) are reported.

	Drift		Diffusion	
	Infeasible 1-step	Feasible 2-step	Infeasible 1-step	Feasible 2-step
	1.2184	3.3333	0.0536	0.6269
$\delta = 1/2$	1.7283	2.2352	0.0393	0.0841
	2.9466	5.5685	0.0929	0.7110
	2.6679	0.2113	0.1250	0.0709
$\delta = 1$	1.0454	1.1474	0.0500	0.0614
	3.7132	1.3587	0.1749	0.1323
	5.7818	4.1220	0.2703	0.1309
$\delta = 2$	0.6394	0.6235	0.0613	0.0628
	6.4211	4.7455	0.3316	0.1937

Table 2: Performance of infeasible and feasible nonparametric drift and diffusion estimators, $\Delta = 1/(24 \times 12)$. In each cell, integrated squared bias ($\times 10^{-4}$), variance ($\times 10^{-4}$) and MSE ($\times 10^{-4}$) are reported.

	α		β		κ^2	
	Infeasible	Feasible	Infeasible	Feasible	Infeasible	Feasible
	0.0021	0.0023	0.3029	24.7633	0.0361	4.4861
$\delta = 1/12$	0.6568	0.6650	19.0296	4.4351	0.0830	0.0815
	0.6588	0.6673	19.3325	29.1984	0.1191	4.5676
	0.0022	0.0026	2.3990	15.5280	0.1579	3.6615
$\delta = 1/6$	0.6576	0.6627	18.4518	12.3283	0.1547	0.1514
	0.6597	0.6652	20.8508	27.8563	0.3126	3.8129
	0.0021	0.0026	6.2912	6.8481	0.3493	0.3609
$\delta = 1/4$	0.6560	0.6623	17.3818	17.1802	0.2377	0.2497
	0.6581	0.6650	23.6731	24.0283	0.5870	0.6106

Table 3: Performance of infeasible and feasible parametric drift and diffusion estimators, $\Delta = 1/(24 \times 60)$. In each cell, squared bias ($\times 10^{-4}$), variance ($\times 10^{-4}$) and MSE ($\times 10^{-4}$) are reported.

	α		β		κ^2	
	Infeasible	Feasible	Infeasible	Feasible	Infeasible	Feasible
	0.0021	0.0024	0.3029	23.9588	0.0361	4.5830
$\delta = 1/12$	0.6568	0.6625	19.0296	7.2365	0.0830	0.1885
	0.6588	0.6650	19.3325	31.1953	0.1191	4.7716
	0.0022	0.0025	2.3990	14.4633	0.1579	21.2663
$\delta = 1/6$	0.6576	0.6614	18.4518	22.1306	0.1547	0.3595
	0.6597	0.6639	20.8508	36.6940	0.3126	21.6258
	0.0021	0.0025	6.2912	24.5878	0.3493	58.6862
$\delta = 1/4$	0.6560	0.6613	17.3818	33.3411	0.2377	0.5172
	0.6581	0.6638	23.6731	57.9289	0.5870	59.2034

Table 4: Performance of infeasible and feasible parametric drift and diffusion estimators, $\Delta = 1/(24 \times 12)$. In each cell, squared bias ($\times 10^{-4}$), variance ($\times 10^{-4}$) and MSE ($\times 10^{-4}$) are reported

C Figures

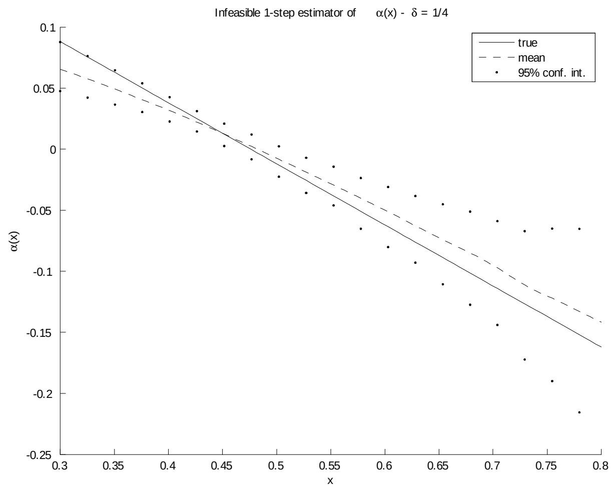

Figure 1: Infeasible 1-step estimator of $\alpha(x)$, $\Delta = 1/(24 \times 60)$ and $\delta = 1/4$.

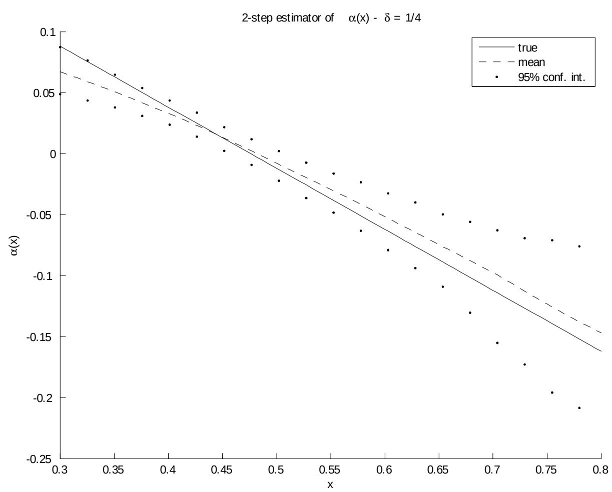

Figure 2: Feasible 2-step estimator of $\alpha(x)$, $\Delta = 1/(24 \times 60)$ and $\delta = 1/4$.

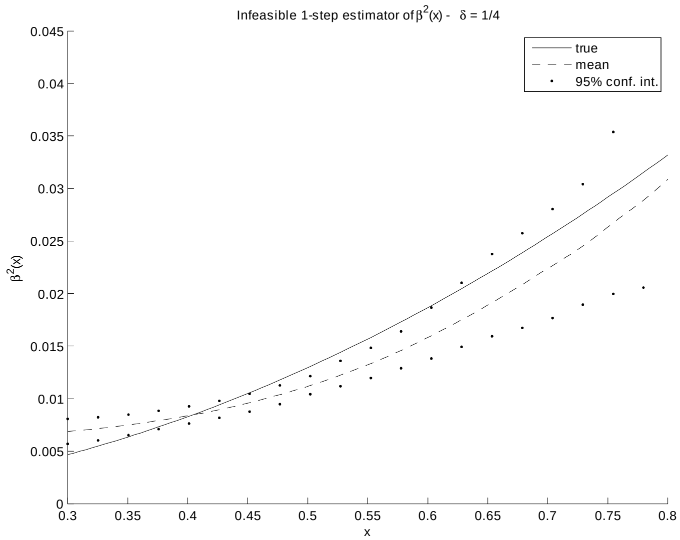

Figure 3: Infeasible 1-step estimator of $\beta^2(x)$, $\Delta = 1/(24 \times 60)$ and $\delta = 1/4$.

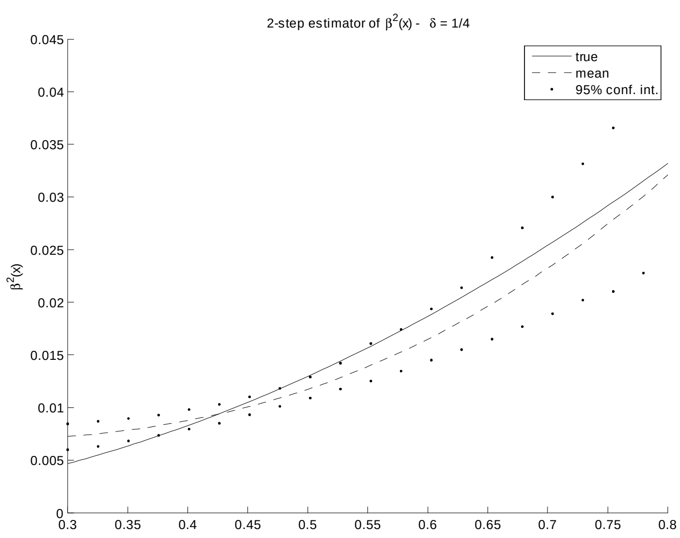

Figure 4: Feasible 2-step estimator of $\beta^2(x)$, $\Delta = 1/(24 \times 60)$ and $\delta = 1/4$.

Research Papers 2010

2010-54: Stefan Holst Bache: Minimax Regression Quantiles

2010-55: Nektarios Aslanidis and Charlotte Christiansen: Sign and Quantiles of the Realized Stock-Bond Correlation

2010-56: Anders Bredahl Kock: Oracle Efficient Variable Selection in Random and Fixed Effects Panel Data Models

2010-57: Charlotte Christiansen, Juanna Schröter Joensen and Jesper Rangvid: The Effects of Marriage and Divorce on Financial Investments: Learning to Love or Hate Risk?

2010-58: Charlotte Christiansen, Maik Schmeling and Andreas Schrimpf: A Comprehensive Look at Financial Volatility Prediction by Economic Variables

2010-59: James G. MacKinnon and Morten Ørregaard Nielsen: Numerical distribution functions of fractional unit root and cointegration tests

2010-60: Bent Jesper Christensen and Paolo Santucci de Magistris: Level Shifts in Volatility and the Implied-Realized Volatility Relation

2010-61: Christian Bach and Bent Jesper Christensen: Latent Integrated Stochastic Volatility, Realized Volatility, and Implied Volatility: A State Space Approach

2010-62: Bent Jesper Christensen and Malene Kallestrup Lamb: The Impact of Health Changes on Labor Supply: Evidence from Merged Data on Individual Objective Medical Diagnosis Codes and Early Retirement Behavior

2010-63: Martin M. Andreasen: How Non-Gaussian Shocks Affect Risk Premia in Non-Linear DSGE Models

2010-64: Tim Bollerslev and Viktor Todorov: Jump Tails, Extreme Dependencies, and the Distribution of Stock Returns

2010-65: Almut E. D. Veraart: How precise is the finite sample approximation of the asymptotic distribution of realised variation measures in the presence of jumps?

2010-66: Ole E. Barndorff-Nielsen, David G. Pollard and Neil Shephard: Integer-valued Lévy processes and low latency financial econometrics

2010-67: Shin Kanaya and Dennis Kristensen: Estimation of Stochastic Volatility Models by Nonparametric Filtering