Estimation of Stochastic Volatility Models by Nonparametric Filtering*

Shin Kanaya†
University of Aarhus and CREATES

Dennis Kristensen‡
UCL, IFS and CREATES

January 2015

*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. They also gratefully acknowledge support from CREATES, Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation (DNRF78). Kristensen gratefully acknowledges financial support from the National Science Foundation (grant no. SES-0961596). Part of this research was conducted while Kanaya was visiting the Institute of Statistical Mathematics and the Institute of Economic Research at Kyoto University (under the Joint Research Program of the KIER), and Kristensen was visiting Princeton University and University of Copenhagen, whose support and hospitality are gratefully acknowledged.

†Department of Economics and Business, University of Aarhus, Fuglesangs Alle 4, Aarhus V, 8210, Denmark. E-mail: skanaya@econ.au.dk.

‡Department of Economics, University College London, Gower Street, London, WC1E 6BT, United Kingdom. E-mail: d.kristensen@ucl.ac.uk.

Proposed running head: Estimation of SV Models

Corresponding author:
Dennis Kristensen

Department of Economics
University College London
Gower Street
London WC1E 6BT
United Kingdom

E-mail: d.kristensen@ucl.ac.uk

Abstract: A two-step estimation method of stochastic volatility models is proposed: In the first step, we nonparametrically estimate the (unobserved) instantaneous volatility process. In the second step, standard estimation methods for fully observed diffusion processes are employed, but with the filtered/estimated volatility process replacing the latent process. Our estimation strategy is applicable to both parametric and nonparametric stochastic volatility models, and can handle both jumps and market microstructure noise. The resulting estimators of the stochastic volatility model will carry additional biases and variances due to the first-step estimation, but under regularity conditions we show that 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; semi-parametric.

JEL codes: C14; C32; C58.

1 Introduction

We propose a general estimation strategy for SV jump-diffusion models that combines a simple, model-free realized volatility estimator with the additional structure imposed by the Markov diffusion model of the volatility process. 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: First, a nonparametric estimator of the spot (or instantaneous) volatility is computed. Second, the spot volatility estimator is plugged into a given existing estimation method for fully observed diffusion models.

The first step takes as input a given spot volatility estimator: A number of spot volatility estimators have been proposed in the literature such as Fan and Wang (2008), Kristensen (2010a), Hoffman, Munk and Schmidt-Hieber (2012), Malliavin and Mancino (2002, 2009), Mancini, Mattiussi and Renò (2012), and Zu and Boswijk (2014), amongst others. We do not restrict ourselves to a specific volatility estimator, and allow for a broad class of spot volatility estimators to be employed in our two-step procedure. In the second step, the volatility model is estimated taking as input the chosen spot volatility estimator. We here consider two leading volatility models with associated estimators: First, we consider a nonparametric Markov model for the volatility with associated kernel estimators as proposed by Bandi and Phillips (2003). As a second example, we analyze (semi-) parametric Markov models with associated least-squares estimators akin to the ones proposed in Prakasa Rao (1988) or Bandi and Phillips (2007). The asymptotic theory that we develop assumes that the volatility process contains no jump component. However, we show how the estimators can be modified to handle jumps in volatility and discuss how our theory can be extended to cover this case.

We show consistency and asymptotic normality for both the nonparametric and parametric two-step estimators of the underlying volatility model. In the nonparametric case, our two-stage estimation problem is similar to the one considered in Sperlich (2009) where kernel regression with generated regressors is considered; see also Newey, Powell and Vella (1999), Xiao, Linton, Carroll and Mammen (2003) and Mammen, Rothe and Schienle (2012). 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. Kristensen (2010b) and Mammen, Rothe and Schienle (2013).

The asymptotic properties of the two-step estimators are established under regularity conditions with a key condition being that the first-step spot volatility estimator is uniformly consistent over a growing time span with a known convergence rate. This is a high-level assumption that needs to be verified for the particular spot volatility estimator being employed. We verify this condition for three particular spot volatility estimators that are consistent under different scenarios as described below. In all three cases, the proof of uniform consistency is technically demanding due to two properties of the object of interest, the realized sample path of the latent volatility process: First, it is not smooth, and second it is potentially 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 uniform rate results over an expanding time interval, including a new result on the global modulus of continuity of stochastic processes.

Four scenarios are considered in the first-step spot volatility estimation: First, the ideal situation where log-prices are observed without market microstructure noise and do not contain jumps. In this case, the kernel-based estimator proposed in Kristensen (2010a) is consistent, and we extend Kristensen's rate results to allow for an expanding time span. Second, noise is introduced and we propose a novel spot volatility estimator based on pre-averaging, similar to Podolskij and Vetter (2009a,b), to handle this case, and derive its uniform rate. Next, we consider the case where jumps, but no noise, are present, and we derive the rate of a kernel-weighted version of the threshold estimator of Mancini (2009). Finally, by combining the estimation strategies from the second and third scenario, we develop a jump and noise-robust spot volatility estimator; the analysis of this estimator proves to be quite complex, and so we do not provide a complete asymptotic theory for this. The estimators in the second and fourth scenarios are both novel, and the uniform rate results of all estimators are new contributions to the literature, and so should be of independent interest.

Our estimators rely on certain nuisance parameters that need to be chosen in the implementation. In particular, bandwidths have 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 estimators through a simulation study with particular emphasis on their 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 parametric Markov SV models, a number of different estimation methods exist. If only low-frequency data is available, the estimation problem is hard due to the volatility process being latent. 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 the latent variable problem (see e.g. Altissimo and Mele, 2009, Andersen and Lund, 1997; Gallant et al.,1997). In the case where high-frequency data is available, a number of studies have proposed to estimate parametric SV models 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), Creel and Kristensen (2014) and Todorov (2009). However, in general, closed form expressions of the moments are not available, and as a result these estimation strategies will in general require the use of simulation-based or other computationally burdensome methods. We also note that the extension of these methods to nonparametric estimation of SV models is not obvious.

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 and do not necessarily provide a complete asymptotic theory. In particular, uniform consistency (and its rate) of the first-step spot volatility estimator over a growing time interval is

not established. Comte et al. (2009) assume that the integrated volatility is observable (if their setting is read in the context of the volatility estimation), while Renò (2006) only provides simulation results. Renò (2008) only establishes consistency of his spot volatility estimator over a fixed time interval, and so can only show results for the estimation of the diffusion coefficient of the volatility model. Furthermore, this consistency result relies on some strong assumptions on the model, including compact support of the volatility process, thereby ruling out all standard models found in the literature. Bandi and Renò (2009) avoid some of these issues by imposing certain high-level assumptions on the volatility process, but these seem difficult to verify in practice. On the other hand, their framework is more general than ours in that they allow for the presence of jumps in the volatility process.

The rest of the paper is organized as follows: In the next section, we outline our proposed estimation method for the nonparametric and fully parametric case. In Section 3, uniform rates of three different spot volatility estimators are derived under regularity conditions. These rate results are then employed in Sections 4 and 5 to establish the asymptotic properties of the estimators of SV model in a nonparametric and parametric setting, respectively. The practical implementation of the estimator is discussed in Section 7. The results of a simulation study investigating finite-sample properties of our estimators are presented in Section 8. Section 9 concludes. Proofs of theorems and lemmas have been collected in Appendices A and B, respectively, while tables and figures can be found in Appendices C and D, respectively. Some details of proofs are provided online at Cambridge Journals Online (journals.cambridge.org/ect) in supplementary material to this article.

We use the following notations throughout: The symbols $\xrightarrow{P}$ and $\xrightarrow{d}$ 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 $B = [b_{i,j}]$, $\|B\|$ denotes $\sum_{i,j} |b_{i,j}|$. 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 semimartingale that is a càdlàg solution to

$$\begin{cases} dX_t = \mu_t dt + \sigma_t dW_t + dJ_t \\ d\sigma_t^2 = \alpha(\sigma_t^2)dt + \beta(\sigma_t^2)dZ_t \end{cases}, \tag{2.1}$$

where $\{W_t\}$ and $\{Z_t\}$ are two (possibly correlated) standard Brownian motions (BM's), while $\{\mu_t\}$ and $\{\sigma_t\}$ are adapted, càdlàg stochastic processes. The process $\left\{\sigma_t^2\right\}$ is usually referred to as the (spot) volatility process of $\{X_t\}$, while $\{\mu_t\}$ is the drift process. The process $\{J_t\}$ is a pure-jump càdlàg process with finite jump activities (i.e., the number of jump occurrences in any finite time interval is finite). Given the finite-jump-activity assumption, we can write $J_t = \sum_{j=1}^{N_t} \kappa_j$, where N_t is the jump arrival process and κ_j, $j = 1, 2, ...$, are the jump sizes. The second part of the model in eq. (2.1), stating the dynamics of the volatility process, is referred to as a stochastic volatility (SV)

model, and restricts the volatility process to be a Markov diffusion process. We discuss in Section 6 how our results can be extended to the case where the volatility process is a jump-diffusion.

We consider two different sampling scenarios: Either we have directly observed X_t at discrete time points $t_1, ..., t_n$, or only noise-contaminated observations of the process are available due to, for example, market microstructure. In the latter case, we only have observed $Y_1, ..., Y_n$ where $Y_i = X_{t_i} + \varepsilon_i$ $(i = 1, ..., n)$ and $\{\varepsilon_i\}$ are the measurement errors. For notational simplicity, we will throughout assume that equidistant observations are available so that the time distance between observations is constant, $\Delta = t_i - t_{i-1}$; all the subsequent results still hold with non-equidistant observations with Δ now being the maximum time distance in the sample.

Given observations of X_t (or the noise-contaminated version of it) at discrete time points, we wish to draw inference about the drift and diffusion terms of the underlying SV model, $\alpha(\cdot)$ and $\beta^2(\cdot)$. Since we have not observed the process $\left\{\sigma_t^2\right\}$, the estimation of these two terms involves a latent stochastic process which we need to learn about from data. To motivate our estimators, consider for the moment the counter-factual situation where $\left\{\sigma_t^2\right\}$ has been observed at discrete time points. In this case, fully nonparametric kernel estimators of $\alpha(\cdot)$ and $\beta^2(\cdot)$ have been developed in Bandi and Phillips (2003), Florens-Zmirou (1993) and Jiang and Knight (1997) amongst others. If parametric forms for drift and/or volatility are specified, a number of estimators offer themselves; see, for example, Florens-Zmirou (1989), Jacod (2006), Sørensen (2009) and Yoshida (1992).

Now, let us return to the actual situation where the volatility is unobserved, in which case all of the above estimators of $\alpha(\cdot)$ and $\beta^2(\cdot)$ are infeasible. Instead, we here suggest a two-step procedure, where in the first step an estimator of the spot volatility is obtained from data which we denote $\tilde{\sigma}_\tau^2$, $\tau \geq 0$. This could, for example, be any of the estimators proposed in the literature which we cited in the Introduction. We can compute $\tilde{\sigma}_\tau^2$ at any given value of τ; in particular, we can evaluate it at a given set of discrete time points τ_j, $j = 1, ..., N$, chosen by us. These time points are under our control and may potentially differ from the actual time points at which X_t (or Y_t) has been observed. We therefore refer to $\{\tau_j : j = 1, 2, ..., N\}$ and $\delta := \tau_{j+1} - \tau_j$ as pseudo-sampling times and time distance, respectively. When deriving the asymptotics of our estimators, we will impose certain restrictions on these.

In the second step, we simply replace the spot volatilities in any of the above estimation methods with the estimates obtained in the first-step. We will here focus on two particular estimation methods. For nonparametric estimation, we employ the kernel estimators of Bandi and Phillips (2003) and obtain the following feasible estimates:

$$\hat{\alpha}(x) = \frac{\sum_{j=1}^{N-1} \mathcal{K}_b(\tilde{\sigma}_{\tau_j}^2 - x)[\tilde{\sigma}_{\tau_{j+1}}^2 - \hat{\sigma}_{\tau_j}^2]}{\delta \sum_{j=1}^{N} \mathcal{K}_b(\tilde{\sigma}_{\tau_j}^2 - x)}; \tag{2.2}$$

$$\hat{\beta}^2(x) = \frac{\sum_{j=1}^{N-1} \mathcal{K}_b(\hat{\sigma}_{\tau_j}^2 - x)[\tilde{\sigma}_{\tau_{j+1}}^2 - \tilde{\sigma}_{\tau_j}^2]^2}{\delta \sum_{j=1}^{N} \mathcal{K}_b(\tilde{\sigma}_{\tau_j}^2 - x)}, \tag{2.3}$$

where $\mathcal{K}_b(x) = \mathcal{K}(x/b)/b$ for some kernel function $\mathcal{K} : \mathbb{R} \to \mathbb{R}$ and some bandwidth $b > 0$. Similarly, for parametric estimators, we simply replace $\sigma_{\tau_j}^2$ by $\tilde{\sigma}_{\tau_j}^2$ in the objective function defining the estimators. 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 some 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):

$$\hat{\theta}_k = \underset{\theta_k \in \Theta_k}{\arg\min} \hat{Q}_k(\theta_k) \quad \text{for} \quad k = 1, 2, \tag{2.4}$$

where

$$\hat{Q}_1(\theta_1) = \sum_{j=1}^{N-1} \left[(\tilde{\sigma}^2_{\tau_{j+1}} - \tilde{\sigma}^2_{\tau_j}) - \alpha(\tilde{\sigma}^2_{\tau_j};\theta_1)\delta \right]^2; \tag{2.5}$$

$$\hat{Q}_2(\theta_2) = \sum_{j=1}^{N-1} \left[(\tilde{\sigma}^2_{\tau_{j+1}} - \tilde{\sigma}^2_{\tau_j})^2 - \beta^2(\tilde{\sigma}^2_{\tau_j};\theta_2)\delta \right]^2. \tag{2.6}$$

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 Spot Volatility Estimation

In the asymptotic analysis of the proposed two-step estimators that is presented in the next section, we need to control the first-step estimation error in $\tilde{\sigma}^2_\tau$. More specifically, we will impose the high-level condition that the chosen spot volatility estimator satisfies $\max_{1 \leq j \leq N} |\tilde{\sigma}^2_{\tau_j} - \sigma^2_{\tau_j}| = O_P(\vartheta_N)$ for some rate parameter $\vartheta_N \to 0$. In this section, we derive such rates for kernel-based spot volatility estimators that takes as starting point the basic estimator proposed in Kristensen (2010a).

The arguments that we employ to establish such rate results are somewhat non-standard since, in general, the target "function" in our case, $\tau \mapsto \sigma^2_\tau$, will be unbounded as $T \to \infty$. This is in contrast to the existing literature on uniform rate results of nonparametric estimators where it is routinely assumed that the function of interest is bounded. Our uniform convergence results may be useful in other applications, and so we do not restrict the volatility process to be a Markov diffusion (as imposed in eq. (2.1)) in this section. Instead, we only require that the drift and volatility processes, μ_t and σ^2_t, 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. The smoothness condition rules out jumps in volatility; we discuss in Section 6, how the spot volatility estimators can be modified to handle this situation.

The specific estimator employed to learn about σ^2_t in the first step depends on whether data is noise-contaminated and/or contains jumps. We consider four different scenarios in the subsequent four subsections: (i) Data contains no market microstructure noise and no jumps; (ii) data is contaminated by noise, but not jumps; (iii) data is contaminated by jumps, but not noise; and finally (iv) data is contaminated by both noise and jumps. In each case, we develop an estimator and analyze its properties.

3.1 Noise and Jump-free Case ($\varepsilon = J = 0$)

In the case of no noise and no jumps ($\varepsilon_t = J_t = 0$), X_t is directly observed and contains no jumps, and so the kernel estimator of Kristensen (2010a) can be employed:

$$\hat{\sigma}^2_\tau = \sum\nolimits_{i=1}^{n} K_h\left(t_{i-1} - \tau\right)\left[X_{t_i} - X_{t_{i-1}}\right]^2, \tag{3.1}$$

where $K_h(z) = K(z/h)/h$, K is another kernel, and $h > 0$ is another bandwidth; see also Fan and Wang (2008). That is, in the two-step estimation procedure outlined in the previous section, we set $\tilde{\sigma}^2_\tau = \hat{\sigma}^2_\tau$ where $\hat{\sigma}^2_\tau$ is given above. To analyze the asymptotic properties of $\hat{\sigma}^2_\tau$, we impose the following conditions on K:

$\mathbb{K}$.1 The kernel function $K : \mathbb{R} \to \mathbb{R}$ satisfies $\int_{-\infty}^{\infty} K(x)\,dx = 1$ and $\int_{-\infty}^{\infty} |x|^m |K(x)|\,dx < \infty$ for $m(> 1)$; and there exist some constants $\bar{K}, C < (0, \infty)$ such that $\sup_{x\in\mathbb{R}} |K(x)| \leq \bar{K}$, $\sup_{x,y\in\mathbb{R}} |K(x) - K(y)| \leq \bar{K}|x - y|$, and $|K(x)|$ is not decreasing on $(-\infty, -C]$ and not increasing on $[C, \infty)$.

Many standard kernels satisfy these conditions, including the Gaussian one. The monotone tail condition imposed in $\mathbb{K}$.1 may be unfamiliar but is actually satisfied by many kernels (e.g., the Gaussian kernel and any kernel with compact support). This is useful in order to obtain sharp convergence rates. We allow for one-sided kernels as discussed in Kristensen (2010a). The continuity and differentiability conditions imposed on K simplify various parts of our subsequent proofs (see, e.g., derivation of (B.7)), but excludes, for example, the uniform kernel.

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

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

A.2 (i) There exist constants $q > 0$ and $l_2 \geq 0$: $\sup_{t\leq T} E[|\sigma_t|^{2+q}] = O\left(T^{l_2}\right)$ as $T \to \infty$. (ii) There exist constants $\lambda > 0$, $\rho > 0$ and $C > 0$ such that $E[\left|\sigma^2_t - \sigma^2_s\right|^\lambda] \leq C|t - s|^{1+\rho}$.

The uniform moment conditions imposed in Assumptions A.1 and A.2(i) are used to extend the uniform convergence results over the interval $[0, T]$ of Kristensen (2010a) from the case where $T = \bar{T} < \infty$ is fixed to the case where $T \to \infty$. If we only wanted to show convergence for fixed $T < \infty$, these moment conditions 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. Kristensen (2010a, Theorem 5).

If the drift is zero, $\mu_t = 0$ for all t, we can set $l_1 = -\infty$ in Assumption A.1. If $\{\mu_t\}$ is stationary, we can choose $l_1 = 0$ in Assumption A.1. The condition is however also satisfied for non-stationary processes; an instructive example of this is a standard BM, say $\{B_t\}$: if $\mu_t = B_t$, we can choose $(p, l_1) = (2, 2)$ (or $(p, l_1) = (\bar{p}, 1 + \bar{p}/2)$ for any constant $\bar{p} > 0$). Similarly, A.2(i) holds in great generality: If $\left\{\sigma^2_t\right\}$ is recurrent, Assumption A.2(i) can be easily satisfied as long as the relevant moments exist (e.g., $l_2 = 0$ for stationary cases). The recurrent case includes most parametric diffusion models found in the literature, including Ornstein-Uhlenbeck (OU) and

CIR/Feller's square-root models, which have finite moments of any order. Even null recurrent processes are included; for example, if $\{\sigma_t\}$ is a diffusion process whose drift function has compact support and diffusion term is (uniformly) bounded (see e.g. Has'minskiĭ, 1980, Chapter IV), then it is null recurrent and A.2(i) holds with $l_2 \leq 1 + q/2$ for any $q > 0$.

Assumption A.2(ii) is a smoothness condition of $\left\{\sigma_t^2\right\}$ 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 given by

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

where we recall that $\Delta > 0$ denotes the fixed time distance between observations. The properties of $\omega_{[0,T]}(\Delta)$ when $\{\sigma_t^2\}$ is a diffusion process are well-known for $T = \bar{T} < \infty$ fixed, 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 $T \to \infty$. We therefore establish a new result showing that the standard rate for the modulus of continuity can be extended to hold over an infinite time interval $[0, \infty)$; see Lemma A.1. In particular, we show that $\omega_{[0,\infty)}(\Delta) = O_{a.s.}(\Delta^\gamma)$ for any $\gamma \in [0, \rho/\lambda)$ as $\Delta \to 0$. This result is often needed when one considers nonparametric estimators for continuous-time processes under long span asymptotics, and should be of independent interest; see Kanaya (2014) for related results. Assumption A.2(ii) is automatically satisfied with $\rho = \lambda/2 - 1$ if $\left\{\sigma_t^2\right\}$ 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.

We restrict the set of feasible bandwidth sequences that can be used to estimate the trajectory of $\left\{\sigma_t^2\right\}$:

B.1 The bandwidth $h \to 0$ is chosen such that, as $T/\Delta(= n) \to \infty$ and $\Delta \to 0$:

$$\Delta^{p/2}[\log(1/\Delta)]T^{2+l_1}/h = O(1); \tag{3.2}$$

$$T^{(1+l_1)(2+q)-(3+l_2)(2+p)} h^{p(2+q)/2} \left[h\Delta^2 \log(1/\Delta)\right]^{2+p} = O(1); \tag{3.3}$$

$$h^{m-2\gamma} T^{2+4l_2/(2+q)} = O(1), \tag{3.4}$$

where (p, l_1), (q, l_2) and m were introduced in Assumption A.1, A.2 and $\mathbb{K}$.1, respectively.

Eqs. (3.2) and (3.3) are used to control the bias of $\hat{\sigma}_t^2$ due to the presence of the drift term $\{\mu_t\}$. They imply that the bias incurred from this term has negligible impact in the estimation uniformly as $T \to \infty$; and the existence of higher order moments of μ_t (i.e., a larger value of p) allows for a more flexible choice of h or a less frequent sampling. They can be thought of as 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 often need to require Δ to shrink faster than in the parametric case. The condition in eq. (3.3) involves q and l_2 which is due to interactions between the two

components μ_t and σ_t (see the decomposition into five terms in the proof of Theorem 3.1). If μ_t has sufficiently high moments ($p \geq (1+l_1)(2+q)/(3+l_2)$), this condition is always satisfied. Note also that if the drift is not present then eqs. (3.2) and (3.3) are automatically satisfied.

The last condition (3.4) in (B.1) is used to control smoothing biases near the boundaries $t = 0$ and T, where m regulates the tail behavior of the kernel K. This is far from restrictive; for example, Gaussian and compactly-supported kernels satisfy $\int_{-\infty}^{\infty}|x|^m|K(x)|dx < \infty$ for any $m \geq 0$.

We are now able to establish a convergence rate of the spot volatility estimator $\hat{\sigma}_\tau^2$:

Theorem 3.1 *Suppose that Assumptions A.1-A.2, B.1 and $\mathbb{K}$.1 hold. Then, for any $\gamma \in (0, \rho/\lambda)$:*

$$\sup_{\tau\in[\sqrt{h},T-\sqrt{h}]} |\hat{\sigma}_\tau^2 - \sigma_\tau^2| = O_P(\bar{\vartheta}_{T,\Delta}) \tag{3.5}$$

as $T/\Delta\,(=n) \to \infty$ with $\Delta \to 0$, where

$$\bar{\vartheta}_{T,\Delta} := h^\gamma + \sqrt{h^{-1}\Delta\log(1/\Delta)} \times T^{2l_2/(2+q)} \left[T^2/h\Delta\log(1/\Delta)\right]^{2/(2+q)}.$$

The first term of $\bar{\vartheta}_{T,\Delta}$, h^γ, is the rate of the kernel smoothing bias which depends on the degree of the continuity of $\{\sigma_t^2\}$; it coincides with the bias rate in Kristensen (2010a, Theorem 3) for fixed $T = \bar{T} < \infty$. In the standard kernel estimation case, such biases may be remedied by using higher-order kernels. However, $\{\sigma_t^2\}$ is modelled as a general stochastic process here, which in general does not have differentiable sample paths, and so higher order kernels would not reduce this bias.

The second term of $\bar{\vartheta}_{T,\Delta}$ is the rate of the variance component. The first part, $\sqrt{h^{-1}\Delta\log(1/\Delta)}$, is the usual term found in many other studies deriving uniform rates of kernel regression estimators (see, e.g., Kristensen, 2010a, Theorem 3), while the second part, $T^{2l_2/(2+q)}\left[T^2/h\Delta\log(1/\Delta)\right]^{2/(2+q)}$, is non-standard. The second part owes to the fact that we here employ a Bernstein-type exponential inequality for bounded martingales. Since the martingale component of X_t, $\int_0^t \sigma_s dW_s$, is unbounded, we truncate the process. Unfortunately, the martingale property is not readily preserved under truncation and so the precise argument is quite involved and leads to the additional, non-standard term. Exponential inequalities combined with truncation are a standard tool for deriving uniform rates; see, e.g., Hansen (2008), Kristensen (2009), Gao, Kanaya, Li and Tjøstheim (2014) and Kanaya (2014). However, these papers assume a mixing (or an i.i.d.) condition which makes the arguments simpler since truncated mixing processes remain mixing. A more closely related paper is Wang and Chan (2014) who derive uniform convergence results for kernel regression estimators with martingale difference errors. If additional mixing and moment conditions were imposed on $\{\sigma_t^2\}$ or $\{X_t\}$, the rate of the variance component can be shown to be $\sqrt{h^{-1}\Delta\log(1/\Delta)}$. However, in order to allow for nonstationary and strongly dependent volatility processes, we do not impose these. When l_2 is small and q is large (implying stronger moment conditions on σ_t^2), the rate is close to $\sqrt{h^{-1}\Delta\log(1/\Delta)}$. Similarly, if $\{\sigma_t\}$ is uniformly bounded over $[0,T]$, which can be understood as $q = \infty$, the second term again reduces to $\sqrt{h^{-1}\Delta\log(1/\Delta)}$. We note that this rate can be also obtained when $\sup_{t\in[0,T]}\sigma_t^2$ is of stochastically bounded (i.e., $O_{a.s.}(1)$), which holds when the time span is fixed.

The result is uninformative about how to choose h in finite samples for good performance of the estimator since the uniform rate depends on l_2 and q, which in general are unknown. However, this is not special to our setting. For example, the rate derived in Hansen (2008, Theorem 5) depends on the mixing rate and the number of moments, while the one of Wang and Chan (2014) involves certain tail properties of the regressor.

3.2 Noise Contaminated Case ($\varepsilon \neq 0$)

We here consider the case where data are contaminated by market microstructure noise and therefore the estimator $\hat{\sigma}^2_\tau$ may no longer consistent. We slightly change notation and assume that we have observed $M \geq 1$ observations given by

$$Y_i = X_{s_i} + \varepsilon_i, \quad i = 1, ..., M, \tag{3.6}$$

where $\{\varepsilon_i\}$ are measurement errors, $s_i = iT/M$ are the sample time points, and $\Delta_s := s_i - s_{i-1} = T/M$ is the time distance between observations. Note the differences in notation relative to the no-noise case where we had observed n observations at time points $t_0, t_1, ..., t_n$. The reason for this change in notation is that it allows for a simpler comparison of the asymptotic properties of the noise-free and noise-robust estimators. As before, we assume observations are equidistant in time; this is imposed only for notational simplicity and can be relaxed.

A number of different approaches have been developed in the estimation of integrated volatility to handle noise contamination. We can in principle localize any of these methods to obtain a noise-robust spot volatility estimator, and we here choose to focus on a localized version of the pre-averaging procedures developed in, amongst others, Jacod, Li, Mykland, Podolskij and Vetter (2009) and Podolskij and Vetter (2009a,b): First, pre-whiten (pre-average) data using a kernel filter:

$$\hat{X}_t = \frac{T}{M} \sum\nolimits_{i=1}^{M} L_a\left(s_i - t\right) Y_i, \tag{3.7}$$

where $L_a\left(z\right) = L\left(z/a\right)/a$, L is a kernel function, and $a > 0$ is another bandwidth. Second, replace the unobserved process X_t by $\hat{X}_t$ in eq. (3.1) yielding the following noise-robust (NR) volatility estimator:

$$\hat{\sigma}^2_{\mathrm{NR},\tau} = \sum\nolimits_{i=1}^{n} K_h\left(t_{i-1} - \tau\right) [\hat{X}_{t_i} - \hat{X}_{t_{i-1}}]^2. \tag{3.8}$$

Note that in this setting, $s_1, ..., s_M$ are the actual observation times, while now both $t_0, t_1..., t_n$ and $\tau_0, \tau_1..., \tau_N$ are pseudo-sampling time points chosen by the econometrician. Through this notation we can conveniently decompose the over-all estimation error of $\hat{\sigma}^2_{\mathrm{NR},\tau}$ as $\hat{\sigma}^2_{\mathrm{NR},\tau} - \sigma^2_\tau = [\hat{\sigma}^2_{\mathrm{NR},\tau} - \hat{\sigma}^2_\tau] + [\hat{\sigma}^2_\tau - \sigma^2_\tau]$, where $\hat{\sigma}^2_\tau$ is the infeasible estimator given in eq. (3.1) assuming that we had directly observed X_t at the pseudo-sampling points $t_1, ..., t_n$.

As an alternative to the estimator in eq. (3.8), one could develop localized versions of the two-scale realized variance estimator (Zhang, Mykland and Aït-Sahalia, 2005), or the realised-kernel estimator (Barndorff-Nielsen, Hansen, Lunde and Shephard, 2008). For example, Zu and Boswijk (2014) analyze a localized version of the two-scale realized variance estimator. We note that up to some approximation (of first order), the two-scale estimator can be re-written as the realised-kernel

one with the Bartlett-type kernel, and the realised-kernel estimator can be seen as a member of the class of pre-averaging estimators.[1] Accordingly, localized versions of the two-scale and realized-kernel estimators can be also re-written, up to some approximation, as our localized pre-averaging estimator in eq. (3.8).

We impose the following conditions on L and the measurement errors:

$\mathbb{K}$.2 $L : \mathbb{R} \to \mathbb{R}$ satisfies $\int_{-\infty}^{\infty} L(x)\,dx = 1$, has compact support, and is continuously differentiable.

A.3 $\{\varepsilon_i\}$ are mutually independent and independent of $\{X_t\}$ with $E[\varepsilon_i] = 0$ and $\sup_{i\geq 1} E[|\varepsilon_i|^{1+d_s}] < \infty$ for some constant $d_s > 0$.

The compact-support condition on L in $\mathbb{K}$.2 excludes some kernels, such as the Gaussian one, but simplifies some of the theoretical arguments (see, e.g., derivations of (B.1) and (B.2) in the proof of Theorem 3.2). It should be possible to replace this assumption by some tail decay conditions on L, but this will complicate the proofs, and we maintain this condition. The assumption of no autocorrelation in the errors can be relaxed to allow for $\{\varepsilon_i\}$ to be weakly dependent (such as α-mixing). By controlling the degree of dependence appropriately, the following rate results should carry over to the weakly dependent case. However, we rule out autocorrelation here to avoid too lengthy proofs. The existence of the higher-order moments of ε_i are used when applying exponential inequalities; see our previous discussion on the exponential inequality and truncation. Finally, we impose the following conditions on Δ_s and a:

B.2 The bandwidth $a \to 0$ is chosen such that as $\Delta_s \to 0$:

$$a\Delta_s^{1-4/d_s}\left[\log(1/\Delta_s)\right]T^{4/d_s} = O(1); \tag{3.9}$$

$$\sqrt{a^{-1}\Delta_s \log(1/\Delta_s)}\left[T^{1+l_1/(2+p)} + T^{(1/2)+l_2/(2+q)}\right] = O(1); \tag{3.10}$$

$$a^p T^{2(1+l_1)} = O(1); \tag{3.11}$$

$$a^q T^{2(1+l_2)}[\log(1/\Delta_s)]^{-(2+q)} = O(1), \tag{3.12}$$

where (p, l_1), (q, l_2) and d_s were defined in Assumptions A.2-A.3.

Eqs. (3.10)-(3.12) are slightly stronger than necessary but allow us to obtain a relatively simple expression for the convergence rate of the noise-robust estimator. Eqs. (3.10) and (3.11) are used to control the effect of the drift term $\{\mu_t\}$. Eq. (3.10) may be regarded as a strengthening of the rapidly increasing experimental design as discussed earlier. If $T = \bar{T} < \infty$, eqs. (3.11) and (3.12) are trivially satisfied. If $\{\varepsilon_i\}$, $\{\mu_t\}$ or $\{\sigma_t\}$ is uniformly bounded, we can set corresponding parameters (d_s, p or q) as $+\infty$, in which case eqs. (3.11) and (3.12) are trivially satisfied.

Given these conditions, we are able to derive the following rate result for the noise-robust estimator:

[1] See Theorem 6 of Barndorff-Nielsen, Hansen, Lunde and Shephard (2004), and discussions in page 2251 and Remark 1 of Jacod et al. (2009).

Theorem 3.2 *Suppose that Assumptions A.1-A.3, B.1-B.2 and* $\mathbb{K}$*.1-*$\mathbb{K}$*.2 are satisfied. Then,*

$$\sup_{\tau\in[\sqrt{h},T-\sqrt{h}]}\left|\hat{\sigma}^2_{\mathrm{NR},\tau}-\hat{\sigma}^2_{\tau}\right| = \Delta^{-1/2}(T^{(1+l_2)}/a)^{1/(2+q)}\left[\sqrt{a}(T^{1+l_2}/a)^{1/(2+q)}+\sqrt{a^{-1}\Delta_s\log(1/\Delta_s)}\right].$$

as $\Delta,\Delta_s\to 0$*, and* $T/\Delta, T/\Delta_s\to\infty$*, where* $\hat{\sigma}^2_{\tau}$ *is given in eq. (3.1). In particular,*

$$\sup_{\tau\in[\sqrt{h},T-\sqrt{h}]}\left|\hat{\sigma}^2_{\mathrm{NR},\tau}-\sigma^2_{\tau}\right| = O_P(\vartheta^{\mathrm{NR}}_{T,\Delta,\Delta_s}), \tag{3.13}$$

where

$$\vartheta^{\mathrm{NR}}_{T,\Delta,\Delta_s} := \bar{\vartheta}_{T,\Delta}+\Delta^{-1/2}(T^{(1+l_2)}/a)^{1/(2+q)}\left[\sqrt{a}(T^{1+l_2}/a)^{1/(2+q)}+\sqrt{a^{-1}\Delta_s\log(1/\Delta_s)}\right].$$

The rate $\vartheta^{\mathrm{NR}}_{T,\Delta,\Delta_s}$ consists of two parts: The first is the same as for the noise-free estimator, $\bar{\vartheta}_{T,\Delta}$, when we observe X_t at sampling frequency Δ, while the second component is due to the first-step filter $\hat{X}_t$ which generates additional errors. More specifically, we show in Lemma A.2 that $\sup_{\tau\in[\sqrt{h},T-\sqrt{h}]}|\hat{X}_t-X_t| = O_P\left(\sqrt{a}(T^{1+l_2}/a)^{1/(2+q)}+\sqrt{a^{-1}\Delta_s\log(1/\Delta_s)}\right)$. Not surprisingly, this rate result is similar to the one for $\hat{\sigma}^2_{\tau}$ stated in Theorem 3.1 and the discussion following this theorem carries over to $\hat{X}_t$ and its stated rate. The expression of $\vartheta^{\mathrm{NR}}_{T,\Delta,\Delta_s}$ suggests that, for a given Δ (as chosen by the econometrician) and T, we should choose $h = h(\Delta,T)$ to minimize $\bar{\vartheta}_{T,\Delta}$ as discussed previously, while $a = a(\Delta_s,T)$ and $\Delta = \Delta(\Delta_s,T)$ should be chosen to minimize the second component of $\vartheta^{\mathrm{NR}}_{T,\Delta,\Delta_s}$; precise guidelines for how to choose a and Δ seem difficult to derive though.

Remarks similar to those made for Theorem 3.1 apply here: For example, if $\{\sigma^2_t\}$ is uniformly bounded over $t\in[0,\infty)$ or $T=\bar{T}<\infty$, we may set $q=\infty$ and convergence rates in the theorem are simplified, e.g., the second term of $\vartheta^{\mathrm{NR}}_{T,\Delta,\Delta_s}$ simplifies to $\Delta^{-1/2}[\sqrt{a}+\sqrt{a^{-1}\Delta_s\log(1/\Delta_s)}]$. In particular, when $T=\bar{T}$ fixed and σ_t is a diffusion process, $\sup_{\tau\in[\sqrt{h},\bar{T}-\sqrt{h}]}\left|\hat{\sigma}^2_{\mathrm{NR},\tau}-\sigma^2_{\tau}\right| = O_P(\Delta_s^{1/12})$ by choosing $a=O(\sqrt{\Delta_s})$, $h=O(\sqrt{\Delta})$ and $\Delta=O(\Delta_s^{1/3})$. This is identical to the pointwise rate derived in Zu and Boswijk (2014) for their alternative noise-robust spot volatility estimator.

3.3 Jump Case ($J\neq 0$)

We here consider the case where jumps on the form $J_t=\sum_{j=1}^{N_t}\kappa_j$ are present, but X_t is not contaminated by noise. Given the same sampling scheme and notation as in the no-noise case, we propose the following jump-robust (JR) estimator of σ^2_{τ}:

$$\hat{\sigma}^2_{\mathrm{JR},\tau} := \sum\nolimits_{i=1}^{n} K_h(t_{i-1}-\tau)\,[X_{t_i}-X_{t_{i-1}}]^2\mathbf{1}\{[X_{t_i}-X_{t_{i-1}}]^2\le r(\Delta,T)\}, \tag{3.14}$$

where $r(\Delta,T)$ is a thresholding parameter chosen by the econometrician. This is a kernel-smoothed version of Mancini's (2009) threshold estimator of the integrated volatility; see also Mancini, Mattiussi, and Renò (2012). Through a suitable choice of the thresholding parameter, Mancini (2009) shows that the effect of jumps can be eliminated by the thresholding device so that the integrated volatility (over a finite interval) can be consistently estimated. The same idea applies here.

To derive the uniform convergence result of $\hat{\sigma}^2_{\mathrm{JR},\tau}$, we make the following assumptions regarding the jump component, which closely follows Mancini (2009). We here let N_{t-} denote the left limit

of the (realized) path of the counting process at t, so that $N_t - N_{t-} = 1$ means a jump occurred at time t.

A.4 $\{N_t\}$ is a Poisson process with bounded intensity, $\lambda_t \leq \bar{\lambda}$, which is independent of σ_t, μ_t and W_t. Furthermore, (i) $\Pr[N_t - N_{t-} = 1 \ \& \ \kappa_{N_t} = 0] = 0$ for any $t \geq 0$; (ii) there exist random variables C_μ and C_{σ^2}, and a deterministic function ξ_T so that

$$\limsup_{\Delta \to 0} \frac{\sup_{1 \leq i \leq n} |\int_{t_{i-1}}^{t_i} \mu_s ds|}{\sqrt{\Delta \log(1/\Delta)}} \leq C_\mu \xi_T; \quad \text{and} \quad \limsup_{\Delta \to 0} \frac{\sup_{1 \leq i \leq n} |\int_{t_{i-1}}^{t_i} \sigma_s^2 ds|}{\Delta} \leq C_{\sigma^2} \xi_T^2,$$

almost surely, as $\Delta \to 0$ and $T/\Delta \to \infty$; (iii) the tresholding parameter $r(T, \Delta)$ satisfies $r(T, \Delta) \to 0$ and $[\Delta \log(1/\Delta)] \times \xi_T / r(T, \Delta) \to 0$, as $\Delta \to 0$ and $T/\Delta \to \infty$.

Assumption A.4 includes the case of a compound Poisson process with bounded jump intensity, and the jump sizes $\{\kappa_j\}$ being an i.i.d. sequence independent of $\{N_t\}$, but more general jump behavior is allowed for: $\{\kappa_j\}$ may not necessarily be i.i.d., nor independent of $\{N_t\}$. A.4 is very similar to the assumptions used to establish Theorem 1 of Mancini (2009). What is distinct is the introduction of the sequence $\{\xi_T\}$, which enables us to control the behavior of $\{\mu_t\}$ and $\{\sigma_t^2\}$ when T tends to ∞. Such a sequence is not required in Mancini's (2009) setup, where only the fixed span case is considered. If T is fixed, we can set $\xi_T = 1$, $C_\mu = \sup_{t \in [0,T]} |\mu_t|$ and $C_{\sigma^2} = \sup_{t \in [0,T]} |\sigma_t^2|$, both of which are almost surely bounded by the càdlàg condition, in which case we can set $r(T, \Delta) = \Delta^{\bar{\alpha}}$ for any $\bar{\alpha} \in (0, 1)$, as discussed in Mancini (2009, page 273).

For the case with $T \to \infty$, knowledge of ξ_T is needed in order to choose $r(T, \Delta)$ to satisfy A.3(iii). This is similar to the issue of choosing the bandwidth(s) employed in the estimation. For example, as discussed after Theorem 3.1, h should be chosen relative to the behavior of certain higher-order moments of μ_t and σ_t^2 which are unknown. As pointed out there, this issue is not special to our setting and is also found when kernel smoothers are employed in other settings where data are dependent.

If $|\mu_t|$ and σ_t^2 are uniformly bounded by some C_0, we can choose $\xi_T = 1$, $C_\mu = C_{\sigma^2} = C_0$. More generally, if we know the growth rates of the extremal/maximal processes $\sup_{t \in [0,T]} |\mu_t|$ and $\sup_{t \in [0,T]} |\sigma_t^2|$ (as $T \to \infty$), the triplet is easily chosen. However, this seems to be a difficult task in general. The behavior of extremal processes of diffusion processes has been investigated in the literature (e.g., Borkovec and Klüpperlberg, 1998), but existing results provide O_P rates (not a.s. rates) of extremal processes, and so do not seem to be directly applicable. However, it is often possible to verify the condition for particular models. As an instructive example, we can show that $\sup_{t \in [0,T]} |B_t| = o_{a.s.}(\sqrt{T} \log T)$ as $T \to \infty$, where $\{B_t\}$ is a BM; see the online supplemental material for a proof of this. Therefore, if $\{\mu_t\}$ and $\{\sigma_t^2\}$ are transformation of BM's, for example, $\mu_t = B_t$ and $\sigma_t^2 = c_0 + \tilde{B}_t^2$ with $c_0 > 0$ and $\{\tilde{B}_t\}$ being another BM, then the condition holds with $\xi_T = \sqrt{T} \log T$ and $C_\mu = C_{\sigma^2} = c_0 + 1$, and so we can choose $r(T, \Delta) = \{\Delta \log(1/\Delta) \sqrt{T} (\log T)\}^{c_1}$ for any $c_1 \in (0, 1)$.

Assumption A.4 allows us to identify occurences of jumps from data by thresholding:

Lemma 3.1 *Suppose that Assumption A.4 holds. Then, for any* ω *(*$\in \Omega^*$*;* Ω^* *is an event with* $\Pr[\Omega^*] = 1$*), there exists some random variable* $\bar{\Delta}(\omega) > 0$ *such that for any* $\Delta \leq \bar{\Delta}(\omega)$*,*

$$\mathbf{1}\{[X_{t_i} - X_{t_{i-1}}]^2 \leq r(\Delta, T)\} = \mathbf{1}\{N_{t_i} - N_{t_{i-1}} = 0\}.$$

This lemma tells us that we can identify jump occurrences through the threshold parameter $r(\Delta, T)$. If T is fixed, the above result is simply Theorem 1 of Mancini (2009), but we here allow for $T \to \infty$. This in turn is used to derive the following uniform rate result:

Theorem 3.3 *Suppose that Assumptions A.1-A.2, A.4, B.1 and* $\mathbb{K}$*.1 hold. Then,*

$$\sup_{\tau \in [\sqrt{h}, T-\sqrt{h}]} |\hat{\sigma}^2_{\mathrm{JR},\tau} - \sigma^2_\tau| = O_P\left(\vartheta^{\mathrm{JR}}_{T,\Delta}\right),$$

as $\Delta \to 0$*, and* $T/\Delta \to \infty$*, where*

$$\vartheta^{\mathrm{JR}}_{T,\Delta} := \bar{\vartheta}_{T,\Delta} + O_P(\Delta(T/h)\left\{\Delta^{p/(2+p)} T^{2(1+l_1)/(2+p)} + [\log(1/\Delta)]\, T^{2(1+l_2)/q}\right\}).$$

If T is fixed, we can show that the second term of $\vartheta^{\mathrm{JR}}_{T,\Delta}$ is reduced to $O_P(h^{-1}\Delta\log(1/\Delta))$ (we omit the proof of this claim for brevity; see discussions that follows Theorem 3.1, and also the proof of Theorem 1 of Mancini, 2009) and is $o_P\left(\bar{\vartheta}_{T,\Delta}\right)$, which means the uniform rate of the noise-free and non-jump-robust estimator $\hat{\sigma}^2_\tau$ coincides with that of the jump-robust estimator $\hat{\sigma}^2_{\mathrm{JR},\tau}$. If $\{\mu_t\}$ and $\left\{\sigma^2_t\right\}$ are uniformly bounded as $T \to \infty$, $\bar{\vartheta}_{T,\Delta} = h^\gamma + \sqrt{h^{-1}\Delta\log(1/\Delta)}$ while the second term of $\vartheta^{\mathrm{JR}}_{T,\Delta}$ takes the form $O_P(Th^{-1}\Delta\log(1/\Delta))$. The presence of "$T$" in $O_P(Th^{-1}\Delta\log(1/\Delta))$ comes from the fact that the number of Poisson jump events over $[0, T]$ is $O_P(T)$.

3.4 Jump and Noise Case ($\varepsilon \neq 0$, $J \neq 0$)

In the case where both jumps and market microstructure noise are present in data, we may combine the ideas of the noise- and jump-robust spot volatility estimators developed in the previous two subsections. A naive approach would be to simply take the jump-robust estimator in eq. (3.14) and then replace X_t by the noise-filtered process $\hat{X}_t$ given in eq. (3.7) using, as before, additional pseudo sampling time points t_i with $\Delta = t_i - t_{i-1}$ being the time distance between these. This approach still provides a pointwise consistent estimator of X_t if the kernel L used in the computation of $\hat{X}_t$ is a forward looking kernel; that is, L has support on $(0, +\infty)$. This is due to the fact that, even with jumps, X_t is càdlàg. However, even with a forward-looking kernel, $\hat{X}_t$ will smooth out jumps and so will not be uniformly consistent. More specifically, the estimator is not stochastically equicontinuous which is required for it to be uniformly consistent (see Newey and McFadden, 1994, Section 2.7). To see this, recall that for $\hat{X}_t$ to be stochastically equicontinuous on $(0, T)$, then, for any $t_0 \in (0, T)$ and any sequence $t_M \to t_0$, it has to satisfy $\hat{X}_{t_M} \xrightarrow{P} X_{t_0}$. Now, suppose that a jump occurred at time t_0 ($N_{t_0^+} - N_{t_0^-} \neq 0$): Choosing $t_M = t_0 - ca_M$ where $a_M \to 0$ is the bandwidth sequence used in the computation of $\hat{X}_{t_M}$ and $c > 0$ is a constant, we then have, conditionally on $\{X_t\}$,

$$\hat{X}_{t_M} = \int_0^c X_{t_0 + a_M(u-c)} L(u)\, du + \int_c^T X_{t_0 + a_M(u-c)} L(u)\, du + o_P(1),$$

as $Ma_M \to \infty$, where this expression can be derived in the same way as in Proof of Lemma A.2) (recalling eq. (3.6) and the independence between $\{X_t\}$ and $\{\varepsilon_i\}$) Here, $X_{t_0+a_M(u-c)} \to X_{t_0^-}$ if $u \in (0,c)$ while $X_{t_0+a_M(u-c)} \to X_{t_0^+}$ if $u \in [c,T)$. Thus, $\hat{X}_{t_M} \xrightarrow{P} \int_0^c L(u)\,du \times X_{t_0^-} + \int_c^T L(u)\,du \times X_{t_0^+} \neq X_{t_0}$, and so $\hat{X}_t$ cannot be stochastically equicontinuous. Given that our asymptotic analysis relies on uniform consistency of $\hat{X}_t$, the naive approach is not viable.[2]

Instead, to detect jumps uniformly over $[0,T]$, we import techniques developed for nonparametric detection of jumps in regression functions. More specifically, we adopt a similar strategy to the one in Gijbels et al. (2007), amongst others, and introduce a backward and forward looking filter,

$$\hat{X}_t^- = (T/M)\sum_{i=1}^M L_a^-(s_i - t)\,Y_i, \quad \hat{X}_t^+ = (T/M)\sum_{i=1}^M L_a^+(s_i - t)\,Y_i,$$

where L^- is a backward-looking kernel with support on $(-\infty, 0)$ and L^+ is a forward-looking kernel with support on $(0, +\infty)$. These two estimators satisfy for any $t \in (0,T)$ and any sequences $t_n^- \to t^-$ and $t_n^+ \to t^+$, $\hat{X}_{t_n^-}^- \xrightarrow{P} X_{t^-}$ and $\hat{X}_{t_n^+}^+ \xrightarrow{P} X_{t^+}$. In particular, with $t_1^* < ... < t_{N_T}^*$ denoting the time points where jumps occurred on $(0,T)$, and, with $t_{n,j}^- = i_j^* T/n$ and $t_{n,j}^+ = (i_j^*+1)T/n$ being the two nearest pseudo time points such that $t_j^* \in [t_{n,j}^-, t_{n,j}^+]$, $(\hat{X}_{t_{n,j}^+}^+ - \hat{X}_{t_{n,j}^-}^-)^2 \xrightarrow{P} \kappa_j^2$ for $j = 1, ..., N_T$, while for all other $i \notin \{i_1^*, ..., i_{N_T}^*\}$, $(\hat{X}_{t_{i+1}}^+ - \hat{X}_{t_i}^-)^2 \xrightarrow{P} 0$. We therefore expect the following generalization of Lemma 3.1 to hold as $\sqrt{a}(T^{1+l_2}/a)^{1/(2+q)} \to 0$ and $\sqrt{a^{-1}\Delta_s \log(1/\Delta_s)} \to 0$:

$$\mathbf{1}\{[\hat{X}_{t_i}^+ - \hat{X}_{t_{i-1}}^-]^2 \leq r(\Delta, T)\} = \mathbf{1}\{N_{t_i} - N_{t_{i-1}} = 0\} \text{ with probability approaching 1.}$$

Note that this is a weaker result compared to Lemma 3.1, where the latter holds almost surely. However, this should suffice in order to show that

$$\hat{\sigma}^2_{\text{JNR},\tau} := \sum_{i=1}^n K_h(t_{i-1} - \tau)\,[\hat{X}_{t_i}^+ - \hat{X}_{t_{i-1}}^-]^2 \mathbf{1}\{[\hat{X}_{t_i}^+ - \hat{X}_{t_{i-1}}^-]^2 \leq r(\Delta, T)\}$$

is a uniformly consistent estimator. A formal proof of this claim is left for future research.

4 Nonparametric Estimation of the SV Model

We here derive the asymptotic properties of the two-step nonparametric estimators of the SV model given in eqs. (2.2)-(2.3). As noted earlier, these estimators could in principle be implemented using any nonparametric spot volatility estimator in the first step, such as the ones analyzed in the previous section. To establish a general result that cover all these, and other, estimators, we here abstract away from the particular features of the first-step estimators analyzed in the previous section and only assume that the chosen estimator $\tilde{\sigma}_\tau^2$ satisfies

$$\max_{1 \leq j \leq N} |\tilde{\sigma}_{\tau_j}^2 - \sigma_{\tau_j}^2| = O_P(\vartheta_N), \quad \text{as } N \to \infty \text{ (and } \delta \to 0), \tag{4.1}$$

[2] Note that what indeed matters in our subsequent analysis is the uniformity over any pseudo (discrete) sampling time points the number of which is finite but increasing as $n \to \infty$ and $\Delta \to 0$, rather than the uniformity over *any* t in a continuum set $(0,T)$ as discussed here; however, we can see the failure of uniform consistency even over such discrete points by the same reason as here.

for some error bound $\vartheta_N \to 0$ which is specific to the estimator (note that ϑ_N may depend on T, Δ and some other quantities such as Δ and Δ_s). For the noise- and jump-free, noise-robust (NR), and jump-robust (JR) estimators analyzed in the previous section, we can choose $\vartheta_N = \bar{\vartheta}_{T,\Delta}$, $\vartheta^{\mathrm{NR}}_{T,\Delta,\Delta_s}$, and $\vartheta^{\mathrm{JR}}_{T,\Delta}$, respectively. Observe that for these three estimators, $[\tau_1, \tau_N] \subset [\sqrt{h}, T - \sqrt{h}]$ by letting $\tau_j = j\delta$ and $N \leq T - \delta$ since $\sqrt{h} < \delta$ under the assumptions imposed on δ below.

The estimation problem is similar to the one of kernel estimation with errors-in-variables. The implications of this for kernel regression have been analyzed in Mammen, Rothe and Schienle (2012) and Sperlich (2009) in a cross-sectional framework. We follow a similar strategy: We split up the total estimation error into two components: One component due to the estimation of $\left\{\sigma_t^2\right\}$ 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. (2.2), we write the total estimation error as

$$\hat{\alpha}(x) - \alpha(x) = \left[\hat{\alpha}(x) - \tilde{\alpha}(x)\right] + \left[\tilde{\alpha}(x) - \alpha(x)\right], \tag{4.2}$$

where $\tilde{\alpha}(x)$ is the infeasible drift estimator based on observations of $\left\{\sigma_t^2\right\}$. The asymptotic properties of the second term follow from arguments as in Bandi and Phillips (2003) under regularity conditions stated below. What remains to be shown is 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. If the rate can be chosen so that the first term is asymptotically negligible, the feasible estimator will be asymptotically equivalent to the infeasible one.

For a given error bound ϑ_N, we constrain the set of feasible bandwidths and pseudo-sampling points used in the second step to control the error arising from the first step:

B-NDR Given ϑ_N in eq. (4.1), δ and b are chosen such that: (i) $\vartheta_N/\delta \to 0$, $\delta^\gamma/b \to 0$ and $Tb \to \infty$; (ii) $\vartheta_N^2 N\left(b^{-1} + b\delta^{-1}\right) \to 0$, $b^5 T = O(1)$ and $Tb\delta^{2\gamma} \to 0$.

B-NDI Given ϑ_N in eq. (4.1), δ and b are chosen such that: (i) $\vartheta_N/\delta^{1-\gamma} \to 0$, $\delta^\gamma/b \to 0$ and $Nb \to \infty$; (ii) $\vartheta_N N\left(b^{-1} + b\delta^{-2+2\gamma}\right) \to 0$, $b^5 N = O(1)$ and $Nb\delta^{2\gamma} \to 0$.

Assumption B-NDR and B-NDI are used to derive the asymptotic properties of the drift and diffusion estimator, respectively. The parts of Assumptions B-NDR and B-NDI that do not involve ϑ_N are similar to the ones imposed in Bandi and Phillips (2003) for the case of stationary diffusion processes. In particular, observations over a growing time span ($T \to \infty$) is required for the drift estimation, but not necessarily so for the diffusion estimation. The additional assumptions involving ϑ_N are introduced to ensure that the error due to the preliminary estimation of $\left\{\sigma_t^2\right\}$ does not affect the asymptotic properties. If we use $\hat{\sigma}_\tau^2$ as a preliminary estimator, 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 impose the following additional assumptions on the volatility dynamics:

A.2' The process $\left\{\sigma_t^2\right\}$ has range $I = (0, \bar{\sigma})$, where $\bar{\sigma} \leq \infty$, and satisfies: (i) $\alpha(x)$ and $\beta^2(x)$ are 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 constant $c \in I$, satisfies $S(x) \to -\infty$ (resp. $+\infty$) as $x \to 0$ (resp. $\bar{\sigma}$) and $\int_0^r \beta^{-2}(x)\,s(x)\,dx < \infty$; (iv) $E\left[\sigma_t^4\right] < \infty$, $E[\left|\alpha\left(\sigma_t^2\right)\right|^\lambda] < \infty$ and $E[\left|\beta\left(\sigma_t^2\right)\right|^\lambda] < \infty$ for some $\lambda > 2$.

Assumption A.2' is a strengthening of Assumption A.2 with $q \geq 2$. 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), Assumption A.2'(iii) is sufficient for the process to be nonexplosive, 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 $\left\{\sigma_t^2\right\}$, and assume that the process has been initialized at this distribution and so is stationary. We can then set $l_2 = 0$ in Assumption A.2 and in the expressions for the uniform rates derived in the previous section.

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 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. Assumption A.2'(iv) imposes two moment conditions on the volatility process. The condition is satisfied by many models, including CIR and GARCH-diffusion models. If one is only interested in estimating the drift of the volatility, and not its diffusion coefficient, A.2'(iv) can be weakened to $E\left[\left|\alpha\left(\sigma_t^2\right)\right|\right] < \infty$ and $E\left[\beta\left(\sigma_t^2\right)\right] < \infty$.

Finally, we impose the following conditions on the kernel $\mathcal{K}$ used in the second step:

$\mathbb{K}$.3 $\mathcal{K} : \mathbb{R} \to \mathbb{R}$ satisfies $\int_{-\infty}^{\infty} \mathcal{K}(x)\,dx = \int_{-\infty}^{\infty} x^2\mathcal{K}(x)\,dx = 1$, $\int_{-\infty}^{\infty} x\mathcal{K}(x)\,dx = 0$ and $\int_{-\infty}^{\infty} \mathcal{K}^2(x)\,dx < \infty$; it is continuously differentiable; and there exist some constants $\bar{K}, C < (0, \infty)$ such that $\sup_{x\in\mathbb{R}} |\mathcal{K}(x)| \leq \bar{K}$, $\sup_{x\in\mathbb{R}} |\mathcal{K}'(x)| \leq \bar{K}$, and $|\mathcal{K}'(x)|$ is not decreasing on $(-\infty, -C]$ and not increasing on $[C, \infty)$.

The conditions imposed in $\mathbb{K}$.3 are almost identical to the ones found in $\mathbb{K}$.1, and the discussion of the latter also applies here.

Theorem 4.1 *Let $\tilde{\sigma}_\tau^2$ be an estimator of σ_τ^2 satisfying eq. (4.1). Assume that Assumptions A.2' and B-NDR(i) hold, and $\mathcal{K}$ satisfies $\mathbb{K}$.3. Then, $\hat{\alpha}(x) \xrightarrow{P} \alpha(x)$. If additionally B-NDR(ii) holds, then*

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

where

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

Theorem 4.2 *Let $\tilde{\sigma}_\tau^2$ be an estimator of σ_τ^2 satisfying eq. (4.1). Assume that Assumptions A.2' and B-NDI(i) hold, and $\mathcal{K}$ satisfies $\mathbb{K}$.3. Then $\hat{\beta}(x) \xrightarrow{P} \beta(x)$. If additionally B-NDI(ii) holds, then*

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

where

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

If the condition $b^5T = O(1)$ in B-NDR(ii) is strengthened to $b^5T \to 0$, the bias component in Theorem 4.1 vanishes fast enough to have no impact on the asymptotic distribution. Similarly in Theorem 4.2, if $b^5N \to 0$ then the bias term can be ignored.

The above results show that the feasible estimators is first-order asymptotically equivalent to the infeasible ones based on actual observations of $\{\sigma_t^2\}$ 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. This is due to Assumptions 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 effects 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 distance δ of $\{\hat{\sigma}_t^2\}$ (in estimating $\alpha(x)$ and $\beta^2(x)$) to be larger than the actual time distance between observations, Δ. A realistic scenario would be that intra-daily observations of $\{X_t\}$ are available. Then by choosing δ corresponding to, for example, sampling at a daily frequency, we expect the first-step estimation error to be negligible. This is supported by Jiang and Knight (1999) and Phillips and Yu (2005) where it is demonstrated that Nadaraya-Watson type estimators for (observable) diffusion processes exhibit good performance even for relatively large choices of δ. In total, by choosing δ 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 7.

5 Parametric Estimation of the SV Model

We here give results for the parametric estimators of the SV model given in eq. (2.4). 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 $\{\sigma_t^2\}$, is shown to be negligible

asymptotically under suitable conditions on the bandwidth and pseudo-sampling points. As for the fully nonparametric estimators, we give results for any first-step spot volatility estimator satisfying eq. (4.1).

We impose the following conditions for the analysis of our estimators of the drift parameters:

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

$$\int_I \left[\alpha\left(x;\theta_1\right) - \alpha\left(x\right)\right]^2 \pi\left(x\right) dx = 0 \Leftrightarrow \theta_1 = \theta_1^*;$$

(ii) $\alpha\left(x;\theta_1\right)$ is twice continuously differentiable in θ_1; there exists $A_k\left(\cdot\right)$ satisfying $E\left[A_k^2\left(\sigma_s^2\right)\right] < \infty$, $k = 1, 2$, such that uniformly over $\theta_1 \in \Theta_1$,

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

(iii) $\alpha\left(x;\theta_1\right)$, $\partial_{\theta_1}\alpha\left(x;\theta_1\right)$ and $\partial_{\theta_1\theta_1^{\bigstar}}\alpha\left(x;\theta_1\right)$ are differentiable w.r.t. x for all $\theta_1 \in \Theta_1$; there exists constants $C > 0$ and $v_1 > 0$ such that and $E[\left|\sigma_t^2\right|^{2v_1}] < \infty$ and, uniformly over $\theta_1 \in \Theta_1$:

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

Assumptions A-SDR(i)-(ii) are standard for parametric diffusion estimation, and are similar to those imposed in, for example, Jacod (2006) and Yoshida (1992). A-SDR(i) ensures identification of θ_1 while 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 used to ensure that the variance of the estimator is well-defined. A-SDR(iii) is somewhat atypical; it is used to demonstrate that the error from replacing σ_t by $\tilde{\sigma}_t$ 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 imply both consistency and asymptotic normality of the estimator. If only consistency is of interest, the conditions could be weakened considerably, but for simplicity we maintain A-SDR throughout.

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

B-SDR Given ϑ_N in eq. (4.1), δ is chosen such that: (i) $\vartheta_N/\delta \to 0$; (ii) $\sqrt{T}\left[\delta + \vartheta_N/\delta\right] \to 0$.

The conditions on the shrinking rates of the bandwidth h and the sampling time δ in Assumption B-SDR are simpler than the ones in B-NDR used 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 5.1 *If Assumptions A-SDR and B-SDR(i) hold, then* $\hat{\theta}_1 \xrightarrow{P} \theta_1^*$. *If additionally B-SDR(ii) holds, then*

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

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

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.

Next, we the derive properties of the estimator of the diffusion parameters under the following conditions:

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

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

(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 $\theta_2 \in \Theta_2$,

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

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 $\partial_{\theta_2\theta_2^{\bigstar}}\beta^2(x;\theta_2)$ are differentiable in x for each $\theta_2 \in \Theta_2$ There exist constants $C > 0$ and $v_2 > 0$ such that $E[\left|\sigma_t^2\right|^{2v_2}] < \infty$, and, uniformly over $\theta_2 \in \Theta_2$,

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

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

B-SDI Given ϑ_N in eq. (4.1), δ is chosen such that: (i) $\vartheta_N/\delta^{1-\gamma} \to 0$; (ii) $\sqrt{N}\left[\delta + \vartheta_N/\delta^{1-\gamma}\right] \to 0$.

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 5.2 *If Assumptions A-SDI and B-SDI(i) hold, then* $\hat{\theta}_2 \xrightarrow{P} \theta_2^*$. *If additionally B-SDI(ii) holds, then*

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

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

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, however, that the conditions imposed on T and δ (N) for the drift and diffusion cases are different. As discussed previously, the drift estimation requires $T \to \infty$. This is stronger than the requirement $N \to \infty$ for estimation of the diffusion coefficient. On the other hand, the requirement $T\delta^2 (= N\delta^3) \to 0$ for the drift may be seen as weaker than the requirement for the diffusion, $T\delta (= N\delta^2) \to 0$, where the latter is interpreted as a requirement of finer sample observations.

The above theoretical results are similar to the ones obtained in Todorov (2009), where estimators of integrated volatility are 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.

6 Extension to SV Models with Volatility Jumps

We have so far focused on the situation where the volatility process contains no jumps. We here outline how the estimation procedure can be extended to handle the case where the volatility process is a jump-diffusion; a formal theory for the resulting estimators is left for future research. Suppose that the volatility process solves

$$d\sigma_t^2 = \alpha\left(\sigma_t^2\right)dt + \beta\left(\sigma_t^2\right)dZ_t + dJ_t^\sigma,$$

where $\{J_t^\sigma\}$ is a finite-activity jump process of the form $J_t^\sigma = \sum_{i=1}^{N_t^\sigma}\kappa_i^\sigma$, and therefore has the same structure as the jump component J_t of X_t. To handle this situation, we first have to modify the first-step estimators of σ_t^2 developed in Section 3 since these estimators will smooth out jumps contained in σ_t^2; the reason for this is similar to the spot volatility estimation with noise and jumps where we had to modify the first-step smoother $\hat{X}_t$ in order to handle jumps in X_t (see discussions in Section 3.4). More specifically, when K is chosen as a forward-looking kernel, the proposed spot volatility estimators remain pointwise consistent but are not stochastically equicontinuous. Consider, for example, the spot volatility estimator in eq. (3.1) when jumps and noise are not present in X_t: By choosing $\tau_n = \tau_0 - ch_n$ for any time point τ_0 where a jump occurred and for any constant $c > 0$, we find $\hat{\sigma}^2_{\tau_n} = \int_0^c K(u)\,du \times \sigma^2_{\tau_0^-} + \int_c^T K(u)\,du \times \sigma^2_{\tau_0^+} + o_P(1)$ as $h_n \to 0$. Thus, $\hat{\sigma}^2_\tau$

cannot be uniformly consistent. Given that our asymptotic results for the two-step estimators rely on uniform consistency of the first-step estimator, modified spot volatility estimators are needed in order to extend the arguments employed to establish the theory in Sections 4-5 to the case of jumps in the volatility.

We use the same idea as in Section 3.4, and construct a backward and forward looking version of either of the four spot volatility estimators developed in Section 3. Again, for simplicity, we focus on the case without noise and jumps in X_t, and introduce

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

where K^+ and K^- are forward and backward looking kernels. These will satisfy for any $\tau \in (0,T)$ and any sequences $\tau_n^- \to \tau^-$ and $\tau_n^+ \to \tau^+$, $\hat{\sigma}^{2,-}_{\tau_n^-} \xrightarrow{P} \sigma^2_{\tau^-}$ and $\hat{\sigma}^{2,+}_{\tau_n^+} \xrightarrow{P} \sigma^2_{\tau^+}$. Thus, by the same arguments given in Section 3.4, we expect

$$\begin{aligned}
\hat{\alpha}_{\mathrm{JR}}\left(x\right) &= \frac{\sum_{j=1}^{N-1} \mathcal{K}_b(\hat{\sigma}^{2,+}_{\tau_{j-1}} - x)[\hat{\sigma}^{2,-}_{\tau_j} - \hat{\sigma}^{2,+}_{\tau_{j-1}}]\mathbf{1}\{[\hat{\sigma}^{2,+}_{\tau_j} - \hat{\sigma}^{2,-}_{\tau_{j-1}}]^2 \leq r_\sigma\left(\Delta,T\right)\}}{\delta \sum_{j=1}^{N} \mathcal{K}_b(\hat{\sigma}^{2,+}_{\tau_{j-1}} - x)\mathbf{1}\{[\hat{\sigma}^{2,+}_{\tau_j} - \hat{\sigma}^{2,-}_{\tau_{j-1}}]^2 \leq r_\sigma\left(\Delta,T\right)\}}, \\
\hat{\beta}^2_{\mathrm{JR}}\left(x\right) &= \frac{\sum_{j=1}^{N-1} \mathcal{K}_b(\hat{\sigma}^{2,+}_{\tau_{j-1}} - x)[\hat{\sigma}^{2,-}_{\tau_j} - \hat{\sigma}^{2,+}_{\tau_{j-1}}]^2\mathbf{1}\{[\hat{\sigma}^{2,+}_{\tau_j} - \hat{\sigma}^{2,-}_{\tau_{j-1}}]^2 \leq r_\sigma\left(\Delta,T\right)\}}{\delta \sum_{j=1}^{N} \mathcal{K}_b(\hat{\sigma}^{2,+}_{\tau_{j-1}} - x)\mathbf{1}\{[\hat{\sigma}^{2,+}_{\tau_j} - \hat{\sigma}^{2,-}_{\tau_{j-1}}]^2 \leq r_\sigma\left(\Delta,T\right)\}},
\end{aligned}$$

to be consistent estimators of the drift and volatility functions, where $\mathbf{1}\{[\hat{\sigma}^{2,+}_{\tau_j} - \hat{\sigma}^{2,-}_{\tau_{j-1}}]^2 \leq r_\sigma\left(\Delta,T\right)\}$ removes the parts of the volatility trajectory that contain jumps as $r_\sigma\left(\Delta,T\right)$ shrinks to zero at a suitable rate.

Note that the above estimators do not impose any Markov structure on the jump component. If we do assume that the jump intensity of J^σ_t is a function of σ^2_t, this can be also estimated by using the estimators developed in Bandi and Nguyen (2003) for the case where σ^2_t is observed, except that we here replace the unobserved component $[\sigma^2_{\tau_j} - \sigma^2_{\tau_{j-1}}]$ by $[\hat{\sigma}^{2,-}_{\tau_j} - \hat{\sigma}^{2,+}_{\tau_{j-1}}]$.

7 Bandwidth Selection and Sampling

All of the estimators analyzed in Sections 4 and 5 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 4.1-5.2 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 multi-step estimators.[3]

To compute the first-step volatility estimator, a natural, data-driven bandwidth selection method is cross-validation. For the basic estimator in eq. (3.1), there is only one bandwidth, h, to be

[3]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.

chosen in the first step. Kristensen (2010a) argues that the cross-validated bandwidth, $h_{CV} = \arg\min_{h>0} CV((\Delta X)^2/\Delta, \hat{\sigma}^2, h)$, should be asymptotically optimal where

$$CV((\Delta X)^2/\Delta, \hat{\sigma}^2, h) := \sum_{i=1}^{n} 1\{T_l \leq t_i \leq T_u\}[(\Delta X_i)^2 - \hat{\sigma}^2_{-i,t_i}]^2,$$

for some $0 \leq T_l < T_u \leq T$, with $(\Delta X_i)^2 := \left(X_{t_{i+1}} - X_{t_i}\right)^2$ and $\hat{\sigma}_{-i,t_i}$ being the leave-one-out version of $\hat{\sigma}^2_{t_i}$. This criterion is tailored to minimize the integrated squared error of the volatility estimator, $\int_{T_u}^{T_l}[\sigma^2_t - \hat{\sigma}^2_t]^2 dt$. Since the end goal is to obtain precise estimates of the SV model, we should rather 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.

In the case of noise-robust estimator, we first have to choose two bandwidths, a and h. One simple way of doing so is to first choose a as $a_{CV} = \arg\min_{a>0} CV(Y, \hat{X}, a)$ and then $h_{CV} = \arg\min_{h>0} CV((\Delta\hat{X})^2/\Delta, \hat{\sigma}^2_{\mathrm{NR}}, h)$. For the jump-robust estimator, we have to choose the threshold parameter r. There appears to be no theory for how to choose this, but Mancini, Mattiussi, and Renò (2012) show that simple rule-of-thumb selection rules work well. Finally, for the noise- and jump-robust estimator, we propose to combine the selection rules for the noise-robust and jump-robust estimators, respectively, as described above.

Once $\left\{\hat{\sigma}^2_t\right\}$ 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 nonparametric 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. Moreover, by choosing δ to correspond to daily observations, we hope that the additional time series dependence in $\left\{\hat{\sigma}^2_t\right\}$ 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 (2015) 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 $\{\hat{\sigma}^2_{i\delta} : i = 1, ..., N\}$ 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. First, the MSE for the estimator of the drift parameter, $\tilde{\theta}_k$, is given by $\mathrm{MSE}_{\theta_k} := E[||\tilde{\theta}_k - \theta^*_k||^2]$ for $k = 1, 2$. By a standard Taylor expansion, $\tilde{\theta}_k - \theta^*_k = -[H^*_k + o_P(1)]^{-1}\hat{S}_k(\theta^*_k, \sigma^2)$, for $k = 1, 2$, where $\hat{S}_k$ is the

score function (see Appendix A.3 for its expression) and H_k^* is the limit of the Hessian function evaluated at the true value θ_k^* as defined in Theorems 5.1 and 5.2. Since it is not easy to directly analyze MSE_{θ_k}, we introduce an approximate version,

$$\mathrm{MSE}_{\theta_k}^* := \mathrm{tr}\left\{H_k^{*-1} E\left[\hat{S}_k\left(\theta_k^*, \sigma^2\right) \hat{S}_k\left(\theta_k^*, \sigma^2\right)^{\bigstar}\right] H_k^{*-1}\right\}, \tag{7.1}$$

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 (see Kanaya and Kristensen, 2015), 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, 2015), and therefore we can decompose $\mathrm{MSE}_{\theta_k}^*$ into three terms:

$$\mathrm{MSE}_{\theta_k}^* = \mathrm{tr}\{\mathcal{B}_{\theta_k}\mathcal{B}_{\theta_k}^{\bigstar}\} + \mathrm{tr}\{\mathcal{V}_{\theta_k}\} + \mathrm{tr}\{\mathcal{C}_{\theta_k}\} \tag{7.2}$$

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. We then derive first-order asymptotic approximations of these terms under the following assumptions:

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

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

Theorem 7.1 *Suppose that Assumption A.2' holds. (i) If Assumptions A-SDR and C-SDR are satisfied, then* $\mathcal{B}_{\theta_1} = \delta\bar{\mathcal{B}}_{\theta_1} + o(\delta)$, $\mathcal{C}_{\theta_1} = O(\delta^2)$, *and* $\mathcal{V}_{\theta_1} = \mathrm{tr}\{H_1^{*-1}\Omega_1^* H_1^{*-1}\}/T + o(1/T)$, *where*

$$\bar{\mathcal{B}}_{\theta_1} := E\left[\partial_{\theta_1}\alpha\left(\sigma_t^2;\theta_1^*\right)\left[\alpha'\left(\sigma_t^2\right)\alpha\left(\sigma_t^2\right) + \alpha''\left(\sigma_t^2\right)\beta^2\left(\sigma_t^2\right)/2\right]\right]. \tag{7.3}$$

(ii) If Assumptions A-SDI and C-SDI are satisfied, then $\mathcal{B}_{\theta_2} = \delta\bar{\mathcal{B}}_{\theta_2} + o(\delta)$, $\mathcal{C}_{\theta_2} = O(\delta^2)$, *and* $\mathcal{V}_{\theta_2} = \mathrm{tr}\{H_2^{*-1}\Omega_2^* H_2^{*-1}\}/n + o(1/n)$, *where*

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

An immediate consequence of the above theorem is that $\mathrm{MSE}_{\theta_1}^* = O(\delta^2) + O(1/T)$ and $\mathrm{MSE}_{\theta_2}^* = O(\delta^2) + O(1/n) = O(\delta^2) + O(\delta/T)$. From the above expressions, 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.

8 A Simulation Study

We here examine the finite-sample performance of the implementation of our non- and semi-parametric estimators proposed in the previous section; this is done in the ideal setting where

jumps and noise are not present. We choose as data-generating process the following stochastic volatility model:

$$\begin{cases} dX_t = \sigma_t dW_t \\ d\sigma_t^2 = \beta\left(\alpha - \sigma_t^2\right) dt + \kappa\sigma_t^2 dZ_t \end{cases}, \tag{8.1}$$

where $\{W_t\}$ and $\{Z_t\}$ are independent standard Brownian motions. This is the continuous-time limit version of the GARCH model (see Drost and Werker, 1996), and satisfies the conditions imposed in Sections 3-5. 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 discretization 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 the previous section 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 $\sigma_\delta^2, \sigma_{2\delta}^2, \sigma_{3\delta}^2, ...$

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 $\text{BIAS}^2 = \int_{0.3}^{0.8} \left[\alpha(x) - \bar{\alpha}(x)\right]^2 dx$ 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 $\text{VAR} = \frac{1}{S}\sum_{s=1}^{S} \int_{0.3}^{0.8} \left[\hat{\alpha}_s(x) - \bar{\alpha}(x)\right]^2 dx$ and $\text{MSE} = \frac{1}{S}\sum_{s=1}^{S} \int_{0.3}^{0.8} \left[\hat{\alpha}_s(x) - \alpha(x)\right]^2 dx = \text{BIAS}^2 + \text{VAR}$.

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 columns 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 columns 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 Figures 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 $\{\sigma_t^2\}$ 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 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 Nadaraya-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. (8.1) as the DGP with the same parameter values. For this model, the parametric least-squares estimators can be written in the closed form: $\hat{\alpha} = -a_{\text{OLS}}/b_{\text{OLS}}$, $\hat{\beta} = -b_{\text{OLS}}$, and $\hat{\kappa}^2 = c_{\text{OLS}}$, where, with $X_{\tau_j} = (1, \hat{\sigma}^2_{\tau_j})^{\bigstar}$,

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

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 frequen-

cies, $\delta = 1/12$, $1/6$ 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.

Before concluding this section, we note that Zu (2014) presents some simulation results with various parameter settings, comparing our method (for parametric models) with the one based on the estimation of integrated volatility. His investigation suggests good performances of our method.

9 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. 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 (2012) in a cross-sectional setting.

It would also be of interest to provide a complete asymptotic analysis of the jump- and noise robust spot volatility estimator proposed in Section 3.4 and the estimators proposed Section 6 allowing for jumps in the volatility.

References

Aït-Sahalia, Y. and J.Y. Park (2013) Bandwidth Selection and Asymptotic Properties of Local Nonparametric Estimators in Possibly Nonstationary Continuous-Time Models. Working paper, Indiana University.

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

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

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

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

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

Barndorff-Nielsen, O.E. and 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.

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

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

Borkovec, M. and Klüpperlberg, C. (1998) Extremal Behavior of Diffusion Models in Finance. *Extremes* 1, 47-80.

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

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

Comte, F. and E. Renault (1996) Long Memory Continuous Time Models, *Journal of Econometrics* 73, 101-149.

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

Creel, M. and D. Kristensen (2014) ABC of SV: Limited Information Likelihood Inference in Stochastic Volatility Jump-Diffusion Models. CREATES Research Paper 2014-30, University of Aarhus.

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

Dzhaparidze, K. and J.H. van Zanten (2001) On Bernstein-type Inequalities for Martingales. *Stochastics Processes and their Applications* 93, 109-117.

Fan, J. and Y. Wang (2008) Spot Volatility Estimation for High-Frequency Data, *Statistics and its Interface* 1, 279-288.

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, and G. Tauchen (1997) Estimation of Stochastic Volatility Models with Diagnostics. *Journal of Econometrics* 81, 159–192.

Gao, J., S. Kanaya, D. Li and D. Tjøstheim (2015) Uniform Consistency for Nonparametric Estimators in Null Recurrent Time Series. Forthcoming in *Econometric Theory*.

Gijbels, I., Lambert, A., and Qiu, P. (2007) Jump-preserving Regression and Smoothing Using Local Linear Fitting: A Compromise. *Annals of the Institute of Statistical Mathematics* 59, 235–272.

Hansen, B.E. (2008) Uniform Convergence Rates for Kernel Estimation with Dependent Data, *Econometric Theory* 24, 726-748.

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

Hoffman, M., A. Munk and J. Schmidt-Hieber (2012) Adaptive wavelet estimation of the diffusion coefficient under additive error measurements. *Annales de l'Institut Henri Poincaré- Probabilités et Statistiques* 48, 1186-1216.

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

Jacod, J., Y. Li, P. Mykland, M. Podolskij and M. Vetter (2009) Microstructure Noise in the Continuous Xase: The Pre-Averaging Approach, *Stochastic Processes and their Applications* 119, 2249-2276.

Jiang, G.J. and J.L. Knight (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.

Jiang, G.J. and 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.

Kanaya, S. (2014) Uniform Convergence Rates of Kernel-Based Nonparametric Estimators for Continuous Time Diffusion Processes: A Damping Function Approach. Manuscript, University of Aarhus.

Kanaya, S. and D. Kristensen (2015) Optimal Sampling and Bandwidth Selection for Nonparametric Kernel Estimators of Diffusion Processes. Manuscript, University of Aarhus.

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

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

Kloeden, P.E. and 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 and M. Schienle (2012) Nonparametric Regression with Nonparametrically Generated Covariates. *Annals of Statistics* 40, 1132-1170.

Mammen, E., C. Rothe and M. Schienle (2013) Semiparametric Estimation with Generated Covariates, Manuscript, Columbia University.

Malliavin, P. and M. Mancino (2002) Fourier Series Method for Measurement of Multivariate Volatilities, *Finance & Stochastics* 6, 49-61.

Malliavin, P. and M. Mancino (2009) A Fourier Transform Method for Nonparametric Estimation of Volatility, *Annals of Statistics* 37, 1983-2010.

Mancini, C. (2009) Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps. *Scandinavian Journal of Statistics* 36, 270-296.

Mancini, C., V. Mattiussi, and R. Renò (2012) Spot Volatility Estimation Using Delta Sequences. DiMaD Working Paper, No 2012-10, Universita' degli Studi di Firenze.

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

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

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

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

Podolskij, M. and M. Vetter (2009a) Estimation of Volatility Functionals in the Simultaneous Presence of Microstructure Noise and Jumps, *Bernoulli* 15, 634-658.

Podolskij, M. and M. Vetter (2009b) Bipower-type Estimation in a Noisy Diffusion Setting, *Stochastic Processes and their Applications* 119, 2803-2831.

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

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

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

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

Wang, Q. and N. Chan (2014) Uniform Convergence Rates for a Class of Martingales with Application in Non-linear Cointegrating Regression. *Bernoulli* 20, 1-393.

Xiao, Z. O.B. Linton, R.J. Carroll and 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, L., P.A. Mykland and Y. Aït-Sahalia (2005) A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data, *Journal of the American Statistical Association* 100, 1394-1411.

Zu, Y. and H.P. Boswijk (2014) Estimating Spot Volatility with High-frequency Financial Data. *Journal of Econometrics* 181, 117–135.

Zu, Y. (2014) Estimating Stochastic Volatility Models with High-Frequency Data: A Monte Carlo Investigation, Working paper, City University London.

A Proofs of Theorems

A.1 Proofs for Section 3

We first state a result on almost sure Hölder continuity of the process $t \mapsto \sigma_t^2$ uniformly over the infinite time interval $[0, \infty)$:

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

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

Proof of Theorem 3.1. 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 setting $\mu_t = \sigma_t^2 = 0$ for $t < 0$. By Ito's lemma for continuous semimartingales,

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

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

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

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

We show below that

$$
\begin{aligned}
R_1 &= O_P\left(\Delta T^{(2+2l_1)/(2+p)}h^{-2/(2+p)}\right); &\text{(A.2)}\\
R_2 &= O_P\left(\Delta^{1/2}T^{\eta_0}\left(T/h\right)^{f_0}\right); &\text{(A.3)}\\
R_3 &= O_P\left(\sqrt{h^{-1}\Delta\log(1/\Delta)}\Delta^{1/2}T^{\eta_0}\left[T^2/h\Delta\log\left(1/\Delta\right)\right]^{f_0}\right); &\text{(A.4)}\\
R_4 &= O_P\left(\sqrt{h^{-1}\Delta\log(1/\Delta)}T^{2l_2/(2+q)}\left[T^2/h\Delta\log(1/\Delta)\right]^{2/(2+q)}\right); &\text{(A.5)}\\
R_5 &= O_P\left(h^{\gamma}\right), &\text{(A.6)}
\end{aligned}
$$

where $\eta_0 := l_1/\left(2+p\right)+l_2/\left(2+q\right)$ and $f_0 := 1/\left(2+p\right)+1/\left(2+q\right)$. We note that R_1 is of smaller order than R_2, R_3 is of smaller order than R_4 by eq. (3.2), and that R_2 is of the same order as R_4 since

$$
\begin{aligned}
R_2 = O_P(R_4)\times\left\{T^{(1+l_1)(2+q)-(3+l_2)(2+p)}h^{p(2+q)/2}\left[h\Delta^2\log(1/\Delta)\right]^{(2+p)}\right\}^{1/(2+q)(2+p)}\\
\times\left[\log(1/\Delta)\right]^{-q/2(2+q)} = o_P(R_4),
\end{aligned}
$$

where the last equality is due to eq. (3.3). As a result, R_4 and R_5 are the dominant terms, yielding the claimed result.

Proof of eq. (A.2). By Jensen's inequality and $\max_{1\le i\le n}\max_{s\in[t_{i-1},t_i]}|s-t_{i-1}|\le\Delta$,

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

Thus, for a given sequence $\{\phi_T\}$,

$$
\begin{aligned}
R_1 &\le \Delta\sup_{\tau\in[0,T]}\sum\nolimits_{i=2}^{n}\left|K_h\left(t_{i-1}-\tau\right)\right|\int_{t_{i-1}}^{t_i}\left|\mu_s\right|^2 1\left\{\left|\mu_s\right|\le\phi_T\right\}ds\\
&\quad+\Delta\sup_{\tau\in[0,T]}\sum\nolimits_{i=2}^{n}\left|K_h\left(t_{i-1}-\tau\right)\right|\int_{t_{i-1}}^{t_i}\left|\mu_s\right|^2 1\left\{\left|\mu_s\right|>\phi_T\right\}ds =: R_{11}+R_{12}. \qquad\text{(A.7)}
\end{aligned}
$$

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

$$
\begin{aligned}
R_{11} &= \Delta\sup_{\tau\in[0,T]}\sum_{i=2}^{n}\frac{1}{h}\int_{t_{i-1}}^{t_i}\left|K\left(\frac{s-\tau}{h}+\frac{t_{j-1}-s}{h}\right)\right|\left|\mu_s\right|^2 1\left\{\left|\mu_s\right|\le\phi_T\right\}ds\\
&= \Delta\sup_{\tau\in[0,T]}\int_{-\tau/h}^{(T-\tau)/h}\left|K\left(u+O\left(\Delta/h\right)\right)\right|\left|\mu_{uh+\tau}\right|^2 1\left\{\left|\mu_{uh+\tau}\right|\le\phi_T\right\}du\\
&\le \Delta\phi_T^2\times\int_{-\infty}^{\infty}\left|K\left(u+O\left(\Delta/h\right)\right)\right|du = O\left(\Delta\phi_T^2\right). \qquad\text{(A.8)}
\end{aligned}
$$

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 bounded convergence theorem. As for R_{12},

$$E[R_{12}] \leq E\left[\Delta \bar{K} \frac{1}{h} \int_0^T |\mu_s|^2 1\{|\mu_s| > \phi_T\}\, ds\right] \leq \frac{\bar{K}}{h\phi_T^p} \Delta T \sup_{s \leq T} E\left[|\mu_s|^{2+p}\right] = O\left(\frac{\Delta T^{1+l_1}}{h\phi_T^p}\right), \quad \text{(A.9)}$$

where the last equality follows from Assumption A.1. Now, choose $\phi_T = T^{(1+l_1)/(p+2)} h^{-1/(2+p)}$. Then, eqs. (A.7)-(A.9) establish that $R_1 = O_P\left(\Delta T^{(2+2l_1)/(2+p)} h^{-2/(2+p)}\right)$.

Proof of eq. (A.3). By an application of Hölder's inequality, we have $R_2 \leq \sqrt{R_{21} \times R_{22}}$, where

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

With $\{\phi_T\}$ being as before: $R_{21} = O\left(\Delta \phi_T^2\right) + O_P\left(\Delta T^{(1+l_1)} / \left(h\phi_T^p\right)\right)$ by the same arguments as for R_1. Let $\Lambda_{t_i}^2 := \max_{s \in [t_{i-1}, t_i]} \left|\int_{t_{i-1}}^{s} \sigma_u dW_u\right|^2 \leq \tilde{\phi}_T^2 + \Lambda_{t_i}^2 1\left\{\Lambda_{t_i} > \tilde{\phi}_T\right\}$ for any positive sequence $\{\tilde{\phi}_T\}$. Then, analogously to the analysis of R_{12},

$$\begin{aligned} E[R_{22}] &= \Delta^{-1} \tilde{\phi}_T^2 \sup_{\tau \in [0,T]} \sum\nolimits_{i=2}^{n} |K_h(t_{i-1} - \tau)|\, \Delta + \frac{\bar{K}}{h\tilde{\phi}_T^q} \sum\nolimits_{i=2}^{n} E\left[\Lambda_{t_i}^{2+q}\right] \\ &= O\left(\Delta^{-1} \tilde{\phi}_T^2\right) + O\left(\frac{\Delta^{q/2} T^{1+l_2}}{h\tilde{\phi}_T^q}\right), \end{aligned}$$

where the last equality uses that

$$\sup_{\tau \in [0,T]} \sum\nolimits_{i=2}^{n} |K_h(t_{i-1} - \tau)|\, \Delta \leq \int_{-\infty}^{\infty} |K(u + O(\Delta/h))|\, du = O(1); \text{ and}$$

$$\max_{i \in \{2,\ldots,n\}} E\left[\Lambda_{t_i}^{2+q}\right] \leq \max_{i \in \{2,\ldots,n\}} CE\left[\left|\int_{t_{i-1}}^{t_i} \sigma_u^2 du\right|^{(2+q)/2}\right] \leq C\Delta^{(q/2)+1} \sup_{s \leq T} E\left[\sigma_s^{2+q}\right] = O(\Delta^{(q/2)+1} T^{l_2}).$$

The latter bound is due to the Burkholder-Davis-Gundy (BDG) and Jensen inequalities, and Assumption A.2. With $\phi_T = T^{(1+l_1)/(2+p)} h^{-1/(2+p)}$ and $\tilde{\phi}_T = \Delta^{1/2} T^{(1+l_2)/(2+q)} h^{-1/(2+q)}$,

$$\begin{aligned} R_2 &= \left\{O\left(\Delta \phi_T^2\right) + O_P\left(\frac{\Delta T^{(1+l_1)}}{h\phi_T^p}\right)\right\}^{1/2} \left\{O_P\left(\Delta^{-1} \tilde{\phi}_T^2\right) + O_P\left(\frac{\Delta^{q/2} T^{1+l_2}}{h\tilde{\phi}_T^q}\right)\right\}^{1/2} \\ &= O_P\left(\Delta^{1/2} T^{[(1+l_1)/(2+p) + (1+l_2)/(2+q)]} h^{-[1/(2+p) + 1/(2+q)]}\right) = O_P\left(\Delta^{1/2} T^{\eta_0} (T/h)^{f_0}\right). \quad \text{(A.10)} \end{aligned}$$

Proof of eq. (A.4). 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 $\mathcal{T}_k$. We also let $\Gamma(s, t_{i-1}) := (\int_{t_{i-1}}^{s} \mu_u du)\sigma_s$ and $\bar{\Gamma}_T(s, t_{i-1}) = \Gamma(s, t_{i-1}) 1\{|\Gamma(s, t_{i-1})| \leq \varphi_T\}$ for a sequence of positive real

numbers $\{\varphi_T\}_{T\geq 1}$ that are specified below. Then, R_3 is bounded by

$$\begin{aligned} R_3 &\leq \max_{k\in\{1,\ldots,\nu_T\}} \sup_{\tau\in\mathcal{T}_k} \left| \sum_{i=2}^{n} \left[K_h\left(t_{i-1}-\tau\right) - K_h\left(t_{i-1}-\tau_k\right)\right] \int_{t_{i-1}}^{t_i} \bar{\Gamma}_T\left(s,t_{i-1}\right) dW_s \right| \\ &\quad + \max_{k\in\{1,\ldots,\nu_T\}} \left| \sum_{i=2}^{n} K_h\left(t_{i-1}-\tau_k\right) \int_{t_{i-1}}^{t_i} \bar{\Gamma}_T\left(s,t_{i-1}\right) dW_s \right| \\ &\quad + \sup_{\tau\in[0,T]} \left| \sum_{i=2}^{n} K_h\left(t_{i-1}-\tau\right) \int_{t_{i-1}}^{t_i} \Gamma\left(s,t_{i-1}\right) 1\left\{\left|\Gamma\left(s,t_{i-1}\right)\right| > \varphi_T\right\} dW_s \right| \\ &= : R_{31} + R_{32} + R_{33}. \end{aligned}$$

Using Hölder's inequality, Lipschitz continuity of K and $E[\sum_{i=2}^{n}(\int_{t_{i-1}}^{t_i} \bar{\Gamma}_T\left(s,t_{i-1}\right) dW_s)^2] \leq T\varphi_T^2$,

$$\begin{aligned} R_{31} &\leq \max_{k\in\{1,\ldots,\nu_T\}} \sup_{\tau\in\mathcal{T}_k} \left\{ \sum_{i=2}^{n} \left[K_h\left(t_{i-1}-\tau\right) - K_h\left(t_{i-1}-\tau_k\right)\right]^2 \right\}^{1/2} O_P\left(\sqrt{T}\varphi_T\right) \\ &= O_P\left(\frac{T^2\varphi_T}{h^2\Delta^{1/2}\nu_T}\right). \end{aligned} \tag{A.11}$$

Next, define $\varrho\left(s\right) := t_{i-1}$ if $s \in [t_{i-1}, t_i)$, and

$$M_T^k\left(r\right) := \sum_{i=2}^{[nr]} K\left(\frac{t_{i-1}-\tau_k}{h}\right) \int_{t_{i-1}}^{t_i} \bar{\Gamma}_T\left(s,t_{i-1}\right) dW_s = \int_0^{rT} K\left(\frac{\varrho\left(s\right)-\tau_k}{h}\right) \bar{\Gamma}_T\left(s,\varrho\left(s\right)\right) dW_s.$$

Note that $\left\{M_T^k\left(r\right)\right\}_{r\in[0,1]}$ is a continuous martingale for each (k,T) which vanishes at zero and has quadratic variation $\left\langle M_T^k\right\rangle_r = \int_0^{rT} K^2\left(\frac{\varrho(s)-\tau_k}{h}\right) \bar{\Gamma}_T^2\left(s,\varrho\left(s\right)\right) ds$. Thus, for any $c > 0$ and any sequence $\{V_T\}$,

$$\Pr\left(\left|M_T^k\left(1\right)\right| \geq c\right) \leq \Pr\left(\left|M_T^k\left(1\right)\right| \geq c,\ \left\langle M_T^k\right\rangle_1 \leq V_T\right) + \Pr\left(\left\langle M_T^k\right\rangle_1 > V_T\right), \tag{A.12}$$

Applying the exponential inequality for continuous martingales (see, e.g., Dzhaparidze and van Zanten, 2001 or Exercise 3.16 in Ch. IV of Revuz and Yor, 1999), $\Pr\left(\left|M_T^k\left(1\right)\right| \geq c,\ \left\langle M_T^k\right\rangle_1 \leq V_T\right) \leq 2\exp\left\{-c^2/\left(2V_T\right)\right\}$, while

$$\left\langle M_T^k\right\rangle_1 \leq \int_0^T K^2\left(\frac{s-\tau_k}{h} + O\left(\Delta/h\right)\right) ds \times \varphi_T^2 \leq h\int_{-\infty}^{\infty} K\left(u + O\left(\Delta/h\right)\right) du \times \varphi_T^2 \leq C_M h\varphi_T^2, \tag{A.13}$$

for some constant $C_M > 0$. With $V_T = C_M h\varphi_T^2$, this yields $\Pr\left(\left|M_T^k\left(1\right)\right| \geq c\right) \leq 2\exp\left\{-c^2/\left(2V_T\right)\right\}$. Thus, setting $c = Jb_T$, for some constant $J > 0$ and some sequence $\{b_T\}$ tending to zero,

$$\Pr\left(R_{32} \geq Jb_T\right) \leq \sum_{k=1}^{\nu_T} \Pr\left(\left|M_T^k\left(1\right)\right| \geq Jb_T h\right) \leq 2\nu_T \exp\left\{-\frac{J^2 b_T^2 h}{2C_M\varphi_T^2}\right\}.$$

With $\nu_T = T^2\varphi_T/h^2\Delta^{1/2}b_T$ and $\varphi_T = b_T\sqrt{h/\log\left(1/\Delta\right)}$, we obtain $R_{31} = O_P\left(b_T\right)$ and $\Pr\left(R_{32} \geq Jb_T\right) \leq 2\nu_T\Delta^{J^2/2C_M}$. Since $\nu_T = T^{3/2}/h\sqrt{\Delta\log\left(1/\Delta\right)} \leq \Delta^{-\bar{c}}$, for some $\bar{c} > 0$, which can be checked by the

rate condition in (B.1), we have $\Pr(|R_{32}| > Jb_T) \to 0$ for J large enough, implying $R_{32} = o_P(b_T)$. Finally, by using the BDG, Jensen and Hölder inequalities,

$$\begin{aligned} E[R_{33}] &\leq \frac{n\bar{K}\Delta^{1/2}}{h\varphi_T^{r_0/2}} \left[\Delta \left\{\sup_{u\leq T} E[|\mu_u|^{2+p}]\right\}^{1/(2+p)} \left\{\sup_{s\leq T} E[|\sigma_s|^{2+q}]\right\}^{1/(2+q)}\right]^{(2+r_0)/2} \\ &= O\left(\Delta^{(1+r_0)/2} T^{1+\eta_0(2+r_0)/2} \Big/ h\varphi_T^{r_0/2}\right), \end{aligned} \tag{A.14}$$

where we have set $r_0 = (1/f_0) - 2 = [1/(2+p) + 1/(2+q)]^{-1} - 2$, and f_0 and η_0 were defined previously.Therefore, for the term R_{33} to have same rate, b_T, as R_{31} and R_{32}, we require

$$\frac{\Delta^{(1+r_0)/2} T^{1+\eta_0(2+r_0)/2}}{h\varphi_T^{r_0/2}} = b_T \Leftrightarrow b_T = \sqrt{h^{-1}\Delta\log(1/\Delta)}\Delta^{1/2}T^{\eta_0}\left[T^2/h\Delta\log(1/\Delta)\right]^{1/(2+r_0)},$$

where the equivalence is due to the above choice of φ_T. This establishes the desired result.

Proof of eq. (A.5). The convergence rate of R_4 can be derived analogously to that of R_3. Construct a covering of $[0,T]$, $\{\mathcal{U}_k\}_{k=1}^{v_T}$, where each $\mathcal{U}_k$ has the radius T/v_T from the center τ_k. Let $\Pi(s,t_{i-1}) := (\int_{t_{i-1}}^{s} \sigma_u dW_u)\sigma_s$ and $\bar{\Pi}_T(s,t_{i-1}) := \Pi(s,t_{i-1})\,1\{|\Pi(s,t_{i-1})| \leq \psi_T\}$ for some sequence $\{\psi_T\}_{T\geq 1}$. Then,

$$\begin{aligned} R_4 \leq & \max_{k\in\{1,\ldots,v_T\}} \sup_{\tau\in\mathcal{U}_k} \left|\sum\nolimits_{i=2}^{n} \left[K_h(t_{i-1}-\tau) - K_h(t_{i-1}-\tau_k)\right] \int_{t_{i-1}}^{t_i} \bar{\Pi}_T(s,t_{i-1})\, dW_s\right| \\ & + \max_{k\in\{1,\ldots,v_T\}} \left|\sum\nolimits_{i=1}^{n} K_h(t_{i-1}-\tau_k) \int_{t_{i-1}}^{t_i} \bar{\Pi}_T(s,t_{i-1})\, dW_s\right| \\ & + \sup_{\tau\in[0,T]} \left|\sum\nolimits_{i=1}^{n} K_h(t_{i-1}-\tau_k) \int_{t_{i-1}}^{t_i} \Pi(s,t_{i-1})\, 1\{|\Pi(s,t_{i-1})| > \psi_T\}\, dW_s\right| \\ =: & R_{41} + R_{42} + R_{43}, \end{aligned} \tag{A.15}$$

By the same arguments as for R_{31}, $R_{41} = O_P\left(\frac{T^2\psi_T}{h^2\Delta^{1/2}v_T}\right)$; with $v_T = T^2\psi_T/h^2\Delta^{1/2}b_T$, $R_{41} = O_P(b_T)$. Define a continuous martingale $N_T^k(r) := \int_0^{rT} K\left(\frac{\varrho(s)-\tau_k}{h}\right)\bar{\Pi}_T(s,\varrho(s))\, dW_s$ whose quadratic variation process $\langle N_T^k\rangle_r$ is computed analogously to $\langle M_T^k\rangle_r$ and satisfies $\langle N_T^k\rangle_1 \leq C_N h\psi_T^2$ for some constant $C_N > 0$. Therefore, by the same arguments as for R_{32} and with $\psi_T = b_T\sqrt{h/\log(1/\Delta)}$,

$$\Pr(|R_{42}| > Jb_T) \leq 2v_T \exp\left\{-\frac{J^2 b_T^2 h}{2C_N\psi_T^2}\right\} \leq 2v_T\Delta^{J^2/2C_N}.$$

Since $v_T = T^2/h^{3/2}\sqrt{\Delta\log(1/\Delta)}$ and it is bounded by $\Delta^{-\tilde{c}}$ for some $\tilde{c} > 0$, $\Pr(|R_{42}| > Jb_T) \to 0$ for J large enough (as $\Delta \to 0$) so that $R_{42} = o_P(b_T)$. Finally, similarly to the moment bound for R_{33} in eq. (A.14), $E[R_{43}] = O\left(\Delta^{\rho_0/4}T^{1+l_2/2}/\left(h\psi_T^{\rho_0/2}\right)\right)$, with $\rho_0 := (2+q)/2 - 2$. Therefore, given the above choice of ψ_T, for $R_{43} = O_P(b_T)$, we set $b_T = \sqrt{h^{-1}\Delta\log(1/\Delta)}\left[T^{(2+l_2)}/h\Delta\log(1/\Delta)\right]^{1/(2+\rho_0)}$. This establishes the desired result.

Proof of eq. (A.6). With $\kappa(s) := (\varrho(s) - s)/h$ and $\varrho(s) = t_{i-1}$ for $s \in [t_{i-1}, t_i)$,

$$\begin{aligned} R_5 &\leq \sup_{\tau\in[0,T]} \left| \int_{-\tau/h}^{(T-\tau)/h} K(u + \kappa(uh+\tau)) \left[\sigma^2_{uh+\tau} - \sigma^2_\tau\right] du \right| \\ &\quad + \sup_{\tau\in[\sqrt{h},T-\sqrt{h}]} \left| \int_{-\tau/h}^{(T-\tau)/h} K(u + \kappa(uh+\tau))\, du - \int_{-\infty}^{\infty} K(u + \kappa(uh+\tau))\, du \right| \sup_{\tau\in[0,T]} \sigma^2_\tau \\ &= : R_{51} + R_{52}. \end{aligned} \tag{A.16}$$

Assumption A.2 and Lemma A.1 imply that there exists $\bar{\Delta} > 0$ such that for any $\Delta \leq \bar{\Delta}$ and $|t-s| \leq \Delta$, $|\sigma_t - \sigma_s| \leq D|t-s|^\gamma$ a.s. Therefore, $R_{51} \leq Dh^\gamma \int_{-\infty}^{\infty} |K(u + O(\Delta/h))|\,|u|^\gamma\, du = O_{a.s.}(h^\gamma)$ since $\int_{-\infty}^{\infty} |K(u + O(\Delta/h))|\,|u|^\gamma\, du \leq \int_{-\infty}^{\infty} K^*(u)\,|u|^\gamma\, du < \infty$, where

$$K^*(u) := \begin{cases} \bar{K} & \text{if } |u| \leq C + \bar{\varepsilon}, \\ \frac{|K(C)| - \bar{K}}{\bar{\varepsilon}} [u - C - \bar{\varepsilon}] + \bar{K} & \text{if } |u| \in (C+\bar{\varepsilon}, C+2\varepsilon], \\ |K(u - 2\bar{\varepsilon})| & \text{if } |u| > C + 2\bar{\varepsilon}, \end{cases} \tag{A.17}$$

for some constant $\bar{\varepsilon} > 0$. To bound R_{52}, we first bound $\sup_{\tau\in[0,T]} \sigma^2_\tau$: Let $\{\mathcal{V}_k\}_{k=1}^{w_T}$ be a finite covering of $[0,T]$, where $\mathcal{V}_k$ has center τ_k and radius T/w_T. Then, with $w_T := T\log(1/h)$,

$$\begin{aligned} &\sup_{\tau\in[0,T]} \sigma^2_\tau = \max_{k\in\{1,\ldots,w_T\}} \sup_{\tau\in\mathcal{V}_k} \left|\sigma^2_\tau - \sigma^2_{\tau_k}\right| + \max_{k\in\{1,\ldots,w_T\}} \sigma^2_{\tau_k} \leq O_{a.s.}(|T/w_T|^\gamma) + \textstyle\sum_{k=1}^{w_T} \sigma^2_{\tau_k} \\ &= O_{a.s.}\left(|\log(1/h)|^{-\gamma}\right) + T[\log(1/h)] \times O_P(T^{l_2}) = O_P\left(T^{1+2l_2/(2+q)} \log(1/h)\right), \end{aligned} \tag{A.18}$$

where we have used $\sup_{s\leq T} E[\sigma^2_t] = O\left(T^{2l_2/(2+q)}\right)$ (see Assumption A.2). Next, observe that

$$\begin{aligned} &\sup_{\tau\in[\sqrt{h},T-\sqrt{h}]} \left| \int_{-\tau/h}^{(T-\tau)/h} K(u + k(uh+\tau))\, du - \int_{-\infty}^{\infty} K(u + k(uh+\tau))\, du \right| \\ &\leq \int_{1/\sqrt{h}}^{\infty} |K(u + k(uh+\tau))|\, du + \int_{-\infty}^{1/\sqrt{h}} |K(u + k(uh+\tau))|\, du. \end{aligned} \tag{A.19}$$

The two terms on the RHS are both $O\left(h^{(m+c)/2}\right)$ for some $c > 0$. To see this, note that there exists some $c > 0$ such that $|K(x)| \leq |x|^{-m-1-c}$ for $|x|$ large enough since $\int_{-\infty}^{\infty} x^m K(x)\, dx < \infty$.and that $\kappa(r) = O(\Delta/h) = o(1)$. Thus, $\int_{1/\sqrt{h}}^{\infty} |K(u + \kappa(uh+\tau))|\, du = O(|1/\sqrt{h}|^{-m-c}) = O\left(h^{(m+c)/2}\right)$. The term $\int_{-\infty}^{1/\sqrt{h}} |K(u + k(uh+\tau))|\, du$ can be treated in the same way. In total, $R_{52} = O_P\left(h^{(m+c)/2} T^{1+2l_2/(2+q)} \log(1/h)\right) = o_P(h^\gamma)$, where we have used $\log(1/h)\, h^{c/2} = o(1)$ and $h^{m-2\gamma} T^{2+4l_2/(2+q)} = O(1)$ as imposed in Assumption B.1. This yields the desired result. ■

For the proof of Theorem 3.2, we first derive the uniform convergence rate of $|\hat{X}_t - X_t|$:

Lemma A.2 *Suppose that Assumptions A.1-A.3 and B.1-B.2 and $\mathbb{K}$.1-$\mathbb{K}$.2 are satisfied. Then, as $\Delta, \Delta_s \to 0$, and $T/\Delta, T/\Delta_s \to \infty$, $\hat{X}_t$ defined in eq. (3.7) satisfies*

$$\sup_{\tau\in[\sqrt{h},T-\sqrt{h}]} |\hat{X}_t - X_t| = O_P\left(\sqrt{a}(T^{1+l_2}/a)^{1/(2+q)} + \sqrt{a^{-1}\Delta_s \log(1/\Delta_s)}\right), \tag{A.20}$$

Proof of Theorem 3.2. Observe that

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{NR},\tau} &= \sum_{i=1}^n K_h\left(t_{i-1}-\tau\right)\left[\left(X_{t_i}-X_{t_{i-1}}\right)+\left(\hat{X}_{t_i}-\hat{X}_{t_{i-1}}\right)-\left(X_{t_i}-X_{t_{i-1}}\right)\right]^2 \\ &= \hat{\sigma}^2_\tau+\sum_{i=1}^n K_h\left(t_{i-1}-\tau\right)\left[\left(\hat{X}_{t_i}-\hat{X}_{t_{i-1}}\right)-\left(X_{t_i}-X_{t_{i-1}}\right)\right]^2 \\ &\quad -2\sum_{i=1}^n K_h\left(t_{i-1}-\tau\right)\left(X_{t_i}-X_{t_{i-1}}\right)\left[\left(\hat{X}_{t_i}-\hat{X}_{t_{i-1}}\right)-\left(X_{t_i}-X_{t_{i-1}}\right)\right] \\ &= : \hat{\sigma}^2_\tau+U_1\left(\tau\right)-2U_2\left(\tau\right),\end{aligned}$$

where

$$\begin{aligned}\sup_{\tau\in[\sqrt{a},T-\sqrt{a}]}\left|U_1\left(\tau\right)\right| &\leq \underbrace{\sup_{\tau\in[0,T]}\sum_{i=1}^n\left|K_h\left(t_{i-1}-\tau\right)\right|\Delta}_{=O(1)}\times\left(4/\Delta\right)\sup_{t\in[\sqrt{a},T-\sqrt{a}]}\left|\hat{X}_t-X_t\right|^2; \\ \sup_{\tau\in[\sqrt{a},T-\sqrt{a}]}\left|U_2\left(\tau\right)\right| &\leq \sup_{\tau\in[0,T]}\sum_{i=1}^n\left|K_h\left(t_{i-1}-\tau\right)\right|\left[\int_{t_{i-1}}^{t_i}|\mu_s|ds+\left|\int_{t_{i-1}}^{t_i}\sigma_s dW_s\right|\right] \\ &\quad \times 2\sup_{t\in[\sqrt{a},T-\sqrt{a}]}\left|\hat{X}_t-X_t\right|.\end{aligned}$$

By the same arguments as those for R_1 and R_2 in the proof of Theorem 3.1, and by eq. (3.12),

$$\sup_{\tau\in[0,T]}\sum_{i=1}^n\left|K_h\left(t_{i-1}-\tau\right)\right|\left[\int_{t_{i-1}}^{t_i}|\mu_s|ds+\left|\int_{t_{i-1}}^{t_i}\sigma_s dW_s\right|\right]=O_P(\Delta^{-1/2}(T^{(1+l_2)}/a)^{1/(2+q)}).$$

Given this expression and Lemma A.2, $\sup_{\tau\in[\sqrt{a},T-\sqrt{a}]}\left|U_1\left(\tau\right)\right|=o_P\left(\sup_{\tau\in[\sqrt{a},T-\sqrt{a}]}\left|U_2\left(\tau\right)\right|\right)$ and so (3.13) holds. Finally, to derive the rate of $\sup_{\tau\in[\sqrt{h},T-\sqrt{h}]}\left|\check{\sigma}^2_\tau-\sigma^2_\tau\right|$, note that $\Delta^{-1}a\to 0$ and $h^{-1}\Delta\to 0$ since $\vartheta^{\mathrm{NR}}_{T,\Delta,\Delta_s}\to 0$. These in turn imply that $a/h\to 0$ and $[\sqrt{h},T-\sqrt{h}]\subset[\sqrt{a},T-\sqrt{a}]$. Therefore, the result of Theorem 3.1 and the triangular inequality imply the desired result. ■

Proof of Theorem 3.3. We write the continuous component of X_t as $X^*_t=\int_0^t\mu_s ds+\int_0^t\sigma_s dW_s$. Thus, we have $X_t=X^*_t+\sum_{j=1}^{N_t}\kappa_j$. By Lemma 3.1, we have for Δ small enough,

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{JR},\tau}-\sigma^2_\tau &= \sum_{i=1}^n K_h\left(t_{i-1}-\tau\right)\left[X^*_{t_i}-X^*_{t_{i-1}}\right]^2\mathbf{1}\left\{N_{t_i}-N_{t_{i-1}}=0\right\}-\sigma^2_\tau \\ &= \left\{\hat{\sigma}^2_\tau-\sigma^2_\tau\right\}-\sum_{i=1}^n K_h\left(t_{i-1}-\tau\right)\left[X^*_{t_i}-X^*_{t_{i-1}}\right]^2\mathbf{1}\left\{N_{t_i}-N_{t_{i-1}}\neq 0\right\},\end{aligned}$$

where $\hat{\sigma}^2_\tau$ is the no-jump and no-noise estimator based on observing X^*_t, and so $\sup_{\tau\in[\sqrt{h},T-\sqrt{h}]}\left|\hat{\sigma}^2_\tau-\sigma^2_\tau\right|=O_P(\vartheta_{\Delta,T})$. The second term is bounded by

$$\begin{aligned}&N_T\max_{i\in\{2,\ldots,n\}}K_h\left(t_{i-1}-\tau\right)\left[\int_{t_{i-1}}^{t_i}\mu_s ds+\int_{t_{i-1}}^{t_i}\sigma_s dW_s\right]^2 \\ \leq\ &\frac{2N_T\bar{K}}{h}\left[\Delta\max_{i\in\{2,\ldots,n\}}\int_{t_{i-1}}^{t_i}\left|\mu_s\right|^2 ds+\max_{i\in\{2,\ldots,n\}}\left|\int_{t_{i-1}}^{t_i}\sigma_s dW_s\right|^2\right], \qquad (A.21)\end{aligned}$$

where we have used $(A+B)^2\leq 2\left(A^2+B^2\right)$ and Jensen's inequality. We note that $E\left[N_T\right]\leq\bar{\lambda}T$ and can show $\max_{i\in\{2,\ldots,n\}}\int_{t_{i-1}}^{t_i}\left|\mu_s\right|^2 ds=O_P(\Delta^{p/(2+p)}T^{2(1+l_1)/(2+p)})$ in the same way as eq. (B.6) was derived. To find a bound of $\max_{i\in\{2,\ldots,n\}}\left|\int_{t_{i-1}}^{t_i}\sigma_s dW_s\right|$, we use the exponential inequality for

continuous martingales (Exercise 3.16 in Ch. IV of Revuz and Yor, 1999),

$$\begin{aligned}
& \Pr\left(\max_{i\in\{2,...,n\}} |\textstyle\int_{t_{i-1}}^{t_i} \sigma_s dW_s| \geq J\right) \\
\leq & \sum_{i=2}^{n} \left\{ \Pr\left(|\textstyle\int_{t_{i-1}}^{t_i} \sigma_s dW_s| \geq J,\ \textstyle\int_{t_{i-1}}^{t_i} \sigma_s^2 ds \leq \eta_T \right) + \Pr\left(\textstyle\int_{t_{i-1}}^{t_i} \sigma_s^2 ds > \eta_T \right) \right\} \\
\leq & 2n \exp\left\{ -J^2/2\eta_T \right\} + \underbrace{n\Delta^{(2+q)/2} \sup_{s\leq T} E\left[|\sigma_s|^{2+q} \right] \eta_T^{-q/2}}_{O(\Delta^{q/2} T^{1+l_1} \eta_T^{-q/2})}.
\end{aligned}$$

By letting $\eta_T = \bar{J}^2 \Delta T^{2(1+l_1)/q}$ and $J = \bar{J}^{(1+q)} \sqrt{[\Delta \log(1/\Delta)] T^{2(l_2+1)/q}}$ for any $\bar{J} > 0$, we have

$$\Pr\left(\max_{i\in\{2,...,n\}} |\textstyle\int_{t_{i-1}}^{t_i} \sigma_s dW_s| \geq \bar{J}^{(1+q)} \sqrt{[\Delta \log(1/\Delta)] T^{2(1+l_2)/q}} \right) \leq 2n\Delta^{(1/2)\bar{J}^{2q}} + O\left(\bar{J}^{-q}\right).$$

Having $\bar{J}$ large enough, the majorant side can be made arbitrarily small as $\Delta \to 0$ and $T/\Delta \to \infty$ (by the rate condition in (B.1)), and thus

$$\max_{i\in\{2,...,n\}} |\textstyle\int_{t_{i-1}}^{t_i} \sigma_s dW_s| = O_P(\sqrt{[\Delta \log(1/\Delta)] T^{2(1+l_2)/q}}).$$

■

A.2 Proofs for Section 4

The following lemmas will be used in the proofs:

Lemma A.3 *Assume $\tilde{\sigma}_\tau^2$ satisfies eq. (4.1) and Assumptions A.2' and $\mathbb{K}$.3 are satisfied. If there exists some $\bar{q} > 0$ such that $(\vartheta_N/b)^{\bar{q}} = O(b)$, then, as $T \to \infty$, $\delta \to 0$ and $b \to 0$,*

$$\hat{\alpha}(x) - \tilde{\alpha}(x) = O_P\left(\vartheta_N / b\delta^{1/2}\right) + O_P(\vartheta_N/\delta). \tag{A.22}$$

Lemma A.4 *Assume that Assumptions A.2' and $\mathbb{K}$.3 hold. If $\delta^\gamma/b \to 0$ 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*

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

where $bias_\alpha(x)$ is given in Theorem 4.1.

Lemma A.5 *Assume $\tilde{\sigma}_\tau^2$ satisfies eq. (4.1) and Assumptions A.2' and $\mathbb{K}$.3 are satisfied. If there exists some $\bar{q}\,(>0)$ such that $(\vartheta_N/b)^{\bar{q}} = O(b)$, then, as $N \to \infty$ and $b, \delta \to 0$ with $\vartheta_N/\delta^{1-\gamma} \to 0$,*

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

Lemma A.6 *Assume that Assumptions A.2' and $\mathbb{K}$.3 hold. If $N \to \infty$, $b \to 0$ 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*

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

where $bias_{\beta^2}(x)$ is given in Theorem 4.2.

Proof of Theorem 4.1. As in eq. (4.2), we split up $\hat{\alpha}(x) - \alpha(x)$ into two terms. The first term of the RHS of eq. (4.2) converges to zero by Lemma A.3 under Assumption B-NDR(i). Note that since $\gamma \in (0, 1/2)$ and $\delta^\gamma / b \to 0$, we have $\delta^{1/2} < \delta^\gamma < b$ (for δ and h small enough), and $\vartheta_N/\delta > \vartheta_N/\delta^\gamma \delta^{1/2} > \vartheta_N / b\delta^{1/2}$. Therefore, $\vartheta_N/\delta \to 0$ and $\delta^\gamma/b \to 0$ lead to $\vartheta_N/b\delta^{1/2} \to 0$. Noting also that $\vartheta_N/b\delta^{1/2} \to 0$ and $\delta^\gamma/b \to 0$ respectively imply $\vartheta_N/b < \delta^{1/2}$ and $\delta^{1/2} < b^{1/2\gamma}$, the condition of Lemma A.3 below is satisfied with $\bar{q} = 1/2\gamma$. The convergence of the second term is investigated in Lemma A.4 below. Assumption 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 $\tilde{\alpha}(x) - \alpha(x)$. ■

Proof of Theorem 4.2. We follow the same strategy as in the proof of Theorem 4.1: First, write $\hat{\beta}^2(x) - \beta^2(x) = [\hat{\beta}^2(x) - \tilde{\beta}^2(x)] + [\tilde{\beta}^2(x) - \beta^2(x)]$, where the two terms on the RHS are analyzed in Lemmas A.5 and A.6 (see below), respectively. Note that $\vartheta_N/\delta^{1-\gamma} \to 0$ and $\delta^\gamma/b \to 0$ respectively imply $\vartheta_N < \delta^{1-\gamma}$ and $\delta^{1-\gamma} < b^{\frac{1-\gamma}{\gamma}}$, and thus $\vartheta_N/b < b^{\frac{1-\gamma}{\gamma}-1} = b^{\frac{1-2\gamma}{\gamma}} \to 0$ since $\gamma \in (0, 1/2)$. Therefore, the condition of Lemma A.5 is satisfied with $\bar{q} = (1-2\gamma)/\gamma$. Assumption 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 $\sqrt{Nb}\left[\tilde{\beta}^2(x) - \beta^2(x)\right]$. ■

A.3 Proofs for Sections 5-7

To derive the asymptotic results for the proposed estimators, we re-define the objective functions. Instead of $\hat{Q}_k(\theta_k)$ (for $k = 1, 2$) in eq. (2.5), we consider the following objective functions:

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

The maximizer of $R_k(\theta_k, \tilde{\sigma}^2)$ is equal to the original estimator $\hat{\theta}_k$ defined in eq. (2.4) as the maximizer of $\hat{Q}_k(\theta_k)$. Moreover, in contrast to $\hat{Q}_k(\theta_k)$, $\hat{R}_k(\theta_1, \sigma^2)$ has a well-defined asymptotic limit, $k = 1, 2$, given by $R_1(\theta_1) := \int_I \alpha(y; \theta_1)[\alpha(y; \theta_1) - 2\alpha(y)]\pi(y)\,dy$ and $R_2(\theta_2) := \int_I \beta^2(y; \theta_2)\left[\beta^2(y; \theta_2) - 2\beta^2(y)\right]\pi(y)\,dy$. The first and second order derivatives of $\hat{R}_k(\theta_1, \sigma^2)$, $k = 1, 2$, are

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

and

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

We first state the asymptotic distribution of the infeasible estimators:

Lemma A.7 *Suppose that Assumption A.2' is satisfied.*

(i) If A-SDR holds, then

$$\sup_{\theta_k\in\Theta_k}\left|\hat{R}_k(\theta_k,\sigma^2)-R_k(\theta_k)\right|=o_P\left(1\right);\quad \textit{and}\quad \sup_{\theta_k\in\Theta_k}\left\|\hat{H}_k(\theta_k,\sigma^2)-H_k\left(\theta_k\right)\right\|=o_P\left(1\right), \tag{A.23}$$

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

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

where Ω_1^* and H_1^* are given in Theorem 5.1.

(ii) If A-SDI holds, then the results (A.23) hold with $k=2$ and thus $\tilde{\theta}_2\xrightarrow{P}\theta_2^*$. Moreover, if $T\delta\to 0$, then

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

where Ω_2^* and H_2^* are given in Theorem 5.2.

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

Lemma A.8 *Let $\tilde{\sigma}^2_\tau$ be an estimator of σ^2_τ satisfying eq. (4.1). Suppose that Assumptions A.1, A.2', B.1' and A-SDR(iii) are satisfied. Then,*

$$\sup_{\theta_1\in\Theta_1}\left|\hat{R}_1\left(\theta_1,\tilde{\sigma}^2\right)-\hat{R}_1\left(\theta_1,\sigma^2\right)\right| = O_P(\vartheta_N/\delta), \tag{A.24}$$

$$\sqrt{T}\left\|\hat{S}_1\left(\theta_1^*,\tilde{\sigma}^2\right)-\hat{S}_1\left(\theta_1^*,\sigma^2\right)\right\| = O_P(T^{1/2}\vartheta_N/\delta), \tag{A.25}$$

$$\sup_{\theta_1\in\Theta_1}\left\|\hat{H}_1\left(\theta_1,\tilde{\sigma}^2\right)-\hat{H}_1\left(\theta_1,\sigma^2\right)\right\| = O_P(\vartheta_N/\delta). \tag{A.26}$$

Lemma A.9 *Let $\tilde{\sigma}^2_\tau$ be an estimator of σ^2_τ satisfying eq. (4.1). Suppose that Assumptions A.1, A.2', B.1' and A-SDI(iii) are satisfied. Then,*

$$\sup_{\theta_2\in\Theta_2}\left|\hat{R}_2\left(\theta_2,\tilde{\sigma}^2\right)-\hat{R}_2\left(\theta_2,\sigma^2\right)\right| = O_P(\vartheta_N/\delta^{1-\gamma}), \tag{A.27}$$

$$\sqrt{N}\left\|\hat{S}_2\left(\theta_2^*,\tilde{\sigma}^2\right)-\hat{S}_2\left(\theta_2^*,\sigma^2\right)\right\| = O_P(\sqrt{N}\vartheta_N/\delta^{1-\gamma}), \tag{A.28}$$

$$\sup_{\theta_2\in\Theta_2}\left\|\hat{H}_2\left(\theta_2,\tilde{\sigma}^2\right)-\hat{H}_2\left(\theta_2,\sigma^2\right)\right\| = O_P(\vartheta_N/\delta^{1-\gamma}). \tag{A.29}$$

We are now ready to prove Theorems 5.1-5.2:

Proof of Theorem 5.1. 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

$$\hat{R}_1\left(\theta_1, \tilde{\sigma}^2\right) - R_1\left(\theta_1\right) = [\hat{R}_1\left(\theta_1, \tilde{\sigma}^2\right) - \hat{R}_1\left(\theta_1, \sigma^2\right)] + [\hat{R}_1\left(\theta_1, \sigma^2\right) - R_1\left(\theta_1\right)], \tag{A.30}$$

where the two terms on the RHS converge uniformly by Lemma A.8 and A.7 respectively. By a Taylor expansion, $\sqrt{T}[\hat{\theta}_1 - \theta_1^*] = \hat{H}_1^{-1}\left(\bar{\theta}_1, \tilde{\sigma}^2\right)\sqrt{T}\hat{S}_1\left(\theta_1^*, \hat{\sigma}^2\right)$, where $\bar{\theta}_1$ is on the line segment connecting $\hat{\theta}_1$ to θ_1^*. By (A.25) in Lemma A.9 and Assumption B-SDR, $\sqrt{T}[\hat{S}_1\left(\theta_1^*, \tilde{\sigma}^2\right) - \hat{S}_1\left(\theta_1^*, \sigma^2\right)] = o_P\left(1\right)$ while the Hessian satisfies $\hat{H}_1\left(\theta_1, \tilde{\sigma}^2\right) - H_1\left(\theta_1\right) = [\hat{H}_1\left(\theta_1, \tilde{\sigma}^2\right) - \hat{H}_1\left(\theta_1, \sigma^2\right)] + [\hat{H}_1\left(\theta_1, \sigma^2\right) - H_1\left(\theta_1\right)] \xrightarrow{P} 0$ uniformly θ_1 by Lemmas A.8 and A.7. Thus, $\hat{\theta}_1$ has the same asymptotic distribution as the infeasible estimator $\tilde{\theta}_1$, which is given in Lemma A.7. ■

Proof of Theorem 5.2. This follows along the same lines as the proof of Theorem 5.1, and so is omitted. ■

Proof of Theorem 7.1. The details of the proof are provided at Cambridge Journals Online in supplementary material to this article. Readers may refer to the supplementary material associated with this article, available online. ■

B Proofs of Lemmas

Proof of Lemma A.1. From Assumption A.3 and Lemma D.1 (the details of Lemma D.1 are provided online at Cambridge Journals Online in supplementary material to this article), there exists a continuous modification $\{\bar{\sigma}_t^2\}$ of $\left\{\sigma_t^2\right\}$ which is a.s. Hölder globally over $[0, \infty)$. Identifying $\left\{\sigma_t^2\right\}$ with $\{\bar{\sigma}_t^2\}$, we have eq. (A.1). ■

Proof of Lemma A.2. Use that $X_t = \int_0^t \mu_u du + \int_0^t \sigma_u dW_u$ to write $\sup_{t\in[\sqrt{a}, T-\sqrt{a}]} |\hat{X}_t - X_t| \leq S_1 + S_2 + S_3 + S_4 + S_5 + S_6$, where

$$\begin{aligned}
S_1 &: = \sup_{t\in[\sqrt{a}, T-\sqrt{a}]} \left| \int_0^T L_a\left(s-t\right)\left(\int_0^s \mu_u du\right) ds - \int_0^t \mu_u du \right|; \\
S_2 &: = \sup_{t\in[\sqrt{a}, T-\sqrt{a}]} \left| \int_0^T L_a\left(s-t\right)\left(\int_0^s \sigma_u dW_u\right) ds - \int_0^t \sigma_u dW_u \right|; \\
S_3 &: = \sup_{t\in[0,T]} \left| \sum\nolimits_{i=1}^M L_a\left(s_i - t\right) \int_{(i-i)\Delta_s}^{i\Delta_s} \left(\int_{(i-i)\Delta_s}^s \mu_u du\right) ds \right|;
\end{aligned}$$

$$
\begin{aligned}
S_4 &: = \sup_{t\in[0,T]} \left| \sum\nolimits_{i=1}^{M} L_a\left(s_i - t\right) \int_{(i-i)\Delta_s}^{i\Delta_s} \left(\int_{(i-i)\Delta_s}^{s} \sigma_u dW_u \right) ds \right|; \\
S_5 &: = \sup_{t\in[0,T]} \left| \sum\nolimits_{i=1}^{M} \int_{(i-i)\Delta_s}^{i\Delta_s} \left[L_a\left(s_i - t\right) - L_a\left(s - t\right) \right] X_s ds \right|; \\
S_6 &: = \sup_{t\in[0,T]} \left| (T/M) \sum\nolimits_{i=1}^{M} L_a\left(s_i - t\right) \varepsilon_i \right|.
\end{aligned}
$$

By following the same steps as those for eqs. (A.2) and (A.3), $S_3 = O_P(\Delta_s T^{(1+l_1)/(2+p)} a^{-1/(2+p)})$ and $S_4 = O_P(\Delta_s^{1/2} T^{(1+l_2)/(2+q)} a^{-1/(2+q)})$, while we show below that:

$$
\begin{aligned}
S_1 &= O_P(a^{(1+p)/(2+p)} T^{(1+l_1)/(2+p)}); && \text{(B.1)} \\
S_2 &= O_P(\sqrt{a}(T^{1+l_2}/a)^{1/(2+q)}); && \text{(B.2)} \\
S_5 &= O_P(\Delta_s a^{-1} \left[T^{1+l_1/(2+p)} + T^{1/2+l_2/(2+q)} \right]); && \text{(B.3)} \\
S_6 &= O_P(\sqrt{a^{-1}\Delta_s \log(1/\Delta_s)}). && \text{(B.4)}
\end{aligned}
$$

The first rate condition in eq. (3.10) of B.2(ii) yields $S_5 = O_P(S_6)$. It also implies that $T^{1+l_1/(2+p)} = O(1/\sqrt{\Delta_s/a\,(\log T)})$, and so $S_3 = O_P(S_4) \times \frac{T^{-(1+p)/(2+p)} a^{1/(2+q)}}{T^{(1+l_2)/(2+q)} a^{1/(2+p)}} \sqrt{a\,(\log T)} = o_P(S_4)$. Given the rates of S_4 and S_6, $S_4 = O_P(S_6) \times \{a^q T^{2(1+l_2)}/[\log(1/\Delta_s)]^{2+q}\}^{1/2(2+q)}$. This expression, together with eq. (3.12) of B.2(ii), implies that $S_4 = O_P(S_6)$. Thus, S_6 is dominant among the four terms S_3, S_4, S_5 and S_6. Finally, by eq. (3.11) of B.2(ii), $S_1 = O_P(S_2) \times (a^p T^{2(1+l_1)})^{1/2(2+p)} \times (a/T^{1+l_2})^{1/(2+q)} = o_P(S_2)$. This yields the desired result of eq. (A.20).

Proof of eq. (B.1). Since L has compact support, $\int_{-t/a}^{(T-t)/a} L(v)\,dv - 1 = 0$ for a small enough. Thus,

$$
\begin{aligned}
S_1 &= \sup_{t\in[\sqrt{a},T-\sqrt{a}]} \left| \int_{-t/a}^{(T-t)/a} L(v) \left(\int_t^{t+va} \mu_u du \right) dv + \left(\int_{-t/a}^{(T-t)/a} L(v)\,dv - 1 \right) \int_0^t \mu_u du dv \right| \\
&\leq \int_{-\infty}^{\infty} |L(v)|\,dv \sup_{t\in[0,T]} \int_{t-c_0 a}^{t+c_0 a} |\mu_u|\,du. && \text{(B.5)}
\end{aligned}
$$

With $\phi_T = (T^{1+l_1}/a)^{1/(2+p)}$,

$$
\begin{aligned}
\sup_{t\in[0,T]} \int_{t-c_0 a}^{t+c_0 a} |\mu_u|\,du &\leq \sup_{t\in[0,T]} \int_{t-c_0 a}^{t+c_0 a} |\mu_u|\,1\{|\mu_u| \leq \phi_T\}\,du + \int_0^T |\mu_u|\,1\{|\mu_u| > \phi_T\}\,du \\
&\leq 2c_0 a \phi_T + O_P(T \times T^{l_1}/\phi_T^{1+p}) = O_P(a^{(1+p)/(2+p)} T^{(1+l_1)/(2+p)}). && \text{(B.6)}
\end{aligned}
$$

Proof of eq. (B.2). Let $\Gamma_s := \sup_{|r|\leq c_0 a} \left| \int_s^{s+r} \sigma_u dW_u \right|$, where the integral $\int_t^s \sigma_u dW_u$ should be

interpreted as $-\int_s^t \sigma_u dW_u$ if $s > t$. Then,

$$\begin{aligned} E[S_2] &\leq E\left[\sup_{t\in[0,T]}\left|\int_0^T L_a(s-t)\,\Gamma_s ds\right|\right] \\ &\leq \varphi_T \sup_{t\in[0,T]}\int_0^T |L_a(s-t)|\,ds + E\left[\sup_{t\in[0,T]}\left|\int_0^T L_a(s-t)\,\Gamma_s 1\{\Gamma_s > \varphi_T\}\,ds\right|\right] \\ &\leq \varphi_T \int_{-\infty}^{\infty} |L(v)|\,dv + (\bar{K}T/a)\sup_{s\in[0,T]} E[\Gamma_s^{2+q}]/\varphi_T^{q+1} = O(\varphi_T + a^{q/2}T^{1+l_2}/\varphi_T^{q+1}), \end{aligned}$$

where the first equality holds for a small enough, c.f. the analysis of eq. (B.5), the second inequality holds since $\Gamma_s(a) \leq \varphi_T + \Gamma_s 1\{\Gamma_s(a) > \varphi_T\}$, and the last equality uses the BDG and Jensen inequalities. With $\varphi_T = \sqrt{a}(T^{1+l_2}/a^2)^{1/(2+q)}$, $S_2 = O_P(\sqrt{a}(T^{1+l_2}/a)^{1/(2+q)})$.

Proof of eq. (B.3). By the differentiability of L, Assumptions A.1-A.2 and the BDG inequality,

$$\begin{aligned} S_5 &\leq \Delta_s a^{-2}\sup_{t\in[0,T]}\int_0^T L'\left(\frac{s-t}{a} + O(\Delta_s/a)\right)ds\left[\int_0^T |\mu_u|\,du + \sup_{s\leq T}\left|\int_0^s \sigma_u dW_u\right|\right] \\ &\leq \Delta_s a^{-1}\int_{-\infty}^{\infty} L'(u + O(\Delta_s/a))\,du \times O_P\left(T^{1+l_1/(2+p)} + T^{1/2+l_2/(2+q)}\right) \\ &= O_P(\Delta_s a^{-1}\left[T^{1+l_1/(2+p)} + T^{1/2+l_2/(2+q)}\right]). \end{aligned} \tag{B.7}$$

Proof of eq. (B.4). For grid points on the interval $[0,T]$: $u_0 < u_1 < \cdots < u_{J_T}$ (with $u_0 = 0$, $u_{J_T} = T$, $u_{j+1} - u_j = T/J_T$, and $(T/J_T) \to 0$), and for $\bar{\eta} > 0$, define

$$\begin{aligned} \bar{Z}_{i,T}(u_j) &:= L_a(s_i - u_j)\,\varepsilon_i \mathbf{1}\{|\varepsilon_i| \leq \eta_T\} - E[L_a(s_i-u_j)\,\varepsilon_i \mathbf{1}\{|\varepsilon_i| \leq \eta_T\}]; \\ \tilde{Z}_{i,T}(u_j) &:= L_a(s_i - u_j)\,\varepsilon_i \mathbf{1}\{|\varepsilon_i| > \eta_T\} - E[L_a(s_i-u_j)\,\varepsilon_i \mathbf{1}\{|\varepsilon_i| > \eta_T\}]; \text{ and} \\ \eta_T &:= \bar{\eta}[M^2/\sqrt{a\Delta_s \log(1/\Delta_s)}]^{1/(d_s+1)}. \end{aligned} \tag{B.8}$$

Since $E[\varepsilon_i] = 0$,

$$\begin{aligned} S_6 \leq \max_{1\leq j\leq J_T}\sup_{t\in[u_{j-1},u_j]} (T/M)\sum_{i=1}^M |L_a(s_i-u_j) - L_a(s_i-t)|\,|\varepsilon_i| \\ + (T/M)\max_{0\leq j\leq J_T}\left|\sum_{i=1}^M \bar{Z}_{i,T}(u_j)\right| + \max_{0\leq j\leq J_T}(T/M)\sum_{i=1}^M \left|\tilde{Z}_{i,T}(u_j)\right| =: S_{61} + S_{62} + S_{63}. \end{aligned} \tag{B.9}$$

Here, $|L_a(s_i-u_j) - L_a(s_i-t)| \leq \left(T/J_T a^2\right) L'(s_j/a + O(T/J_T a))$ and $\max_{1\leq i\leq M}|\varepsilon_i| = O_P(M)$, and thus

$$\begin{aligned} S_{61} &\leq (1/J_T a)\int_{-\infty}^{\infty} L'(v + O(T/Ma + T/J_T a))dv \times \max_{1\leq i\leq M}|\varepsilon_i| \\ &= O_P(M/J_T a) = O_P(\sqrt{a^{-1}\Delta_s \log(1/\Delta_s)}), \end{aligned} \tag{B.10}$$

where we have set $J_T = M/\sqrt{a\Delta_s \log(1/\Delta_s)}$. We wish to apply Bernstein's inequality on S_{62} (see p. 102 of Van der Vaart and Wellner, 1996) and to this end, compute the variance bound

$$\begin{aligned}
&\text{Var}\left[\sum_{i=1}^{M} \bar{Z}_{i,T}(u_j)\right] \leq (2/\Delta_s) \sum_{i=1}^{M} L_a^2(s_i - u_j)\,\Delta_s E\left[\varepsilon_i^2\right] \\
\leq\ & \left(2/\Delta_s a^2\right) \int_0^T L^2\left((s-u_j)/a + O(\Delta_s/a)\right) ds \sup_{i\geq 1} E\left[\varepsilon_i^2\right] \\
\leq\ & O(1/\Delta_s a) \int_{-\infty}^{\infty} L^2\left(v + O(\Delta_s/a)\right) du \leq C/\Delta_s a,
\end{aligned}$$

for some constant $C > 0$. Given this and the fact that $\left|\bar{Z}_{i,T}(u_j)\right| \leq 2\bar{K}\eta_T/a$, Bernstein's inequality implies that

$$\begin{aligned}
&\Pr\left(S_{62} \geq \bar{\eta}\sqrt{a^{-1}\Delta_s \log(1/\Delta_s)}\right) \leq \sum_{j=1}^{J_T} \Pr\left[\left|\sum_{i=1}^{M} \bar{Z}_{i,T}(u_j)\right| \geq (M/T)\,\bar{\eta}\sqrt{a^{-1}\Delta_s \log(1/\Delta_s)}\right] \\
\leq\ & \sum_{j=1}^{J_T} 2\exp\left\{-\frac{\bar{\eta}^2 (M/T)^2 [a^{-1}\Delta_s \log(1/\Delta_s)]/2}{\sum_{i=1}^{M} \text{Var}\left[\bar{Z}_{i,T}(u_j)\right] + (1/3)\left(2\bar{K}\eta_T/a\right) c\,(M/T)\sqrt{a^{-1}\Delta_s \log(1/\Delta_s)}}\right\} \\
\leq\ & 2J_T \exp\left\{-\frac{\bar{\eta}^2 [\log(1/\Delta_s)]/2}{C + \left(2\bar{K}c/3\right) \times \eta_T \sqrt{\Delta_s \log(1/\Delta_s)/a}}\right\} \to 0. \qquad \text{(B.11)}
\end{aligned}$$

The convergence in the last line occurs for $\bar{\eta}$ large enough (as $T \to \infty$), since $J_T = T/\sqrt{a\Delta_s^3 \log(1/\Delta_s)}$ grows at most at polynomial order of $1/\Delta_s$ (note that $a^{-1} \leq \Delta^{-1/2} T^{c_1}$ and $T \leq \Delta^{-c_2}$ for some $c_1, c_2 > 0$, both of which follow from eq. (3.10)), and since $\eta_T \sqrt{a\Delta_s \log(1/\Delta_s)} = \bar{\eta} \times O(1)$ by eq. (3.9). Next, consider S_{63}: First,

$$\begin{aligned}
&E\left[(T/M) \sum_{i=1}^{M} |L_a(s_i - u_j)\,\varepsilon_i 1\{|\varepsilon_i| > \eta_T\}|\right] \leq (T/M) \sum_{i=1}^{M} |L_a(s_i - u_j)| \sup_{i\geq 1} E[|\varepsilon_i|^{d_s+1}]/\eta_T^{d_s} \\
&\leq \int_{-u_j/a}^{(T-u_j)/a} |L(v + O(1/aM + T/aM))|\, dv \times O(\eta_T^{-d_s}) = O(\eta_T^{-d_s}) \text{ uniformly over } j,
\end{aligned}$$

where $\eta_T^{-d_s} = \bar{\eta}^{-d_s} \times o(\sqrt{a^{-1}\Delta_s \log(1/\Delta_s)})$ since $\eta_T^{-d_s}/\sqrt{a^{-1}\Delta_s \log(1/\Delta_s)} = o(1)$ by eq. (B.8). Therefore, for any $\bar{\eta}$ and for any T large enough,

$$\begin{aligned}
&\Pr\left(S_{63} > \sqrt{a^{-1}\Delta_s \log(1/\Delta_s)}\right) \\
&\leq \sum_{j=1}^{J_T} \Pr\left[(T/M) \sum_{i=1}^{M} |L_a(s_i - u_j)\,\varepsilon_i \mathbf{1}\{|\varepsilon_i| > \eta_T\}| > (1/2)\sqrt{a^{-1}\Delta_s \log(1/\Delta_s)}\right] \\
&\leq \sum_{j=1}^{J_T} \sum_{i=1}^{M} \Pr\left[|L_a(s_i - u_j)|\,|\varepsilon_i \mathbf{1}\{|\varepsilon_i| > \eta_T\}| > (1/2T)\sqrt{a^{-1}\Delta_s \log(1/\Delta_s)}\right] \\
&\leq \sum_{j=1}^{J_T} \sum_{i=1}^{M} \Pr\left[|\varepsilon_i| > \eta_T\right] \leq J_T M \sup_{i\geq 1} E[|\varepsilon_i|^{1+d_s}]/\eta_T^{1+d_s} = O(J_T M/\eta_T^{1+d_s}) = O(1/\bar{\eta}^{1+d_s}),
\end{aligned} \qquad \text{(B.12)}$$

which can be made arbitrarily close to zero by choosing $\bar{\eta}$ large enough. This completes the proof. ■

Proof of Lemma 3.1. The proof proceeds along the same lines as that of Theorem 1 of Mancini (2009). The only difference between ours and her setup is that we allow for $T \to \infty$, and so we only

point out the modifications. Her original proof relies on three preliminary results/assumptions: (i) the modulus of continuity of the Brownian motion $\{W_t\}$ over $[0,T]$ with $T<\infty$ fixed; (ii) the Brownian time-change technique of continuous martingales, i.e., the Dambis-Dubins-Schwarz theorem; and (iii) a certain uniform boundedness condition, which is stated as eq. (14) in Mancini (2009), which is implied by her conditions (1)-(3) of Theorem 1. We extend (i)-(iii) to allow for $T\to\infty$. Re. (i): The Brownian modulus-of-continuity result can be extended in the following manner:

$$\Pr\left[\limsup_{\Delta\searrow 0}\sup_{t,s\in[0,\infty);\,|t-s|\leq\Delta}|W_t-W_s|/\sqrt{2\Delta\log(1/\Delta)}=1\right]=1,$$

as shown in Kanaya (2014). Re. (ii): The Brownian time-change arguments hold irrespectively of whether T is fixed or diverges (see Theorems 1.6, 1.9 and 1.10 in Ch. V of Revuz and Yor, 1999). Re. (iii): The sequence $\{\xi_T\}$ in A4(ii)-(iii) controls the behavior of $\{\mu_t\}$ and $\{\sigma_t^2\}$ as $T\to\infty$, and allows us to obtain a uniform boundedness condition analogous to her ineq. (14). The rest of Mancini's proof remains unchanged. ■

Proof of Lemma A.3. Write $\hat{\alpha}(x)-\tilde{\alpha}(x)=A_1+A_2+A_3$, where

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

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

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

We show below that $A_1=O_P(\vartheta_N/b)$, $A_2=O_P\left(\vartheta_N/b\delta^{1/2}\right)$ and $A_3=O_P(\vartheta_N/\delta)$, which establish the result stated in eq. (A.22). The proofs of these bounds will use the following result repeatedly:

$$\frac{\delta}{T}\sum_{j=1}^{N}\left[\mathcal{K}_b(\sigma^2_{\tau_j}-x)-\mathcal{K}_b(\tilde{\sigma}^2_{\tau_j}-x)\right]=O_P\left(\frac{\vartheta_N}{b}\right). \tag{B.13}$$

Proof of eq. (B.13). Observe that

$$\begin{aligned}
&\text{the LHS of eq. (B.13)}\\
&=\frac{1}{N}\sum_{j=1}^{N}\frac{1}{b}\mathcal{K}'\left((\sigma^2_{\tau_j}-x)b^{-1}-w_j(\sigma^2_{\tau_j}-\tilde{\sigma}^2_{\tau_j})b^{-1}\right)\times w_j(\sigma^2_{\tau_j}-\tilde{\sigma}^2_{\tau_j})b^{-1}\\
&\leq\sup_{1\leq j\leq N}w_j\left|\sigma^2_{\tau_j}-\tilde{\sigma}^2_{\tau_j}\right|b^{-1}\\
&\times\left\{\frac{1}{N}\sum_{j=1}^{N}\frac{1}{b}\left|\mathcal{K}'\left((\sigma^2_{\tau_j}-x)b^{-1}-w_j(\sigma^2_{\tau_j}-\tilde{\sigma}^2_{\tau_j})b^{-1}\right)\right|\mathbf{1}\{w_j|\sigma^2_{\tau_j}-\tilde{\sigma}^2_{\tau_j}|b^{-1}\leq\bar{\epsilon}\}\right.\\
&\left.+\frac{1}{N}\sum_{j=1}^{N}\frac{1}{b}\left|\mathcal{K}'\left((\sigma^2_{\tau_j}-x)b^{-1}-w_j(\sigma^2_{\tau_j}-\tilde{\sigma}^2_{\tau_j})b^{-1}\right)\right|\mathbf{1}\{w_j|\sigma^2_{\tau_j}-\tilde{\sigma}^2_{\tau_j}|b^{-1}>\bar{\epsilon}\}\right\}
\end{aligned}$$

$$\leq O_P\left(\vartheta_N/b\right)$$
$$\times\left\{\frac{1}{N}\sum_{j=1}^{N}\frac{1}{b}\mathcal{K}^*\left((\sigma^2_{\tau_j}-x)b^{-1}\right)+\frac{1}{N}\sum_{j=1}^{N}\frac{\bar{K}}{b}\times\left(\frac{\sup_{1\leq j\leq N}w_j|\sigma^2_{\tau_j}-\tilde{\sigma}^2_{\tau_j}|b^{-1}}{\bar{\epsilon}}\right)^{\bar{q}}\right\} \tag{B.14}$$
$$=O_P\left(\vartheta_N/b\right)\times\left\{O_P\left(1\right)+\frac{\bar{K}}{b}\times O_P\left(\frac{b}{\bar{\epsilon}^{\bar{q}}}\right)\right\}, \tag{B.15}$$

where the first equality holds by the mean-value theorem for some random variables $0\leq w_j\leq 1$, $j=1,\ldots,N$. The inequality (B.14) holds by eq. (4.1) and

$$\left|\mathcal{K}'\left((\sigma^2_{\tau_j}-x)b^{-1}+w_j(\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j})b^{-1}\right)\right|\mathbf{1}\{w_j|\sigma^2_{\tau_j}-\hat{\sigma}^2_{\tau_j}|b^{-1}\leq\bar{\epsilon}\}\leq\mathcal{K}^*\left((\sigma^2_{\tau_j}-x)b^{-1}\right) \tag{B.16}$$

where $\mathcal{K}^*$ is some function and $\bar{\epsilon}>0$, such that $\sup_{|\epsilon|\leq\bar{\epsilon}}|\mathcal{K}'((\sigma^2_{\tau_j}-x)b^{-1}+\epsilon)|\leq\mathcal{K}^*((\sigma^2_{\tau_j}-x)b^{-1})$. By the last condition in $\mathbb{K}.3$, such $\mathcal{K}^*$ and $\bar{\epsilon}$ exist by the same arguments used in the proof of Theorem 3.1 with $\mathcal{K}^*(u)$ being constructed in the same way as $K^*(u)$ in eq. (A.17). The equality in eq. (B.15) holds since $N^{-1}\sum_{j=1}^{N}\mathcal{K}_b^*(\sigma^2_{\tau_j}-x)=O_P(1)$ and $w_j(\sigma^2_{\tau_j}-\tilde{\sigma}^2_{\tau_j})b^{-1}=O_P(\vartheta_N/b)$ (uniformly over j), where the former follows from the positive recurrence of the process $\{\sigma^2_t\}$ and standard arguments for kernel estimators given the properties of $\mathcal{K}^*$.

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

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

By eq. (B.13), and the two following equations, we have $A_1=O_P(\vartheta_N/b)$:

$$\frac{1}{T}\sum\nolimits_{j=1}^{N-1}\mathcal{K}_b(\sigma^2_{\tau_j}-x)(\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j})=\alpha(x)\pi(x)+o_P(1); \tag{B.17}$$
$$\frac{\delta}{T}\sum\nolimits_{j=1}^{N}\mathcal{K}_b(\sigma^2_{\tau_j}-x)=\pi(x)+o_P(1). \tag{B.18}$$

Convergence of A_2. Write $\frac{1}{T}\sum_{j=1}^{N-1}\left[\mathcal{K}_b(\tilde{\sigma}^2_{\tau_j}-x)-\mathcal{K}_b(\sigma^2_{\tau_j}-x)\right](\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j})=A_{21}^{nu}+A_{22}^{nu}$ where

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

Similarly to eq. (B.15) and the proof of eq. (B.13),

$$A_{21}^{nu}=O_P\left(\vartheta_N/b\right)\times\left\{O_P\left(1\right)+\frac{\bar{K}}{b}\times O_p\left(1\right)\times O_P\left(\frac{b}{\bar{\epsilon}^{\bar{q}}}\right)\right\}=O_P\left(\vartheta_N/b\right), \tag{B.19}$$

$$A_{22}^{nu}=O_P\left(\vartheta_N/b\right)\times\left\{O_P\left(\delta^{-1/2}\right)+\frac{\bar{K}}{b}\times O_P\left(\delta^{-1/2}\right)\times O_P\left(\frac{b}{\bar{\epsilon}^{\bar{q}}}\right)\right\}.$$

The denominator of A_2 is $\pi(x) + o_P(1)$ which, combined with eqs. (B.13) and (B.18), yields

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

Convergence of A_3. By analogous arguments, we have

$$A_3 \leq \frac{(1/T)\sum_{j=1}^{N-1} |\mathcal{K}_b(\tilde{\sigma}^2_{\tau_{j-1}} - x)| \times \left(2\sup_{1\leq j\leq N} |\sigma^2_{\tau_j} - \tilde{\sigma}^2_{\tau_j}|\right)}{(\delta/T)\sum_{j=1}^{N} \mathcal{K}_b(\tilde{\sigma}^2_{\tau_j} - x)} = O_P(\vartheta_N/\delta).$$

■

Proof of Lemma A.4. This follows from Aït-Sahalia and Park (2013, Theorem 2). ■

Proof of Lemma A.5. As in the proof of Lemma A.3, $\hat{\beta}^2(x) - \tilde{\beta}^2(x) = B_1 + B_2 + B_3$, where

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

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

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

By the same arguments as in the proof of Lemma A.3, $B_1 = O_P(\vartheta_N/b)$. The numerator of B_2 is bounded by

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

where we have used the mean-value theorem, the same arguments as those for eq. (B.14), $\frac{1}{T}\sum_{j=1}^{N-1} \mathcal{K}_b^*(\sigma^2_{\tau_j} - x)(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j})^2 = O_P(1)$, and $\frac{1}{T}\sum_{j=1}^{N-1}(\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j})^2 = O_P(1)$. The denominator of B_2 can be shown to be $\pi(x) + o_P(1)$ by eqs. (B.13) and (B.18)). Thus, $B_2 = O_P(\vartheta_N/b)$. Finally,

$$\begin{aligned}
&\sup_{1\leq j\leq N-1} \left|(\tilde{\sigma}^2_{\tau_{j+1}} - \tilde{\sigma}^2_{\tau_j})^2 - (\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j})^2\right| \\
&\leq 2 \sup_{1\leq j\leq N} |\tilde{\sigma}^2_{\tau_j} - \sigma^2_{\tau_j}| \times \left[2 \sup_{1\leq j\leq N} |\tilde{\sigma}^2_{\tau_j} - \sigma^2_{\tau_j}| + 2 \sup_{1\leq j\leq N-1} |\sigma^2_{\tau_{j+1}} - \sigma^2_{\tau_j}|\right] \\
&= O_P\left(\vartheta_N^2\right) + O_P(\vartheta_N) \times O_{a.s.}(\delta^\gamma), \qquad \text{(B.20)}
\end{aligned}$$

where we have used eq. (4.1) and Lemma A.1. Note that $T^{-1}\sum_{j=1}^{N-1}\mathcal{K}_b(\tilde{\sigma}^2_{\tau_{j-1}}-x)=O_P\left(\delta^{-1}\right)$, and $\vartheta_N/\delta^{1-\gamma}\to 0$ implies $\vartheta_N/\delta^{\gamma}\to 0$. Thus, $\vartheta_N^2\delta^{-1}=\vartheta_N/\delta^{1-\gamma}\times\vartheta_N/\delta^{\gamma}\leq\vartheta_N/\delta^{1-\gamma}$, which, together with eq. (B.20), establishes $B_3=O_P\left(\vartheta_N/\delta^{1-\gamma}\right)$. ■

Proof of Lemma A.6. This follows from Aït-Sahalia and Park (2013, Theorem 4). ■

Proof of Lemma A.7. This follows along the same lines as in Sørensen (2009) or Yoshida (1992). ■

Proof of Lemma A.8. We only prove eq. (A.24) since the proofs of eqs. (A.25)-(A.26) are analogous. Write $\hat{R}_1\left(\theta_1,\tilde{\sigma}^2\right)-\hat{R}_1\left(\theta_1,\sigma^2\right)=\mathbf{a}\left(\theta_1\right)-2\mathbf{b}\left(\theta_1\right)$, where

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

We below derive the convergence rate of each term. By the mean-value theorem,

$$\left|\mathbf{a}\left(\theta_1\right)\right|\leq\frac{1}{N}\sum\nolimits_{j=1}^{N-1}C\left[1+|\sigma^2_{\tau_j}|^{2v_1}+|\sigma^2_{\tau_j}-\tilde{\sigma}^2_{\tau_j}|^{2v_1}\right]\times O_P\left(\vartheta_N\right)=O_P\left(\vartheta_N\right),$$

where we have used Assumption A-SDR(iii), $(A+B)^{v_1}\leq C_{v_1}(|A|^{2v_1}+|B|^{2v_1})$ for some constant $C_{v_1}(>0)$, and that $N^{-1}\sum_{j=1}^{N-1}|\sigma^2_{\tau_j}|^{2v_1}=O_p\left(1\right)$ which follows by $E[\left|\sigma^2_s\right|^{2v_1}]<\infty$, and $|\sigma^2_{\tau_j}-\tilde{\sigma}^2_{\tau_j}|=O_P\left(\vartheta_N\right)$ uniformly over j. Similarly

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

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

$$\begin{aligned}\sup_{1\leq j\leq N-1}|\tilde{\sigma}^2_{\tau_{j+1}}-\tilde{\sigma}^2_{\tau_j}| &\leq 2\sup_{1\leq j\leq N}|\tilde{\sigma}^2_{\tau_j}-\sigma^2_{\tau_j}|+\sup_{1\leq j\leq N-1}|\sigma^2_{\tau_{j+1}}-\sigma^2_{\tau_j}|\\ &= O_P\left(\vartheta_N\right)+O_{a.s.}\left(\delta^{\gamma}\right).\end{aligned}\tag{B.21}$$

■

Proof of Lemma A.9. We only prove eq. (A.27). Eqs. (A.28) and (A.29) can be shown analogously. Write $\hat{R}_2\left(\theta_2,\tilde{\sigma}^2\right)-\hat{R}_2\left(\theta_2,\sigma^2\right)=\mathbf{c}\left(\theta_2\right)-2\mathbf{d}\left(\theta_2\right)$, where

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

Similar to the analysis of $\mathbf{a}(\theta_1)$ in the proof of Lemma A.8, $\mathbf{c}(\theta_2) = O_P(\vartheta_N)$, while

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

where we have employed eqs. (B.21) and (B.20). ■

C Tables

	Drift		Diffusion	
	Infeasible 1-step	Feasible 2-step	Infeasible 1-step	Feasible 2-step
$\delta = 1/8$	0.6036	5.0377	0.0110	0.2038
	1.6332	0.6911	0.0147	0.0073
	2.2368	5.7288	0.0257	0.2111
$\delta = 1/4$	0.8503	0.6218	0.0258	0.0145
	1.5776	1.3627	0.0220	0.0201
	2.4279	1.9845	0.0478	0.0347
$\delta = 1/2$	1.3417	0.7511	0.0607	0.0304
	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
$\delta = 1/2$	1.2184	3.3333	0.0536	0.6269
	1.7283	2.2352	0.0393	0.0841
	2.9466	5.5685	0.0929	0.7110
$\delta = 1$	2.6679	0.2113	0.1250	0.0709
	1.0454	1.1474	0.0500	0.0614
	3.7132	1.3587	0.1749	0.1323
$\delta = 2$	5.7818	4.1220	0.2703	0.1309
	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

D Figures

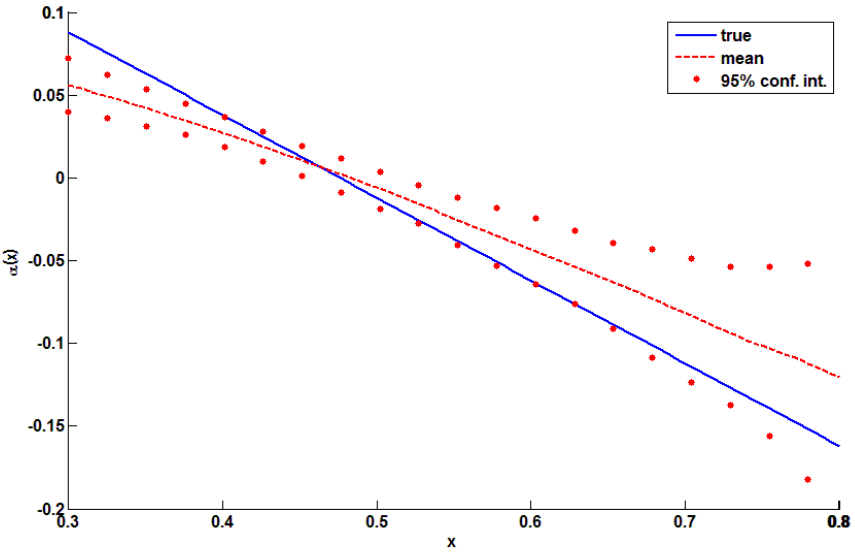

Figure 1: Infeasible 1-step estimator of $\alpha(x)$, $\Delta = 1/(24 \times 60)$ and $\delta = 1/4$.

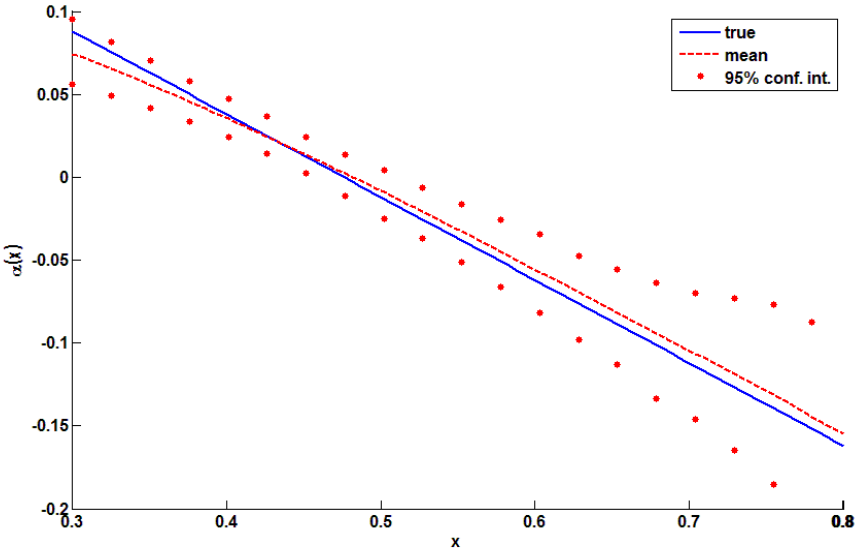

Figure 2: Feasible 2-step estimator of $\alpha(x)$, $\Delta = 1/(24 \times 60)$ and $\delta = 1/4$.

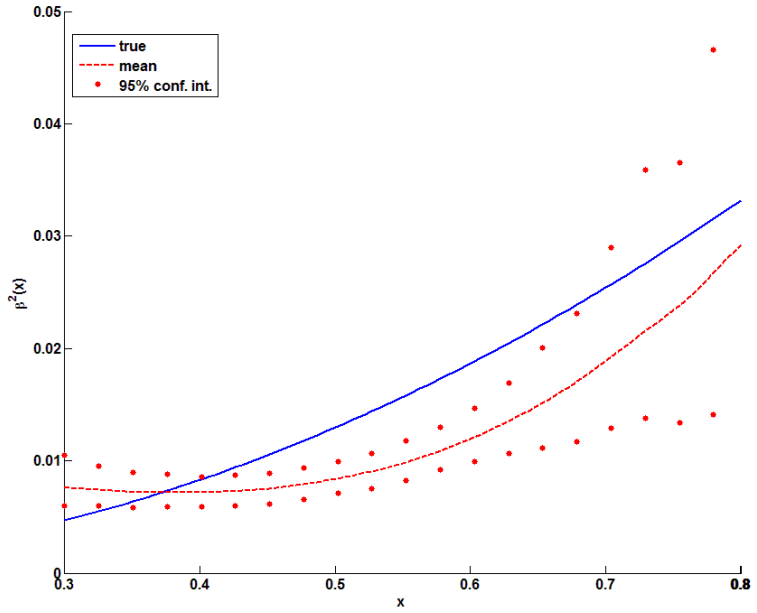

Figure 3: Infeasible 1-step estimator of $\beta^2(x)$, $\Delta = 1/(24 \times 60)$ and $\delta = 1/4$.

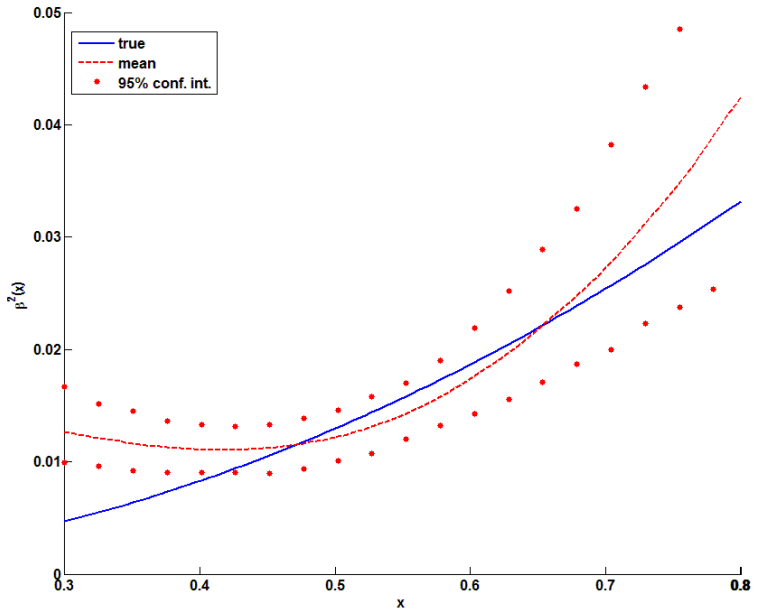

Figure 4: Feasible 2-step estimator of $\beta^2(x)$, $\Delta = 1/(24 \times 60)$ and $\delta = 1/4$.