

# Optimal Iterative Threshold-Kernel Estimation of Jump Diffusion Processes

José E. Figueroa-López \*

Cheng Li †

Jeffrey Nisen ‡

April 4, 2020

## Abstract

In this paper, we propose a new threshold-kernel jump-detection method for jump-diffusion processes, which iteratively applies thresholding and kernel methods in an approximately optimal way to achieve improved finite-sample performance. As in Figueroa-López and Nisen (2013), we use the expected number of jump misclassifications as the objective function to optimally select the threshold parameter of the jump detection scheme. We prove that the objective function is quasi-convex and obtain a new second-order infill approximation of the optimal threshold in closed form. The approximate optimal threshold depends not only on the spot volatility  $\sigma_t$ , but also the jump intensity and the value of the jump density at the origin. Estimation methods for these quantities are then developed, where the spot volatility is estimated by a kernel estimator with thresholding and the value of the jump density at the origin is estimated by a density kernel estimator applied to those increments deemed to contain jumps by the chosen thresholding criterion. Due to the interdependency between the model parameters and the approximate optimal estimators built to estimate them, a type of iterative fixed-point estimation algorithm is developed to implement them. Simulation studies for a prototypical stochastic volatility model, show that it is not only feasible to implement the higher-order local optimal threshold scheme but also that this is superior to those based only on the first order approximation and/or on average values of the parameters over the estimation time period.

## 1 Introduction

In this work, we study a jump diffusion process of the form

$$X_t := \int_0^t \gamma_u du + \int_0^t \sigma_u dW_u + \sum_{j=1}^{N_t} \zeta_j,$$

where  $W$  is a Wiener process,  $N$  is an independent Poisson process with local intensity  $\{\lambda_t\}_{t \geq 0}$ , and  $\{\zeta_j\}_{j \geq 1}$  are i.i.d. variables independent of  $W$  and  $N$ . With the presence of jumps, several statistical inference problems, including volatility estimation and jump detection, can be addressed by the thresholding approach developed by Mancini (2001, 2004, 2009). The basic idea is to introduce a threshold tuning parameter  $B$  so that whenever the absolute value of an increment  $\Delta X := X_{t_i} - X_{t_{i-1}}$  exceeds  $B$ , we conclude that an unusual event (aka a “jump”) has happened during the interval  $(t_{i-1}, t_i]$ , based on which we can then proceed to estimate the volatility and other parameters. Many works have been conducted to further extend the threshold method to various statistical inference problems. For an Itô semimartingale with finite or infinite jump activity, jump detection and integrated volatility estimation was studied by Mancini (2009) and Jacod (2007, 2008). We also refer to Corsi et al. (2010), Aït-Sahalia

---

\*Department of Mathematics and Statistics, Washington University in St. Louis, St. Louis, MO, 63130, USA.  
figueroa-lopez@wustl.edu.

†Citadel Securities, New York, NY, 10022, USA. Cheng.Li@citadelsecurities.com.

‡Quantitative Analytics, Barclays, New York, NY, 10019, USA. jeffrey.nisen@barclayscapital.com.

and Jacod (2009b,a, 2010), Cont and Mancini (2011), Figueroa-López (2012), Jing et al. (2012), and others for further applications of the threshold method.

One of the key issues that we have to address in order to have a good performance of the jump detection procedure is the selection of the threshold  $B$ . Ideally, we hope to select the best possible threshold under a suitable criterion. Such a problem was studied by Figueroa-López and Nisen (2013) using the expected number of jump misclassifications as the estimation loss function and, more recently, by Figueroa-López and Mancini (2018) using the mean-square error of the threshold realized quadratic variation. Under the assumption of zero drift, constant volatility  $\sigma$ , and constant jump intensity, Figueroa-López and Nisen (2013) showed that the first-order approximation of the optimal threshold is given by  $\sqrt{3\sigma^2h\log(1/h)}$  (cf. Theorems 4.2 and 4.3 therein), when  $h$ , the time span between observations, shrinks to 0 (i.e., infill or high-frequency asymptotics). Based on this result, Figueroa-López and Nisen (2013) proposed a method to estimate time-dependent deterministic volatilities and, by simulation, showed that its performance is good for smooth volatilities. In this work, we generalize this framework in three directions. We first prove that the loss function is quasi-convex and admits a global minimum in the more general case of non-homogeneous drift, volatility, and jump intensity. A simpler version of this result was stated without proof in Figueroa-López and Nisen (2013). We then proceed to obtain a second-order asymptotic approximation of the optimal localized threshold, in closed form, which depends on the spot volatility  $\sigma_t$ , the local jump intensity  $\lambda_t$ , and the value of the jump density at the origin. We find out that, as expected, if the spot volatility is high, then it is more preferable to have a larger threshold. However, when the jump intensity or the jump density at the origin is large, the possibility of having smaller jumps is higher, which favors a smaller threshold to detect such jumps. Although an explicit formula for the second-order approximation is derived, the method is not feasible unless we are able to estimate all the unknown parameters appearing in this formula: the spot volatility, the jump intensity, and the jump density at the origin. To this end, we apply kernel estimation techniques, as described below, to devise feasible plug-in type estimators for the optimal threshold.

Kernel estimation has a long history and has been applied to a large range of statistical problems. In our work, we use it to estimate the jump density at the origin. The problem we are facing differs from the usual density kernel estimation in several ways. Firstly, the data we have is contaminated by noise, and to make things even worse, part of the data may not contain any information at all about the density we want to estimate. Moreover, due to the usage of a threshold, the data we have is at best drawn from a truncated distribution and, the point at which we hope to estimate the density, is not even inside the support of the truncated data. Due to these reasons, we have to adjust the standard method of kernel density estimation and select the threshold appropriately so that we can get a satisfactory estimation of the jump density at the origin. It turns out that the optimal threshold that we should use in such a situation is larger than the one we use for optimal jump detection (see Section 2.4 for the intuition behind this).

Another quantity we have to estimate is the spot volatility, which can also be estimated by the kernel estimator. One earlier research on this topic is Foster and Nelson (1996), where a rolling window estimator is analyzed, which is similar to the idea of the kernel estimation with a uniform kernel. The kernel-based estimation of the spot volatility, *with general kernel*, was studied by Fan and Wang (2008), Kristensen (2010), Mancini et al. (2015) and, more recently, Figueroa-López and Li (2017). See also the excellent monographs of (Jacod and Protter, 2012, Ch. 13) and (Aït-Sahalia and Jacod, 2014, Ch. 8) for a general treatment of the problem of spot volatility estimation of Itô semimartingales via uniform kernels (though Remark 8.10 in Aït-Sahalia and Jacod (2014) also briefly mentions the case of a general kernel with support on  $[0, 1]$ ). One of the key issues related to kernel estimators of spot volatility is how to select the bandwidth. Kristensen (2010) proposed a leave-one-out cross-validation method, which is a general method, but suffers from the loss of accuracy and computational inefficiency. In this work, we adapt and extend the approach of Figueroa-López and Li (2017) by applying a threshold-kernel estimator of the spot volatility rather than just kernel estimation. The leading order terms of the MSE of the estimator are explicitly derived, based on which we propose a procedure for optimal bandwidth and kernel selection. The CLT of the estimation error is also given.

As explained above, the approximated optimal threshold depends on the spot volatility, jump intensity, and the value of the jump density at the origin, while the approximated optimal estimators of these three quantities depend on the threshold. Such an interdependency immediately suggests an iterative algorithm that starts with an initial guess of these parameters and gradually converges to a fixed point result. Due to the nature of the

threshold estimator, the result is purely determined by whether the absolute value of each data increment exceeds the threshold, so we can conclude convergence without any ambiguity based on whether each data increment is included by the threshold or not.

The rest of the paper is organized as follows. Section 2.1 introduces the framework and assumptions. In Section 2.2, we analyze the optimal threshold and obtain the second order approximation thereof. The bias and variance of the estimator are derived in Section 2.3. In Section 2.4, we consider the kernel estimation of the jump density at the origin. The threshold-kernel estimation of the spot volatility is studied in Section 3. The three estimators are then combined into an iterative algorithm presented in Section 4. Finally, the performance of the proposed methods are analyzed through several simulations in Section 5. Conclusions and some thoughts about future work are provided in Section 6. The proofs of the main results are deferred to an Appendix section.

## 2 The Optimal Threshold of TRV

In this section we extend the modelling framework and optimal thresholding results of Figueroa-López and Nisen (2013). Specifically, we will allow non-constant drift, volatility, and jump intensity, though we keep the jump density constant through time. In the first subsection, we introduce all the assumptions that we need for the optimal threshold results. However, we temporarily set the drift, volatility, and intensity to be deterministic, which would subsequently be relaxed when we discuss the kernel threshold estimation of spot volatility. All the results can be generalized to stochastic drift and volatility, and doubly stochastic Poisson process  $N$ , as long as we assume that the Brownian motion and jumps of the semimartingale are independent from all these processes, since we can always condition on the paths of the drift, the volatility, and the jump intensity of  $N$ . It is also important to point out that, though our results in this section are derived under the just mentioned independence assumption, our simulation experiments show that this is not essential as the proposed estimators perform well under prototypical stochastic volatility models with leverage.

### 2.1 The Framework and Assumptions

Throughout, we consider an Itô semimartingale of the form:

$$X_t := \left( \int_0^t \gamma_u du + \int_0^t \sigma_u dW_u \right) + \sum_{j=1}^{N_t} \zeta_j =: X_t^c + J_t, \quad (1)$$

where  $W = \{W_t\}_{t \geq 0}$  is a Wiener process,  $\{\zeta_j\}_{j \geq 1}$  are i.i.d. variables with density  $f$ ,  $N = \{N_t\}_{t \geq 0}$  is a non-homogeneous Poisson process with intensity function  $\{\lambda_t\}_{t \geq 0}$ , and the continuous component  $\{X_t^c\}_{t \geq 0}$  and jump component  $\{J_t\}_{t \geq 0}$  are independent. The processes  $\gamma$  and  $\sigma$  satisfy standard conditions for the integrals in (1) to be well-defined. In this section, we shall additionally assume the following conditions on  $\gamma$ ,  $\sigma$ , and  $\lambda$ <sup>1</sup>:

**Assumption 1.** *The functions  $\gamma : [0, \infty) \rightarrow \mathbb{R}$ ,  $\sigma : [0, \infty) \rightarrow \mathbb{R}^+$ , and  $\lambda : [0, \infty) \rightarrow \mathbb{R}^+$  are deterministic such that, for any given fixed  $t > 0$ ,*

$$\begin{aligned} \underline{\sigma}_t &:= \inf_{0 \leq s \leq t} \sigma_s > 0, & \bar{\sigma}_t &:= \sup_{0 \leq s \leq t} \sigma_s < \infty, \\ \underline{\gamma}_t &:= \inf_{0 \leq s \leq t} \gamma_s > 0, & \bar{\gamma}_t &:= \sup_{0 \leq s \leq t} \gamma_s < \infty, \\ \underline{\lambda}_t &:= \inf_{0 \leq s \leq t} \lambda_s > 0, & \bar{\lambda}_t &:= \sup_{0 \leq s \leq t} \lambda_s < \infty. \end{aligned} \quad (2)$$

Furthermore, we assume that  $t \mapsto \sigma_t$  is continuous.

The following notation will be needed:

$$\bar{\sigma}_{t,h}^2 := \frac{1}{h} \int_t^{t+h} \sigma_u^2 du, \quad \bar{\gamma}_{t,h} := \frac{1}{h} \int_t^{t+h} \gamma_u du, \quad \bar{\lambda}_{t,h} := \frac{1}{h} \int_t^{t+h} \lambda_u du. \quad (3)$$

---

<sup>1</sup>In Section 3, we will consider stochastic processes  $\gamma$  and  $\sigma$ .

Note that with these notations, our model assumptions imply that, for any  $t, h \geq 0$  and  $k \in \mathbb{N}$ ,

$$X_{t+h}^c - X_t^c =_D N(h\bar{\gamma}_{t,h}, h\bar{\sigma}_{t,h}^2), \quad \mathbb{P}(X_{t+h} - X_t \in dx | N_{t+h} - N_t = k) = \phi_{t,h} * f^{*k}(x)dx,$$

where  $\phi_{t,h}$  is the density of  $X_{t+h}^c - X_t^c$ , i.e.  $\phi_{t,h}(x) := \frac{1}{\bar{\sigma}_{t,h}\sqrt{h}}\phi\left(\frac{x-h\bar{\gamma}_{t,h}}{\bar{\sigma}_{t,h}\sqrt{h}}\right)$ . For these types of processes, the associated local characteristics are of the form  $(\gamma, \sigma, \nu)$ , where the density of the local Lévy measure is given by  $\nu_t(x) = \lambda_t f(x)$ .

**Assumption 2.** *The jump density  $f$  has the form*

$$f(x) = pf_+(x)\mathbf{1}_{[x \geq 0]} + qf_-(x)\mathbf{1}_{[x < 0]}, \quad (4)$$

where  $p \in [0, 1]$  and  $q := 1 - p$ , and  $f_+ : [0, \infty) \rightarrow [0, \infty)$  and  $f_- : (-\infty, 0] \rightarrow [0, \infty)$  are bounded functions such that  $\int_0^\infty f_+(x)dx = \int_{-\infty}^0 f_-(x)dx = 1$ . Furthermore, we assume that

$$f_\pm(0) = \lim_{x \rightarrow 0^\pm} f_\pm(x) \in (0, \infty).$$

The following notations will also be needed:

$$\mathcal{C}_0(f) := \lim_{\varepsilon \rightarrow 0^+} \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} f(x)dx = pf_+(0) + qf_-(0), \quad \mathcal{C}_d(f) := |pf_+(0) - qf_-(0)|, \quad \mathcal{C}_m(f) := \min\{f_+(0), f_-(0)\}. \quad (5)$$

Note that  $\mathcal{C}_0(f) = f(0)$  and  $\mathcal{C}_d(f) = 0$  if  $f$  is continuous at the origin. For some results, we also need the following assumption:

**Assumption 3.**  *$f_+ \in C^1([0, b])$ ,  $f_- \in C^1((a, 0])$ , for some  $a \in (-\infty, 0)$ ,  $b \in (0, \infty)$  and  $f'_\pm(0) := \lim_{x \rightarrow 0^\pm} f'_\pm(x)$  exists.*

Throughout, we assume that we observe the process  $X$  at evenly spaced times,

$$t_i := ih_n, \quad i = 0, \dots, n, \quad (6)$$

where  $h_n$  is the time span between observations and  $T := T_n := nh_n$  is the time horizon. We will also use  $\Delta_i^n X := X_{t_i} - X_{t_{i-1}}$  to denote the increment of the underlying process over  $[t_{i-1}, t_i]$ , and when no ambiguity can be brought, we will drop the superscript  $n$ . Finally, we introduce the jump detection procedure we consider in this work. We first specify a vector of thresholds  $[\mathbf{B}]_T^n = (B_1^n, \dots, B_n^n)$ , where we often drop the superscript  $n$  when no confusion can be generated. Given  $[\mathbf{B}]_T^n$ , we would conclude that a jump had occurred during  $[t_{i-1}, t_i]$  whenever  $|\Delta_i X| > B_i$ . As a byproduct of this jump detection criterion, we can then devise the following natural estimators of  $N_T$ ,  $J_T$ , and the integrated variance  $IV_T := \int_0^T \sigma_s^2 ds$ :

$$\widehat{N}_T = \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i X| > B_i\}}, \quad \widehat{J}_T = \sum_{i=1}^n (\Delta_i X) \mathbf{1}_{\{|\Delta_i X| > B_i\}}, \quad \widehat{IV}_T = TRV(X)[\mathbf{B}]_T^n = \sum_{i=1}^n (\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \leq B_i\}}. \quad (7)$$

These estimators were first studied in Mancini (2001), Mancini (2004). The estimator  $\widehat{IV}_T$  has extensively been studied in the literature and is commonly called the truncated or thresholded realized quadratic variation (TRV) of  $X$ .

## 2.2 Optimal Threshold and Its Approximation

In this subsection, we formulate the problem of optimal threshold selection. We adopt the approach in Figueroa-López and Nisen (2013), which we now briefly review for completeness. We seek to find a threshold  $[\mathbf{B}]_T = (B_1, \dots, B_n) \in \mathbb{R}_+^n$  to minimize the loss function:

$$L([\mathbf{B}]_T) := \mathbb{E} \left( \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i X| > B_i, \Delta_i N = 0\}} + \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i X| \leq B_i, \Delta_i N \neq 0\}} \right). \quad (8)$$

The above loss function represents the expected number of “jump” mis-classifications (i.e., subintervals erroneously classified as having jumps when in fact they do not, or not having jumps when in fact they do). The previous formulation gives the same weight to both types of error, while a more general loss function is given by:

$$L([\mathbf{B}]_T; w) := \mathbb{E} \left( \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i X| > B_i, \Delta_i N = 0\}} + w \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i X| \leq B_i, \Delta_i N \neq 0\}} \right). \quad (9)$$

For our purpose, (8) is enough, but in certain applications, (9) may be useful. For instance, it is more likely that market participants become more conservative when they erroneously identify a price change as an unusual event, i.e., a jump. In this case, one may prefer to take  $w < 1$ .

In both (8) and (9), the loss function is additive. Therefore, we can optimize each  $B_i$  separately. Indeed, we define the following loss function for given  $t$  and  $h$ :

$$L_{t,h}(B; w) := \mathbb{P}(|X_{t+h} - X_t| > B, N_{t+h} - N_t = 0) + w \mathbb{P}(|X_{t+h} - X_t| \leq B, N_{t+h} - N_t \neq 0). \quad (10)$$

If we were able to devise a method to find  $B^* = \operatorname{argmin}_B L_{t,h}(B; w)$  for any  $t$  and  $h$ , then, by setting  $t = t_{i-1}$  and  $h = t_i - t_{i-1}$ , we would be able to specify the whole optimal  $[\mathbf{B}]_T$ . Obviously, the first issue that we have to address is whether or not there is a global minimum point  $B^*$ . As it turns out, the loss function (10) is quasi-convex<sup>2</sup> in  $B$ , when  $h$  is small enough. This property was established in Figueroa-López and Nisen (2013) for a *driftless* Lévy processes (i.e.,  $\gamma \equiv 0$  and  $\sigma$  and  $\lambda$  are constants). Nonzero drifts create some nontrivial subtleties that are resolved in the following theorem, which was stated without proof in Figueroa-López and Nisen (2013).

**Theorem 2.1 (Uniform Quasi-Convexity of the Loss Functions).** *Assume that we have model (1), and Assumptions 1-3 are satisfied. Then, for any fixed  $T > 0$ , there exists  $h_0 := h_0(T) > 0$ , such that, for all  $t \in [0, T]$ ,  $h \in (0, h_0]$ , and  $w > 0$ , the function  $L_{t,h}(B; w)$  is **quasi-convex** in  $B$ , and possesses a unique global minimum point  $B_{t,h}^*$ .*

We proceed to give a fixed-point formulation of the optimal threshold  $B_{t,h}^*$ , which in turn enables us to find a second-order asymptotic expansion for  $B_{t,h}^*$  in a high-frequency asymptotic regime ( $h \rightarrow 0$ ). This characterization will equip us with the theoretical basis for developing feasible estimation algorithms later. In what follows, we focus on the case of  $w = 1$  and for easiness of notation, we drop the variable  $w$  in  $L_{t,h}(B; w)$ .

**Theorem 2.2 (Characterizations of the Optimal Threshold).** *Assume that we have model (1), and Assumptions 1-3 are satisfied. For each fixed  $T > 0$ , there exists  $h_0 := h_0(T) > 0$  such that, for any  $t \in [0, T]$  and  $h \in (0, h_0)$ , the optimal threshold  $B_{t,h}^*$ , based on the increment  $X_{t+h} - X_t$ , is such that,*

$$B_{t,h}^* = h \bar{\gamma}_{t,h} + \sqrt{2h\bar{\sigma}_{t,h}^2} \left[ \ln \left( 1 + \exp \left( \frac{-2B_{t,h}^* \bar{\gamma}_{t,h}}{\bar{\sigma}_{t,h}^2} \right) \right) - \ln \left( \sqrt{2\pi h \bar{\sigma}_{t,h}^2} \sum_{k=1}^{\infty} \frac{(h \bar{\lambda}_{t,h})^k}{k!} [\phi_{t,h} * f^{*k}(B_{t,h}^*) + \phi_{t,h} * f^{*k}(-B_{t,h}^*)] \right) \right]^{1/2}. \quad (11)$$

Furthermore, as  $h \rightarrow 0$ , we have the asymptotics:

$$B_{t,h}^* = \sqrt{h} \bar{\sigma}_{t,h} \left[ 3 \log(1/h) - 2 \log \left( \sqrt{2\pi} \mathcal{C}_0(f) \bar{\sigma}_{t,h} \bar{\lambda}_{t,h} \right) \right]^{1/2} + o(h^{1/2+\alpha}), \quad (12)$$

for any  $\alpha \in (0, 1/2)$ . If, furthermore,  $\sigma_t^2, \lambda_t \in C^1((0, T))$  and continuous on  $[0, T]$ , then the asymptotics in (12) remains true if we replace  $\bar{\sigma}_{t,h}$  and  $\bar{\lambda}_{t,h}$  with  $\sigma_t$  and  $\lambda_t$ , respectively.

**Remark 2.3.** *The last assertion of Theorem 2.2 remains true if  $t \rightarrow \sigma_t^2$  is Hölder continuous for any exponent  $\chi \in (0, 1/2)$ . In particular, this is the case for any volatility model driven by a Brownian motion (see (Revuz and Yor, 1998, Ch. V, Exercise 1.20)). Recall that if a function is Hölder continuous with exponent  $\chi$ , then it is Hölder continuous for any exponent  $\chi' < 1/2$ .*

<sup>2</sup>A mapping  $g : D \rightarrow \mathbb{R}$ , for convex  $D$ , is quasi-convex if for any  $\lambda \in [0, 1]$  and  $x, y \in D$ ,  $g(x\lambda + y(1 - \lambda)) \leq \max\{g(x), g(y)\}$ .

The previous result extends the first-order approximation  $\sqrt{3\sigma_{t,h}^2 h \log(1/h)}$  of Figueroa-López and Nisen (2013), whose remainder is just of order  $O(h^{1/2} \log^{-1/2}(1/h))$ . However, with the second order approximation, the remainder is  $o(h^{1-\epsilon})$  for any  $\epsilon \in (1/2, 1)$ . It is convenient to introduce the following notations for the first- and second-order optimal threshold approximations, respectively:

$$B_{t,h}^{*1} = \bar{\sigma}_{t,h} [3h \log(1/h)]^{1/2}, \quad B_{t,h}^{*2} = \sqrt{h} \bar{\sigma}_{t,h} \left[ 3 \log(1/h) - 2 \log \left( \sqrt{2\pi} \mathcal{C}_0(f) \bar{\sigma}_{t,h} \bar{\lambda}_{t,h} \right) \right]^{1/2}. \quad (13)$$

These tell us that, in a high-frequency sampling setting, the single most important parameter to determine a suitable threshold level  $B$  is the spot volatility  $\sigma_t$ , followed by the parameter  $\nu_t(0) := \lambda_t \mathcal{C}_0(f)$ , which broadly determines the likelihood of a small jump occurrence around time  $t$ . It is interesting to note that the optimal threshold  $B_{t,h}^{*2}$  can differ substantially from  $B_{t,h}^{*1}$  when  $\sigma_t \lambda_t \mathcal{C}_0(f)$  is large. This is intuitive since, for instance, if  $\sigma_t$  and  $\mathcal{C}_0(f)$  are fixed, as the jump rate  $\lambda_t$  increases, the optimal threshold should decrease in order to account for an increase in the appearance of “small” jumps. If the threshold is not adjusted, there would be more “false-negatives”, i.e., missed jumps. On the other hand, as  $\lambda_t$  decreases, the optimal threshold should be larger in order to offset an increment in the likelihood of false-positives (namely, wrongly concluding the occurrence of a jump during the small interval  $[t, t+h]$ ). Similarly, for fixed  $\sigma_t$  and  $\lambda_t$  as the likelihood for small jumps, approximately parameterized by  $\mathcal{C}_0(f)$ , increases (decreases) the optimal threshold decreases (increases) accordingly.

Although we have proved the asymptotic properties of (13), these optimal thresholds are not yet feasible, since we still need to estimate the spot volatility  $\sigma_t^2$ , jump intensity  $\lambda_t$ , and the mass concentration of the jump density at the origin,  $\mathcal{C}_0(f)$ . We will introduce estimators to these quantities in Subsection 2.4 and Section 3, respectively.

**Remark 2.4.** *Although the criterion (10) provides a reasonable approach for threshold selection, there is no guarantee that the resulting optimal threshold is the one that minimizes the mean-square error of the truncated realized quadratic variation  $\widehat{IV}_T$  introduced in (7). We refer to Figueroa-López and Mancini (2018) for some results regarding the latter problem.*

### 2.3 Bias and Variance

We conclude with the following asymptotic result of the estimation error of the TRV, which generalizes a result of Figueroa-López and Nisen (2019) to non-homogeneous drift, volatility, and jump intensities. As usual, the notation  $a_h \sim b_h$ , as  $h \rightarrow 0$ , means that  $\lim_{h \rightarrow 0} a_h/b_h = 1$ .

**Proposition 2.5.** *Suppose that the assumptions of Theorem 2.2 are enforced and that  $\mathbf{B} = (B_n)_{n \geq 1}$  is set to be  $B_n^{*1} = \sqrt{3\sigma_{t_i}^2 h_n \log(1/h_n)}$ . Then, as  $n \rightarrow \infty$ ,*

$$\mathbb{E}[TRV(X)[\mathbf{B}]_T^n] - \int_0^T \sigma_s^2 ds \sim h_n \int_0^T (\gamma_s^2 - \lambda_s \sigma_s^2) ds, \quad \text{Var}(TRV(X)[\mathbf{B}]_T^n) \sim 2h_n \int_0^T \sigma_s^4 ds.$$

Furthermore, the asymptotic behavior above also holds with any threshold sequence of the form

$$\check{B}_{n,i} = \sqrt{c_{n,i} \sigma_{t_i}^2 h_n \log(1/h_n)} + o(\sqrt{h_n \log(1/h_n)}),$$

provided that  $c := \liminf_{n \rightarrow \infty} \inf_i c_{n,i} \in (2, \infty)$ .

*Proof.* Let us write  $B_{n,i}$  of the form  $\sqrt{3\sigma_{t_i}^2 h_n \log(1/h_n)}$ . The bias of the TRV estimator can be decomposed as the following:

$$\begin{aligned} & TRV(X)[\mathbf{B}]_{T_n}^n - \int_0^T \sigma_s^2 ds \\ &= \sum_{i=1}^n \left( |\Delta_i^n X|^2 \mathbf{1}_{[\Delta_i^n N=0]} - h_n \bar{\sigma}_{t_{i-1}, h_n}^2 \right) + \sum_{i=1}^n |\Delta_i^n X|^2 \mathbf{1}_{[\Delta_i^n X \leq B_n, \Delta_i^n N \neq 0]} - \sum_{i=1}^n |\Delta_i^n X|^2 \mathbf{1}_{[\Delta_i^n X > B_n, \Delta_i^n N=0]}. \end{aligned} \quad (14)$$

Using Lemmas C.1 and C.2 in Figueroa-López and Nisen (2019) as well as Assumption 1, for any  $0 < \epsilon < 1/2$ , we have:

$$\mathbb{E} [|\Delta_i^n X|^2 \mathbf{1}_{[|\Delta_i^n X| \leq B_{n,i}, \Delta_i^n N \neq 0]}] = O(B_{n,i}^3 h_n) = O(h_n^{\frac{5}{2}} [\log(1/h_n)]^{\frac{3}{2}}), \quad (15)$$

$$\mathbb{E} [|\Delta_i^n X|^2 \mathbf{1}_{[|\Delta_i^n X| > B_{n,i}, \Delta_i^n N = 0]}] = O(\sqrt{h_n} B_{n,i} \phi(B_{n,i}/\bar{\sigma}_{t_i, h_n} \sqrt{h_n})) = O\left(h_n^{\frac{5}{2}-\epsilon} [\log(1/h_n)]^{\frac{1}{2}}\right), \quad (16)$$

where the  $O(\cdot)$  terms are uniform in  $i$ . These would imply that the second and third terms of (14) are of orders  $O_P(h^{3/2} [\log(1/h)]^{3/2})$  and  $O_P(h^{3/2-\epsilon} [\log(1/h)]^{1/2})$ , respectively. Both of these terms are then  $o(h_n)$ . For the first term therein, note that

$$\begin{aligned} \mathbb{E} [|\Delta_i^n X|^2 \mathbf{1}_{[\Delta_i^n N = 0]} - h_n \bar{\sigma}_{t_{i-1}, h_n}^2] &= \mathbb{P}(\Delta_i^n N \neq 0) h_n \bar{\sigma}_{t_{i-1}, h_n}^2 + \mathbb{P}(\Delta_i^n N = 0) h_n^2 \bar{\gamma}_{t_{i-1}, h_n}^2 \\ &= h_n^2 (\bar{\gamma}_{t_{i-1}, h_n}^2 - \bar{\lambda}_{t_{i-1}, h_n} \bar{\sigma}_{t_{i-1}, h_n}^2) + O(h_n^3), \\ \mathbb{E} \left[ \left( |\Delta_i^n X|^2 \mathbf{1}_{[\Delta_i^n N = 0]} - h_n \bar{\sigma}_{t_{i-1}, h_n}^2 \right)^2 \right] &= \mathbb{P}(\Delta_i^n N \neq 0) h_n^2 \bar{\sigma}_{t_{i-1}, h_n}^4 + \mathbb{P}(\Delta_i^n N = 0) 2h_n^2 \bar{\sigma}_{t_{i-1}, h_n}^4 \\ &= 2h_n^2 \bar{\sigma}_{t_{i-1}, h_n}^4 + O(h_n^3). \end{aligned}$$

Calculating the summation of the above and noticing the independence of different terms, we conclude the first part of the desired result. For  $\check{B}_{n,i}$ , the term (16) will instead be of order  $O_P(h^{1+c/2-\epsilon} [\log(1/h)]^{1/2})$ . Therefore, as long as  $c > 2$ , the asymptotic behavior does not change. This proves the second part of the desired result.  $\square$

**Remark 2.6.** *The motivation for considering the threshold  $\check{B}_{n,i}$  in Proposition 2.5 comes from the fact that the true value of  $\sigma^2$  is not available and, in practice, we have to use an estimate  $\hat{\sigma}^2$  of it. Suppose we have an estimator of  $\sigma_{t_i}^2$  denoted by  $\hat{\sigma}_{t_i}^2$ , and we use the corresponding estimated threshold  $\hat{B}_n^{*1} = \sqrt{3\hat{\sigma}_{t_i}^2 h_n \log(1/h_n)}$ . The second part of Proposition 2.5 tells us that if, for instance, the estimator is such that  $\liminf_{n \rightarrow \infty} \hat{\sigma}_{t_i}^2 / \sigma_{t_i}^2 = c > 2/3$ , we would have  $\hat{B}_n^{*1} = \sqrt{3\hat{\sigma}_{t_i}^2 h_n \log(1/h_n)} \geq \sqrt{3(c-\epsilon)\sigma_{t_i}^2 h_n \log(1/h_n)}$ , for  $n$  large enough and  $\epsilon \in (0, c-2/3)$ . This will result in an estimator such that the asymptotics of the expectation and variance of Proposition 2.5 hold.*

## 2.4 A Threshold-Kernel Estimation of the Jump Density at 0

In this section, we investigate the estimation of the jump density at the origin, which is needed in order to implement the second order optimal threshold  $B_{t,h}^{*2}$  given by (13). We propose a method based on kernel estimators. For a related method, but for a more general class of Itô semimartingales, see Ueltzhöfer (2013). The main difference between the method proposed below and the one proposed in that paper is the thresholding technique.

We impose the following regularity conditions, which in particular imply that  $\mathcal{C}_0(f) = f(0)$ .

**Assumption 4.**  $f \in C^2([a, b])$  for some  $a < 0 < b$ . Also,  $f(0) \neq 0$  and  $f''(0) \neq 0$ .

**Remark 2.7.** *It is possible to relax the previous assumption. For instance, if the density  $f$  merely satisfies Assumption 2, the estimation of  $f(0^+)$  and  $f(0^-)$  would have to be done separately using one-sided kernel estimators. The basic idea is the same as what we present below, but the convergence rate and the choice of bandwidth will be different.*

As mentioned above, we wish to construct a consistent estimator for  $\mathcal{C}_0(f) = f(0)$ , which is not feasible during a fixed time interval  $[0, T]$ . Hence, in this part, we consider a high-frequency/long-run sampling setting, where simultaneously

$$h_n = t_i - t_{i-1} \rightarrow 0, \quad T_n = t_n \rightarrow \infty,$$

as  $n \rightarrow \infty$ . Throughout, we also assume that  $\gamma$ ,  $\sigma$ , and  $\lambda$  are constant so that the distribution of  $\Delta_i X$  does not depend on  $i$ .

In the spirit of threshold estimation, the basic idea is to treat the “large” increments  $\Delta_i X$ , whose absolute values exceed an appropriate threshold, as proxies of the process’ jumps. These large increments can then be plugged into

a standard kernel estimator of  $f(0)$ . Concretely, we consider the estimator:

$$2\hat{f}(0) := \frac{1}{|\{i : |\Delta_i X| > B\}|} \sum_{\{i : |\Delta_i X| > B\}} K_\delta(|\Delta_i X| - B), \quad (17)$$

under the convention that  $0/0 = 0$  in the case that  $\{i : |\Delta_i X| > B\} = \emptyset$ . As usual,  $K_\delta(x) := K(x/\delta)/\delta$ , where  $K : [0, \infty) \rightarrow [0, \infty)$  is a right-sided kernel function such that  $\int_0^\infty K(x)dx = 1$  and  $\delta$  is the bandwidth parameter. We also use  $|A|$  to denote the number of elements in a set  $A$ . We expect that the estimator (17) will have poor performance if  $|\{i : |\Delta_i X| > B\}|$  is small, but, since we assume that  $T \rightarrow \infty$  and  $f(x) \neq 0$  in a neighborhood of  $\{0\}$ , for large-enough  $n$ , we have  $P(\{|\Delta_i X| > B\} = \emptyset) \approx e^{-\lambda T} \rightarrow 0$ . For our implementation of (17) in the Monte Carlo studies of Section 5, we will set  $\hat{f}(0) = 0$  if  $|\{i : |\Delta_i X| > B\}| \leq 5$ , which simply makes the second order threshold to be the first order threshold.

In what follows,  $f^*$  stands for the density of  $|\Delta_i X|$ , which depends on  $n$ , while  $f_{|\Delta X| > B}^*$  stands for the density of  $|\Delta_i X|$  conditioning on  $|\Delta_i X| > B$ . To analyze the performance of the estimator (17) and choose a suitable thresholding level  $B$  and bandwidth  $\delta$ , we decompose the estimation error into the following two terms:

- (i)  $E_1 = \frac{1}{|\{|\Delta_i X| > B\}|} \sum_{|\Delta_i X| > B} K_\delta(|\Delta_i X| - B) - f_{|\Delta X| > B}^*(B)$ ,
- (ii)  $E_2 = f_{|\Delta X| > B}^*(B) - 2f(0)$ .

Next, we follow a “greedy” strategy to determine suitable values for the threshold  $B$  and bandwidth  $\delta$ . Specifically, we minimize  $E_2$  to obtain an “optimal” threshold  $B$ , and with that given, we minimize  $E_1$  to obtain an “optimal” bandwidth  $\delta$ . Minimizing  $E_1 + E_2$  directly will be a much more involved problem, and requires more assumptions. However, we believe solving such a problem does not significantly improve the performance of the proposed estimator. Therefore, we leave it as an open problem.

Minimizing  $E_1$  over  $\delta$  given  $B$  is closely related to the standard theory of kernel density estimation, so we can directly apply the general theory for such a problem. We only need to ensure that  $|\{|\Delta_i X| > B\}| \rightarrow \infty$ , which follows from Proposition 2.8 below with the additional assumption that  $T \rightarrow \infty$ . Two widely used methods are plug-in method and cross-validation, which both have pros and cons. These methods are beyond the scope of this paper and, for simplicity, we instead use the well-known Silverman’s (1986) rule of thumb for bandwidth selection:

$$\delta = 1.06L^{-1/5} \text{sd}, \quad (18)$$

where “sd” is the standard deviation of  $\{|\Delta_i X| : |\Delta_i X| > B\}$  and  $L$  is the number of observations, i.e.  $|\{|\Delta_i X| : |\Delta_i X| > B\}|$ . Such a rule of thumb works the best with Gaussian kernel function and Gaussian density function. However, the method is known to be robust for other kernel and density functions.

We now proceed to show that  $B^* = \sqrt{4h\sigma^2 \log(1/h)}$  minimizes the leading order terms of the second error  $E_2$ . The proof of the following two results are given in Appendix A.

**Proposition 2.8.** *Suppose that Assumption 4 is satisfied and  $\gamma$ ,  $\sigma$ , and  $\lambda$  are constant. Further assume that  $B \rightarrow 0$  and  $B/\sqrt{h} \rightarrow \infty$ . Then,  $E_2$  converges to 0 as  $h \rightarrow 0$  if and only if  $h^{-3/2} \exp\left(-\frac{B^2}{2h\sigma^2}\right) \rightarrow 0$ . Under this condition, we have*

$$E_2 = \frac{2}{\lambda\sqrt{2\pi h^3\sigma^2}} \exp\left(-\frac{B^2}{2h\sigma^2}\right) + 2f(0)B + o(B) + o(h^{-3/2}e^{-\frac{B^2}{2h\sigma^2}}). \quad (19)$$

Furthermore, if  $E_2$  converges to 0, then  $\mathbb{P}(|\Delta_i X| > B) = \lambda h + o(h)$ , as  $h \rightarrow 0$ .

In addition to providing us conditions for the error  $E_2$  to vanish, Proposition 2.8 implies that, in that case,  $\mathbb{E}[\{i : |\Delta_i X| > B\}] = \lambda T + o(T)$ , as  $T \rightarrow \infty$  and  $h \rightarrow 0$ . Therefore, the average sample size that can be used for the estimation of  $f(0)$  is approximately constant with respect to  $B$ . Heuristically, this suggests that the selection of  $B$  will not affect significantly the selection of  $\delta$  that minimizes  $E_1$ . We are now ready to obtain an approximate optimal threshold  $B$ , which minimizes the leading order terms of  $E_2$ .

**Corollary 2.9.** *The approximate optimal threshold  $\tilde{B}^*$  that minimizes the leading order term of  $E_2$  given by (19) is such that*

$$\tilde{B}^* = \sqrt{4h\sigma^2 \log(1/h)} + O(\sqrt{h \log \log(1/h)}), \quad (20)$$

It is interesting to notice that the ‘‘optimal’’ threshold here is not the same as the one identified in the previous section. Indeed, if we do use the optimal threshold  $B^{*1}$  or  $B^{*2}$  in (13),  $E_2$  would diverge. It is interesting and important to get some sense why the optimal thresholds differ from each other. Indeed, in the previous section, we optimize the expected number of jump misclassification. In that case, we are minimizing the sum of unconditional false positive (mistakenly claim a jump) and unconditional false negative (miss a jump). However, since the probability that a jump occurs is so small, proportional to the length of the time increments, the probability of having a false negative, by nature, cannot be too large. Therefore, by having the expected number of misclassification as the objective function, we would choose a threshold in favour of having a much smaller unconditional false positive rate. As it turns out, if we choose  $B^{*1}$  or  $B^{*2}$ , conditioning on  $|\Delta X| > B$ , the probability that no jump occurs is comparable to the probability that a jump occurs, both  $O(h)$ . That is, the conditional false negative rate does not vanish. Such a situation would minimize the expected number of misclassifications, but would not enable us to distinguish the distribution of the jump from the noise. Using  $\sqrt{4h\sigma^2 \log(1/h)}$ , on the other hand, makes the conditional false negative vanishing and, thus, enables us to get consistent estimation of jump density.

### 3 Threshold-Kernel Estimation of Spot Volatility

In this section, we consider the estimation of the spot volatility of a jump-diffusion process, which is needed to implement the approximate optimal threshold formula (12), but is also an important problem on its own. Unlike Section 2, here we also work with certain stochastic volatility models. The precise conditions are given below.

The idea of kernel estimation of spot volatility is to take a weighted average of the squared increments (see, e.g., Foster and Nelson (1996) and Fan and Wang (2008)):

$$\hat{\sigma}_\tau^2 := KW(\tau, n, \delta) := \sum_{i=1}^n K_\delta(t_{i-1} - \tau)(\Delta_i X)^2. \quad (21)$$

Here,  $K(\cdot)$  is a kernel function with  $\int K(x)dx = 1$ ,  $K_\delta(x) = K(x/\delta)/\delta$ , and  $\delta > 0$  is the bandwidth. However, when jumps do occur, the estimator above becomes inaccurate. A natural idea is to combine (21) with the threshold method. Concretely, given a threshold vector  $[\mathbf{B}]_T^n = (B_1^n, \dots, B_n^n)$ , we consider the local threshold-kernel estimator:

$$\hat{\sigma}_\tau^2 := TKW(\tau, n, \delta) := \sum_{i=1}^n K_\delta(t_{i-1} - \tau)(\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \leq B_i\}}. \quad (22)$$

In what follows we will investigate the properties of (22). In order to do this, we will have to deal with the randomness of the volatility, for which we extend some of the results in Figueroa-López and Li (2018). We will mention the assumptions on  $\{\sigma_t\}_{t \geq 0}$  and  $K$  in Subsection 3.1, and then discuss the asymptotic properties of (22) in subsequent subsections.

#### 3.1 Assumptions on the Volatility Process

The first assumptions are some non-leverage and boundedness conditions, which enable us to condition on the whole path of the volatility and drift and use estimates from Figueroa-López and Nisen (2019):

**Assumption 5.** *In (1),  $(\gamma, \sigma)$  are locally bounded càdlàg independent of the Brownian motion  $W$  and the jump component  $J$ . Furthermore, there exists a deterministic  $M_T < \infty$  for which  $\bar{\gamma}_T$  and  $\bar{\sigma}_T^2$  defined in (2) satisfy  $\bar{\gamma}_T < M_T$  and  $\bar{\sigma}_T^2 < M_T$ . The intensity  $\lambda$  is still assumed to be deterministic such that  $\underline{\lambda}_t := \inf_{0 \leq s \leq t} \lambda_s > 0$  and  $\bar{\lambda}_t := \sup_{0 \leq s \leq t} \lambda_s < \infty$ .*

We now introduce the key assumption on the volatility process.

**Assumption 6.** *Suppose that for  $\varpi > 0$  and certain functions  $L : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ ,  $C_\varpi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ , such that  $C_\varpi$  is not identically zero and*

$$C_\varpi(hr, hs) = h^\varpi C_\varpi(r, s), \quad \text{for } r, s \in \mathbb{R}, h \in \mathbb{R}_+, \quad (23)$$

the variance process  $V := \{V_t = \sigma_t^2 : t \geq 0\}$  satisfies

$$\mathbb{E}[(V_{t+r} - V_t)(V_{t+s} - V_t)] = L(t)C_\varpi(r, s) + o((r^2 + s^2)^{\varpi/2}), \quad r, s \rightarrow 0. \quad (24)$$

An additional assumption on the kernel function  $K$  is the following:

**Assumption 7.** Given  $\varpi > 0$  and  $C_\varpi$  as defined in Assumption 6, the kernel function  $K : \mathbb{R} \rightarrow \mathbb{R}$  satisfies the following conditions:

- (1)  $\int K(x)dx = 1$ ;
- (2)  $K$  is Lipschitz and piecewise  $C^1$  on its support  $(A, B)$ , where  $-\infty \leq A < 0 < B \leq \infty$ ;
- (3) (i)  $\int |K(x)| |x|^\varpi dx < \infty$ ; (ii)  $K(x)x^{\varpi+1} \rightarrow 0$ , as  $|x| \rightarrow \infty$ ; (iii)  $\int |K'(x)|dx < \infty$ , (iv)  $V_\infty^\infty(|K'|) < \infty$ , where  $V_\infty^\infty(\cdot)$  is the total variation;
- (4)  $\iint K(x)K(y)C_\varpi(x, y)dxdy > 0$ .

We refer to Figueroa-López and Li (2018) for more details on Assumptions 5, 6 and 7. We just mention here that Assumption 6 covers a wide range of frameworks such as deterministic and smooth volatility, Brownian motion and fractional Brownian motion driven volatility, etc. In the following subsection, we will establish asymptotic properties of (22) based on Assumption 5, 6 and 7.

### 3.2 Asymptotic Properties of Threshold-Kernel Estimator

Figueroa-López and Li (2018) proves the following result under Assumption 5, 6, and 7 (c.f. Section 3 therein):

$$\begin{aligned} & \mathbb{E} \left[ \left( \sum_{i=1}^n K_\delta(t_{i-1} - \tau)(\Delta_i X^c)^2 - \sigma_\tau^2 \right)^2 \right] \\ &= 2 \frac{h}{\delta} \mathbb{E}[\sigma_\tau^4] \int K^2(x)dx + \delta^\varpi L(\tau) \iint K(x)K(y)C_\varpi(x, y)dxdy + o\left(\frac{h}{\delta}\right) + o(\delta^\varpi), \end{aligned} \quad (25)$$

where  $X^c$  is the continuous part of  $X$  defined in (1). The key result to extend the theory of kernel estimators, as developed in Figueroa-López and Li (2018), to the threshold-kernel estimators (22) is the following.

**Proposition 3.1.** Suppose that Assumptions 2, 5, 6, and 7 are satisfied, and take a bandwidth sequence  $\delta_n$  such that  $h_n/\delta_n \rightarrow 0$ . Let  $B_i := B_{n,i}(c) := \sqrt{c\bar{\sigma}_{t_i, h}^2 h \log(1/h)} + o(\sqrt{h \log(1/h)})$ , with  $c > 0$ . Then, we have:

$$\bar{\mathcal{E}}_n := \sum_{i=1}^n K_\delta(t_{i-1} - \tau) [(\Delta_i X^c)^2 - (\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \leq B_i\}}] = O_P \left( \max\{h, h^{c/2} \log^{1/2}(1/h)\} \right). \quad (26)$$

Furthermore,

$$\mathbb{E}(\bar{\mathcal{E}}_n^2) = O\left(\frac{h^2}{\delta}\right) + O\left(\frac{h^{1+\frac{c}{2}}}{\delta} [\log(1/h)]^{\frac{3}{2}}\right) + O(h^c \log(1/h)). \quad (27)$$

*Proof.* Let  $\mathcal{E}_i := (\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \leq B_i\}} - (\Delta_i X^c)^2$  and observe that

$$\mathcal{E}_i = -(\Delta_i X^c)^2 \mathbf{1}_{\{|\Delta_i X| \neq 0\}} + (\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \leq B_i, \Delta_i N \neq 0\}} - (\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| > B_i, \Delta_i N = 0\}} =: \mathcal{E}_{i,1} + \mathcal{E}_{i,2} + \mathcal{E}_{i,3}. \quad (28)$$

Now, conditioning on the paths of  $\sigma$  and  $\gamma$  and applying Lemmas C.1-C.2 in Figueroa-López and Nisen (2019), the following holds:

$$\begin{aligned} & (\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \leq B_i, \Delta_i N \neq 0\}} = O_P(B_i^3 h) = O_P(h^{5/2} [\log(1/h)]^{3/2}), \\ & (\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| > B_i, \Delta_i N = 0\}} = O_P(\sqrt{h_n} B_{n,i} \phi(B_{n,i}/\bar{\sigma}_{t_i, h_n} \sqrt{h_n})) = O_P(h^{1+c/2} [\log(1/h)]^{1/2}), \\ & (\Delta_i X^c)^2 \mathbf{1}_{\{|\Delta_i X| \neq 0\}} = O_P(h^2). \end{aligned} \quad (29)$$

From Assumption 5, the above holds uniformly over  $1 \leq i \leq n$ . Therefore, by Assumption 7, we have:

$$\sum_{i=1}^n K_\delta(t_{i-1} - \tau) [(\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \leq B_i\}} - (\Delta_i X^c)^2] = O_P \left( \max\{h, h^{c/2} \log^{1/2}(1/h)\} \right).$$

For the second assertion of the theorem, first note that

$$\begin{aligned} \mathbb{E}(\bar{\mathcal{E}}_n) &= \sum_{i=1}^n K_\delta(t_{i-1} - \tau) \mathbb{E}[\mathcal{E}_{i,1} + \mathcal{E}_{i,2} + \mathcal{E}_{i,3}] \\ &= \sum_{i=1}^n K_\delta(t_{i-1} - \tau) \left[ O(h^2) + O\left(h^{\frac{5}{2}} [\log(1/h)]^{\frac{3}{2}}\right) + O\left(h^{1+\frac{c}{2}} [\log(1/h)]^{\frac{1}{2}}\right) \right] \\ &= O(h) + O\left(h^{\frac{c}{2}} [\log(1/h)]^{\frac{1}{2}}\right). \end{aligned}$$

Similarly,

$$\begin{aligned} \text{Var}(\bar{\mathcal{E}}_n) &= \sum_{i=1}^n K_\delta^2(t_{i-1} - \tau) \text{Var}((\Delta_i X^c)^2 - (\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \leq B_i\}}) \\ &\leq 4 \sum_{i=1}^n K_\delta^2(t_{i-1} - \tau) [\mathbb{E}(\mathcal{E}_{i,1}^2) + \mathbb{E}(\mathcal{E}_{i,2}^2) + \mathbb{E}(\mathcal{E}_{i,3}^2)] \\ &= \sum_{i=1}^n K_\delta^2(t_{i-1} - \tau) \left[ O(h^3) + O\left(h^{2+\frac{c}{2}} [\log(1/h)]^{\frac{3}{2}}\right) \right] \\ &= O\left(\frac{h^2}{\delta}\right) + O\left(\frac{h^{1+\frac{c}{2}}}{\delta} [\log(1/h)]^{\frac{3}{2}}\right). \end{aligned}$$

We then conclude the result.  $\square$

With Proposition 3.1, we get the following proposition, which characterizes the leading order terms of the MSE of the threshold-kernel estimator (22). This allows us to perform bandwidth and kernel function selection.

**Proposition 3.2.** *Assume that Assumptions 2, 5, 6, and 7 are satisfied, and take the threshold vector to be  $B_{n,i}(c) = \sqrt{c\bar{\sigma}_{t_i,h}^2 h \log(1/h)} + o(\sqrt{h \log(1/h)})$  for any  $c \in (\frac{\varpi}{\varpi+1}, \infty)$ . Then, we have that, for each  $\tau \in (0, T)$ ,*

$$\mathbb{E} \left[ (TKW(\tau, n, \delta) - \sigma_\tau^2)^2 \right] = 2 \frac{h}{\delta} \mathbb{E}[\sigma_\tau^4] \int K^2(x) dx + \delta^\varpi L(\tau) \iint K(x) K(y) C_\varpi(x, y) dx dy + o\left(\frac{h}{\delta}\right) + o(\delta^\varpi). \quad (30)$$

*Proof.* We consider the following decomposition:

$$\begin{aligned} TKW(\tau, n, \delta) - \sigma_\tau^2 &= \sum_{i=1}^n K_\delta(t_{i-1} - \tau) [(\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \leq B_i\}} - (\Delta_i X^c)^2] + \left[ \sum_{i=1}^n K_\delta(t_{i-1} - \tau) (\Delta_i X^c)^2 - \sigma_\tau^2 \right] \\ &=: (I) + (II). \end{aligned} \quad (31)$$

From (25), we have that the second moment of (II) above converges with rate  $O\left(\frac{h}{\delta}\right) + O(\delta^\varpi)$ . The optimal rate of (II) is given by  $h^{\varpi/(1+\varpi)}$  and is attained with  $\delta \sim h^{1/(\varpi+1)}$ . Therefore, by Proposition 3.1, as long as  $c > \varpi/(1+\varpi)$ , (I) is of higher order than (II), in which case, (I) will be either of  $o\left(\frac{h}{\delta}\right)$  or  $o(\delta^\varpi)$ . This completes the proof.  $\square$

**Remark 3.3.** *The leading order term of the MSE of (22) does not depend on the threshold. However, by selecting the optimal threshold or its approximations, we are able to optimize the sub-order part of the error, which enhances the performance of the estimator in practice. Also, since taking  $c \in (2, \infty)$  does not change the asymptotic rate of convergence, we have a certain degree of robustness of this method.*

With some further assumptions, we are also able to obtain the CLT of the threshold-kernel estimator. The proof of the following result is similar to that Proposition 3.2, but taking advantage of Theorems 6.1 and 6.2 in

Figueroa-López and Li (2018) which deal with the analogous results without jumps.

**Theorem 3.4.** *Assume that Assumption 1, 2, 5, 6 and 7 are satisfied, and take the threshold vector to be  $B_{n,i}(c) = \sqrt{c\bar{\sigma}_{t_i,h}^2 h \log(1/h)} + o(\sqrt{h \log(1/h)})$  for any  $c \in (\frac{\varpi}{\varpi+1}, \infty)$ . Then, for each  $\tau \in (0, T)$ ,*

$$\left(\frac{h}{\delta}\right)^{-1/2} \left[ TKW(\tau, n, \delta) - \int_0^T K_\delta(t - \tau) \sigma_t^2 dt \right] \rightarrow_D \delta_1 N(0, 1), \quad (32)$$

where  $\delta_1^2 = 2\sigma_\tau^4 \int K^2(x) dx$ . Furthermore, suppose that either one of the following conditions holds:

- (1)  $\{\sigma_t^2\}_{t \geq 0}$  is an Itô process given by  $\sigma_t^2 = \sigma_0^2 + \int_0^t f_s ds + \int_0^t g_s dB_s$ , where  $B$  is a Brownian motion independent of  $W$  and we further assume that  $\sup_{t \in [0, T]} \mathbb{E}[|f_t|] < \infty$ ,  $\sup_{t \in [0, T]} \mathbb{E}[g_t^2] < \infty$ , and  $\mathbb{E}[(g_{\tau+h} - g_\tau)^2] \rightarrow 0$  as  $h \rightarrow 0$ ;
- (2)  $\sigma_t^2 = f(t, Z_t)$ , for a deterministic function  $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  such that  $f \in C^{1,2}(\mathbb{R})$ , and a Gaussian process  $\{Z_t\}_{t \geq 0}$  satisfying Assumption 6 and some mild additional conditions<sup>3</sup>.

Then, on an extension  $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$  of the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , equipped with a standard normal variable  $\xi$  independent of  $\{\sigma_t\}_{t \geq 0}$ , we have, for each  $\tau \in (0, T)$ ,

$$\delta^{-\varpi/2} \left( \int_0^T K_\delta(t - \tau) (\sigma_t^2 - \sigma_\tau^2) dt \right) \rightarrow_D \delta_2 \xi, \quad (33)$$

where, under the condition (1) above,  $\delta_2^2 = g(\tau, \omega)^2 \iint K(x) K(y) C(x, y) dx dy$ , while, under the condition (2),  $\delta_2^2 = [f_2(\tau, Z_\tau)]^2 L^{(Z)}(\tau) \iint K(x) K(y) C_\varpi^{(Z)}(x, y) dx dy$ . Here,  $f_2(t, z) = \frac{\partial f}{\partial z}(t, z)$ .

It is interesting to realize the difference between the range of  $c$  allowed here and the one allowed for the integrated volatility. Indeed, for  $\varpi \in (0, \infty)$ , the range for spot volatility estimation is strictly larger than the range for the integrated volatility estimation. The reason is that the estimation of spot volatility is much less accurate than the integrated volatility. Therefore, we may conclude that even with a bad estimation of spot volatility, we are still able to get a threshold that is accurate enough for us to apply the threshold estimation and obtain another estimation of the spot volatility.

### 3.3 Bandwidth and Kernel Selection

With the leading order approximation we obtained from the previous subsection, we are now able to develop a feasible plug-in type bandwidth selection method. Furthermore, we can derive the optimal kernel function when the volatility is driven by Brownian motion. In this subsection, we describe all related results, which are direct consequences of Proposition 3.2, and are parallel to results given by Figueroa-López and Li (2018). We refer to Figueroa-López and Li (2018) for the details of the proofs.

The first result is the theoretical approximated optimal bandwidth, which can be obtained by taking the derivatives of the leading order terms in (30) with respect to the bandwidth  $\delta$ .

**Proposition 3.5.** *With the same assumptions as Proposition 3.2, the approximated optimal bandwidth, denoted by  $\delta_n^{a,opt}$ , which is defined to minimize the leading order term of MSE in (30), is given by*

$$\delta_n^{a,opt} = n^{-1/(\varpi+1)} \left[ \frac{2T\mathbb{E}[\sigma_\tau^4] \int K^2(x) dx}{\varpi L(\tau) \iint K(x) K(y) C_\varpi(x, y) dx dy} \right]^{1/(\varpi+1)}, \quad (34)$$

while the attained global minimum of the approximated MSE is given by

$$MSE_n^{a,opt} = n^{-\varpi/(1+\varpi)} \frac{1+\varpi}{\varpi} \left( 2T\mathbb{E}[\sigma_\tau^4] \int K^2(x) dx \right)^{\varpi/(1+\varpi)} \left( \varpi L(\tau) \iint K(x) K(y) C_\varpi(x, y) dx dy \right)^{1/(1+\varpi)}. \quad (35)$$

<sup>3</sup>We refer the reader to Figueroa-López and Li (2018) for more details. In Figueroa-López and Li (2018), we assume  $\sigma_t^2 = f(Z_t)$ , but it is actually trivial to generalize to the case that  $\sigma_t^2 = f(t, Z_t)$  for  $f \in C^{1,2}(\mathbb{R})$ .

As shown in Figueroa-López and Li (2018), the resulting bandwidth obtained by replacing  $\mathbb{E}[\sigma_\tau^4]$  and  $L(\tau)$  in the formula (34) with their integrated versions,  $\int_0^T \mathbb{E}[\sigma_\tau^4]d\tau$  and  $\int_0^T L(\tau)d\tau$ , is asymptotically equivalent to the optimal bandwidth that minimizes the integrated MSE,  $\int_0^T \mathbb{E}[(\hat{\sigma}_t^2 - \sigma_t^2)^2] dt$ . In the case of a volatility process driven by Brownian motion, as in the setup (1) of Theorem 3.4, Figueroa-López and Li (2018) showed that  $\varpi = 1$ ,  $C_1(x, y) = \min\{|x|, |y|\} \mathbf{1}_{xy \geq 0}$ , and  $L(t) = \mathbb{E}(g_t^2)$ , which leads to the formula:

$$\delta_n^{a, \text{opt}} = n^{-1/2} \left[ \frac{2T \mathbb{E}[\int_0^T \sigma_t^4 dt] \int K^2(x) dx}{\mathbb{E}[\int_0^T g_t^2 dt] \iint K(x) K(y) C_1(x, y) dx dy} \right]^{1/2}. \quad (36)$$

Furthermore, since, at best, we only have one realization of the path of  $\sigma$  and we are working with a nonparametric setting for  $\sigma$ , it is natural to use  $\int_0^T \sigma_t^4 dt$  and  $\int_0^T g_t^2 dt$  as proxies of  $\mathbb{E}[\int_0^T \sigma_\tau^4 d\tau]$  and  $\mathbb{E}[\int_0^T g_t^2 dt]$ , respectively. These considerations suggest the following bandwidth selection method:

$$\delta_n^{a, \text{opt}} = n^{-1/2} \left[ \frac{2T \int_0^T \sigma_t^4 dt \int K^2(x) dx}{\int_0^T g_t^2 dt \iint K(x) K(y) C_1(x, y) dx dy} \right]^{1/2}. \quad (37)$$

Alternatively, by virtue of the independence condition in Assumption 5, we can see (37) as an approximation of the optimal bandwidth that minimizes the conditional integrated MSE,  $\mathbb{E} \left[ \int_0^T (\hat{\sigma}_t^2 - \sigma_t^2)^2 dt | \sigma_s, \gamma_s : 0 \leq s \leq T \right]$ .

However, the bandwidth (37) is not yet feasible, since it depends on the unknown random quantities  $\int_0^T \sigma_t^4 dt$  and  $\int_0^T g_t^2 dt$ . A well-known estimator of  $\int_0^T \sigma_t^4 dt$  is the truncated realized quarticity, which is defined by  $\widehat{IQ} = (3h)^{-1} \sum_{i=1}^n (\Delta_i X)^4 \mathbf{1}_{\{|\Delta_i X| < B_{n,i}\}}$  (see Proposition 1 in Mancini (2009) for consistency). The estimation of  $\int_0^T g_t^2 dt$  is more involved. This quantity is sometimes called the integrated vol of vol (or vol vol for short) and is essentially the quadratic variation of the volatility process. Figueroa-López and Li (2018) introduced an estimator based on the Two-time Scale Realized Quadratic Variation introduced in Zhang et al. (2005). Concretely, let  $\hat{\sigma}_{l,t_i}^2$  and  $\hat{\sigma}_{r,t_i}^2$  be the left and right side estimator of  $\sigma_{t_i}^2$ , respectively, defined as the following:

$$\hat{\sigma}_{l,t_i}^2 = \frac{\sum_{j>i} K_\delta(t_{j-1} - t_i)(\Delta_j^n X)^2 \mathbf{1}_{\{|\Delta_j^n X| \leq B_j\}}}{h \sum_{j>i} K_\delta(t_{j-1} - \tau) \mathbf{1}_{\{|\Delta_j^n X| \leq B_j\}}}, \quad \hat{\sigma}_{r,t_i}^2 = \frac{\sum_{j \leq i} K_\delta(t_{j-1} - t_i)(\Delta_j^n X)^2 \mathbf{1}_{\{|\Delta_j^n X| \leq B_j\}}}{h \sum_{j \leq i} K_\delta(t_{j-1} - \tau) \mathbf{1}_{\{|\Delta_j^n X| \leq B_j\}}}. \quad (38)$$

Next, we define the following two finite differences:  $\Delta_i \hat{\sigma}^2 = \hat{\sigma}_{r,t_{i+1}}^2 - \hat{\sigma}_{l,t_i}^2$ ,  $\Delta_i^{(k)} \hat{\sigma}^2 = \hat{\sigma}_{r,t_{i+k}}^2 - \hat{\sigma}_{l,t_i}^2$ . Finally, we can construct the following estimator:

$$\widehat{IVV}_T^{(\text{tsrvv})} = \frac{1}{k} \sum_{i=b}^{n-k-b} (\Delta_i^{(k)} \hat{\sigma}^2)^2 - \frac{n-k+1}{nk} \sum_{i=b+k-1}^{n-k-b} (\Delta_i \hat{\sigma}^2)^2. \quad (39)$$

Here,  $b$  is a small enough integer, when compared to  $n$ . The purpose of introducing such a number  $b$  is to alleviate the boundary effect of the one sided estimators, since, for instance, it is expected that  $\hat{\sigma}_{l,t_i}^2$  will be more inaccurate as  $i$  gets smaller. The consistency of the TSRVV estimator can be proved by Proposition 3.2 and the corresponding results from Figueroa-López and Li (2018).

The final result that we will mention in this subsection is about the optimal kernel function. Indeed, as was proved in Figueroa-López and Li (2018), when the volatility is driven by Brownian motion, the optimal kernel function is given by the double exponential function.

**Theorem 3.6.** *With the same assumptions as Proposition 3.2 and assuming  $C_\varpi(r, s) = \min\{|r|, |s|\} \mathbf{1}_{\{rs > 0\}}$ , we have that the optimal kernel function that minimizes the approximated optimal MSE given by (35) is the double exponential kernel function:*

$$K^{opt}(x) = \frac{1}{2} e^{-|x|}, \quad x \in \mathbb{R}.$$

## 4 Full Implementation Scheme of The Threshold-Kernel Estimation

In this section, we propose a complete data-driven threshold-kernel estimation scheme. We consider several versions, depending on whether we treat the volatility to be constant or not and whether we use the first- or second-order approximation formula. One of our main interests is to investigate whether or not *local* and/or *second-order* thresholding can improve the performance of threshold estimation.

Let us recall that the key problem at hand is jump detection; i.e., we hope to determine whether  $\Delta_i N = 0$  or not. We are, of course, also interested in estimating the volatility, jump intensity, and jump density, but we are operating under the premise that effective jump detection leads to good estimation of the other model features. In Section 2.2, we introduced the expected number of jump misclassification as the objective function and obtained the theoretical first and second order infill approximations of the optimal threshold, respectively given by

$$B_i^{*1} = [3\sigma_i^2 h \log(1/h)]^{1/2}, \quad B_i^{*2} = \sqrt{h} \sigma_i \left[ 3 \log(1/h) - 2 \log \left( \sqrt{2\pi} \mathcal{C}_0(f) \sigma_i \lambda_i \right) \right]^{1/2}, \quad (40)$$

where, with certain abuse of notation, we denote  $\sigma_i^2 := \sigma_{t_i}^2$  and  $\lambda_i := \lambda_{t_i}$ . Although we have assumed that  $\mathcal{C}_0(f)$  remains constant as the time evolves, we do allow non-constant volatility  $\sigma_t$  and jump intensity  $\lambda_t$ .

Since estimating spot values is typically less accurate than estimating average values, a simple first approach to implement (40) is to substitute  $\sigma_i^2$  and  $\lambda_i$  by their average values,  $\bar{\sigma}^2 := \int_0^T \sigma_s^2 ds / T$  and  $\bar{\lambda} := \int_0^T \lambda_s ds / T$ , respectively. This simplification leads us to consider the following threshold sequences:

$$B_i^{c1} = [3\bar{\sigma}^2 h \log(1/h)]^{1/2}, \quad B_i^{c2} = \sqrt{h} \bar{\sigma} \left[ 3 \log(1/h) - 2 \log \left( \sqrt{2\pi} \mathcal{C}_0(f) \bar{\sigma} \bar{\lambda} \right) \right]^{1/2}, \quad (41)$$

where the superscript  $c$  above is used to denote “constant” volatility and jump intensity. In light of (7), natural estimates of  $\bar{\lambda}$  and  $\bar{\sigma}^2$  are given by

$$\hat{\lambda} = \frac{1}{T} \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i X| > B_i\}}, \quad \hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^n (\Delta_i X)^2 \mathbf{1}_{\{|\Delta_i X| \leq B_i\}}, \quad (42)$$

respectively. The estimator of  $\mathcal{C}_0(f) = f(0)$ , as developed in Section 2.4, is given by

$$\widehat{\mathcal{C}_0(f)} := \frac{1}{2|\{|\Delta_i X| > B_i\}|} \sum_{|\Delta X_i| > B_i} K_\delta(|\Delta_i X| - B_i), \quad (43)$$

where the bandwidth  $\delta$  is set according to Silverman’s rule of thumb (18) and, for the threshold  $B_i$ , we could use the same threshold as in (42) or an estimate of  $\tilde{B}_i = \sqrt{4h\sigma^2 \log(1/h)}$  as suggested in Corollary 2.9. In the algorithms below and in the simulations of Section 5, we use the former threshold. Putting all together, the Algorithms 1 and 2 below detail the implementation of the 1st and 2nd order constant thresholds (41). Algorithm 1 is the same as that proposed in Figueroa-López and Nisen (2013) and, because it generates a nonincreasing sequence of thresholds and volatility estimates, is guaranteed to finish in finitely many steps. See the end of this section for more information about the stopping criteria for Algorithm 2.

We now consider the implementation of the local or non-constant thresholds (40). First of all, since Theorem 2.2 establishes that  $\sigma_i^2$  has a much greater effect on the approximated optimal threshold than that of  $\lambda_i$ , we simplify the problem by estimating  $\lambda_i$  with  $\hat{\lambda}$  as defined in (42). The estimation of  $\sigma_i^2$ , per our discussion in Section 3, is given by the kernel estimator:

$$\hat{\sigma}_i^2 := \sum_{j=1}^n K_\delta(t_{j-1} - t_i) (\Delta_j X)^2 \mathbf{1}_{\{|\Delta_j X| \leq B_j\}}. \quad (44)$$

Above, we could try to calibrate the bandwidth  $\delta$  using an approach similar to that described in Section 3.3. However, for simplicity, in the simulations we set  $\delta = h_n^{1/2}$ , which, per (36), is rate optimal at first order. Based on the  $\hat{\sigma}_i^2$ ’s,  $\hat{\lambda}$ , and  $\widehat{\mathcal{C}_0(f)}$  as defined in (43), we can then compute estimates of the first and second order approximation

---

**Algorithm 1** Iterative (Constant) 1st-Order Threshold Kernel Algorithm

---

Calculate  $\hat{\sigma}_{Old}^2$  by (42) setting  $B_i = \infty$ ;  
 Initialize  $B_i^{c1} = [3\hat{\sigma}^2 h \log(1/h)]^{1/2}$ , for  $i = 1, \dots, n$ ;  
 Calculate  $\hat{\sigma}_{New}^2$  as in (42) with  $B_i$  replaced with  $B_i^{c1}$ ;  
**while**  $\hat{\sigma}_{New}^2 \neq \hat{\sigma}_{Old}^2$  **do**  
 $\hat{\sigma}_{Old}^2 = \hat{\sigma}_{New}^2$ ;  
 Update  $B_i^{c1} = [3\hat{\sigma}_{Old}^2 h \log(1/h)]^{1/2}$ , for  $i = 1, \dots, n$ ;  
 Calculate  $\hat{\sigma}_{New}^2$  as in (42) with  $B_i$  replaced with  $B_i^{c1}$ ;  
**end while**  
 Use final  $B_i^{c1}$  for jump detection;

---

**Algorithm 2** Iterative (Constant) 2nd-Order Threshold Kernel Algorithm

---

Calculate  $\hat{\sigma}^2$  by (42) setting  $B_i = \infty$ ;  
 Initialize  $B_i^{c2} = [3\hat{\sigma}^2 h \log(1/h)]^{1/2}$ , for  $i = 1, \dots, n$ ;  
**while** “Stopping Criteria” not satisfied **do**  
 Calculate  $\hat{\lambda}$  and  $\hat{\sigma}^2$  as in (42) with  $B_i$  replaced with  $B_i^{c2}$ ;  
 Estimate  $\widehat{\mathcal{C}_0(f)}$  by (43) with  $B_i = B_i^{c2}$ ;  
 Update  $B_i^{c2}$  by (41) with  $\bar{\sigma}^2 = \hat{\sigma}^2$ ,  $\bar{\lambda} = \hat{\lambda}$ , and  $C_0(f) = \widehat{\mathcal{C}_0(f)}$ , based on newly estimated parameters;  
**end while**  
 Use final  $B_i^{c2}$  for jump detection.

---

of the optimal thresholds as follows:

$$B_i^{n1} = [3\hat{\sigma}_i^2 h \log(1/h)]^{1/2}, \quad B_i^{n2} = \sqrt{h} \hat{\sigma}_i \left[ 3 \log(1/h) - 2 \log \left( \sqrt{2\pi} \widehat{\mathcal{C}_0(f)} \hat{\sigma}_i \hat{\lambda} \right) \right]^{1/2}, \quad (45)$$

where above the superscript  $n$  stands for non-constant volatility estimation. Algorithm 3 below gives the details of the implementation of the non-constant thresholds (40). Therein, the initial threshold is taken as  $B_i^{n1} = [3\hat{\sigma}_0^2 h \log(1/h)]^{1/2}$ , where  $\hat{\sigma}_0^2$  is an initial estimate of  $\bar{\sigma}^2 := \int_0^T \sigma_s^2 ds / T$  such as those obtained from the previous Algorithms. In the simulations of Section 5, we take that from Algorithm 1. See also below for more details about the “stopping criteria” of the algorithm.

---

**Algorithm 3** Iterative Threshold Kernel Algorithm

---

Initialize  $B_i^{n1} = [3\hat{\sigma}_0^2 h \log(1/h)]^{1/2}$  (or  $B_i^{n2} = [3\hat{\sigma}_0^2 h \log(1/h)]^{1/2}$  when using 2nd order approx.), for  $i = 1, \dots, n$ ;  
**while** “Stopping Criteria” not satisfied **do**  
 Calculate  $\hat{\sigma}_i^2$  as in (44) with  $B_i$  replaced with  $B_i^{n1}$  (or  $B_i^{n2}$  when using 2nd order approximation);  
 Calculate  $\hat{\lambda}$  and  $\widehat{\mathcal{C}_0(f)}$  by (42)-(43) with  $B_i$  replaced with  $B_i^{n1}$  (or  $B_i^{n2}$  when using 2nd order approximation);  
 Update  $B_i^{n1}$  (or  $B_i^{n2}$  when using 2nd order approximation) by (45) based on newly estimated parameters;  
**end while**  
 Use  $B_i^{n1}$  (or  $B_i^{n2}$ ) as the final threshold value.

---

Note that, in (41) and (45),  $B^{c2}$ , and  $B^{n2}$  may not be well defined, under a finite sample setting. Indeed, for a fixed time period and a fixed sample size, it is possible to have  $3 \log(1/h) < 2 \log(\sqrt{2\pi} \widehat{\mathcal{C}_0(f)} \hat{\sigma}_i \hat{\lambda})$ , in which case the square root in (45) is not well defined. Of course, asymptotically this is never an issue since we only need to consider a small enough  $h$ . As to implementation, however, it is natural to use  $B^{*2}$  whenever  $3 \log(1/h) > 2 \log(\sqrt{2\pi} \widehat{\mathcal{C}_0(f)} \hat{\sigma}_i \hat{\lambda})$ , and use  $B^{*1}$ , otherwise.

We now briefly discuss some stopping criteria for the Algorithms 1 and 3. Typically, most iterative algorithms

are stopped when the updated value is “close” enough to the old value. However, for the threshold estimator, we note that there are only  $2^n$  possible threshold vectors after the initial set up. Therefore, there are only two possible situations for the “while” loop in Algorithm 3:

1. After a few iterations, the algorithm comes to a fixed threshold vector  $[\mathbf{B}_T]$ .
2. After a few iterations, the algorithm comes to a loop of threshold vectors given by  $[\mathbf{B}_T^1], \dots, [\mathbf{B}_T^k]$ .

As we will see at the end of Subsection 5.1, generally the threshold vector converges within 2 iterations.

## 5 Monte Carlo Study

In this section, we investigate the performance of our proposed methods. Specifically, in Section 5.1, we will compare the four different threshold methods given by (41) and (45) and detailed in Algorithms 1-3. In Section 5.2, we investigate the performance of the threshold-kernel estimation of the jump density at the origin.

Throughout, we consider the jump-diffusion model given by (1), with the continuous part  $\{X_t^c\}_{t \geq 0}$  following a Heston model:

$$\begin{aligned} dX_t^c &= \mu_t dt + \sqrt{V_t} dB_t, \\ dV_t &= \kappa(\theta - V_t)dt + \xi \sqrt{V_t} dW_t. \end{aligned} \tag{46}$$

Here,  $V_t = \sigma_t^2$  is the variance process. The parameters of (46) are selected according to the following setting also used in Zhang et al. (2005):

$$\kappa = 5, \quad \theta = 0.04, \quad \xi = 0.5, \quad \mu_t = 0.05 - V_t/2. \tag{47}$$

As to the initial values, we use  $X_0^c = 1$  and  $V_0 = \sigma_0^2 = 0.04$ . The unit of time in this study is 1 year and, thus, the parameter values above are annualized. Although the properties of the threshold-kernel estimators studied in this work were derived under a non-leverage setting (i.e.,  $\rho = 0$ , where  $\rho$  is the correlation between  $B_t$  and  $W_t$ ), we run simulations on both the non-leverage setting and a negative leverage setting ( $\rho = -0.5$ ) in order to check the robustness of the method against the leverage effect.

As to the jump component, we consider Merton type of jumps:

$$f_{normal}(x) = \frac{1}{\sqrt{2\pi\vartheta^2}} \exp\left(-\frac{x^2}{2\vartheta^2}\right). \tag{48}$$

The intensity of the jump component is set to be a constant value, i.e.,  $\lambda_t \equiv \lambda$  for all  $t \geq 0$ . For the values of  $\lambda$  and  $\vartheta$ , we consider the following scenarios:

1.  $\lambda = 50$  and  $\vartheta = 0.03$ , which gives an average annualized volatility of about  $\sqrt{0.04 + 50(0.03)^2} \approx 0.29$ ;
2.  $\lambda = 100$  and  $\vartheta = 0.03$ , which gives an average annualized volatility of about  $\sqrt{0.04 + 100(0.03)^2} \approx 0.36$ ;
3.  $\lambda = 200$  and  $\vartheta = 0.03$ , which gives an average annualized volatility of about  $\sqrt{0.04 + 200(0.03)^2} \approx 0.46$ ;
4.  $\lambda = 1000$  and  $\vartheta = 0.01$ , which gives an annualized volatility of about  $\sqrt{0.04 + 1000(0.01)^2} \approx 0.37$ .

The reason for choosing these  $\lambda$ ’s is to investigate how large levels of jump intensity can affect the performance of the estimators, while  $\vartheta$  is selected accordingly so that the annualized volatility is reasonable.

We assume that there are 252 trading days in a year and 6.5 trading hours in each day. We focus on 5-minute data, which is standard in the literature to avoid microstructure noise effects. Furthermore, the length of the data is set to be 1 month (21 trading days), 3 month (63 trading days), and 1/2 year.

### 5.1 Comparison of Different Thresholds

We now proceed to examine how the different “optimal” threshold approximation methods introduced in Section 4 affect the number of jump misclassifications. In Tables 1, we report the average total number of jump misclassifications corresponding to the four threshold approximation methods  $B^{c1}$ ,  $B^{c2}$ ,  $B^{n1}$  and  $B^{n2}$ , as well as an

#OfDays	#Obs./Hr	$\rho$	$\lambda$	sd( $f$ )	$\bar{\mathcal{L}}^{c1}$	$\bar{\mathcal{L}}^{c2}$	$\bar{\mathcal{L}}^{n1}$	$\bar{\mathcal{L}}^{n2}$	$\bar{\mathcal{L}}^{*2}$
21	12	0	50	0.03	0.848	0.864	0.835	0.795	0.678
21	12	-0.5	50	0.03	0.844	0.884	0.856	0.829	0.690
21	12	0	100	0.03	1.669	1.591	1.623	1.382	1.259
21	12	-0.5	100	0.03	1.628	1.643	1.584	1.381	1.272
21	12	0	200	0.03	3.384	2.967	3.318	2.603	2.529
21	12	-0.5	200	0.03	3.372	2.882	3.284	2.577	2.487
21	12	0	1000	0.01	51.301	32.673	49.700	31.087	30.218
21	12	-0.5	1000	0.01	51.547	32.937	49.895	31.361	30.480
63	12	0	50	0.03	2.660	4.217	2.531	2.174	2.098
63	12	-0.5	50	0.03	2.590	4.137	2.466	2.125	2.051
63	12	0	100	0.03	4.952	6.688	4.741	3.876	3.739
63	12	-0.5	100	0.03	4.914	6.822	4.737	3.937	3.820
63	12	0	200	0.03	10.195	11.518	9.842	7.651	7.491
63	12	-0.5	200	0.03	10.001	11.144	9.658	7.515	7.339
63	12	0	1000	0.01	148.661	107.325	143.339	89.477	87.434
63	12	-0.5	1000	0.01	149.923	107.895	144.979	90.393	88.293
126	12	-0.5	100	0.03	10.106	18.243	9.636	7.890	7.624
126	12	-0.5	200	0.03	20.129	27.433	19.353	15.036	14.605
126	12	-0.5	1000	0.01	298.656	241.770	285.045	177.588	173.745

Table 1: Average total number of jump mis-classifications for Normal Jumps based on 1000 Samples.

oracle threshold, where we use the second order approximation  $B_i^{*2}$  in (40) with all the true parameter values plugged in. In each case, we compute the average number of jump misclassifications:

$$\bar{\mathcal{L}}^a := \frac{1}{m} \sum_{j=1}^m \left( \sum_{i=1}^n \mathbf{1}_{\{|X_{t_i}^{(j)} - X_{t_{i-1}}^{(j)}| > B_{i,j}^a, N_{t_i}^{(j)} - N_{t_{i-1}}^{(j)} = 0\}} + \sum_{i=1}^n \mathbf{1}_{\{|X_{t_i}^{(j)} - X_{t_{i-1}}^{(j)}| \leq B_{i,j}^a, N_{t_i}^{(j)} - N_{t_{i-1}}^{(j)} \neq 0\}} \right), \quad (49)$$

where  $m$  is the number of simulations,  $X^{(j)}$  and  $N^{(j)}$  are the  $j$ th simulated paths of  $X$  and  $N$ , respectively, and  $a \in \{c1, c2, n1, n2, *2\}$ , depending on the used thresholding method. For the non-constant methods, we use an exponential kernel  $K(x) = e^{-|x|}/2$  to estimate the spot volatility, which, as shown in Theorem 3.6, is optimal. We ran only 4 iterations of the iterative algorithms described in Section 4. As shown below, this typically suffices to reach convergence.

The conclusion is that the jump detection method based on the second order approximation with non-constant volatility estimation (“n2” method) performs the best among all the four methods. Although, as it should be expected, this is slightly worse than the oracle one, it is remarkably close to the latter. Even for a relatively low value of  $\lambda = 50$ , which is typically hard to estimate  $\lambda$  and  $C_0(f)$  because of relatively few jumps, the 2nd order local method is still a bit better than those based on constant threshold. For instance, for a time horizon of 1 month, the n2 method only misses about 1 jump out of the expected 4 jumps during the month. The difference between the constant and local thresholds becomes more crucial as the intensity of jumps increases. For an intensity of 200, the method will only miss about 3 of the expected 16 jumps.

As mentioned above, the results of Table 1 were based on 4 iterations of the Algorithms of Section 4. To assess the convergence of the algorithm, in Table 2, we show the results of the average number of jump misclassifications  $\bar{\mathcal{L}}^a$ , as defined in (49), for each of the first 4 iterations of Algorithm 3 based on the 2nd order approximation. As it can be seen, convergence is typically reached after the 2nd iteration.

#OfDays	#Obs./Hr	$\rho$	$\lambda$	$sd(f)$	$\bar{\mathcal{L}}^{n2, \text{Iter1}}$	$\bar{\mathcal{L}}^{n2, \text{Iter2}}$	$\bar{\mathcal{L}}^{n2, \text{Iter3}}$	$\bar{\mathcal{L}}^{n2, \text{Iter4}}$
21	12	0	50	0.03	0.795	0.795	0.795	0.795
21	12	-0.5	50	0.03	0.830	0.829	0.829	0.829
21	12	0	100	0.03	1.383	1.382	1.382	1.382
21	12	-0.5	100	0.03	1.384	1.382	1.381	1.381
21	12	0	200	0.03	2.602	2.603	2.603	2.603
21	12	-0.5	200	0.03	2.586	2.577	2.577	2.577
21	12	0	1000	0.01	31.855	31.216	31.121	31.087
21	12	-0.5	1000	0.01	32.080	31.482	31.385	31.361
63	12	0	50	0.03	2.183	2.174	2.174	2.174
63	12	-0.5	50	0.03	2.126	2.125	2.125	2.125
63	12	0	100	0.03	3.875	3.874	3.876	3.876
63	12	-0.5	100	0.03	3.952	3.938	3.937	3.937
63	12	0	200	0.03	7.680	7.653	7.651	7.651
63	12	-0.5	200	0.03	7.544	7.517	7.515	7.515
63	12	0	1000	0.01	91.283	89.747	89.526	89.477
63	12	-0.5	1000	0.01	92.324	90.722	90.411	90.393
126	12	-0.5	100	0.03	7.929	7.892	7.889	7.890
126	12	-0.5	200	0.03	15.097	15.036	15.035	15.036
126	12	-0.5	1000	0.01	181.403	178.022	177.677	177.588

Table 2: Average total number of jump mis-classifications for the first 4 iterations of the nonhomogeneous Algorithm 3 based on 2nd order approximations. Again, we use 1000 Samples for all values of  $T$ .

## 5.2 Estimation of Jump Density at the Origin and Spot Volatility

We now study the performance of the kernel estimator of the jump density at the origin that we proposed in Section 2.4. Since we have already confirmed that the second order approximation of the optimal threshold with non-constant volatility estimation outperforms other thresholds, we will only consider this threshold in this and later subsections.

The results are shown in Table 3. These basically confirm what we expect that the performance of the estimator improves as the time-horizon and intensity become larger (for the same level of jump variance). It is hard to compare the performance of the estimators when  $\vartheta = 0.03$  to those when  $\vartheta = 0.01$  and  $\lambda = 1000$  because, though we expect more jumps in the latter case, those will also be much harder to detect since  $\vartheta$  is smaller. Finally, an interesting phenomenon is that we usually underestimate the jump density at the origin. This is acceptable for our purpose. Indeed, if we denote  $\widehat{B}^2$  as the estimated second order threshold, we generally have  $B^1 > \widehat{B}^2 > B^2$ . This is better than having  $\widehat{B}^2 < B^2$ , in which case we might suffer significantly from false positives (i.e., mis-classifying the increments of the continuous component as jumps).

We finally give some illustrations about the performance of the the kernel/threshold spot volatility estimator (44). We apply 4 iterations of the local Algorithm 3 based on the 2nd order approximation of the optimal threshold. In Figure 1, we show a prototypical realization of the variance process  $\{V_t\}_{t \geq 0}$  defined in (46) together with the estimated spot variance process resulting from the 1st iteration (red dotted), from the final iteration 4 (long-dashed blue), and from the oracle (green double-dashed), which uses  $B_i^{*1}$  in (40) with the true values of  $\sigma_i^2 = V_{t_i}$ ,  $C_0(f)$ , and  $\lambda$ . We take  $\lambda = 200$ ,  $\vartheta = 0.03$ ,  $T = 6$  months, and  $h = 5$  minutes. The three spot variance estimates are close to each other and are able to fit well the overall level of the volatility through time. The Sum Of Square Errors,

$$SSE = \sum_{i=1}^n (\hat{\sigma}_{t_i}^2 - \sigma_{t_i}^2)^2,$$

#OfDays	#Obs./Hr	$\rho$	$\lambda$	$\vartheta = \text{sd}(f)$	$f(0)$	$E(\hat{f}^{n^2}(0))$	$\text{sd}(\hat{f}^{n^2}(0))$	$\sqrt{MSE(\hat{f}^{n^2}(0))}$
21	12	0	100	0.03	13.30	8.6965	5.4797	7.1567
21	12	-0.5	100	0.03	13.30	8.8117	5.6950	7.2510
63	12	0	100	0.03	13.30	11.6005	2.1668	2.7537
63	12	-0.5	100	0.03	13.30	11.4145	2.2175	2.9107
126	12	-0.5	100	0.03	13.30	11.9558	1.6286	2.1116
21	12	0	200	0.03	13.30	11.2759	2.6362	3.3236
21	12	-0.5	200	0.03	13.30	11.1900	2.7161	3.4393
63	12	0	200	0.03	13.30	11.9714	1.6997	2.1582
63	12	-0.5	200	0.03	13.30	11.9234	1.6539	2.1518
126	12	-0.5	200	0.03	13.30	12.4776	1.3081	1.5451
21	12	0	1000	0.01	39.89	37.9363	4.5286	4.9321
21	12	-0.5	1000	0.01	39.89	37.8582	4.2041	4.6693
63	12	0	1000	0.01	39.89	41.4335	2.9176	3.3007
63	12	-0.5	1000	0.01	39.89	41.5071	3.0726	3.4722
126	12	-0.5	1000	0.01	39.89	41.8874	2.4255	3.1420

Table 3: MSE of Jump Density Estimation at the origin 0 for Normal Jumps based on 1000 Samples.

for the 1st, 4th, and oracle estimates are respectively given by 1.6525, 1.4457, and 1.4450. Figure 2 shows the same results corresponding to  $\lambda = 1000$  and  $\vartheta = 0.01$ . The SSE are in this case 2.0015, 1.5069, and 1.4006 for the 1st, 4th, and oracle estimates, respectively.

## 6 Conclusion and Future Work

In this paper, we study the problem of jump detection via the thresholding method, which is obviously closely related to the problem of spot volatility estimation. We extend the approximated optimal threshold of Figueroa-López and Nisen (2013) by considering a second-order approximation and a non-homogeneous parameter setting. The result is of theoretical interest since the remainder of the second order approximation is much smaller and, at the same time, the resulting threshold estimator is time-invariant, which makes more sense in reality. Monte Carlo studies also demonstrate the superior performance of the second-order approximation.

The higher accuracy comes with the price of more parameters to estimate. We first managed to build a threshold-kernel estimator of the jump density at the origin. We propose a different “optimal” threshold for this purpose and demonstrate the reason why this should be different from the original “optimal” threshold. The intuition is that we have to be more accurate when claiming that an increment contains a jump in order to have a good estimation of its density at the origin. We also put forward a modified version of the threshold-kernel estimator of spot volatility where increments that exceed the threshold are filtered out.

In order to implement the proposed methods, we need to resolve some key obstacles. Concretely, estimates of the optimal threshold, the jump density at the origin, and the spot volatility depend on each other. To resolve the issue, we propose an iterative threshold-kernel estimation scheme. Although we are not guaranteed that the iterative algorithm always converges, Monte Carlo studies show that this rarely creates any problem in reality.

The spirit of jump detection by threshold method is to claim that a jump occurs whenever the absolute value of the increment of the process exceeds the threshold, which, by definition, is a binary outcome. In this case, when an increment is close to the threshold, a small difference in the increment can lead to totally different results. One way to alleviate such a problem is to estimate the probability that a jump happens during a specific time interval, which is similar to the idea of Logistic regression. This suggests an alternative approach to threshold-based classification. Given a non-decreasing function  $F : [0, \infty) \rightarrow [0, 1]$  and an increment  $|\Delta_i X|$ , we can postulate that the probability that a jump occurs during  $[t_{i-1}, t_i]$  is  $F(|\Delta_i X|)$ . We can then adopt the following loss function, that is frequently

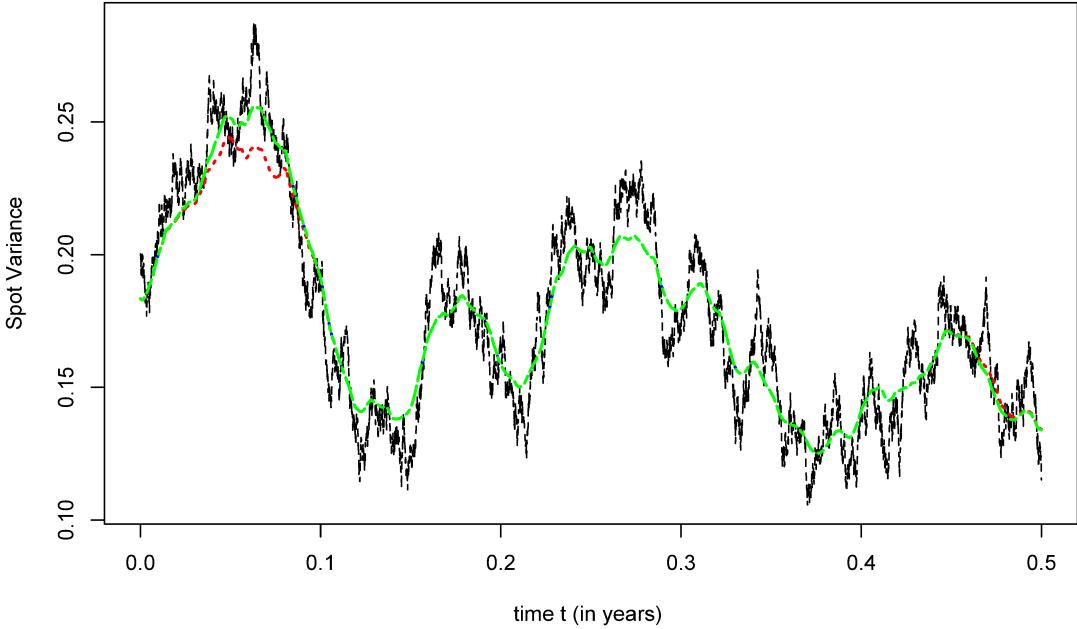


Figure 1: Variance process  $\{V_t\}_{t \geq 0}$  (jiggling dotted black curve) and the estimated spot variance process (44) resulting from the 1st iteration (red dotted) and 4th iteration 4 (long-dashed blue) of Algorithm 3 based on  $B^{n^2}$ . We also plot the oracle variance process (44) (green double-dashed) replacing  $B_i$  with the true  $B_i^*$  in (40). The oracle and the 4th iteration variance estimates overlap. We take the parameter values in (47) as well as  $\rho = -0.5$ ,  $\lambda = 200$ , and the Merton jumps (48) with  $\vartheta = 0.03$ . Estimation based on 5-minute observations during 6 months.

used in classification problems:

$$L_{t,h}(F) = \mathbb{E} (F(|X_{t+h} - X_t|) \mathbf{1}_{\{N_{t+h} - N_t = 0\}}) + \mathbb{E} ([1 - F(|X_{t+h} - X_t|)] \mathbf{1}_{\{N_{t+h} - N_t \neq 0\}}).$$

Indeed, it could be cumbersome to optimize over all continuous functions  $F$ . However, we can try to limit ourself to a suitable, relatively small, class of possible functions  $F$ . One possible direction is to consider  $F_B(x) = F(x/B)$ , which is a generalization of what we have done in this paper. Another possible direction is to consider  $F(x) = F_{1n}(x) \mathbf{1}_{\{0 < x < B\}} + F_{2n}(x) \mathbf{1}_{\{x \geq B\}}$ , where  $F_{1n}$  and  $F_{2n}$  are two functions that can depend on  $n$ . This can potentially provide insight on how the shape of  $F$  should look like around the “optimal” threshold.

## A Proof of the Main Results

Let us start by giving a lemma necessary for the proof of Theorem 2.1.

**Lemma A.1.** *For  $i = 1, 2$  let  $f_i \in \mathcal{C}([0, \infty))$  be strictly positive and differentiable on  $(0, \infty)$ . Further suppose that  $f_1$  is non-increasing while  $f_2$  is non-decreasing and  $\lim_{x \rightarrow 0^+} [f'_1(x) + f'_2(x)]$  exists. If there exists  $x_0 \in (0, \infty)$  such that*

$$(a) \quad |f'_1(x)| \geq |f'_2(x)| \quad \text{for all } x \in (0, x_0) \quad (b) \quad |f'_2(x)| \geq |f'_1(x)| \quad \text{for all } x \in (x_0, \infty), \quad (50)$$

then,  $f := f_1 + f_2$  is quasi-convex on  $[0, \infty)$ .

*Proof.* From (a) in (50) and since  $f'_1(x) \leq 0$  and  $f'_2(x) \geq 0$ ,  $f'(x) = f'_1(x) + f'_2(x) \leq 0$  for all  $x \in (0, x_0)$ . Furthermore, this implies  $\lim_{x \rightarrow 0^+} f'(x) \leq 0$ . On the other hand, from (b) in (50)  $f'(x) = f'_1(x) + f'_2(x) \geq 0$  for all  $x \in (x_0, \infty)$ .

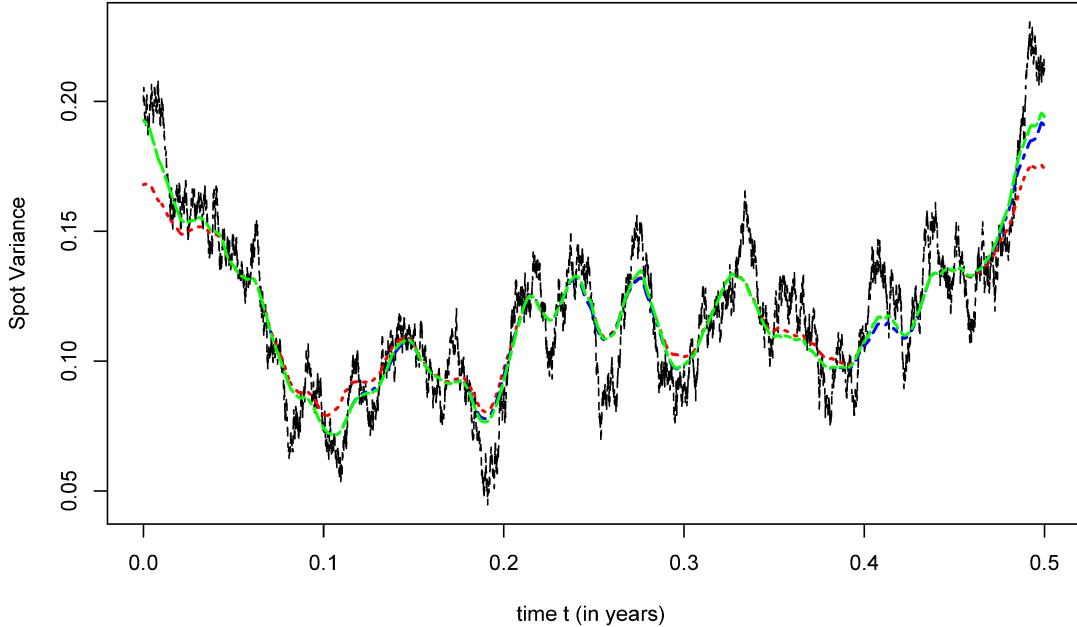


Figure 2: Variance process  $\{V_t\}_{t \geq 0}$  (jiggling dotted black curve) and the estimated spot variance process (44) resulting from the 1st iteration (red dotted) and 4th iteration 4 (long-dashed blue) of Algorithm 3 based on  $B^{n2}$ . We also plot the oracle variance process (44) (green double-dashed) replacing  $B_i$  with the true  $B_i^*$  in (40). The oracle and the 4th iteration variance estimates almost overlap during the whole domain except at the end. We take the parameter values in (47) as well as  $\rho = -0.5$ ,  $\lambda = 1000$ , and the Merton jumps (48) with  $\vartheta = 0.01$ . Estimation based on 5-minute observations during 6 months.

From the well known sufficient conditions for the quasi-convexity of continuous real-valued functions of a real variable (see Boyd and Vandenberghe (2004) pg. 99 (3.20) therein for more details), it follows that  $f$  is quasi-convex on  $[0, \infty)$ .  $\square$

**Proof of Theorem 2.1.** Throughout, we assume  $w = 1$  and  $\bar{\gamma}_{t,h} > 0$  (the cases of  $\bar{\gamma}_{t,h} \leq 0$  and  $w \neq 0$  can be proved in a similar way). For simplicity we omit the argument  $w$  in  $L_{t,h}(B; w)$ . Let  $F_{t,h}^{*k}$  denotes the distribution of the density  $\phi_{t,h} * f^{*k}$ . Conditioning on the number of jumps  $N_{t+h} - N_t$ , the loss function  $L_{t,h}$  is split as follows:

$$L_{t,h}(B) := L_{t,h}^{(1)}(B) + L_{t,h}^{(2)}(B),$$

where

$$\begin{aligned} L_{t,h}^{(1)}(B) &:= \mathbb{P}(|X_{t+h} - X_t| > B, N_{t+h} - N_t = 0) = e^{-h\bar{\lambda}_{t,h}} \left[ 1 - \Phi \left( \frac{B - h\bar{\gamma}_{t,h}}{\bar{\sigma}_{t,h}\sqrt{h}} \right) + \Phi \left( \frac{-B - h\bar{\gamma}_{t,h}}{\bar{\sigma}_{t,h}\sqrt{h}} \right) \right], \\ L_{t,h}^{(2)}(B) &:= \mathbb{P}(|X_{t+h} - X_t| \leq B, N_{t+h} - N_t \neq 0) = e^{-h\bar{\lambda}_{t,h}} \sum_{k=1}^{\infty} \frac{(h\bar{\lambda}_{t,h})^k}{k!} [F_{t,h}^{*k}(B) - F_{t,h}^{*k}(-B)]. \end{aligned}$$

Here,  $\Phi(\cdot)$  is the cdf of standard normal distribution. Note that by definition,  $L_{t,h}^{(1)}$  is strictly decreasing while  $L_{t,h}^{(2)}$  is strictly increasing. It is also clear that for each  $h > 0$  and  $t \in [0, T]$ ,  $L_{t,h}^{(1)} \in \mathcal{C}^\infty(\mathbb{R}^+)$  and  $\partial_B L_{t,h}^{(1)}(B) < 0$  for all

$B \in \mathbb{R}^+$ . For the differentiability of  $L_{t,h}^{(2)}$ , since

$$\sup_{k \in \mathbb{N}} \sup_{x \in \mathbb{R}} |\phi_{t,h} * f^{*k}(x)| \leq \sup_{x \in \mathbb{R}} f(x) =: M(f) < \infty, \quad (51)$$

it follows that  $\sup_{k \in \mathbb{N}} \sup_{B \in (0,\infty)} |\phi_{t,h} * f^{*k}(B) + \phi_{t,h} * f^{*k}(-B)| \leq 2M(f)$  and, thus, by Bounded Convergence Theorem,  $L_{t,h}^{(2)}$  is differentiable. Similarly, since  $\sup_{m \in \mathbb{N}} \sup_{k \in \mathbb{N}} \sup_{B \in (0,\infty)} |\phi_{t,h}^{(m)} * f^{*k}(B) + \phi_{t,h}^{(m)} * f^{*k}(-B)| \leq 2M(f)$ , we can further prove that  $L_{t,h}^{(2)} \in \mathcal{C}^\infty(\mathbb{R}^+)$  by Bounded Convergence Theorem.

We observe that  $L_{t,h}^{(1)}(B) \neq 0$  and  $L_{t,h}^{(2)}(B) \neq 0$  for all  $B > 0$ , so we now proceed to study the ratio

$$R_{t,h}(B) := \frac{\partial_B L_{t,h}^{(2)}(B)}{-\partial_B L_{t,h}^{(1)}(B)}.$$

Let us start by noting that

$$\partial_B L_{t,h}^{(1)}(B) = -\frac{e^{-h\bar{\lambda}_{t,h}}}{\sqrt{h\bar{\sigma}_{t,h}}} \left[ \phi\left(\frac{B - h\bar{\gamma}_{t,h}}{\bar{\sigma}_{t,h}\sqrt{h}}\right) + \phi\left(\frac{B + h\bar{\gamma}_{t,h}}{\bar{\sigma}_{t,h}\sqrt{h}}\right) \right], \quad (52)$$

$$\partial_B L_{t,h}^{(2)}(B) = e^{-h\bar{\lambda}_{t,h}} \sum_{k=1}^{\infty} \frac{(h\bar{\lambda}_{t,h})^k}{k!} [\phi_{t,h} * f^{*k}(B) + \phi_{t,h} * f^{*k}(-B)]. \quad (53)$$

An immediate consequence is that  $R_{t,h}(B)$  is continuous for  $B \in [0, \infty)$ .  $R_{t,h}$  may now be written as:

$$R_{t,h}(B) = \sum_{k=1}^{\infty} \frac{(h\bar{\lambda}_{t,h})^k}{k!} I_{t,h,k}(B), \quad \text{where } I_{t,h,k}(B) := \frac{\bar{\sigma}_{t,h}\sqrt{h} (\phi_{t,h} * f^{*k}(B) + \phi_{t,h} * f^{*k}(-B))}{\phi\left(\frac{B - h\bar{\gamma}_{t,h}}{\bar{\sigma}_{t,h}\sqrt{h}}\right) + \phi\left(\frac{B + h\bar{\gamma}_{t,h}}{\bar{\sigma}_{t,h}\sqrt{h}}\right)}.$$

By definition of convolution,  $I_{t,h,k}$  can be written as:

$$I_{t,h,k}(B) = \int g_{t,h}(w, B) f^{*k}(w) dw, \quad \text{where } g_{t,h}(w, B) := \frac{\phi\left(\frac{B - h\bar{\gamma}_{t,h} - w}{\bar{\sigma}_{t,h}\sqrt{h}}\right) + \phi\left(\frac{B + h\bar{\gamma}_{t,h} + w}{\bar{\sigma}_{t,h}\sqrt{h}}\right)}{\phi\left(\frac{B - h\bar{\gamma}_{t,h}}{\bar{\sigma}_{t,h}\sqrt{h}}\right) + \phi\left(\frac{B + h\bar{\gamma}_{t,h}}{\bar{\sigma}_{t,h}\sqrt{h}}\right)}.$$

Plugging in the normal p.d.f.,  $g_{t,h}$  can be factorized to be:

$$g_{t,h}(w, B) = \exp\left(-\frac{w^2 + 2wh\bar{\gamma}_{t,h}}{2h\bar{\sigma}_{t,h}^2}\right) \frac{\exp(B(h\bar{\gamma}_{t,h} + w)/h\bar{\sigma}_{t,h}^2) + \exp(-B(h\bar{\gamma}_{t,h} + w)/h\bar{\sigma}_{t,h}^2)}{\exp(B\bar{\gamma}_{t,h}/\bar{\sigma}_{t,h}^2) + \exp(-B\bar{\gamma}_{t,h}/\bar{\sigma}_{t,h}^2)} =: g_{t,h}^{(1)}(w) g_{t,h}^{(2)}(w, B).$$

It is not hard to prove the following properties of  $g_{t,h}^{(1)}$ :

$$1 \leq g_{t,h}^{(1)}(w) \leq e^{h\bar{\gamma}_{t,h}^2/2\bar{\sigma}_{t,h}^2}, \quad \omega \in (-2h\bar{\gamma}_{t,h}, 0), \quad \text{and} \quad 0 < g_{t,h}^{(1)}(w) \leq 1, \quad \omega \in (-2h\bar{\gamma}_{t,h}, 0)^C. \quad (54)$$

$g_{t,h}^{(2)}(\omega, \cdot)$  is a function of type  $t(x) = \frac{e^{ax} + e^{-ax}}{e^{bx} + e^{-bx}}$ , where  $x \in [0, \infty)$ ,  $a = |(h\bar{\gamma}_{t,h} + w)/h\bar{\sigma}_{t,h}^2|$  and  $b = |\bar{\gamma}_{t,h}/\bar{\sigma}_{t,h}^2|$ . Note that the derivative  $t'(x)$  can be written as

$$t'(x) = \frac{e^{ax} + e^{-ax}}{e^{bx} + e^{-bx}} \left[ a \frac{e^{ax} - e^{-ax}}{e^{ax} + e^{-ax}} - b \frac{e^{bx} - e^{-bx}}{e^{bx} + e^{-bx}} \right].$$

When  $a > b > 0$ ,  $t(x)$  is an increasing function from 1 to  $+\infty$  and

$$\begin{aligned} t'(x) &\geq \frac{e^{ax}}{2e^{bx}}(a-b) \frac{e^{bx} - e^{-bx}}{e^{bx} + e^{-bx}} \geq \frac{(a-b)}{4} e^{(a-b)x} (1 - e^{-2bx}) \geq \begin{cases} \frac{a-b}{4}(1 - e^{-1})2bx, & x \leq \frac{1}{2b} \\ \frac{a-b}{4}(1 - e^{-1})(a-b)x, & x > \frac{1}{2b} \end{cases} \\ &\geq \frac{(a-b)(1 - e^{-1}) \min(a-b, 2b)}{4} x. \end{aligned}$$

For the third inequality, when  $0 \leq x \leq 1/2b$ , we use  $1 - e^{-2bx} \geq (1 - e^{-1})2bx$ , and when  $x > 1/2b$ , we use  $e^{(a-b)x} \geq (a-b)x$  and  $1 - e^{-2bx} \geq (1 - e^{-1})$ . Specifically, when  $a > 3b$ , we have

$$t'(x) \geq b^2(1 - e^{-1})x. \quad (55)$$

When  $b > a > 0$ ,  $t(x)$  is a decreasing function from 1 to 0 and

$$|t'(x)| \leq b \frac{e^{bx} - e^{-bx}}{e^{bx} + e^{-bx}} \leq b^2 x, \quad (56)$$

where we use the property that  $\tanh'(x) \leq 1$ .

Here we notice that  $a < b \Leftrightarrow \omega \in (-2h\bar{\gamma}_{t,h}, 0)$ . Based on this, for each fixed  $k \in \mathbb{N}$ , we decompose  $I_{t,h,k}$  into two parts:

$$I_{t,h,k}(B) = \left( \int_{(-2h\bar{\gamma}_{t,h}, 0)} + \int_{(-2h\bar{\gamma}_{t,h}, 0)^c} \right) g_{t,h}^{(1)}(w) g_{t,h}^{(2)}(w, B) f^{*k}(w) dw =: I_{t,h,k}^{(1)}(B) + I_{t,h,k}^{(2)}(B). \quad (57)$$

In what follows, We shall prove that there exists  $h_0 > 0$ , which may depend on  $T$ , such that for all  $t \in [0, T]$  and  $h \in (0, h_0)$ , there exists  $B_{t,h}^* > 0$ , such that

$$R_{t,h}(B) < 1, \quad \text{for } B \in (0, B_{t,h}^*), \quad \text{and} \quad R_n(B) > 1, \quad \text{for } B \in (B_{t,h}^*, \infty).$$

These two conditions, together with the signs of  $\partial_B L_{t,h}^{(1)}$  and  $\partial_B L_{t,h}^{(2)}$ , will imply that  $B \mapsto L_{t,h}(B)$  is quasi-convex (see Lemma A.1 below) for  $h$  small enough. To do this, we will prove the following:

- (i) For any  $h > 0$ ,  $\lim_{B \rightarrow \infty} R_{t,h}(B) = +\infty$ .
- (ii)  $\lim_{h \rightarrow 0} \sup_{t \in [0, T]} R_{t,h}(0) = 0$ .
- (iii) There exists  $h_0 > 0$ , which may depend on  $T$ , such that for all  $t \in [0, T]$  and  $h \in (0, h_0)$ ,  $R_{t,h}(\cdot)$  is strictly increasing.

For (i), it is clear that  $I_{t,h,k}^{(1)} \geq 0$ , and by Fatou's Lemma, for  $k$  large enough <sup>4</sup>,  $I_{t,h,k}^{(2)}$  satisfies

$$\liminf_{B \rightarrow \infty} I_{t,h,k}^{(2)}(B) \geq \int_{(-2h\bar{\gamma}_{t,h}, 0)^c} \liminf_{B \rightarrow \infty} g_{t,h}^{(1)}(w) g_{t,h}^{(2)}(w, B) f^{*k}(w) dw = +\infty.$$

These two relationships imply (i).

For (ii), since  $g_{t,h}^{(2)}(w, 0) = 1$ ,

$$I_{t,h,k}(0) = \int g_{t,h}^{(1)}(w) f^{*k}(w) dw = \sqrt{2\pi h} \bar{\sigma}_{t,h} e^{h\bar{\gamma}_{t,h}^2/2\bar{\sigma}_{t,h}^2} \int \frac{e^{-(w+h\bar{\gamma}_{t,h})^2/2h\bar{\sigma}_{t,h}^2}}{\sqrt{2\pi h} \bar{\sigma}_{t,h}} f^{*k}(w) dw \leq \sqrt{2\pi h} \bar{\sigma}_{t,h} e^{h\bar{\gamma}_{t,h}^2/2\bar{\sigma}_{t,h}^2} M(f).$$

Note that the right-hand side converges to zero as  $h \rightarrow 0$ , and does not depend on  $k$ . By Assumption 1, the convergence is uniformly in  $t$ , so (ii) follows.

Now we proceed to consider (iii). Indeed, for any given  $t \in [0, T]$ , by the upper bound of  $g_{t,h}^{(1)}(w)$  given by (54)

<sup>4</sup> $k$  has to be large, since now we are not assuming small  $h$ , so it is possible that  $f^{*k}(\omega) \equiv 0$  for  $\omega \in (-2h\bar{\gamma}_{t,h}, 0)^c$ .

and the upper bound of  $|\partial_B g_{t,h}^{(2)}(w, B)|$  given by (56), we have

$$\begin{aligned} |I_{t,h,k}^{(1)}(B + \delta) - I_{t,h,k}^{(1)}(B)| &= \int_{(-2h\bar{\gamma}_{t,h}, 0)} g_{t,h}^{(1)}(w) \times |g_{t,h}^{(2)}(w, B + \delta) - g_{t,h}^{(2)}(w, B)| \times f^{*k}(w) dw \\ &\leq \int_{(-2h\bar{\gamma}_{t,h}, 0)} e^{h\bar{\gamma}_{t,h}^2/2\bar{\sigma}_{t,h}^2} \times \frac{\bar{\gamma}_{t,h}^2}{\bar{\sigma}_{t,h}^4} (B + \delta) \delta \times f^{*k}(w) dw \leq 2h \frac{\bar{\gamma}_{t,h}^3}{\bar{\sigma}_{t,h}^4} e^{h\bar{\gamma}_{t,h}^2/2\bar{\sigma}_{t,h}^2} M(f)(B + \delta) \delta. \end{aligned}$$

Furthermore, for  $I_{t,h,k}^{(2)}$ , note that for  $\omega \in (-2h\bar{\gamma}_{t,h}, 0)^C$ ,  $g_{t,h}^{(2)}(w, B)$  is increasing in  $B$ , and for  $\omega \in [-4h\bar{\gamma}_{t,h}, 4h\bar{\gamma}_{t,h}]^C$ , we have  $|(h\bar{\gamma}_{t,h} + w)/h\bar{\sigma}_{t,h}^2| > 3|\bar{\gamma}_{t,h}/\bar{\sigma}_{t,h}^2|$ . Thus, we have

$$\begin{aligned} I_{t,h,k}^{(2)}(B + \delta) - I_{t,h,k}^{(2)}(B) &= \int_{(-2h\bar{\gamma}_{t,h}, 0)^C} g_{t,h}^{(1)}(w) \times (g_{t,h}^{(2)}(w, B + \delta) - g_{t,h}^{(2)}(w, B)) \times f^{*k}(w) dw \\ &\geq \int_{[-4h\bar{\gamma}_{t,h}, 4h\bar{\gamma}_{t,h}]^C} g_{t,h}^{(1)}(w) \times \frac{(1 - e^{-1})\bar{\gamma}_{t,h}^2}{\bar{\sigma}_{t,h}^4} B \delta \times f^{*k}(w) dw \\ &\geq \frac{(1 - e^{-1})\bar{\gamma}_{t,h}^2}{\bar{\sigma}_{t,h}^4} B \delta \left( \int g_{t,h}^{(1)}(w) f^{*k}(w) dw - 8h\bar{\gamma}_{t,h} e^{h\bar{\gamma}_{t,h}^2/2\bar{\sigma}_{t,h}^2} M(f) \right). \end{aligned}$$

Putting these two inequalities together, we have that for any  $B > 0$  and  $0 < \delta < B$ :

$$\frac{1}{B\delta} \sum_{k=1}^{\infty} \left| \frac{(h\bar{\lambda}_{t,h})^k}{k!} \left( I_{t,h,k}^{(1)}(B + \delta) - I_{t,h,k}^{(1)}(B) \right) \right| \leq 4h \left( e^{h\bar{\lambda}_{t,h}} - 1 \right) \frac{\bar{\gamma}_{t,h}^3}{\bar{\sigma}_{t,h}^4} e^{h\bar{\gamma}_{t,h}^2/2\bar{\sigma}_{t,h}^2} M(f) = O(h^2), \quad h \rightarrow 0,$$

and

$$\begin{aligned} &\frac{1}{B\delta} \sum_{k=1}^{\infty} \frac{(h\bar{\lambda}_{t,h})^k}{k!} \left( I_{t,h,k}^{(2)}(B + \delta) - I_{t,h,k}^{(2)}(B) \right) \\ &\geq \frac{(1 - e^{-1})\bar{\gamma}_{t,h}^2}{\bar{\sigma}_{t,h}^4} \left( h\bar{\lambda}_{t,h} \int g_{t,h}^{(1)}(w) f(w) dw - \sum_{k=1}^{\infty} \frac{(h\bar{\lambda}_{t,h})^k}{k!} 8h\bar{\gamma}_{t,h} e^{h\bar{\gamma}_{t,h}^2/2\bar{\sigma}_{t,h}^2} M(f) \right) \\ &\geq h^{3/2} \bar{\lambda}_{t,h} \frac{(1 - e^{-1})\bar{\gamma}_{t,h}^2}{\bar{\sigma}_{t,h}^4} \sqrt{2\pi} \bar{\sigma}_{t,h} \exp \left( \frac{h\bar{\gamma}_{t,h}^2}{2\bar{\sigma}_{t,h}^2} \right) \frac{\mathcal{C}_m(f)}{2} + O(h^2), \quad h \rightarrow 0, \end{aligned}$$

where the last equality can be justified by  $\int g_{t,h}^{(1)}(w) f(w) dw \geq \sqrt{2\pi h} \bar{\sigma}_{t,h} \exp \left( \frac{h\bar{\gamma}_{t,h}^2}{2\bar{\sigma}_{t,h}^2} \right) \frac{\mathcal{C}_m(f)}{2} + O(h)$  for small  $h$ , where  $\mathcal{C}_m(f)$  is defined in (5), since the following holds:

$$g_{t,h}^{(1)}(w) = \exp \left( -\frac{w^2 + 2wh\bar{\gamma}_{t,h}}{2h\bar{\sigma}_{t,h}^2} \right) = \sqrt{2\pi h} \bar{\sigma}_{t,h} \exp \left( \frac{h\bar{\gamma}_{t,h}^2}{2\bar{\sigma}_{t,h}^2} \right) \frac{1}{\sqrt{2\pi h} \bar{\sigma}_{t,h}} \exp \left( -\frac{(w + h\bar{\gamma}_{t,h})^2}{2h\bar{\sigma}_{t,h}^2} \right).$$

Also note that both convergences do not depend on  $B$  and  $\delta$ , and by Assumption 1, both the convergences can all be made uniform in  $t$ . This proves (iii).  $\square$

**Proof of Theorem 2.2.** For simplicity, we use the notation  $f_{t,h}^{*k} := \phi_{t,h} * f^{*k}$ , where recall that  $\phi_{t,h}(x) := \frac{1}{\bar{\sigma}_{t,h}\sqrt{h}}\phi\left(\frac{x-h\bar{\gamma}_{t,h}}{\bar{\sigma}_{t,h}\sqrt{h}}\right)$  is the density of  $X_{t+h}^c - X_t^c$ . We start by demonstrating that the optimal thresholds  $(B_{t,h}^*)_{t,h}$  converge to 0 uniformly on  $t \in [0, T]$ , as  $h \rightarrow 0$ . Let us first note that the loss function (10) can be written as

$$L_{t,h}(B) := e^{-h\bar{\lambda}_{t,h}} \mathbb{P} \left( |h\bar{\gamma}_{t,h} + \bar{\sigma}_{t,h}\sqrt{h}Z| > B \right) + e^{-h\bar{\lambda}_{t,h}} \sum_{k=1}^{\infty} \frac{(h\bar{\lambda}_{t,h})^k}{k!} \mathbb{P} \left( |h\bar{\gamma}_{t,h} + \bar{\sigma}_{t,h}\sqrt{h}Z + \sum_{i=1}^k \zeta_i| \leq B \right).$$

Next, by partitioning  $E := \{|h\bar{\gamma}_{t,h} + \bar{\sigma}_{t,h}\sqrt{h}Z + \sum_{i=1}^k \zeta_i| \leq B\}$  into  $E \cap \{|h\bar{\gamma}_{t,h} + \bar{\sigma}_{t,h}\sqrt{h}Z| \leq B\}$  and  $E \cap \{|h\bar{\gamma}_{t,h} +$

$\bar{\sigma}_{t,h}\sqrt{h}Z| > B\}$  and simplifying,

$$\begin{aligned} L_{t,h}(B) &\leq \mathbb{P}\left(\left|h\bar{\gamma}_{t,h} + \bar{\sigma}_{t,h}\sqrt{h}Z\right| > B\right) + e^{-h\bar{\lambda}_{t,h}} \sum_{k=1}^{\infty} \frac{(h\bar{\lambda}_{t,h})^k}{k!} \mathbb{P}\left(E, \left|h\bar{\gamma}_{t,h} + \bar{\sigma}_{t,h}\sqrt{h}Z\right| \leq B\right) \\ &\leq \mathbb{P}\left(h\gamma_{T+h}^* + \sigma_{T+h}^*\sqrt{h}|Z| > B\right) + \sum_{k=1}^{\infty} \frac{(h\lambda_{T+h}^*)^k}{k!} \mathbb{P}\left(\left|\sum_{i=1}^k \zeta_i\right| \leq 2B\right), \end{aligned}$$

where we have used that  $\gamma_t^* := \sup_{s \leq t} |\gamma_s|$ ,  $\sigma_t^* := \sup_{s \leq t} \sigma_s$ , and  $\lambda_t^* := \sup_{s \leq t} \lambda_s$  are finite for any  $t$ . Next, consider a sequence of thresholds given by  $B_{h,c}^{P_{\text{ow}}} := ch^\alpha$  for  $\alpha \in (0, 1/2)$  and  $c > 0$ . Thus, using that  $\mathbb{P}(|\zeta_1| \leq 2B) \sim 4\mathcal{C}_0(f)B$  and  $\mathbb{P}\left(\left|\sum_{i=1}^k \zeta_i\right| \leq 2B\right) = O(B)$  as  $B \rightarrow 0$ ,

$$\sup_{t \in [0, T]} L_{t,h}(B_{h,c}^{P_{\text{ow}}}) \leq 4c\mathcal{C}_0(f)\lambda_{T+h}^*h^{1+\alpha} + o(h^{1+\alpha}).$$

Now suppose that  $\epsilon := \limsup_{h \rightarrow 0^+} \sup_{t \in [0, T]} B_{t,h}^* > 0$ . Then, there exists subsequences  $(h_n)_n$  and  $(t_n)_n$  such that  $\inf_n B_{t_n, h_n}^* \geq \epsilon/2$ . In that case,

$$L_{t_n, h_n}(B_{t_n, h_n}^*) \geq e^{-h\lambda_{T+h}^*} h\lambda_{T+h} \mathbb{P}\left(\left|h_n\bar{\gamma}_{t_n, h_n} + \bar{\sigma}_{t_n, h_n}\sqrt{h_n}Z + \zeta_1\right| \leq \epsilon/2\right),$$

but, also  $L_{t_n, h_n}(B_{t_n, h_n}^*) \leq L_{t_n, h_n}(B_{h_n, c}^{P_{\text{ow}}})$  and, since  $\mathbb{P}(|h_n\bar{\gamma}_{t_n, h_n} + \bar{\sigma}_{t_n, h_n}\sqrt{h_n}Z + \zeta_1| \leq \epsilon/2) \rightarrow \mathbb{P}(|\zeta_1| \leq \epsilon/2) > 0$ ,<sup>5</sup> as  $n \rightarrow \infty$ , we would have that

$$4c\lambda_{T+h}^*\mathcal{C}_0(f)h^{1+\alpha} + o(h^{1+\alpha}) \geq h\lambda_{T+h} + o(h),$$

which leads to a contradiction. Hence, it is necessary that the optimal thresholds converge to 0 uniformly on  $[0, T]$ .

Now we will show the asymptotic characterization of the optimal thresholds. From Theorem 2.1, there exists  $h_0 > 0$ , depending on  $T$ , such that, for all  $t \in [0, T]$  and  $h \in (0, h_0]$ , the loss functions  $L_{t,h}$  possess a unique critical point. By equating the first-order derivative of the loss function to zero, from (52)-(53) it follows that the unique optimal threshold,  $B_{t,h}^*$ , must satisfy the equation given by

$$\frac{1}{\sqrt{h}\bar{\sigma}_{t,h}} \left[ \phi\left(\frac{B_{t,h}^* - h\bar{\gamma}_{t,h}}{\sqrt{h}\bar{\sigma}_{t,h}}\right) + \phi\left(\frac{B_{t,h}^* + h\bar{\gamma}_{t,h}}{\sqrt{h}\bar{\sigma}_{t,h}}\right) \right] = \sum_{k=1}^{\infty} \frac{(h\bar{\lambda}_{t,h})^k}{k!} [f_{t,h}^{*k}(B_{t,h}^*) + f_{t,h}^{*k}(-B_{t,h}^*)]. \quad (58)$$

A rearrangement of this equation shows

$$\phi\left(\frac{B_{t,h}^* - h\bar{\gamma}_{t,h}}{\sqrt{h}\bar{\sigma}_{t,h}}\right) = \sqrt{h}\bar{\sigma}_{t,h} \left[1 + e^{-2B_{t,h}^*\bar{\gamma}_{t,h}/\bar{\sigma}_{t,h}^2}\right]^{-1} \sum_{k=1}^{\infty} \frac{(h\bar{\lambda}_{t,h})^k}{k!} [f_{t,h}^{*k}(B_{t,h}^*) + f_{t,h}^{*k}(-B_{t,h}^*)]. \quad (59)$$

Upon taking the log on both sides of (59), we arrive at the fixed point equation (11). From (51) together with Assumption 1, we conclude that  $\lim_{h \rightarrow 0^+} B_{t,h}^*/h^{1/2} = \infty$ , uniformly for  $t \in [0, T]$ , i.e.,

$$\lim_{h \rightarrow 0} \inf_{t \in [0, T]} \frac{B_{t,h}^*}{\sqrt{h}} = +\infty.$$

A further modification of this equation indicates that

$$B_{t,h}^* = h\bar{\gamma}_{t,h} + \sqrt{2h}\bar{\sigma}_{t,h} \log^{1/2} \left( \frac{1}{\bar{\sigma}_{t,h}\bar{\lambda}_{t,h}h^{3/2}} \right) \left[ 1 + \frac{\log \left( \frac{\sqrt{2\pi}(f_{t,h}^{*1}(B_{t,h}^*) + f_{t,h}^{*1}(-B_{t,h}^*))}{1 + e^{-2B_{t,h}^*\bar{\gamma}_{t,h}/\bar{\sigma}_{t,h}^2}} \right)}{\log(\bar{\sigma}_{t,h}\bar{\lambda}_{t,h}h^{3/2})} + \frac{\log(1 + S_{t,h}(B_{t,h}^*))}{\log(\bar{\sigma}_{t,h}\bar{\lambda}_{t,h}h^{3/2})} \right]^{1/2},$$

<sup>5</sup>It is necessary to have  $\mathcal{C}_0(f) > 0$ . Otherwise,  $B \rightarrow 0$  is not optimal.

where above, we have defined

$$S_{t,h}(B) := \sum_{k=2}^{\infty} \frac{(h\bar{\lambda}_{t,h})^{k-1}}{k!} \left[ \frac{f_{t,h}^{*k}(B) + f_{t,h}^{*k}(-B)}{f_{t,h}^{*1}(B) + f_{t,h}^{*1}(-B)} \right].$$

From this, a direct consequence is that  $B_{t,h}^* = O(\sqrt{h \log(1/h)})$ , so we have

$$\frac{\sqrt{2\pi} (f_{t,h}^{*1}(B_{t,h}^*) + f_{t,h}^{*1}(-B_{t,h}^*))}{1 + e^{-2B_{t,h}^* \bar{\gamma}_{t,h} / \bar{\sigma}_{t,h}^2}} = \sqrt{2\pi} \mathcal{C}_0(f) + O(B_{t,h}^*), \quad S_{t,h}(B_{t,h}^*) = O(h^2).$$

The second relationship above is because  $f_{t,h}^{*k}$  are bounded by  $M(f)$  and  $f_{t,h}^{*1}(B_{t,h}^*)$  is bounded away from zero. We prove the first relationship above now. Indeed, by our assumption on the smoothness of  $f$ , there exists  $\epsilon > 0$ , such that  $f \in C^1((0, \epsilon))$  and  $f \in C^1((-\epsilon, 0))$ . Then, we have:

$$\begin{aligned} f_{t,h}^{*1}(B_{t,h}^*) &= f_{t,h}^{*1}(0) + O(B_{t,h}^*) \\ f_{t,h}^{*1}(0) - \mathcal{C}_0(f) &= f_{t,h}^{*1}(0) - \left( f(0-) \int_{-\infty}^0 \phi_{t,h}(y) dy + f(0+) \int_0^{+\infty} \phi_{t,h}(y) dy \right) + O(\sqrt{h}) \\ &= \int_{-\infty}^0 (f(y) - f(0-)) \phi_{t,h}(y) dy + \int_0^{+\infty} (f(y) - f(0+)) \phi_{t,h}(y) dy + O(\sqrt{h}) \\ &= \int_{-\epsilon}^0 (f(y) - f(0-)) \phi_{t,h}(y) dy + \int_0^{+\epsilon} (f(y) - f(0+)) \phi_{t,h}(y) dy + O(\sqrt{h}) \\ &= \int_{-\epsilon}^0 \int_0^1 f'(yv) dv y \phi_{t,h}(y) dy + \int_0^{+\epsilon} \int_0^1 f'(yv) dv y \phi_{t,h}(y) dy + O(\sqrt{h}) \\ &= O(\sqrt{h}). \end{aligned} \tag{60}$$

Above, the first equality uses  $\int_0^{+\infty} \phi_{t,h}(y) dy = 1/2 + O(\sqrt{h})$  and  $\int_{-\infty}^0 \phi_{t,h}(y) dy = 1/2 + O(\sqrt{h})$ . The third equality uses  $\int_{-\epsilon}^{+\infty} \phi_{t,h}(y) dy = o(h)$  and  $\int_{-\infty}^{-\epsilon} \phi_{t,h}(y) dy = o(h)$ . From this, we have  $f_{t,h}^{*1}(0) = \mathcal{C}_0(f) + O(\sqrt{h})$ . We then have  $f_{t,h}^{*1}(B_{t,h}^*) + f_{t,h}^{*1}(-B_{t,h}^*) = 2\mathcal{C}_0(f) + O(B_{t,h}^*)$ . Therefore, for any  $\alpha \in (0, 1/2)$ ,

$$\frac{\log \left( \frac{\sqrt{2\pi} (f_{t,h}^{*1}(B_{t,h}^*) + f_{t,h}^{*1}(-B_{t,h}^*))}{1 + e^{-2B_{t,h}^* \bar{\gamma}_{t,h} / \bar{\sigma}_{t,h}^2}} \right)}{\log (\bar{\sigma}_{t,h} \bar{\lambda}_{t,h} h^{3/2})} = \frac{\log (\sqrt{2\pi} \mathcal{C}_0(f))}{\log (\bar{\sigma}_{t,h} \bar{\lambda}_{t,h} h^{3/2})} + o(h^\alpha), \quad \frac{\log \left( 1 + S_{t,h}(B_{t,h}^*) \right)}{\log (\bar{\sigma}_{t,h} \bar{\lambda}_{t,h} h^{3/2})} = o(h^\alpha).$$

For the last assertion of the theorem, if we further note that  $\bar{\sigma}_{t,h}^2 = \sigma_t^2 + O(h)$  and  $\bar{\lambda}_{t,h} = \lambda_t + O(h)$  under the specified smoothness of  $t \rightarrow \sigma_t^2$  and  $t \rightarrow \lambda_t$ , then we conclude the following approximation of  $B_{t,h}^*$ :

$$\begin{aligned} B_{t,h}^* &= \sqrt{2h} \bar{\sigma}_{t,h} \log^{1/2} \left( \frac{1}{\bar{\sigma}_{t,h} \bar{\lambda}_{t,h} h^{3/2}} \right) \left[ 1 + \frac{\log (\sqrt{2\pi} \mathcal{C}_0(f))}{\log (\bar{\sigma}_{t,h} \bar{\lambda}_{t,h} h^{3/2})} \right]^{1/2} + o(h^{\frac{1}{2}+\alpha}) \\ &= \sqrt{h} \sigma_t \left[ 3 \log (1/h) - 2 \log \left( \sqrt{2\pi} \mathcal{C}_0(f) \sigma_t \lambda_t \right) \right]^{1/2} + o(h^{\frac{1}{2}+\alpha}), \end{aligned}$$

for any  $\alpha \in (0, 1/2)$ . □

**Proof of Proposition 2.8.** First, note that

$$P(|\Delta X| > B) = e^{-h\lambda} \mathbb{P}(|h\gamma + \sqrt{h}\sigma Z| > B) + h\lambda e^{-h\lambda} \mathbb{P}(|h\gamma + \sqrt{h}\sigma Z + \zeta| > B) + O(h^2). \tag{61}$$

Let  $\phi_h(x)$  be the density of  $h\gamma + \sqrt{h}\sigma Z$  and note that, for  $k \geq 1$ ,  $h\gamma + \sqrt{h}\sigma Z + \sum_{i=1}^k \zeta_i$  has density  $\phi_h * f^{*k}$ , which is bounded by  $M(f) := \sup_x f(x)$ . Therefore, we have  $\left| \frac{\partial}{\partial B} \mathbb{P}(|h\gamma + \sqrt{h}\sigma Z + \sum_{i=1}^k \zeta_i| > B) \right| < 2M(f)$  and,

furthermore,

$$\frac{\partial}{\partial B} \sum_{k \geq 2} \frac{(h\lambda)^k}{k!} \mathbb{P} \left( \left| h\gamma + \sqrt{h}\sigma Z + \sum_{i=1}^k \zeta_i \right| > B \right) = \sum_{k \geq 2} \frac{(h\lambda)^k}{k!} \frac{\partial}{\partial B} \mathbb{P} \left( \left| h\gamma + \sqrt{h}\sigma Z + \sum_{i=1}^k \zeta_i \right| > B \right) = O(h^2).$$

We then have

$$f^*(B) = -\frac{\partial}{\partial B} P(|\Delta X| > B) = e^{-h\lambda} [\phi_h(B) + \phi_h(-B)] + h\lambda e^{-h\lambda} [g(B) + g(-B)] + O(h^2), \quad (62)$$

where  $g$  denotes the density of  $h\gamma + \sqrt{h}\sigma Z + \zeta$ . Combining (61) and (62), the conditional density is such that

$$f_{|\Delta X| > B}^*(B) = \frac{f^*(B)}{P(|\Delta X| > B)} = \frac{\frac{1}{\lambda} \frac{1}{\sqrt{2\pi h^3 \sigma^2}} \left[ \exp \left( -\frac{(B-h\gamma)^2}{2h\sigma^2} \right) + \exp \left( -\frac{(B+h\gamma)^2}{2h\sigma^2} \right) \right] + g(B) + g(-B)}{\frac{1}{\lambda h} \mathbb{P}(|h\gamma + \sqrt{h}\sigma Z| > B) + \mathbb{P}(|h\gamma + \sqrt{h}\sigma Z + \zeta| > B)} + O(h).$$

Now, by  $g = \phi_h * f$  and the smoothness of  $f$  near 0, we have that if  $x$  is close enough to 0,

$$\begin{aligned} g(x) - f(x) &= \left( \int_{(x-\epsilon, x+\epsilon)} + \int_{(x-\epsilon, x+\epsilon)^c} \right) (f(y) - f(x)) \phi_h(y-x) dy \\ &= \int_{(x-\epsilon, x+\epsilon)} (f'(x)(y-x) + f''(\theta_y)(y-x)^2) \phi_h(y-x) dy + o(h) = O(h), \end{aligned}$$

where  $\theta_y$  is between  $x$  and  $y$  and  $\epsilon$  is a fixed positive number such that  $f \in C^2((x-\epsilon, x+\epsilon))$ . Such an  $\epsilon$  exists due to Assumption 4. Above, we have used the following facts:

$$\begin{aligned} \int_{(x-\epsilon, x+\epsilon)^c} \phi_h(y-x) dy &= o(h), \quad \int_{(x-\epsilon, x+\epsilon)} (y-x) \phi_h(y-x) dy = \gamma h + o(h), \\ \int_{(x-\epsilon, x+\epsilon)} |f''(\theta_y)| (y-x)^2 \phi_h(y-x) dy &\leq M\sigma^2 h. \end{aligned}$$

Note that the above holds uniformly in  $x$  near 0, so we have  $g(B) = f(B) + O(h)$  for  $h$  small enough. This also implies

$$\mathbb{P} \left( \left| h\gamma + \sqrt{h}\sigma Z + \zeta \right| \leq B \right) = 2g(0)B + o(B) = 2f(0)B + o(B).$$

Therefore, we have the following:

$$\begin{aligned} f_{|\Delta X| > B}^*(B) &= \frac{\frac{1}{\lambda} \frac{1}{\sqrt{2\pi h^3 \sigma^2}} \left[ \exp \left( -\frac{(B-h\gamma)^2}{2h\sigma^2} \right) + \exp \left( -\frac{(B+h\gamma)^2}{2h\sigma^2} \right) \right] + 2f(0) + O(h) + O(B^2)}{\frac{1}{\lambda h} \mathbb{P}(|h\gamma + \sqrt{h}\sigma Z| > B) + 1 - 2f(0)B + o(B)} + O(h) \\ &= 2f(0) + \frac{2}{\lambda \sqrt{2\pi h^3 \sigma^2}} \exp \left( -\frac{B^2}{2h\sigma^2} \right) + 2f(0)B + o(B) + o(h^{-3/2} e^{-\frac{B^2}{2h\sigma^2}}), \end{aligned}$$

where we used the following:

$$\frac{1}{h} \mathbb{P}(|h\gamma + \sqrt{h}\sigma Z| > B) \sim \frac{\sigma}{B\sqrt{2\pi h}} \exp \left( -\frac{B^2}{2h\sigma^2} \right), \quad \frac{1}{\sqrt{2\pi h^3 \sigma^2}} \exp \left( -\frac{(B-h\gamma)^2}{2h\sigma^2} \right) \sim \frac{1}{\sqrt{2\pi h^3 \sigma^2}} \exp \left( -\frac{B^2}{2h\sigma^2} \right).$$

This completes the proof.  $\square$

**Proof of Corollary 2.9.** Denote the leading order term of (19) as:

$$F(B) = \frac{1}{\lambda f(0) \sqrt{2\pi h^3 \sigma^2}} \exp \left( -\frac{B^2}{2h\sigma^2} \right) + B.$$

Set  $a = 1/(\lambda f(0)\sqrt{2\pi h^3\sigma^2})$ ,  $b = 1/(2h\sigma^2)$ . For  $h$  small enough, we do have  $a\sqrt{b} > 1/(1 - \exp(-1/2))$ , and  $\log(2ab) < b$ . By the Lemma A.2 below, the minimum of  $F$  is in  $(\sqrt{2h\sigma^2}, \sqrt{2h\sigma^2 \log(1/\sqrt{2\pi h^5\sigma^6})})$  and satisfies  $B \exp\left(-\frac{B^2}{2h\sigma^2}\right) = \sqrt{2\pi h^5\sigma^6}$ . Taking logarithms on both sides and rearranging terms, we get

$$\frac{B^2}{2h\sigma^2} = \log(B) - \frac{5}{2} \log(h) + C,$$

for some constant  $C$ . Note that since  $B$  lies in  $(\sqrt{2h\sigma^2}, \sqrt{2h\sigma^2 \log(1/\sqrt{2\pi h^5\sigma^6})})$ ,  $\log(B) = \frac{1}{2} \log(h) + O(\log \log(1/h))$ . Thus, we get the approximation of the optimal  $B$  as

$$B^* = \sqrt{4h\sigma^2 \log(1/h)} + O(\sqrt{h \log \log(1/h)}).$$

This completes the proof.  $\square$

**Lemma A.2.** Suppose  $a, b > 0$  and  $a\sqrt{b} > 1/(1 - \exp(-1/2))$ , and  $\log(2ab) < b$ . Define  $F(x) = a \exp(-bx^2) + x$  where  $x \geq 0$ . Then, the minimum point of  $F$  is in  $(1/\sqrt{2b}, \sqrt{\log(2ab)/b})$  and satisfies  $2abx \exp(-bx^2) = 1$ .

*Proof.* Taking derivative twice, we get  $F'(x) = -2abx \exp(-bx^2) + 1$  and  $F''(x) = 2ab(2bx^2 - 1) \exp(-bx^2)$ . By studying the sign of  $F''$ , we have that  $F'$  is decreasing in  $(0, 1/\sqrt{2b})$  and increasing in  $(1/\sqrt{2b}, \infty)$ , and we also have  $F'(1/\sqrt{2b}) = -a\sqrt{2b} \exp(-1/2) + 1$ . Now since  $a\sqrt{2b} > 1/(1 - \exp(-1/2)) > \exp(1/2)$ ,  $F'(0) = F'(+\infty) = 1$ , we have that  $F'$  has a root  $r_1$  in  $(0, 1/\sqrt{2b})$  and another root  $r_2$  in  $(1/\sqrt{2b}, \infty)$ . All these further imply that  $F$  is increasing in  $(0, r_1)$  and  $(r_2, \infty)$  and decreasing in  $(r_1, r_2)$ . Notice that  $F'(\sqrt{\log(2ab)/b}) = 1 - \sqrt{\log(2ab)/b} > 0$ , since we have assumed that  $\log(2ab) < b$ , so we have that  $r_2 \in (1/\sqrt{2b}, \sqrt{\log(2ab)/b})$ . Also notice that  $F(1/\sqrt{2b}) = a \exp(-1/2) + 1/\sqrt{2b} < a = F(0)$ , since we have assumed that  $a\sqrt{b} > 1/(1 - \exp(-1/2))$ . Therefore, 0 is not the minimum point. In summary, the minimum point of  $F$  is in  $(1/\sqrt{2b}, \sqrt{\log(2ab)/b})$  and satisfies  $2abx \exp(-bx^2) = 1$ .  $\square$

## Acknowledgements

The first author's research was supported in part by the NSF Grants: DMS-1561141 and DMS-1613016. The authors are sincerely grateful to the Associate Editor and two anonymous referees for their many corrections and suggestions that help to significantly improve the manuscript.

## References

- Y. Aït-Sahalia and J. Jacod. Estimating the degree of activity of jumps in high frequency data. *The Annals of Statistics*, pages 2202–2244, 2009a.
- Y. Aït-Sahalia and J. Jacod. Testing for jumps in a discretely observed process. *The Annals of Statistics*, pages 184–222, 2009b.
- Y. Aït-Sahalia and J. Jacod. Is brownian motion necessary to model high-frequency data? *The Annals of Statistics*, pages 3093–3128, 2010.
- Y. Aït-Sahalia and J. Jacod. *High-Frequency Financial Econometrics*. Princeton University Press, New Jersey, 2014.
- S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge university press, 2004.
- R. Cont and C. Mancini. Nonparametric tests for pathwise properties of semimartingales. *Bernoulli*, 17(2):781–813, 2011.
- F. Corsi, D. Pirino, and R. Renò. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159:276–288, 2010.

J. Fan and Y. Wang. Spot volatility estimation for high-frequency data. *Statistics and its Interface*, 1(2):279–288, 2008.

J.E. Figueroa-López. Statistical estimation of Lévy-type stochastic volatility models. *Annals of Finance*, 8(2):309–335, 2012.

J.E. Figueroa-López and C. Li. Optimal kernel estimation of spot volatility of stochastic differential equations. To appear in *Stochastic Processes and their Applications*. Preprint available at <https://pages.wustl.edu/figueroa/publications>, 2020.

J.E. Figueroa-López and C. Mancini. Optimum thresholding using mean and conditional mean square error. *Journal of Econometrics*, Vol. 208, issue 1, 179–210, 2019.

J.E. Figueroa-López and J. Nisen. Optimally thresholded realized power variations for Lévy jump diffusion models. *Stochastic Processes and their Applications*, 123(7):2648–2677, 2013.

J.E. Figueroa-López and J. Nisen. Second-order properties of thresholded realized power variations of FJA additive processes. *Statistical Inference for Stochastic Processes*, Vol. 22, Issue 3, 431–474, 2019.

D. Foster and D. Nelson. Continuous record asymptotics for rolling sample variance estimators. *Econometrica*, 64(1):139–174, 1996.

J. Jacod. Asymptotic properties of power variations of Lévy processes. *ESAIM: Probability and Statistics*, 11, 173–196, 2007.

J. Jacod. Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Processes and Their Applications*, 118(4), 517–559, 2008.

J. Jacod and P. Protter. *Discretization of Processes*. Springer-Verlag Berlin Heidelberg, 2012.

B. Jing, X. Kong, Z. Liu, and P. Mykland. On the jump activity index for semimartingales. *Journal of Econometrics*, 166(2):213–223, 2012.

D. Kristensen,. Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory*, 26(01), 60–93, 2010.

C. Mancini. Disentangling the jumps of the diffusion in a geometric jumping brownian motion. *Giornale dell'Istituto Italiano degli Attuari*, 64(19-47):44, 2001.

C. Mancini. Estimation of the characteristics of the jumps of a general poisson-diffusion model. *Scandinavian Actuarial Journal*, 2004(1):42–52, 2004.

C. Mancini. Non parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36:270–296, 2009.

C. Mancini, V. Mattiussi, and R. Renò. Spot volatility estimation using delta sequences. *Finance & Stochastics*, 19(2):261–293, 2015.

F. Ueltzhöfer. On non-parametric estimation of the Lévy kernel of Markov processes. *Stochastic Processes and their Applications*, 123:3663–3709, 2013.

D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer-Verlag Berlin Heidelberg, 1998.

L. Zhang, P. Mykland, and Y. Aït-Sahalia. A tale of two time scales. *Journal of the American Statistical Association*, 100(472), 2005.