Threshold estimation for stochastic differential equations with jumps

Yasutaka Shimizu
Osaka University, Osaka, JAPAN
E-mail: yasutaka@sigmath.es.osaka-u.ac.jp

Abstract

In recent years, the statistical inference for discretely observed jump processes is an important issue in finance and insurance. Due to the discreteness, it is unclear that an increment of neighboring data essentially comes from continuous or discontinuous shocks, which causes some difficulties for estimating unknowns in the underlying process. Threshold Estimation is one of the useful techniques to disentangle the continuous shocks and real jumps, where if an increment of data is smaller than a predetermined threshold, then we regard the increment does not include any jump. However, the choice of the threshold has optionality in practice, and the standard way has not been established yet. In this paper, we shall propose how to select some “optimal” thresholds from given data, and study the finite-sample performance by simulations.

Keywords: Jump-type processes, discretely observations, threshold estimation, threshold selection.

1 Introduction

Consider a (1-dim) stochastic process $X = (X_t)_{t \geq 0}$ satisfying a stochastic differential equation

$$dX_t = a(X_t, \mu) dt + b(X_t, \sigma) dW_t + c(X_t, \delta) dZ_t,$$

where $a, b$ and $c$ are suitable functions with parameters $(\mu, \sigma, \delta)$, $W$ is a Wiener process, and $Z$ is a compound Poisson process with intensity $\lambda$ and a jump-distribution $F$. Suppose a situation that the process $X$ is observed at discrete time points $t^n_i = i \Delta_n (i = 0, 1, \ldots, n)$ for $\Delta_n > 0$. In the sequel, we suppose that $\Delta_n \rightarrow 0$ as $n \rightarrow \infty$.

Such a jump-type stochastic process is recently a standard tool, e.g., for modeling asset values in finance and insurance, and as a practical demand, we need to specify parameters $(\mu, \sigma, \delta, \lambda)$ and $F$ from discrete samples of $X$. A useful technique to execute those inference is the threshold estimation, which is proposed by Mancini (2001) and also by Shimizu (2002), independently. The fundamental idea is doing the following judgement: for $\Delta^n_i X := X^n_{t^n_i} - X^n_{t^n_{i-1}}$ and a predetermined threshold $r_n > 0$,

$$|\Delta^n_i X| \leq r_n \quad \Rightarrow \quad \text{No ‘large’ jump occurred in } (t^n_{i-1}, t^n_i);$$

$$|\Delta^n_i X| > r_n \quad \Rightarrow \quad \text{A ‘large’ jump occurred in } (t^n_{i-1}, t^n_i),$$

and use $\{\Delta^n_i X : |\Delta^n_i X| \leq r_n\}$ to estimate a continuous part of $X$, and use $\{\Delta^n_i X : |\Delta^n_i X| > r_n\}$ to estimate jump part. see Shimizu and Yoshida (2006), or Ogihara and Yoshida (2011). The similar idea is also used by various authors in different contexts; see, e.g., Aït-Sahalia et al. (2009a,b,2011,2012), Gobbi and Mancini (2008), Cont and Mancini (2011), among others.

From a viewpoint of asymptotic theory, it is desirable that, e.g., $r_n = O(\Delta^n_n \varpi)$ for $\varpi < 1/2$ as $\Delta_n \rightarrow 0$ to disentangle jumps and diffusion shocks, so most authors consider thresholds of the form $r_n = \alpha \Delta^n_n$ for a constant $\alpha > 0$. However, how to choose constants $(\alpha, \varpi)$ is a big problem because most statistics based on this idea are very sensitive with choosing $r_n$. Some authors propose $(\alpha, \varpi)$, sometimes data-adaptively, via simulations, or ad hoc methods, but a standard way has not established yet. We shall propose an automatic algorithm to choose $r_n$ from given data along the results by Shimizu (2010).
2 Thresholds via a bias correction

Consider the following statistic:

\[ \hat{I}_V(r_n) := \sum_{i=1}^{n} \frac{\Delta_i^2}{|\Delta_i^2| \leq r_n}, \]

which is a threshold estimator of the integrated volatility \( I_0^T b^2(X_t, \sigma) \, dt \); see, e.g., Mancini (2009). The idea to choose \( r = r_n \) is to make the bias of \( \hat{I}_V(r) \) as small as possible. Let

\[ B_n(r) := \mathbb{E}\left[ \hat{I}_V(r) - \int_0^{T_n} b^2(X_t) \, dt \right] \]

We say that \( \hat{I}_V(r_n) \) is asymptotically unbiased if \( B_n(r_n) \to 0 \) as \( n \to \infty \).

**Theorem 1.** Suppose that \( a \) and \( b \) are of linear growth, and that \( \lim_{n \to \infty} \sup_{t \leq T_n} \mathbb{E}|X_t|^4 < \infty \). Moreover suppose that

\[ \lim_{n \to \infty} \sqrt{T_n} \left( r_n + \frac{\Delta_n}{r_n} \right) = 0. \tag{2} \]

Then \( \hat{I}_V(r_n) \) is asymptotically unbiased.

**Remark 1.** The condition (2) implies that, in finite activity case, taking \( r_n = O(\Delta_n^{-\omega}) \) with \( \omega < 1/2 \) is enough as \( T_n = T \) (fixed). As \( T_n \to \infty \), we further need \( n\Delta_n^{1+\delta} \to 0 \) for some \( \delta \in (0, 1/2) \). These facts are consistent with the results in earlier works; e.g., Mancini (2009), or Shimizu and Yoshida (2006).

**Theorem 2.** Suppose the same assumptions as in Theorem 1, and that \( \Delta_n \to 0 \). Then, for each \( r > 0 \),

\[ \left| B_n(r) - e^{-\lambda \Delta_n} \mathbb{E}[\hat{B}_n(r)] \right| = O(n\Delta_n^{3/2}), \quad n \to \infty, \]

where

\[ \hat{B}_n(r) := \sum_{i=1}^{n} \left[ I_i^n(r) - \int_{|z|>r} |z|^2 \Phi(dz; \Sigma_i^n) \right], \tag{3} \]

\( \Phi(z; \Sigma) \) is a Gaussian distribution function with mean zero and variance \( \Sigma \), and

\[ \Sigma_i^n := \int_{t_{i-1}}^{t_i} b^2(X_s, \sigma) \, ds, \quad I_i^n(r) := \lambda \Delta_n \int_{|c(X_{t_{i-1}} - X_{t_i})| \leq r} |c(X_{t_{i-1}}, \delta)|^2 F(dz). \]

Due to the above result, \( B_n \approx 0 \) if \( \hat{B}_n \approx 0 \) when \( n\Delta_n^{3/2} \to 0 \). Now, our aim is to find \( r = r_n \) that minimizes \( \hat{B}_n(r) \) in (3):

\[ r_{opt} := \arg \min_{r \geq 0} |\hat{B}_n(r)| \]

However, \( \hat{B}_n(r) \) still has some unknown parameters \((\mu, \sigma, \delta, \lambda, F)\), which are also to be estimated via a suitable threshold.

Let \( \hat{\Sigma}_i^n(r_n) \) and \( \hat{I}_i^n(r; r_n) \) be estimators of \( \Sigma_i^n \) and \( I_i^n(r) \) with parameters replaced with some threshold estimators: e.g., \( \hat{\Sigma}_i^n(r_n) = \int_{t_{i-1}}^{t_i} b^2(X_s, \hat{\sigma}_n(r_n)) \, ds \) where \( \hat{\sigma}_n(r_n) \) is a threshold estimator of \( \sigma \). How to construct threshold estimators, see, e.g., Shimizu and Yoshida (2006), Mancini (2009), Shimizu (2009), etc. Using such plug-in estimators \( \hat{\Sigma}_i^n(r_n) \) and \( \hat{I}_i^n(r; r_n) \), we propose the following algorithm:
we construct estimators for given (fixed) \( r \) respectively, under some regularities as \( n \) where \( b \) jump-detecting test and Mykland (2008) propose a according to Shimizu and Yōshida (2006). We compare our algorithm with several earlier methods: Lee MLE-type threshold estimator For this model, we can construct follows:

We try the algorithm in Section 2 as well as a modification in Section 3 by a simple O-U model as

\[ e \text{ is estimated as the root of the following estimated equation in } e \in (0, 1]: \text{for given threshold } r_n > 0, \]

\[ e = \int_{|z|>r_n} F_{\hat{\lambda}(r_n)}(dz). \quad (4) \]

\textbf{Plug-in algorithm} (Shimizu, 2010)

\[ \textbf{[Step 0]} \text{ Choose a pilot threshold } r_n^{(0)} < \max_{1 \leq i \leq n} |\Delta_i^n X|, \text{ and calculate } \hat{\Sigma}_i^n (r_n^{(0)}) \text{ and } \hat{T}_i^n (r; r_n^{(0)}); \]

\[ \textbf{[Step k]} \text{ For } k = 1, 2, \ldots, \text{ iterate the following steps:} \]

\[ \text{[Step k]-1: Calculate } \hat{\Sigma}_i^n (r_n^{(k-1)}) \text{ and } \hat{T}_i^n (r; r_n^{(k-1)}); \]

\[ \text{[Step k]-2: Find the root } r = r_n^{(k)} \text{ to the following equation:} \]

\[ \sum_{i=1}^n \left[ \hat{T}_i^n (r; r_n^{(k-1)}) - \int_{|z|>r} |z|^2 \Phi \left( dz; \hat{\Sigma}_i^n (r_n^{(k-1)}) \right) \right] = 0. \]

Iterate [Step k] until \( \{r_n^{(k)}\} \in \mathbb{N} \) ‘converges’ to a constant \( r^* \), which is used as a threshold.

\textbf{Remark 2.} We will numerically check that \( r_n^{(k)} \to r^* \) as \( k, n \to \infty \).

\section{A modification of estimators of \( \lambda \) and \( F \)}

For simplicity, suppose that \( c(x, \delta) \equiv 1 \), and that a distribution \( F \) is parametrized with \( \zeta \in \mathbb{R} : \]

\[ F(dz) = F_{\zeta}(dz), \quad \zeta := \int_{\mathbb{R}} G(z) F_{\zeta}(dz) \quad \text{for a known function } G. \]

Then threshold estimators of \( \lambda \) and \( \zeta \) is given by (e.g., Shimizu, 2009)

\[ \hat{\lambda}(r) = \frac{1}{T_n} \sum_{i=1}^n 1_{\{|\Delta_i^n X|>r\}}, \quad \hat{\zeta}(r) = \frac{1}{\hat{\lambda}(r) T_n} \sum_{i=1}^n G(\Delta_i^n X) 1_{\{|\Delta_i^n X|>r\}}, \]

where \( \hat{\delta}(r) \) is also a threshold estimator of a parameter \( \delta \). Those are consistent estimators of \( \lambda \) and \( \zeta \), respectively, under some regularities as \( n \to \infty \) and \( r_n \to 0 \) with a suitable rate. However, in practice, we construct estimators for given (fixed) \( r = r_n > 0 \), and ‘small’ jumps less than \( r_n \) are cut off. Hence the rate of expected information loss of \( \hat{\lambda}(r_n) \) seems approximately

\[ \mathbb{E} \left[ \hat{\lambda}(r_n)/\lambda \right] \approx \int_{|z|>r_n} F_{\zeta}(dz) =: e_n, \]

that is, \( \hat{\lambda}^* = e_n^{-1} \hat{\lambda}(r_n) \) and \( \hat{\zeta}^* = e_n \hat{\zeta}(r_n) \) would be compensate the information loss. The value of \( e_n \) is estimated as the root of the following estimated equation in \( e \in (0, 1]: \]

\[ e = \int_{|z|>r_n} F_{e \hat{\zeta}(r_n)}(dz). \]

\section{Simulation}

We try the algorithm in Section 2 as well as a modification in Section 3 by a simple O-U model as follows:

\[ dX_t = -\mu X_t dt + \sigma dW_t + dZ_t^{(\lambda, \alpha, \beta)}, \quad X_0 = 0, \]

where \( Z \) is a compound Poisson with intensity \( \lambda \) and Gaussian jumps with mean \( \alpha \) and variance \( \beta \). For this model, we can construct MLE-type threshold estimator for all the parameters \( (\mu, \sigma, \lambda, \alpha, \beta) \) according to Shimizu and Yōshida (2006). We compare our algorithm with several earlier methods: Lee and Mykland (2008) propose a jump-detecting test in each \( (t_{n-1}^n, t_n^n) \) via the extreme value theory. Their
method corresponds to use a data-adaptive threshold \( r = r_t^n \in F_{t^n} \) in \((t_{i-1}^n, t_i^n]\). Aït-Sahalia and Jacod (2009b) propose to use \( r_n = \alpha \tilde{\sigma} \Delta_{n}^{0.48} \) with, e.g., \( 3 \leq \alpha \leq 5 \) and \( \Delta = 0.47 \) or 0.48, and \( \tilde{\sigma} \) is an estimator of \( \sigma \) without using a threshold, for example, use a bi-power variation etc.; see, e.g., Barndorff-Nielsen and Shephard (2004). Shimizu (2008) proposes an algorithm to choose \( r_n \) via a bias correction of \( \tilde{\lambda}(r_n) \). This is also an automatic way to choose the threshold. In the end, we shall compare the following candidates. The symbol * means the result with a modification as in Section 3.

- **LM**: Test by Lee and Mykland (2008) with significance level of 5%\(^1\).
- **S08**: By Shimizu (2008). The threshold is chosen so that the bias of \( \hat{\lambda}(r_n) \) is as small as possible.
- **AS*/AS*+**: \( r_n = 3\tilde{\sigma}_B \Delta_{n}^{0.48} \) without/with the modification, where \( \tilde{\sigma}_B^2 \) is the realized bi-power variation. This is the minimum threshold used by Aït-Sahalia and Jacod (2009b).
- **AS+/AS++**: \( r_n = 5\tilde{\sigma}_B \Delta_{n}^{0.47} \) without/with the modification, which is the maximum threshold used by Aït-Sahalia and Jacod (2009b).
- **S10/S10+**: Our method without/with the modification. The threshold is chosen so that the bias of \( \hat{IV}(r_n) \) is as small as possible.

We set the true value as \( (\mu, \sigma, \lambda, \alpha, \beta) = (0.2, 0.7, 10.0, 0.0, 0.5) \). Table 1 shows means and standard deviations (s.d.) of each estimator through 5000 times experiments in the case where \( n = 3000, \Delta_n = 1/250 \), and Figure 1 shows relative errors of estimators of \( \sigma \) by \( AJ \pm \) and our proposal \( S10 \pm \) for different values of \( \sigma = 0.05 \sim 0.8 \). We see that our proposal \( S10 \pm \) always choose a suitable threshold such that the estimation of \( \sigma \) is so nice. From those results, we see a superiority of our automatic ‘plug-in’ algorithm.

### 5 Conclusions

- Threshold estimators are very sensitive for a choice of a threshold. They are some ad hoc proposals to select a threshold, which highly depend on the true model. Therefore we need to establish a way to choose it suitably from given data without optionality.

- Our method can automatically determine a suitable threshold that can give us accurate estimation of parameters, if needed, via a suitable modification in Section 3.

- If the diffusion component is \( b(x, \sigma) \equiv 0 \), the estimator \( \hat{IV}(r_n) \) with a threshold given by our method becomes close to zero, which would be available for testing the existence of a jump from discrete data to observe an estimator of a diffusion term.

- In this paper, we used a model with compound Poisson jumps, but our method can be applied to models with infinite activity jumps in principle; Shimizu (2010) reports the details with numerical experiments.

### References


\(^1\)We used \( \beta^* = -\log(-\log(0.95)) \) and \( K = \lfloor n^{1/2} \rfloor \) in their notation.
Table 1: Joint estimation using “Threshold Estimation”. The bold type means the best estimate. The mean (s.d.) of $\hat{\sigma}$ through the experiments was 0.8891 (s.d. = 0.0620), which has a large bias from the truth 0.7.

<table>
<thead>
<tr>
<th></th>
<th>LM</th>
<th>S08</th>
<th>AJ+/AJ++</th>
<th>AJ−/AJ−−</th>
<th>S10/S10*</th>
<th>true</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.139</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>NA</td>
<td>0.114</td>
<td>0.332</td>
<td>0.199</td>
<td>0.141</td>
<td>(r_{opt})</td>
</tr>
<tr>
<td>$\mu$</td>
<td></td>
<td>0.209</td>
<td>0.198</td>
<td>0.215</td>
<td>0.214</td>
<td>0.210</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(0.093)</td>
<td>(0.082)</td>
<td>(0.095)</td>
<td>(0.087)</td>
<td>(0.085)</td>
<td>0.2</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.776</td>
<td>0.674</td>
<td>0.782</td>
<td>0.717</td>
<td>0.699</td>
<td>0.7</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(0.033)</td>
<td>(0.014)</td>
<td>(0.027)</td>
<td>(0.012)</td>
<td>(0.010)</td>
<td>0.7</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(0.667)</td>
<td>(1.690)</td>
<td>(0.624/1.229)</td>
<td>(0.724/1.023)</td>
<td>(0.897/1.077)</td>
<td>10</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.002</td>
<td>0.0016</td>
<td>0.0017/0.0015</td>
<td>0.0011/0.0016</td>
<td>0.0009/0.0014</td>
<td>0.0</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(0.093)</td>
<td>(0.055)</td>
<td>(0.0997/0.0627)</td>
<td>(0.0826/0.0639)</td>
<td>(0.0731/0.0598)</td>
<td>0.0</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.713</td>
<td>0.459</td>
<td>0.767/0.489</td>
<td>0.644/0.508</td>
<td>0.577/0.493</td>
<td>0.5</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(0.092)</td>
<td>(0.080)</td>
<td>(0.100/0.081)</td>
<td>(0.085/0.074)</td>
<td>(0.076/0.071)</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Figure 1: Relative errors of mean of $\hat{\sigma}(r)$ through 5000 simulations: $\approx \frac{E[\hat{\sigma}(r)]-\sigma}{\sigma} \times 100$ (%), via AJ± and our proposal S10*. Although AJ* has a wide range of values, our proposal S10* is always accurate.


